Calhoun: The NPS Institutional Archive

DSpace Repository

Reports and Technical Reports	Faculty and Researchers' Publications

2022

Performance Evaluation Trait Validation

Helzer, Erik; Bacolod, Marigee

Monterey, California: Naval Postgraduate School

https://hdl.handle.net/10945/71857

NPS-DDM-22-005

NAVAL POSTGRADUATE SCHOOL

## MONTEREY, CALIFORNIA

PERFORMANCE EVALUATION TRAIT VALIDATION

by

Erik G. Helzer, PhD

Marigee Bacolod, PhD

October 2022

Distribution Statement A: Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE October 2022	2. REPORT TYPE Technical Report	3. DATES COVERED	
		START DATE 10/24/2021	END DATE 10/22/2022

4. TITLE AND SUBTITLE
Performance Evaluation Trait Validation

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER 0605853N/2098
5d. PROJECT NUMBER NPS-22-N137-A	5e. TASK NUMBER	5f. WORK UNIT NUMBER

6. AUTHOR(S)
Erik G. Helzer, PhD
Marigee Bacolod, PhD

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School, Monterey, CA Naval Research Program 1 University Circle Monterey, CA 93943	8. PERFORMING ORGANIZATION REPORT NUMBER NPS-DDM-22-005

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) N1 Manpower, Personnel, Training & Education Millington, TN	10. SPONSOR/MONITOR'S ACRONYM(S) NRP; OPNAV N1	11. SPONSOR/MONITOR'S REPORT NUMBER(S) NPS-DDM-22-005 NPS-22-N137-A

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A: Approved for public release, Distribution: Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this report, we present the results of a year-long study supporting the Navy's efforts to transform its existing Performance Evaluation System (PES) with an evaluation tool that has been updated to reflect current Navy values and priorities. Our team validated and refined trait and value statements (TVS) developed for rating Sailors on a variety of dimensions, including Leadership, Teamwork, Communication, Resilience, and other character strengths. In addition, our team developed rating scales for assessing Sailors' future potential for performance at the next paygrade, as well as a comparative assessment tool for evaluating Sailors against their peers. We then conducted a large-scale prototyping study, which involved the development of a performance evaluation prototype instrument, recruitment of Sailors-reported-on and performance raters, and statistical analysis of the properties of the prototype instrument – in particular, the measurement validity of the TVS. Using data from 606 performance raters, we find that ratings of Sailors based on past performance and future potential predict comparative assessments, with some traits (ex. Leadership) better distinguishing the top from middle performers. Few of the trait ratings distinguished among bottom performers, although such performers were relatively rare in our sample. Ratings based on past performance and future potential were correlated with one another; while they convey unique information about the Sailor-reported-on, both sets of ratings predicted workplace behaviors. We close with four concrete recommendations for action. First, we recommend amplifying the comparative assessment in future PES. Second, consider adopting ratings of future potential for developmental/coaching purposes using actionable, concrete developmental feedback. Third, focus on a subset of performance and/or future potential traits in assessing job performance to keep pes simple and useful. Finally, we recommend assessing the predictive validity of these new measures using multi-source data.

15. SUBJECT TERMS
performance evaluation system, performance assessment, traits, job performance, measurement validity

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 87
a. REPORT U	b. ABSTRACT U	C. THIS PAGE U		

19a. NAME OF RESPONSIBLE PERSON Erik G. Helzer	19b. PHONE NUMBER *(Include area code)* 831-656-2471

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

Ann E. Rondeau
President

Scott Gartner
Provost

Distribution Statement A: Approved for public release. Distribution is unlimited.

This report was prepared by:

_____
 Erik G. Helzer
 Associate Professor

_____
 Marigee Bacolod
 Associate Professor

Reviewed by:

Released by:

_____
 Raymond Jones, Chairman
 Department of Defense Management

_____
 Kevin B. Smith
 Vice Provost for Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In this report, we present the results of a year-long study supporting the Navy's efforts to transform its existing Performance Evaluation System (PES) with an evaluation tool that has been updated to reflect current Navy values and priorities. Our team validated and refined trait and value statements (TVS) developed for rating Sailors on a variety of dimensions, including Leadership, Teamwork, Communication, Resilience, and other character strengths. In addition, our team developed rating scales for assessing Sailors' future potential for performance at the next paygrade, as well as a comparative assessment tool for evaluating Sailors against their peers. We then conducted a large-scale prototyping study, which involved the development of a performance evaluation prototype instrument, recruitment of Sailors-reported-on and performance raters, and statistical analysis of the properties of the prototype instrument – in particular, the measurement validity of the TVS. Using data from 606 performance raters, we find that ratings of Sailors based on past performance and future potential predict comparative assessments, with some traits (ex. Leadership) better distinguishing the top from middle performers. Few of the trait ratings distinguished among bottom performers, although such performers were relatively rare in our sample. Ratings based on past performance and future potential were correlated with one another; while they convey unique information about the Sailor-reported-on, both sets of ratings predicted workplace behaviors. We close with four concrete recommendations for action. First, we recommend amplifying the comparative assessment in future PES. Second, consider adopting ratings of future potential for developmental/coaching purposes using actionable, concrete developmental feedback. Third, focus on a subset of performance and/or future potential traits in assessing job performance to keep pes simple and useful. Finally, we recommend assessing the predictive validity of these new measures using multi-source data.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

I.   INTRODUCTION

A.   BACKGROUND

A valid and credible performance evaluation system (PES) is critical for identifying and managing talent in the US Navy. The Navy's *Sailor 2025* initiative called for an updating of the Navy's personnel management system to reflect the Navy's current goals with respect to recruitment, retention, and advancement. In addition, concerns expressed by Task Force One Navy point to the need for a PES that is both fair and objective. Major efforts to generate updated performance trait statements that are consistent with Navy doctrine, instructions on performance appraisal, and Sailor values were conducted in 2002 (under the Task Force Excel 5-Vector Model) and 2019 (by researchers associated with the performance evaluation transformation initiative). In 2019 a working group identified trait and value statements (TVS) to serve as the basis of revised performance evaluation metrics. This set of TVS modernizes the performance criteria against which Sailors would potentially be evaluated; however, a systematic study on the validity of this set of TVS has not been conducted.

Herein, we report the results of a year-long study supporting the Navy's efforts to transform the existing PES with an evaluation tool that has been updated to reflect current Navy values and priorities. We describe the process by which we validated and refined the TVS, benchmarking items against current Navy doctrine and performance evaluation materials from other military services. We then present the results of a large-scale prototyping study, which involved the development of a performance evaluation prototype instrument, recruitment of Sailors-reported-on and performance raters, and statistical analysis of the properties of the prototype instrument – in particular, the measurement validity of the TVS. We close by offering conclusions and recommendations for updates to the existing FITREP/EVAL instrument based on our research findings.

1.   Transforming the Navy's Performance Evaluation System

In an effort to modernize and improve the existing PES, the Navy stood up *Sailor 2025*, which encompasses a set of human resources initiatives aimed at aligning its personnel management system with accepted current best practices and Navy needs (Serbu, 2017;

Weatherspoon, 2016). The initiative outlines widespread modernization goals for the PES, to include a transformation and modernization of its performance evaluation instruments.

Currently, the Navy uses three separate tools in its PES to evaluate service member performance. Evaluations (EVALs), Chief Evaluations (CHIEFEVALs), and Fitness Reports (FITREPs) are used to rate the performance of enlisted Sailors, Chief Petty Officers, and officers, respectively. On the current FITREP instrument, for example, Officers are rated on a set of performance traits including tactical performance, leadership, and professionalism. The current FITREP also documents notable job achievement, command duties performed, and reporting seniors' recommendations for meeting career milestones and promotions. Our focus in this research is on assessing the validity of proposed revisions and updates to the performance traits section of the FITREP.

B.   ASSESSING THE MEASUREMENT VALIDITY OF REFINED TVS

In psychology and related disciplines, new and existing measurement tools are assessed for quality based in part on their *measurement validity*. Sound measures are those that have been shown to be *valid* across several different indicators, or criteria, of validity. Table 1 displays the validity criteria that we focused on as we were developing, refining, and ultimately analyzing the data obtained using the updated, refined TVS. The first two criteria, construct and face validity, were assessed using content analysis and benchmarking procedures that compared the TVS to current Navy doctrine and performance evaluation instruments used in other branches of the military. The TVS would be considered construct and face valid if the broad traits, sub-traits, and value statements, or a subset thereof, are conceptually related to the traits, attributes, and/or values that the Navy espouses and deems essential to advancement within its ranks.

The next two criteria, predictive and convergent validity, were assessed using data from our prototype field study. Predictive and convergent validity pertain to the degree to which our new measures based on the TVS predict important job performance outcomes and correlate with similar attributes or Sailor characteristics. If some or all of the TVS are reliably correlated with indicators of job performance and workplace behavior in a theoretically-predicted direction (e.g., higher ratings of *character* are associated with more ethical job performance), these TVS would have demonstrated predictive validity. If TVS are associated with related aspects of a Sailor's

character or other person-level characteristics, they would satisfy the criterion of convergent validity. The data reported here allowed us to examine aspects of both predictive and convergent validity; however, a more thorough investigation will require merging data from Sailors' personnel records with the prototype testing data to examine correlations between other indicators of job performance (from prior FITREPs or promotion history, for example) and TVS ratings. At present, this merging process is on-going; our team will provide a thorough analysis of these issues with our FY23 research project.

The final criterion, divergent validity, pertains to the degree to which the TVS are statistically <u>un</u>related to (i.e., *diverge from*) characteristics of the Sailor that should be unrelated to job performance. In the context of this study, this took the form of investigating potential adverse impact, such as whether Sailors belonging to certain demographic groups were systematically rated more or less favorably using the TVS items. If TVS ratings are statistically correlated with Sailors' gender or race, for example, the risk of adverse impact is higher than if the TVS ratings are uncorrelated with these characteristics. As with predictive and convergent validity, the best test for divergent validity involves merging data from Sailors' personnel records with the prototype testing data. At present, this merging process is on-going; thus, divergent validity of the TVS will be assessed in our FY23 research project.

Table 1.   Dimensions of measurement validity assessed in this research project

Validity	Criterion	In context of TVS:
Construct validity	Does the assessment tool measure the appropriate/intended theoretical constructs?	Are the TVS reflective of legitimate/desired USN promotion criteria?
Face validity	Does the assessment tool appear to measure what it purports to measure?	Do the TVS seem to reflect the corresponding trait or value?
Predictive validity	Does the assessment tool predict outcomes it ought to predict?	Do TVS ratings correlate with/predict indicators of job performance?
Convergent validity	Does the assessment tool track with theoretically similar constructs/assessments?	Do TVS ratings correlate with similar attributes or Sailor characteristics?

| Divergent validity | Does the assessment tool *not* track with theoretically dissimilar or problematic constructs? | Are the TVS free from adverse impact? |

C.   ORGANIZATION OF TECHNICAL REPORT

The remainder of the technical report is structured as follows. In section 2, we provide a literature review of scholarship pertaining to performance evaluation systems both in general organizations and in the Naval services in particular. In section 3, we describe our methodology for validating and refining the proposed TVS set, developing our performance evaluation prototype instrument, and then testing our prototype in a large field study. In section 4, we present statistical analysis of the data from the field study that speaks to the structure and predictive validity of the items in the prototype instrument. In section 4 we close with conclusions and recommendations based on findings from the research project. The prototype instrument is provided in its entirety in the Appendix, as well as additional statistical analysis supporting section 4.

II. RELEVANT SCHOLARSHIP ON PERFORMANCE EVALUATION SYSTEMS

Performance evaluation or *performance appraisal* entails "a formal process, which occurs infrequently, by which employees are evaluated by some judge (typically a supervisor) who assesses the employee's performance along a given set of dimensions, assigns a score to that assessment, and then usually informs the employee of his or her formal rating" (DeNisi & Murphy, 2017, p. 421). Performance evaluation is one critical aspect in a broader system or performance management, which organizations use for two principal purposes: to aid administrative and personnel decisions and to facilitate employee development (Campbell & Wiernik, 2015).

In defining the playing field, Campbell and Wiernik (2015) carefully distinguish *performance* – the "things that people actually do, actions they take, that contribute to the organization's goals" (p. 48) – from the antecedents (e.g., personality traits, job characteristics) and consequences of performance. Although antecedents and consequences are clearly important to organizations as part of their broader performance management efforts, employees may have much less control over these than over actual performance. Performance evaluation, then, can involve a variety of methods –including objective metrics, supervisor evaluations, as well as peer-, self-, and stakeholder-ratings– to assess the degree to which individuals take actions to contribute to organizational goals, both generally and through their job-specific roles.

PES serve a number of purposes for both organizations and their employees (Campbell & Wiernik, 2015; Cleveland, Murphy, & Williams, 1989). PES are used both to *inform* and *justify* "high stakes" decisions, including promotion, compensation, reassignment, or termination. PES also may be used for developmental purposes, including feedback and self-managed performance improvements. Finally, PES may be used for research purposes, typically as an outcome variable in studies of organizational or individual performance. Although a reliable and valid PES is beneficial to each of these purposes, the specific purpose of the PES may determine what PES-quality indicators are most important.

## A.	EVALUATIVE AND DEVELOPMENTAL FUNCTIONS OF PES

The literature distinguishes between components of performance management that are intended to facilitate *evaluation* and those that are intended to promote *development*. Evaluation and development are distinct both from the perspective of the rater and the perspective of the individual being rated. For this reason, the literature provides clear guidance that evaluative and developmental feedback should be maintained as separate processes in the PES (for a review, see Boswell & Boudreau, 2001).

From the perspective of the rater, evaluative assessments, which are used for high stakes decision-making (promotion, compensation, termination, awards), typically focus raters on how individuals compare (i.e., "rack and stack") against their peers, whereas developmental assessments, which seek to promote growth and improvement, encourage raters to focus on how individuals compare against their potential (Cleveland et al., 1989). Many empirical studies point out the costs associated with conflating these two purposes in PES, including inaccurate, biased, or "gamed" assessments (for a review, see Cleveland et al., 1989). In addition, PES systems that conflate evaluative and developmental assessment may be ill suited for identifying high-potential, still-developing employees.

From the perspective of the individual being rated, evaluative and developmental assessments may prompt distinct cognitive and behavioral responses. Some research suggests that developmental feedback is received more favorably than evaluative feedback. Boswell and Boudrea (2001) found that employee satisfaction with performance assessments, as well as with their assessor, were positively correlated with their perception of that the assessment was developmental in nature. In turn, greater employee satisfaction with performance evaluation has been linked to stronger organizational commitment to and lower turn-over intentions, as well as to greater work performance, particularly among intrinsically motivated employees (Kuvaas, 2006).

More generally, literature suggests that the effects of performance evaluation (of either the evaluative or developmental kind) on work performance, employee attitudes, or other employee outcomes tends to depend on the degree to which recipients feel empowered, engaged, or motivated by the evaluation and the degree to which the system is perceived as fair (Cawley, Keeping, & Levy, 1998; Gruman & Saks, 2011; Pettijohn, Pettijohnn, & d'Amico, 2001).

B.   DEVELOPMENT OF STANDARDS AND ELEMENTS OF PES

Boice and Kleiner (1997) lay out a series of considerations and steps for designing effective PES. At the center of a well-designed PES are organizational objectives – i.e., clarity around what the organization is trying to do and how employees in various roles facilitate those objectives. From there, more specific objectives can be established for departments, units, and individual positions. Appropriate standards based on these objectives provide employees with a set of criteria against which their job performance will be evaluated.

It is not enough, however, to stipulate performance standards and then leave raters to the task of evaluation. Once standards for performance are established and promulgated, managers and other raters must be trained on the PES to ensure that inputs and general use align the intended use of the system and organizational objectives. Rater training may reduce systematic errors in performance evaluations by increasing collective understanding of the criteria and standards of evaluation (Roch, Woehr, Mishra, & Kieszczynska, 2011; Woehr & Huffcutt, 1994). In addition, training can facilitate buy-in on the part of those charged with making consequential decisions on the basis of the PES (Boice & Kleiner, 1997). Research points to Frame of Reference (FOR) training as a particularly effective method of calibrating ratings across different raters (Uggerslev & Sulsky, 2008). To prevent "drift," refresher trainings are recommended.

1.   What aspects of performance should be evaluated?

The question of which aspects of job performance should be targeted by PES has received much attention from researchers and theorists. Empirical efforts to identify broad factors of performance have yielded a variety of taxonomies; however, common themes emerge among them.

Campbell and colleagues used a bottom-up approach to identifying components of performance evaluation. Using data from the Army, they identified eight principal components of demonstrated job performance, including technical proficiency in both job-specific and non-job-specific aspects of work, communication, effort/initiative, personal discipline, team performance, and leadership/management (cited in Campbell, 2012).

Competency models of performance, by comparison, focus on domain-general competencies, or "sets of behaviors that are instrumental in the delivery of desired results or outcomes" (Bartram et al., 2002, p. 7). For example, the "great eight" competencies (Bartram, 2005) include leading/deciding, supporting/cooperating, interacting/presenting, analyzing/interpreting, creating/innovating, organizing/executing, adapting/coping, enterprising/performing.

More recently, Campbell (2012) proposed a revised model that integrates across empirical efforts on this question. Based on this updated taxonomy, job performance can be evaluated along eight revised dimensions: technical performance, communication, initiative/persistence/effort, hierarchical leadership, hierarchical management, team leadership, and team management, as well as a negative contributor, counterproductive work behaviors.

Both Campbell (2012) and Bartram (2005) map their dimensions of job performance onto broad personality traits, including the Big Five. Campbell (2012) notes, "The eight factors are somewhat analogous to the Big Five dimensions of personality. In both domains, higher-order factors with less informational content and specific facets with more information are present above and below the designated level of specificity" (p. 57). This provides an acknowledgement that at least some aspects of job performance are attributable to traits that are relatively stable within individuals and variable across individuals. In some cases, observed correlations between personality traits and job performance are substantial, raising interesting questions about the degree to which PES criterion variables are, or should be, sensitive to changes in performance due to professional development, coaching, or other interventions (vs. aspects of enduring aspects of employees' personalities).

C.   MEASUREMENT ISSUES

Even with regular training, PES ratings can be plagued by the inherent difficulty of observing, assessing, and measuring aspects of employees' performance. A number of measurement issues have been raised in the literature, ranging from those stemming from the foibles of human judgment to questions about how to accommodate the dynamic nature of performance.

1.   **Human subjectivity**

Particularly for competencies like motivation, leadership, teamwork, and other job performance factors discussed in the prior section, assessments based on raters' judgments or inferences about unobservable characteristics are subject to the shortcomings of human subjectivity. Thus, organizations must make trade-offs between performance assessment based on directly observable, quantifiable metrics, which promote quantitative accuracy at the expense of qualitative richness, and assessment based on raters' inferences about targets' traits, values, and other characteristics, which offer qualitative richness at the expense of rater bias.

Even when attempting to minimize rater bias through the use of standardized, behaviorally-based questionnaires, Muckler & Seven (1992) point to several sources of human subjectivity that can bias inputs or outputs of the PES. Human subjectivity can affect the selection of performance metrics, as well as the collection of data, analysis, and interpretation of performance data. Thus, organizations must be on the lookout for subjective biases that influence the design and end-use of their PES, independent of biases introduced by individual raters.

Minimizing the biasing effects of human subjectivity on PES accuracy can take many forms. Evidence-based standardized practices for selecting and interpreting measures lead to more informed and even-handed design and application of PES (Muckler & Seven, 1992). Aggregated performance ratings across independent raters on the same target is another practice that can minimize (or cancel out) biases idiosyncratic to particular raters and provide other benefits to performance assessment (Boice & Kleiner, 1997). Even with best practices in place, it may be impossible and possibly undesirable to remove biases due to human subjectivity from PES. In general, designers of PES must be deliberate about which aspects of performance should be assessed via objective and subjective information, and, where possible, seek to corroborate evaluations based on one type of information with evaluations based on the other type.

2.   **Dynamics and distribution of performance**

Another measurement challenge is that performance within individuals is not stable. Average performance across time may be relatively reliable from one measurement period to another, however, performance from day to day or week to week is variable. Thus, a critical issue for PES designers is how to account for fluctuations in performance over time and circumstance. One specific instantiation of this problem is the question of whether to measure and assess typical, day-in, day-out performance or maximum performance, such as performance

under pressure or stress (Campbell & Wiernik, 2015). Although most PES focus on the former, assessment of the latter may be appropriate for certain jobs requiring strong performance under strain. In general, PES systems that rely upon multiple observation periods or summary evaluations may be best suited for handling dynamics in performance.

PES systems must also be designed around the way performance is distributed within the organization. Both in terms of overall performance and performance on specific tasks, PES outputs should mimic the distribution of actual performance, rather than forcing a distribution that does reflect observed performance across the organization.

3.   Cultural considerations

In global, diverse organizations, PES should account for cultural differences, both between and within countries. The values embedded in PES, including which characteristics are looked to as indicators of good performance, may be culturally-specific and overly-narrow, rewarding those who live up to cultural expectations and penalizing those who do not.

D.   PREVIOUS RESEARCH ON PES ACROSS NAVAL SERVICES

1.   Ellison (2014) Improving the Signal for US Navy Officer Productivity

LCDR Joshua Ellison's (2014) NPS master's thesis is a comparative analysis of the Navy and Marine Corps PES, grounding the comparison in the economics literature on performance appraisals and promotion systems. Ellison argues the Navy is structured as an internal labor market with tournament-style incentives, given its pyramid structure, set number of jobs at each level, and promotion to higher levels based on relative performance at the level below. Tournament labor markets are less concerned with absolute productivity than with promoting the most productive individual relative to others. A good PES must then be able to signal individual workers' productivity so that talent can be differentiated and promoted.

Ellison conducts a textual analysis of USN and USMC performance appraisal rating instruments and instruction documents, comparing them on rating accuracy, talent differentiation, and performance comparison methods.

Ellison's textual analysis finds that compared to the Navy PES, the Marine PES puts a lot more emphasis on rating accuracy. Words relating to accuracy show up 10 times more frequently in Marine PES documents than in the Navy's, with dedicated sections on the importance of

accuracy and exhortations by the Commandant. In the Marine PES, a Marine is assessed by an immediate supervisor, the rating senior (RS), as well as that senior's senior, the reviewing officer (RO). RS ratings are reviewed by the RO, providing an added check for accuracy.

Ellison also finds that the Marine FITREP may be better able to differentiate talent. The Navy FITREP rates 7 performance traits on a scale that ranges from 1.0 to 5.0, allowing for 29 different performance trait average values at 2 decimal places. In contrast, the Marine FITREP appraises 13 traits on a rating scale from 1.0 to 7.0, allowing for 79 different average values. Ellison concludes the Marine FITREP is relatively better able than the Navy FITREP to differentiate talent, given it assesses 6 more traits at a finer scale than the Navy's.

Finally, Ellison notes how the Navy uses performance trait grades, rankings, and promotion recommendations to allow the institution to compare its officers' performance. Marines also use RS performance trait grades, the Comparative Assessment or "Christmas Tree" section for relative performance comparisons by the RO, and the Relative Value. Ellison notes the Navy makes 7 absolute comparisons (on the 7 performance traits) and two implied relative comparisons. A Navy's officer's relative performance can be inferred by comparing the individual's average trait score to those of peers within their unit (Summary Group Average or SGA) and to their Reporting Senior's Cumulative Average (RSCA). The Marine FITREP has the RO's direct assessment of where that officer fits in a distribution of officers of that same grade the RO has observed. Marine Corps also calculates a FITREP Relative Value, where the average across 13 traits is compared to the RS's cumulative average on Marines of the same grade. This difference is then mapped onto a standard curve scaled from 80 to 100. Marine FITREPs calculate a Relative Value at processing and a cumulative Relative Value that changes over time.

Ellison concludes with three recommendations for the Navy PES given his literature review and comparative analysis: (1) Emphasize rating accuracy in the training of raters, educate them on organizational goals and standards so they can accurately measure performance and signal a Navy officer's productivity; (2) Navy evaluations should be weighted more towards relative comparisons to differentiate talent; and (3) Create a cumulative productivity metric.

2.   Small (2020) Successful Practices for Employee Performance Evaluations

LCDR Laura Small's (2020) NPS master's thesis follows Ellison's and employs case analysis and a narrative systematic review to pinpoint shortcomings in the current Navy PES and

identify performance evaluation best practices in the scholarly (industrial and organizational psychology, behavioral sciences) and industry literature that the Navy should consider. Small's thesis is a qualitative literature review of performance evaluations as it pertains to the Navy, describing the evolution of Navy performance evaluations from its implicit origins in 1799 to the eNavFit.

Small identifies five main issues in the current Navy PES: (1) system age, (2) process over performance, (3) past versus future focus, (4) lack of transparency, and (5) inaccurate measures. At the time of thesis writing, the Navy PES uses NAVFIT98A, a system that does not connect to the internet or collect evaluation data that could help with MyNavyHR goals of using predictive analytics. Small notes as the second issue the administrative burdens of Navy FITREPs, which have to begin months in advance of evaluation due dates given how many times comments must be iterated over. The Navy FITREP also focuses only on past performance, unable to assess potential success of the officer in future jobs. Small refers to "lack of transparency" in discussing Sailors' lack of trust in the PES, mainly stemming from how RSs and their subordinates know that ratings can be influenced more by RSCA management rather than that officer's performance.

The key themes Small identifies from the literature as best practices that the Navy ought to consider in its future PES design that could correct these issues are: (1) define the purpose of evaluation; (2) create a culture of communication; (3) foster a perception of fairness; and (4) there is no "one size fits all" solution.

In defining the evaluation purpose, Small cites Cleveland et al (1989) who describe the two main purposes of evaluations as development (feedback, job improvement, identification of training needs) and administration (promotion, firing, salary decisions). Small concludes that "(1) a single system designed to meet both purposes of development and administration does neither well and eventually leads to one purpose overtaking the other […], (2) the quality of feedback diminishes when raters must prioritize administration or development […], and (3) rating purpose impacts both rater observations and recorded results."

Creating a culture of communication is Small's second key identified theme, as the current Navy PES offers minimal opportunities for feedback and prioritizes the administrative process over delivering feedback. Minimal feedback could in turn feed perceptions of a lack of transparency and a lack of fairness in performance evaluations.

Small's last theme encourages future PES transformation efforts to iterate and adapt, as there is no "one size fits all" solution. Successful implementation depends on "(1) focus of the integration between the performance evaluation and how it fits into the overall picture of performance management […] and (2) rigorous testing and application of user feedback through a thoughtful iterative process."

3.   Larger (2017) Effectiveness of the Marine Corps' junior enlisted performance evaluation system: an evaluation of proficiency and conduct marks

Capt Richard Larger Jr (2017)'s NPS master's thesis analyzes the effectiveness of the US Marine Corps' proficiency and conduct marks (pro/con) as measures of job performance of junior enlisted Marines. While this thesis focuses on performance appraisal of junior Marines, the methods used could inform the validation of the draft performance traits for Navy officers.

Pro/con marks are used by USMC for personnel management functions of its junior Marines, such as promotion, retention, selection to special duty assignments, and characterization of service upon discharge. Employing factor analysis and multivariate regressions on data of 360,690 active duty Marines at paygrades of E3 or E4 between 2006 and 2016, Larger's thesis seeks to estimate the reliability, validity, accuracy, and practicality of pro/con marks.

On reliability, Larger examines whether pro/cons are consistent measures of performance over time as well as inter-rater reliability. For validity, Larger turns to two other job performance assessment tools: the Composite Score and FITREP. Lance Corporals and Corporals are promoted based on the Composite Score, which is an overall "quality" score that combines performance-related elements (including pro/cons), seniority, special duty assignments, education, etc. To assess validity of pro/cons, Larger examines which Composite Score variables provide the most information on a Marine's performance and whether pro/cons predict future job performance as indicated by FITREPs when the E3/E4's eventually get promoted. To examine accuracy, Larger examines pro/con distinctiveness, ability to differentiate performance, and rating inflation. Table 2 summarizes his findings.

Table 2.   Table summarizing findings from Larger (2017). This table reproduced from Larger (2017), p. 3

Research Question	Null Hypotheses—Reliability, Validity, and Accuracy	Literature Suggests…	Supporting Evidence?
R1	Pro/con marks are stable	Yes	Yes
R2	Pro/con marks are consistent between raters	No	Inconclusive
V1	Pro/con marks are important contributions to a Marine's composite score	Yes	Yes
V2	Pro/con marks predict future performance	Yes	Yes
A1	Pro/con marks differentiate between levels of performance	No	Yes
A2	Pro/con marks are not inflated	No	No
A3	Pro/con marks are distinct measures of performance	No	No

Data limitations prevent a finer analysis of pro/cons' stability over time, though descriptive statistics (trends in the variance) suggest pro/cons are stable. Regressions indicate small but statistically significant inconsistencies in marks on the same Marine from different raters; Larger attributes this to differences in grading philosophies across communities (e.g. ground vs aviation). Larger suggests further research is needed to see whether promotions are being affected by the low inter-rater reliability.

Larger then uses factor analysis to examine the construct of the junior Marines' Composite Score and see which variables explain a Marine's performance as a Lance Corporal or Corporal. The factor analysis reveals that pro/cons provide the most information on a Marine's performance in the sample of promotion-eligible Corporals. The latent performance variables are then included in the model assessing predictive validity, to examine whether these variables have significant effects on the Marine's FITREP scores up to their first three years as a Sergeant. Larger finds the factor containing pro/cons is the strongest predictor of future performance as measured by FITREPs.

Examining the distribution of scores and estimating univariate regressions, Larger finds pro/con marks, intended to measure proficiency and conduct separately, measure much of the same performance and also tend to be inflated.

Overall, on the question of whether or not pro/cons are effective indicators of performance, Larger finds pro/cons do well at identifying Marines with the most potential to perform as a Sergeant. However, pro/cons' low inter-rater reliability, lack of distinction between proficiency vs conduct, and grading inflation suggest further tweaking is needed.

Indeed, beginning in 2021, USMC replaced the pro/con marks and Composite Score with the Junior Enlisted Performance Evaluation System (JEPES). JEPES was initially created in 2018.

4.   Clemens et al (2012) An Evaluation of the Fitness Report System for Marine Officers

In 2012, the Director of Manpower Management Division at HQMC commissioned a review of the USMC FITREP system from the Center for Naval Analyses (CNA). While this study focuses on the Marine FITREP, there are some common themes and lessons to be learned relevant to the Navy FITREP.

Clemens et al (2012) examine RS and RO rating inflation, whether ratings change as the RS or RO gain experience, whether the ratings carry important information, whether ratings differ by observable characteristics and if there is evidence of bias, to what extent subjective comments correlate with FITREP scores, how Marine officers are taught about the FITREP system, how promotion boards view FITREP results, and how the process of completing and submitting FITREPs might be improved.

To answer these questions, Clemens et al use data from all officer FITREPS from January 1999 to August 2011 matched to personnel records, review the FITREP training curriculum at The Basic School (TBS), and interview various instructors and students at Expeditionary Warfare School and Command and Staff College, general officers who have served on promotion and command boards, and other Marine officers from Manpower Management Division. The insights from the interviews are then compared with or used to interpret statistical analyses of the data.

Overall, Clemens et al finds that the FITREP system works well. There is no evidence of rampant inflation—a prior main concern, and having both RS and RO assessments appears to make ratings more informative. Marine FITREP ratings have face and predictive validity, and are consistent with other officer quality indicators. The interviews reveal that the system usually results in promoting the best and most qualified officers.

Clemens et al do recommend more training for both RSs and ROs. Marines are trained on how to write FITREPs at TBS, before any experience with the system. Marine officers are provided "a handout, three lectures, a case study homework assignment, and a discussion group"

for their FITREP training. The authors recommend expanding education and training on FITREPs to solidify in officers the long-term implications of the ratings they provide.

Clemens et al find both RSs and ROs do not always mark assessments consistent with the intent of the PES manual. For instance, RSs have learned to award "room to grow" marks the first time they evaluate an officer and then give higher scores as they continue to evaluate that officer. Officers learn this at TBS or from their mentors, rather than being instructed to consider each reporting period separately from all others. RO marks are also intended to have a forced distribution and be a relative assessment within a paygrade. Clemens et al finds that the actual distribution is dramatically different, with officers in higher grades receiving higher RO marks on average. The authors also find that presentation of RO marks to promotion boards could be improved.

Clemens et al further highlight potential issues of concern for monitoring. Marine FITREP average scores are trending to becoming less varied (and potentially less informative) over time. Regression analysis also shows that black and Hispanic officers receive lower ratings on average than white officers, even controlling for academic credentials, accession source, and GCT scores. A lot of the gap is driven by differences in TBS standing.

Finally, Clemens et al consider the interaction of demographic characteristics of the RS and Marine Reported On (MRO). White RSs award slightly lower ratings to Black MROs and vice versa, even controlling for TBS standing, commissioning source, and family structure. Male and female officers receive slightly higher scores from RSs of the opposite gender. The regressions also indicate there may be positive biases towards some occupational fields and negative biases toward others. In particular, RS officers in other communities may be biased toward infantry, logistics and military police officers and against aviators. These correlations are all plausible indicators of gender, racial/ethnic, and community biases or other omitted variables.

III. METHODOLOGY

Below, we provide an abbreviated account of our methods for prototype development and prototype testing. Additional information on prototype development is available in Luke's (2021) and Gervato's (2023) MBA capstone projects, which spanned the full project period of performance.

A.   PROTOTYPE DEVELOPMENT

Our efforts to develop a new performance evaluation prototype began with materials the research team inherited from early Navy PES transformation efforts. In 2019, the Navy assembled personnel and performance evaluation Subject Matter Experts (SMEs) to brainstorm new trait and value statements (TVS) that could be used for assessing Sailors under a transformed PES. The resulting list contained 8 broad traits with 39 associated sub-traits, with 2–5 value statements per sub-trait. This nested structure for the 8 broad traits, 39 sub-traits, and 82 value statements is displayed in Table 2.

Our research team was first tasked with assessing this set of TVS on content, construct, and face validity by comparing the traits, values, and statements against current Navy doctrine and PES materials from other military services. The full methodology and results of these validation efforts was the basis of Luke's (2021) thesis, which we summarized below.

Although our preliminary analysis established the content, construct, and face validity of most TVS items, our team judged the full set of TVS items as limited in practicality and usability. The list of 82 statements was too large to use in a testable prototype. From the standpoint of potential respondents, the request to provide ratings of active duty Sailors on 82 different TVS was too arduous and unlikely to yield quality data for assessing the statistical validity of the proposed items. The team also had concerns about the overarching nested structure, including the mapping of statements to sub-traits and sub-traits to broad traits.

To address both issues, the research team completed a modified sorting procedure aimed at reducing the number of sub-traits and value statements associated with each broad trait. For each of the eight broad traits, a member of the research team sorted the corresponding value statements into clusters based on conceptual similarity. The sorting process fully discarded the original TVS sub-traits so that the sort produced clusters arrived at through a bottom-up grouping

process, rather than the top-down structure imposed by the originators of the TVS set. Once these clusters were established, the statements belonging to each cluster were combined into one robust statement per cluster. Finally, these single cluster statements were combined into a holistic trait definition for each of the eight traits. In all, this sorting procedure yielded one trait definition for each of the eight traits, with cluster statements that could be used to provide further detail and/or behaviorally-based anchors for the eventual rating scale used in the prototype.

The research team also conducted a benchmarking exercise, identifying and compiling existing trait definitions and trait descriptions from the current performance evaluation tools of the US Navy, US Marine Corps, and US Coast Guard. Using the trait definitions and cluster statements as our operational definitions of each of the eight broad traits, the team identified conceptually-similar traits and compiled all relevant content from other services' tools into master categories. We added to these categories any trait-relevant content from recent doctrine relevant to the Navy's current PES transformation (e.g., *Get Real, Get Better*, *21st Century Sailor Initiative*).

Together, these two processes yielded a validated, extensive lexicon from which the research team, led by the project topic sponsor at Navy Personnel Command, drew to compose the final trait definitions and behavioral anchors that were used in the performance evaluation prototype instrument.

Table 3.   List of TVS and nested structure developed at NPC PERS-3 PET-TM Summit. Adapted from CNP (2020a)

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
Character Conduct in accordance with the Navy Ethos and Navy Core Values. Includes the combination of traits and moral and ethical qualities that are revealed through an individual's consistent behaviors, on and off duty.	Responsibility	Takes responsibility for actions regardless of consequences
		Acknowledges and corrects mistakes
	Ethics	Adheres to the Navy standards of ethical conduct at all times
		Demonstrates high standard of personal and professional behavior
		Does not misrepresent self or use position or authority for personal gain
		Holds self-accountable to Navy core values
	Integrity	Is honest and forthcoming
		Displays actions that are in-line with stated intent
	Respect	Demonstrates respect for others' values and customs

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
		Treats others with dignity and respect
	Moral Courage	Morally steadfast in the face of opposing pressure
		Does the right thing, even when it is difficult
	Professionalism	Uses discretion when handling the sensitive personal information of others
		Avoids situations and actions considered inappropriate
Leadership The ability to influence and inspire others by providing a shared sense of purpose, direction, and vision. Includes the knowledge and appropriate use of motivational resources for guiding others toward achievement of a goal or objective.	Goals and Expectations	Provides direction in crisis situations
		Ensures all members understand their role and responsibilities
	Personnel Development	Addresses performance issues promptly and corrects poor performance
		Holds others accountable to job performance standards
	Feedback	Provides consistent performance feedback to others
		Creates a culture that encourages feedback
	Inclusion	Creates a positive work environment where all staff are motivated to do their best
		Fosters a culture of respect within the organization
	Wellness	Recognizes and addresses signs of stress in others
		Guides others to seek support through available wellness resources
	Delegation	Delegates tasks and responsibilities appropriately
		Allows others to make decisions or take charge
	Motivation	Motivates others toward achieving desired results
		Provides recognition for superior performance
	Change Management	Clarifies priorities when leading change
		Persuades others to approach issues in an open, constructive, professional manner
Initiative and Drive Takes independent and proactive action to contribute to the accomplishment of objectives and goals. Includes the identification, ownership, and follow-through of activities with little to no direction.	Innovation	Initiates improvements through new methods or practices
		Identifies and recommends innovative ways to address inefficiencies
	Personal Development	Seeks learning opportunities to enhance job performance
		Acquires new competencies, methods, and information to improve efficiency and effectiveness
	Independence	Takes appropriate action in the absence of specific direction
		Proactively addresses problems
	Volunteering	Seeks opportunities to contribute
		Willingly puts in extra time and effort
Teamwork Develops, supports, participates in, and maintains positive work relationships to facilitate the	Team Pride	Demonstrates inclusion and actively supports teamwork and team spirit
		Supports unit cohesion
	Relationships	Develops productive working relationships

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
accomplishment of shared goals. Includes collaboration with others, inside and outside of the organization.		Supports group decisions even when not in total agreement
	Contribution	Collaborates with others in identifying solutions
		Provides assistance to teammates when they need it
Communication The exchange of information and ideas. Includes all messages that an individual sends and receives, through verbal, written, and non-verbal channels.	Listening	Listens attentively to people's ideas and concerns
		Allows others to speak without unnecessarily interrupting them
	Comprehension	Actively listens to ensure comprehension
		Asks for clarification when unsure of what is being said or asked
	Clarity	Communicates clear, well-defined expectations for others' work
		Presents information clearly, concisely, and logically
	Non-verbal	Demonstrates appropriate use of nonverbal communication
		Reads body language, and adjusts tone and style accordingly
	Feedback	Provides open and honest feedback
		Responds positively to feedback
	Conflict Management	Addresses sensitive issues in ways that allow rational and open discussion
		Addresses issues in an open, constructive, professional manner
	Information Sharing	Consults with supervisor, when necessary, to determine priorities
		Keeps leadership informed about progress and problems
Critical Thinking and Decision Making Seeks, identifies, and analyzes information from appropriate sources to understand issues, problems, and opportunities. Uses this information to make timely and informed choices to ensure the optimal course of action is taken.	Risk Assessment	Assesses risk throughout implementation of a course of action
		Considers risk to mission before taking action
	Planning	Consults multiple resources before making a decisive plan
		Assesses potential barriers to new approaches
	Evaluates Alternatives	Switches to a different strategy when an initial one is unsuccessful
		Elevates problems or risks to higher levels of decision-making when necessary
	Problem Solving	Makes sound decisions with best available information
		Makes timely decisions with best available information
Resiliency and Toughness The ability to maintain performance and self-control under pressure. Includes the ability to recover from or adjust to adversity or change.	Coping	Engages in positive coping strategies
		Willing to seek help when dealing with stress
	Persistence	Maintains composure in stressful environments
		Maintains focus under adverse conditions
		Sustains workload during high operational tempo
	Recovery	Responds to setbacks with renewed and increased efforts

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
		Recovers from setbacks or failures to accomplish mission
	Adaptable	Remains flexible in the face of changing needs and demands
		Adjusts to changing requirements
Mission Accomplishment and Productivity Performance in assigned duties, roles, functions, and completion of tasks and assignments in accordance with established standards. Includes the rate of production and the quality of the output and the development, application, and sustainment of job-relevant knowledge, skills, and abilities.	Professional Competence	Demonstrates professional knowledge and technical ability in primary role
		Maintains working knowledge of governing documents affecting assigned areas
	Time Management	Adheres to scheduled timelines for task completion
		Effectively uses time management to complete assigned tasks
	Quality and Attention to Detail	Produces quality work
		Adheres to safety procedures
		Adheres to security procedures

1.   Prototype Instrument

The prototype instrument was programmed through the Qualtrics online survey platform to allow for widespread web-based distribution and private responding for respondents who were stationed across the globe. The full instrument can be viewed in Appendix A of this report. Web links to the instrument are available through direct request to the principal investigator.

The prototype was structured into four blocks. In the initial block, respondents were briefed on the purpose of the research and provided opt-in consent to participate in the prototype testing study. Respondents were told that their ratings of Sailor-reported-on would only be seen by the NPS research team and would not impact the official records or promotion decisions regarding the Sailor-reported-on. Respondents then provided some basic information about the Sailor they would be evaluating, including the Sailor's current rank/rate, the respondent's relationship to the Sailor, and the respondent's confidence level in being able to accurately evaluate the Sailor.

Block 2 of the prototype assessed Sailors' *past performance*. Ratings of past performance were made on the eight broad traits developed and defined through our validation and content analysis procedures. Ratings of the Sailor were made using Behaviorally-Anchored Rating Scales (BARS), which provided behaviorally-based descriptions of performance typified by

individuals who are low, average, and high on each trait of interest (Landy & Farr, 1980). Respondents were implored to be honest and reminded that their "individual responses will NOT be shared with Navy Personnel Command… and will be kept completely separate from any Sailor's military records and separate from any other information used to make career decisions about any specific Sailor."

In Block 3, Sailors were rated on five traits pertaining to their *future potential*. The five future potential traits and their corresponding definitions are shown in Appendix A. Ratings of future potential were elicited using a "readiness" standard. For each future potential trait, Sailors were rated according to their readiness to succeed "at the next paygrade or in key, particularly jobs." Put concretely, the evaluator of a Sailor at the O-3 rank would be asked how ready that Sailor was for success at the O-4 rank on each of the five future potential traits. Responses ranged: *Not ready and is unlikely to be ready in the near- to mid-future*, *Not ready yet, but could be ready in the next 3-5 years with consistent effort*, *Partially ready now and could be fully ready with developmental progress in the next 1-2 years*, *Ready NOW*, *Ready NOW and has been for some time*, and an option for *Not observed*. If respondents indicated one of the *Not ready yet* options, they were further prompted to "briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job."

In Block 4, we asked respondents to offer a comparative ranking of the Sailor against "others of the same paygrade whom you have known in your career." The intent for ratings of past performance (Block 2) and future potential (Block 3) was for Sailors to be rated against an absolute standard, a behaviorally-based anchor for ratings of past performance and a readiness threshold for ratings of future performance. In Block 4, a comparative assessment was made using a modified "Christmas Tree" schematic, borrowing heavily from the assessment tool utilized by the US Marine Corps. The measure imposed a performance distribution by telling and illustrating to raters that only 15% of all Sailors were expected to be rated as "one of the very few best and most qualified" (5%) or "one of the exceptionally qualified" (10%). The distribution indicated that an addition 20% of Sailors were expected to fall into the next ordinal category, "one of the highly qualified," with most Sailors (55%) "one of the many qualified individuals who form the majority of this grade." The distribution stipulated that only 10% of Sailors were expected to fall into one of the bottom two categories, "a progressing Sailor" or "an unqualified Sailor." Respondents placed the Sailor they were rating into one of those

performance categories based upon all relevant interactions with the Sailor. If respondents wished to do so, they were offered a text box to "amplify their assessment." Following that, respondents were asked to formulate an all-things-considered promotion recommendation for the Sailor, ranging, in order: *Promote now*, *Promote with top 20% of peers*, *Promote with peers*, *Promotion potential*, *Retain at current paygrade*, and *Do not retain.* Respondents were also give the options to indicate the Sailor had *already selected to next paygrade* or was *recently promoted (<12 months in rank)*.

In the fifth block of the survey, respondents rated the frequency with which the Sailor-reported-on engaged in six different workplace behaviors. Three behaviors were adapted from a standard research measure of *counterproductive workplace behaviors* (*CWB*); three behaviors were adapted from a standard research measure of *organizational citizenship behaviors* (*OCB*). These categories represent, respectively, day-in-day-out workplace behaviors that are detrimental or beneficial to the broader organization and those who comprise it. These measures served as behavioral proxies for daily job performance. Respondents indicated the frequency with which the Sailor engaged in each behavior (ex. *Musters late without reason or prior notification*; *Takes time to advise, coach, or mentor fellow Sailors*) using a scale: *Never*, *Once or twice*, *Once or twice per month*, *Once or twice per week*, and *Every day*.

In the sixth block of the survey, respondents were thanked for their involvement in the prototype testing and offered the opportunity to make any final comments about the research or the topic under investigation. They were also offered the opportunity to provide contact information for future studies and/or information concerning the Navy's PET transformation efforts.

B.   PROTOTYPE TESTING

1.   Respondent nomination process

Because our research required us to obtain performance measures of Sailors-reported-on, it was important that we utilized an opt-in procedure so that Sailors could provide consent for their performance records to be utilized for the purpose of prototype evaluation. Obtaining opt-in consent from Sailors-reported-on also afforded us the opportunity to solicit names and contact information of potential respondents to our prototype, nominated by the Sailors they would evaluate with the prototype.

We used two procedures to obtain a sufficiently large sample of Sailors-reported-on. First, we coordinated with another research team at NPS that was fielding a survey to support a related N1 project. At the end of this team's survey, participants were informed about the performance evaluation prototype study and invited to support the research effort. Second, in collaboration with N1's Public Affairs Office (PAO), we advertised the broader performance evaluation transformation research project on Navy social media platforms and invited Sailors at all rates and ranks to take part.

Regardless of the procedure used, those Sailors who opted in to the study were linked to a page that prompted them to provide the names and contact information for up to three individuals whom Sailors felt could accurately assess their job performance. Our specific instruction was to "think about up to three Sailors (at least one of whom is senior to you) in your current command who are well-positioned to assess your past performance and your potential for the future."

This process yielded 1823 nominees across all Sailors-reported-on.

2. Nominee recruitment

Each nominee was contacted with a personalized email inviting them to participate in the prototype evaluation study. The recruitment email provided the name of the Sailor-reported-on and informed nominees that they had been identified as potential raters for the Sailor-reported-on. Of the 1823 nominees contacted, 678 advanced to the first page of the prototype. Of those, 606 respondents provided usable data on their Sailor-reported-on. This reflects a response rate between 33% and 37%.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

We collected 606 valid responses rating 416 Sailors. Figure 1 shows the prototype FITREP rated Sailors predominantly at the rank of E6 (n=145, 24%) or E7 (n=115, 19%), but the survey also included Sailors-reported-on at all ranks up to O6. The majority of the raters (57%) were supervisors of the Sailor reported on; 16% were peers and 23% were junior to the Sailor-reported-on. Rater confidence was high: 84% of raters felt very or extremely confident about their ability to rate the Sailor. As a final note, of the 416 Sailors being reported on, 149 had multiple raters, which will allow us to examine inter-rater reliability in a future (FY23) study.

Figure 1.   Characteristics of survey respondents
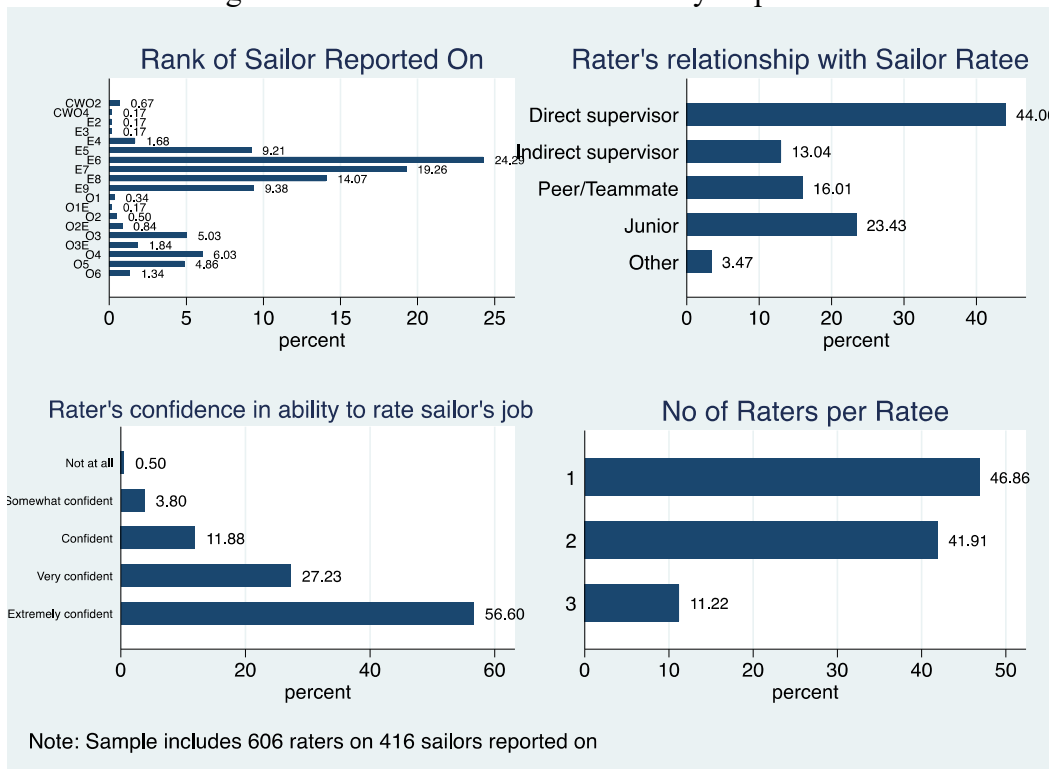
A.   FINDING #1: RESPONDENTS OF THE SURVEY DISPROPORTIONATELY DREW RATINGS ON HIGH PERFORMERS

Table 4 reports the means and standard errors of ratings of past performance and future potential, for the overall sample and separately for officer/enlisted. Ratings on all traits were made on a 5-point scale. The past performance traits yielded an average of 4.33 in this sample. All the past performance and potential traits had means above 4.0, suggesting high performance on average. The means and standard deviations in columns (2) vs (3) also indicate there was no statistically significant difference in ratings for officer vs. enlisted in this sample.

Table 4.   TVS descriptive statistics by enlisted/officer Sailor-reported on.

	(1) All		(2) Enlisted		(3) Officers, incl CWO	
	mean	sd	mean	sd	mean	sd
Past performance						
Character	4.55	0.64	4.54	0.63	4.57	0.67
Leadership	4.23	0.76	4.25	0.75	4.17	0.80
Initiative	4.39	0.74	4.40	0.74	4.37	0.76
Teamwork	4.34	0.76	4.38	0.74	4.22	0.82
Communication	4.17	0.81	4.19	0.81	4.09	0.80
Crit Think	4.25	0.75	4.28	0.73	4.18	0.83
Productivity	4.40	0.71	4.41	0.70	4.36	0.75
Resilience	4.28	0.78	4.29	0.79	4.25	0.76
Trait Average	4.33	0.56	4.34	0.55	4.27	0.59
Future potential						
Learning	4.42	0.71	4.44	0.71	4.34	0.72
Leadership	4.33	0.72	4.34	0.71	4.28	0.75
Character	4.48	0.63	4.45	0.64	4.58	0.58
Judgment	4.31	0.71	4.32	0.73	4.31	0.64
Experience	4.43	0.73	4.44	0.73	4.37	0.74
Motivation	4.53	0.64	4.53	0.64	4.51	0.64
N	606		467		139	

Evidence that our sample is comprised of high performers across multiple metrics is illustrated in Appendix B Figures. Consistent with the summary statistics in Table 4 above, Appendix B Figures A-1 and A-2 show that majority of Sailors received ratings of 4 or 5 on performance and potential traits, respectively. The potential ratings (Appendix B Figure A-2) indicate that across all 6 traits, most Sailors were rated as *Ready NOW* or *Ready NOW and has been for some time*. The frequency of counterproductive workplace behaviors was quite low and

the frequency of organizational citizenship behaviors was rather high (Appendix B Figure A-3). Finally, very few Sailors-reported-on were not recommended for promotion (Appendix B Figure A-4).

Figure 2 further confirms the finding that our sample is comprised of top performers. Depicted in the figure is the distribution of survey responses on the comparative assessment ("Christmas tree") question, with 33.77% rating Sailors in the upper 5% (*few best and most qualified*) and 36.24% in the upper 10% (*exceptionally qualified*). In other words, 70% of Sailors-reported-on were placed into the upper 15% tier of performance across all Sailors in the Navy by their raters.

Figure 2.   Distribution of rankings on the Comparative Assessment ("Christmas Tree" Question
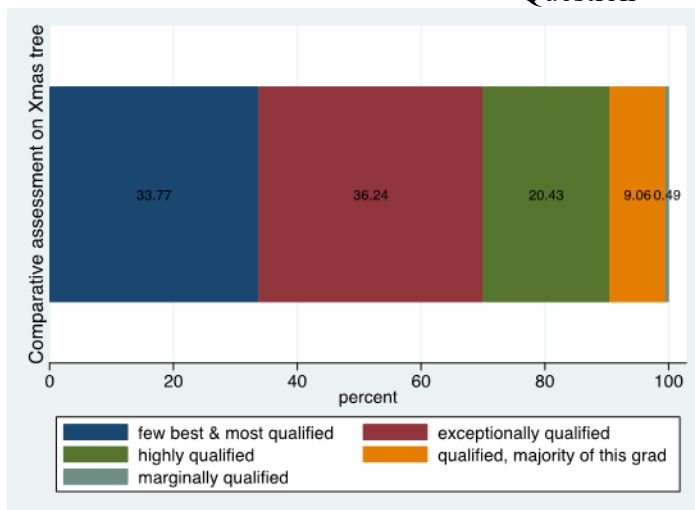
B.   FINDING #2: RATINGS OF PAST PERFORMANCE AND FUTURE POTENTIAL CONVEY RELATED BUT DIFFERENT INFORMATION

The correlation matrix reported in Appendix B Table A-1 shows several patterns regarding the relationship between ratings of past performance and ratings of future potential:

- Ratings of traits related to past performance are positively correlated with each other, ranging from 0.38 to 0.59. These are considered moderate correlations in the behavioral sciences, indicating that Sailors who were rated relatively high (vs. low) on one past performance trait tended to also be rated high on the other past performance traits. However, while trait ratings were correlated, they tended to convey different information.

Of the 8 broad performance traits, *leadership* was the trait most consistently highly correlated with all other past performance traits.

- Ratings of traits related to future potential were also positively correlated with each other and with current performance ratings, ranging from 0.25 to 0.60.

- Ratings of traits related to past performance and future potential were all correlated with the OCBs and CWBs in the expected directions. That is, those rated highly on performance and potential tended to engage in organizational citizenship behaviors more frequently and counterproductive workplace behaviors less frequently than their peers. This indicates that the trait and value statements showed sufficient convergent validity.

To investigate how useful the trait ratings may be in generating an index or average metric of performance, we explored the psychometric properties of the trait ratings using factor analysis. We report these estimates in Appendix B Table A-2.

Panel A of Appendix B Table A-2 shows that the variance across all 8 past performance traits and the six future potential traits can be reduced to 2 dimensions or factors. The cumulative column shows that 100% of the variation in the data can be accounted for by the first 2 factors. The scree plot below further shows that 2 to 3 factors may be retained, depending on the eigenvalue criteria used. The Kaiser criterion suggests retaining those factors with eigenvalues equal to or greater than 1, while alternative criteria set the cut-off at 0.

Figure 3.　　　Scree plot of factor analysis

Scree plot of eigenvalues after factor
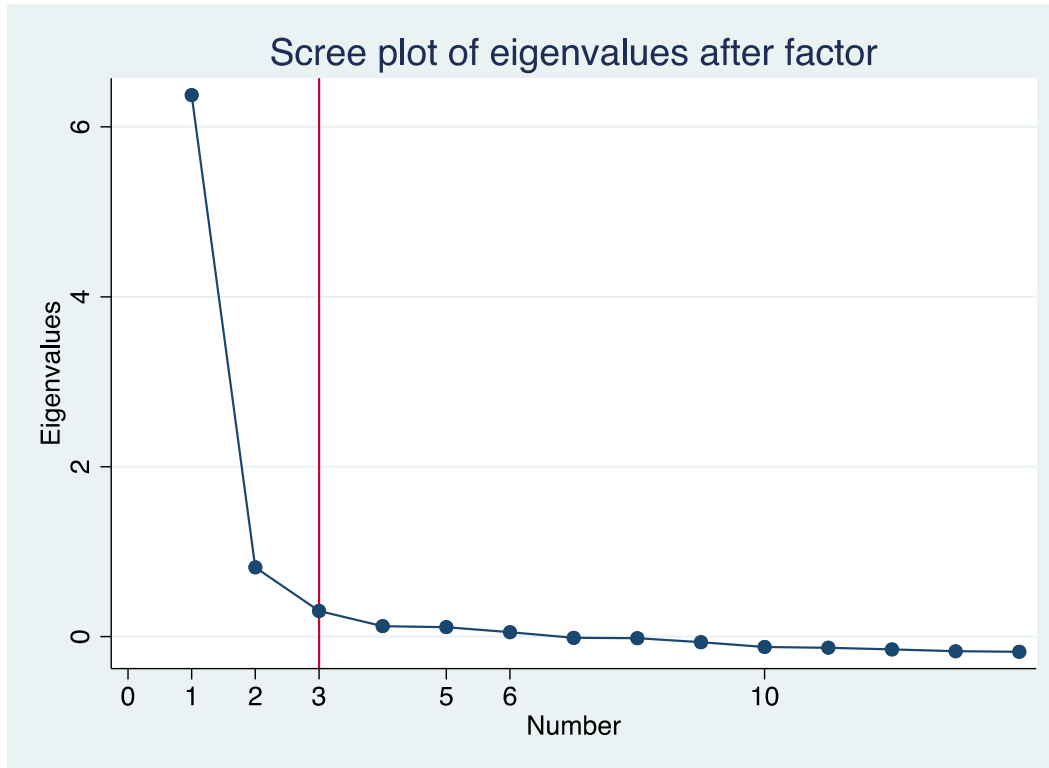
Meanwhile, factor loadings are the weights (or correlation) of each trait with that factor; the higher the load the more relevant in determining that factor. The factor loadings for the first 2 factors shown in Panel A of Table A-2 indicate that both performance and potential traits load high on the first factor, likely because they are positively correlated as shown in Table A-1. However, in examining the factor loadings for factor 2, all the performance traits (character down to resilience) all load negative while potential traits all load positive. This indicates that while performance and potential traits are correlated, they convey different sets of information.

Indeed, in the final rotated factor analysis shown in Panel B of Table A-2, all the performance trait ratings load on the first factor and all the potential trait ratings load on the second factor. While some traits load on the other (e.g. 2 potential traits, learning and leadership, also load on factor 1), their weights or loadings are lower than on their "primary" factor.

In other words, if the goal is to create a summary metric of job performance ratings, statistically speaking there are two separate factors/dimensions or trait averages in the data. The first index (factor 1) is based on past performance ratings; the second index (factor 2) is based on ratings of future potential. For factor 1, leadership, teamwork, and communication enter the

index with the highest weights, while the potential trait of "experience and competence" has the highest weight in factor 2.

C.   FINDING #3: RATINGS OF PAST PERFORMANCE AND FUTURE POTENTIAL PREDICT COMPARATIVE ASSESSMENTS; SOME TRAITS ARE BETTER AT DISTINGUISHING THE TOP FROM MIDDLE PERFORMERS; FEW TRAITS DISTINGUISH BOTTOM PERFORMERS

We assessed whether the trait ratings had predictive validity using survey responses on the comparative assessment as the criterion. To assess whether the multiple past performance and future potential traits (X) predict comparative performance on a forced distribution, we estimate the following multinomial logistic regression model:

$$prob(y_i = j) = \frac{e^{b_0 + b_1 X_{1i} + ... + b_k X_{ki}}}{1 + e^{b_0 + b_1 X_{1i} + ... + b_k X_{ki}}}.$$

where

$$j = \begin{cases} 1 & \text{if sailor } i \text{ one of very few best \& most qualified,} \\ 2 & \text{if } i \text{ exceptionally qualified,} \\ 3 & \text{if } i \text{ highly qualified,} \\ 4 & \text{if } i \text{ qualified similar to majority of this grade,} \\ 5 & \text{if } i \text{ marginally qualified.} \end{cases}$$

Table 5 reports estimates of the exponentiated coefficients or odds ratios relative to the base category of "Highly qualified." The asterisks next to the estimates indicate whether or not it is statistically significant at various p-values, as indicated in the notes at the bottom of the Table. Since these estimates are relative odds ratios, numbers that are greater than one indicates that trait makes that assessment category more likely while numbers less than one make that assessment less likely, relative to being rated "Highly qualified."

For example, turning to the trait Leadership as a predictor, the estimates show a one-unit increase on the leadership behavioral scale rating increases the odds of being rated "Very few best" 3.09 times more than just being rated "Highly qualified," holding everything else constant. The same one-unit increase in leadership rating increases the odds of being rated "Exceptionally qualified" 1.52 times more than just "Highly qualified." On the other hand, ratings on leadership

performance does not statistically significantly distinguish the bottom (Marginally qualified) nor the "Qualified" from the "Highly qualified."

Figure 4 further illustrates how to interpret and use these multinomial logit estimates. The y-axis shows the predicted probabilities of being assessed in that category as a function of the x-axis leadership trait rating. The green and orange lines at the bottom of all these lines indicate that being ranked "Marginally qualified" or even just "Qualified like majority of the grade" is not very likely (lowest probability). The blue line shows that as leadership rating increases, the predicted probability of being rated the "Very few best" or in the upper 5% of the Navy increases. In contrast, the probability of being assessed just "Highly qualified" (red line) decreases as the leadership trait rating increases; the blue and red lines cross at a rating of 3.5 rating. At the highest leadership rating(=5), the predicted probability is highest to be rated as either "Very few best(blue)" or "Exceptionally qualified(purple)" as expected.

Table 5. Multinomial Logistic Regression Relative Risk Ratios.

	Very few best & most qualified	Exceptionally qualified	Highly qualified	Qualified, majority of this grade	Marginally Qualified
Performance Traits					
Character	1.840+	1.387	1	0.443*	0.196
Leadership	3.090***	1.521+	1	1.318	1.115
Initiative	2.053*	2.093**	1	1.586	0.335
Teamwork	1.006	0.941	1	1.005	5.617
Communication	1.740*	1.267	1	0.824	1.257
Crit Think	1.777+	0.712	1	0.813	1.426
Productivity	1.077	1.070	1	0.535+	0.0927
Resilience	0.734	0.969	1	0.481*	0.649
Potential Traits					
Learning	2.467**	2.071**	1	1.240	0.842
Leadership	1.447	1.293	1	0.900	1.247
Character	0.889	0.831	1	0.925	0.488
Judgment	1.971*	1.196	1	1.434	11.94
Experience	2.622**	2.075**	1	0.388**	0.278
Motivation	2.687**	1.292	1	0.577	2.132
Enlisted	0.656	0.533+	1	0.816	0.817
N=559	192	207	110	47	3

Note: Table reports exponentiated coefficients or relative risk ratios from a multinomial logistic regression of comparative assessment on the Christmas Tree forced distribution on TVS ratings. + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Figure 4.   Multinomial logit predicted probabilities on comparative assessment with Leadership
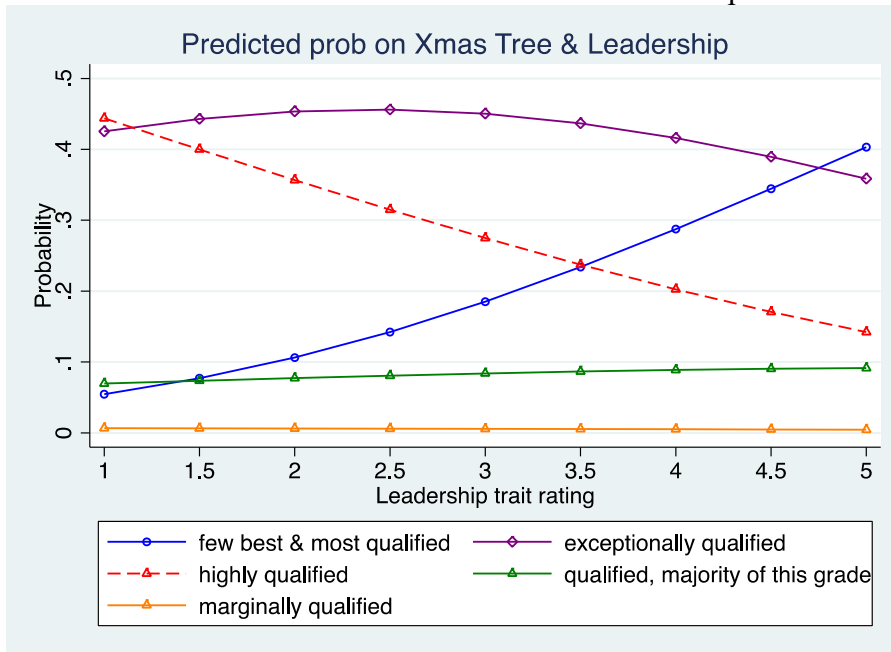
All of the past performance traits had some predictive power, while two of the future potential traits (*leadership* and *character*) did not have statistical significance in predicting any of the categorical comparative assessments. On the whole, these estimates suggest that some traits are better at distinguishing the top from the middle of the performance distribution. For example, ratings of leadership and initiative based on past performance are better able to distinguish the top from the middle, but ratings of character and resilience based on past performance are better at distinguishing the bottom from the middle. On the other hand, the future potential trait of "experience and competence" was able to distinguish both top and bottom performance.

V. CONCLUSIONS & RECOMMENDATIONS

In this research, we developed and refined trait and value statements for potential inclusion in the Navy's updated performance evaluation tool. We incorporated two sets of trait items – one based on Sailors' demonstrated past job performance and the other based on Sailors' future potential – into a prototype performance evaluation instrument. We then fielded this prototype to respondents who rated current Sailors on their job performance using these updated measures. Our objective was to provide Navy Personnel Command with a final set of TVS that are psychometrically valid, useful, simple, and fair to support ongoing efforts at PES transformation.

Based on the analyses reported herein, we offer four recommendations pertaining to the TVS developed, refined, and tested by our research team.

A. RECOMMENDATION 1: AMPLIFY THE COMPARATIVE ASSESSMENT IN FUTURE PES.

To differentiate talent and provide useful information to promotion boards and talent managers, we recommend amplifying the comparative assessment in future PES. Of all the job performance indicators assessed in our prototype, including job performance traits, potential performance traits, promotion recommendations, organizational citizenship and counterproductive workplace behaviors, none was as informative in differentiating performance as the comparative assessment. In the current PES, a Navy officer's relative performance can be inferred by comparing the officer's average trait score to those of peers within their unit (SGA) and to their Reporting Senior's Cumulative Average (RSCA). Navy promotion boards must infer this relative performance as the comparison is not automatically calculated. Adding the comparative assessment in future PES instruments will provide key information useful to promotion boards and other talent managers who are trying to differentiate talent. Our review of the literature shows the Navy is structured as an internal labor market with tournament-style incentives, given its pyramid structure, set number of jobs at each level, and promotion to higher levels based on relative performance at the level below. Tournament labor markets are less concerned with absolute productivity than with promoting the most productive individual

relative to others. A good PES must then be able to signal individual workers' relative productivity so that talent can be differentiated and promoted.

B.   RECOMMENDATION 2: CONSIDER ADOPTING FUTURE POTENTIAL RATINGS FOR DEVELOPMENTAL/COACHING PURPOSES, ESPECIALLY THE WORD BLOCKS FOR ACTIONABLE, CONCRETE DEVELOPMENTAL FEEDBACK.

Our review of the scholarly literature and best practices on job performance assessment particularly highlighted the importance of separating the evaluative purpose of PES (promotion, retention) from their development purpose (feedback, job improvement). Trait items based on future potential provide a unique source of information that may be ideal for the purpose of development and coaching. In this study, Sailors were evaluated on six trait dimensions (e.g., *Character development*, *Leadership and teamwork skills*, *Judgment and decision-making*) based on their readiness to succeed at the next paygrade or in key, particularly demanding jobs. The benefit of this rating system is that it is forward-focused and constructive. When Sailors were rated as "not ready," we prompted respondents to offer concrete, actionable feedback to help the Sailor move toward greater perceived readiness. These trait ratings and qualitative feedback could serve as the basis for coaching sessions with Sailors to identify focus areas for job performance improvements that can advance them toward their next career milestone. This serves the interests of both the Sailors-reported-on and the Navy and reflects the intent of current Navy personnel initiatives such as *Get real, get better*, *21st Century Sailor Initiative*, and *Culture of Excellence*.

C.   RECOMMENDATION 3: FOCUS ON A SUBSET OF PERFORMANCE AND/OR POTENTIAL TRAITS IN ASSESSING JOB PERFORMANCE TO KEEP PES SIMPLE AND USEFUL.

To reach the goal of transforming the PES toward a model that is simple, useful, and fair, we recommend focusing on a subset of the traits rating past performance and/or those rating future potential. Ratings of past performance are potentially simpler and more useful indicators of job performance than ratings of future potential. Our analysis shows that the trait ratings based on past performance are as a group more predictive of other measures of performance than are ratings of future potential. For instance, the statistical relationships between ratings on past performance and comparative assessment on a forced distribution were stronger than relationships between comparative assessment and trait ratings of future potential. Both sets of

items, trait ratings based on past performance and trait ratings of future potential, conveyed unique information about Sailors-reported-on. However, for a future PES that is simple, useful, and serves the evaluative purposes of performance assessment, we recommend focusing on a subset of the traits. The next recommendation should enable further paring down of the trait items relating to past performance and future potential.

D.   RECOMMENDATION 4: ASSESS PREDICTIVE VALIDITY OF TRAIT STATEMENTS USING MULTI-SOURCE DATA.

As indicated in section 3, data from Sailors' personnel records will be merged with data from our prototype testing to expand our view of predictive, convergent, and divergent validity of the TVS. These data and resulting analyses are critical to understanding how well ratings of Sailors' job performance using the refined set of TVS reflect and track with job performance on a longer time horizon across multiple indicators of performance. These data will also allow us to carefully examine potential adverse impact of the refined TVS, including the question of whether some TVS items inappropriately favor certain groups over others. We caution against implementing the TVS items into performance evaluation tools until the results of these analyses are obtained.

E.   LIMITATIONS

The present study offers many insights for performance evaluation transformation efforts, however it is not without limitations.

In our sample, we observed an overrepresentation of high performers. Trait ratings based on past performance tended to skew toward the top of the scale, and comparative assessments placed more than 2/3 of Sailors-reported-on in the top 15% of their peers. This is perhaps not surprising, as Sailors-reported-on were recruited through an opt-in process and nominated their own raters; presumably struggling Sailors would be less likely to opt-in to the research. However, it limits our ability to rigorously examine how well the prototype instrument identifies and distinguishes between low- or even average-performers. Notwithstanding, we did find that some TVS, such as the future potential trait of "experience and competence," were able to distinguish between both top and bottom performers, while other traits (such as leadership) were able to distinguish among top-most performers.

In addition, in this study we opted not to constrain respondents' ratings of Sailors-reported-on with a Reporting Senior's Cumulative Average (RSCA). The RSCA is an important and controversial component of the Navy's existing performance evaluation system because it requires reporting seniors to balance their ratings of individual Sailors with their cumulative average of ratings across all Sailors-reported-on. In our prototype, raters were unconstrained by this second criterion, potentially allowing them to offer more generous ratings of Sailors-reported-on than they might otherwise make. The comparative assessment ("Christmas tree") question *did* force a distribution on Sailor ratings; however, there was no mechanism for enforcing that distribution on respondents' rankings. Moving forward, the Navy will need to carefully examine the issue of RSCA with an eye toward striking a balance between fairness for individual Sailors and fairness and integrity for the PES overall.

APPENDIX A

Performance Evaluation Transformation Prototype Testing

Thank you for agreeing to assist in the effort to improve the Navy performance management system by completing this survey.

To continue, please enter the **four-digit access code** provided with the link in your recruitment email:

_____

Performance Evaluation Transformation Prototype Testing

Authority to request this information is granted under 5 U.S.C. 301, Departmental Regulations; 10 U.S.C. 5031 and 5032. License to administer this survey is granted per OPNAVINST 5300.8C under RCS# NSP1610.01, expiration 6/7/2024.

We are a team based at the Naval Postgraduate School. We are working with the Talent Management Task Force at Navy Personnel Command to support the improvement of the Navy's performance management system, which includes how the Navy does fitness reports (FITREPs) and evaluations (EVALs).

To ensure that any future system is an improvement, we need help from currently serving Sailors. Specifically, we are seeking your help with testing some aspects of a potential future system. In particular, we are asking that you use this system to rate the past performance and future potential of the Sailor identified in your survey invitation email.

PLEASE NOTE that individual responses will not be shared with Navy Personnel Command. Individual responses will also be kept completely separate from any Sailor's military records and separate from any information used to make career decisions about any specific Sailor.

We estimate it will take you about 20 minutes to complete this survey. Thank you for your participation.

PURPOSE: The purpose of this survey is to evaluate and improve the properties of a performance evaluation prototype.

ROUTINE USES: Your responses in this survey will be combined with the responses of all others and will not be attributed to any single individual. The anonymized survey responses will be stored on a password-protected server at the Naval Postgraduate School.

CONFIDENTIALITY: All responses will be kept COMPLETELY confidential. Personal identifiers such as DoD ID number will only be used to obtain rank and demographic data that will be analyzed as part of a group. All the survey data will be statistically summarized and will not be attributed to any single individual. Individual responses will not be shared with Navy Personnel Command. Individual responses will also be kept completely separate from any Sailor's military records and separate from any information used to make career decisions about any specific Sailor.

PARTICIPATION: Completion of this survey is entirely voluntary. Failure to respond to any of the questions will NOT result in any penalties except possible lack of representation of your views in the final results and outcomes. You may withdraw your participation in the survey at any time by simply exiting the survey.

Q1 Please provide your DoD ID number. The reason we are asking for your DoD ID number is so that we can verify that all responses are from currently serving Sailors. It will also allow us to skip demographic questions. Your responses, however, will remain completely confidential and analyses of these data will take place only when DoD ID numbers have been removed from the data set.

_____

Q2 To ensure data fidelity, please enter the last name of the Sailor you will be reporting on in this survey (i.e., the Sailor identified in your recruitment email):

_____

Q3 What is this Sailor's current rank/rate?

○ E-1

○ E-2

○ E-3

○ E-4

○ E-5

○ E-6

○ E-7

○ E-8

○ E-9

○ WO1

○ CWO2

○ CWO3

○ CWO4

○ CWO5

○ O1E

○ O2E

○ O3E

○ O1

○ O2

○ O3

○ O4

○ O5

○ O6

Q4 Which of the following BEST describes your relationship with the person you are rating?

- ○   I directly supervise this person or did so recently
- ○   I indirectly supervise this person or did so recently
- ○   I am a peer to this person; we are on the same level or are teammates
- ○   I am junior to this person
- ○   Other (please specify):

_____

Q5 How confident are you in your ability to rate this person's job performance and career potential?

- ○   Not at all
- ○   Somewhat confident
- ○   Confident
- ○   Very confident
- ○   Extremely confident

End of Block 1: Introduction

DIRECTIONS

For all questions below, **please evaluate the Sailor you identified at the beginning of this survey.** Please use the trait descriptions and scale anchors provided in the questionnaire to guide your ratings.

In this first section, please consider this Sailor's actual workplace behavior. Carefully review each dimension and then provide your rating of this Sailor's performance during this reporting period as listed below. BE HONEST.

Again, individual responses will NOT be shared with Navy Personnel Command. Individual responses will also be kept completely separate from any Sailor's military records and separate from any other information used to make career decisions about any specific Sailor. Your assessment of this Sailor is entirely for the purpose of helping shape the future of Navy EVALs and FITREPs.

CHARACTER

Description: Taking responsibility for actions regardless of consequences; holding self accountable to Navy core values and ethical standards. Upholding the highest degree of integrity in professional and personal life. Doing the right thing, even when it is difficult; using discretion and avoiding inappropriate situations and actions. Being honest and forthcoming; treating others with dignity and respect.

1	2	3	4	5
Struggled to take responsibility for actions. Demonstrated a lack of judgment and alignment between Navy core values and behavior in professional or personal contexts. Acted inappropriately in ways that demonstrated questionable ethics; treated others with a lack of dignity and respect. Concealed or obscured the truth.		Consistently accepted responsibility commensurate with scope of duties and experience. Adhered to Navy core values at all times and held self accountable for actions. Demonstrated appropriate judgment and decision-making skills for paygrade; was always honest and forthcoming. Did the right thing, even when difficult. Used discretion and avoided inappropriate situations and actions. Demonstrated inner strength commensurate with scope of duties and experience; was willing to face moral challenges in pursuit of mission accomplishment. Treated others with complete dignity and respect.		Demonstrated integrity and ethics beyond reproach; was an uncommon role model of character and adherence to Navy core values. Always took responsibility for actions, placing conscience over competing interests, regardless of physical or personal consequences. Treated others with rare level of concern and respect; contributed significantly to a climate in which everyone felt valued and included.

Q6 Based on the criteria above, how would you rate this Sailor on **character**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

LEADERSHIP

Description: Ensuring all members understand their roles and responsibilities as appropriate; maintaining performance standards and holding others accountable for their actions. Delegating tasks and responsibilities appropriately. Embracing the diversity of ideas, experience, and backgrounds of all, creating a positive, motivating work environment. Acting as a leader and encouraging leadership in others. Providing and encouraging feedback appropriate to performance. Guiding others to seek support through available wellness resources. Effectively leading in times of change or crisis; demonstrating courage by intervening when necessary.

1	2	3	4	5
Lacked ownership of own role in driving positive outcomes, failed to encourage others in their roles. Tolerated substandard performance or deviation from necessary procedures. Tolerated or contributed to a negative work environment. Neglected appropriate delegation. Struggled both to correct own performance deficiencies and to deliver constructive feedback to others. Missed opportunities to intervene to correct inappropriate behavior or unsafe conditions. Exhibited discriminatory tendencies toward others. Failed to take responsibility for own words and actions and their impact on others.		Led by example. Set high standards, clearly articulated expectations and held others accountable as appropriate. When appropriate, delegated authority to those directly responsible for the task. Promoted a climate which values fairness, dignity, creativity, diverse perspectives, individual differences, and open communication. Took responsibility for own words and actions and their impact. Sought and provided high-quality performance feedback. Boldly intervened when necessary to support shipmates' well-being or safety. Created or contributed toward a professional work environment where all are motivated toward achieving desired results. Encouraged others to lead and learn.		Demonstrated a rare level of leadership considering current paygrade. Motivated others to achieve extraordinary results. Won people over rather than imposing will. Clearly articulated vision; empowered others to set goals and objectives to accomplish tasks. Modified leadership style to best meet challenging situations. Inspired others to perform at a higher level of leadership and creativity. Exceptionally skilled in creating a high-performance, high-trust environment; demonstrated unusual courage to take action in support of shipmates and command objectives.

Q7 Based on the criteria above, how would you rate this Sailor on **leadership**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

INITIATIVE and DRIVE

Description: Proactively addressing problems in the absence of specific direction; willingly putting in extra time and effort; seeking opportunities to contribute and innovate. "Embracing the red," being curious and taking pride in fixing problems. Seeking learning opportunities to acquire new competencies, methods, and information to enhance job performance; growing personally and professionally every day. Exercising discipline in conduct and performance, striving for continual improvement, self-control, and balance in mental, physical, and spiritual readiness.

1	2	3	4	5
Postponed needed action. Implemented or supported improvements only when directed to do so. Showed little interest in personal or professional development. Feasible improvements in methods, services, or products went unexplored. Demonstrated apathy regarding learning opportunities, personal appearance, physical fitness, and maintaining military bearing. Showed a lack of discipline and motivation toward proactive improvement of self or the team.		Proactively addressed problems in the absence of specific direction. Championed improvement through new ideas, methods, and practices; a self-starter with strong work ethic. Sought opportunities to contribute, innovate, and develop expertise. Pursued productivity gains and enhanced mission performance by applying new ideas and methods. Maintained required military competencies, skills, and professional demeanor/bearing. Has completed or is enrolled in appropriate level of training or education for grade and level of experience. Demonstrated disciplined approach toward physical, mental, and spiritual readiness.		Aggressively sought out additional responsibility. Made worthwhile ideas and practices work when others might have given up. Optimized use of new ideas and methods to improve work processes, decision-making, and service delivery. Displayed exceptional awareness of environment and ability to sense and respond proactively with innovative solutions. Demonstrated rare level of work ethic and commitment to life-long learning and overall readiness. Made time for study, professional reading, and took advantage of learning resources; inspired others to get better every day.

Q8 Based on the criteria above, how would you rate this Sailor on **initiative and drive**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

TEAMWORK

Description: Contributing to team success through actions and attitudes. Demonstrating inclusion and support of teamwork, assisting teammates in identifying solutions, developing productive working relationships. Honoring and valuing team members, recognizing others' supportive behavior. Actively supporting unit cohesion and group decisions even when not in complete agreement. Working collaboratively, building trust, and creating opportunities for the team to progress.

1	2	3	4	5
Used teams ineffectively or at wrong times. Mismanaged or ignored conflict. Excluded team members from vital information. Stifled group discussions or did not contribute productively. Inhibited cross-functional cooperation to the detriment of unit or service goals. Worked independently when the task required others. Made other feel unvalued or excluded, eroded trust, or claim credit when it should have been shared.		Built and supported unit cohesion through high-quality working relationships. Skillfully used teams to increase unit effectiveness, quality, and service. Resolved or managed group conflict, enhanced cooperation, and involved team members in decision processes. Valued and included team members. Effectively negotiated work across functional boundaries to enhance support of broader mutual goals. Shared credit, built trust, and helped others succeed.		Exceptionally skilled in facilitating collaborative solutions. Insightfully used teamwork to raise unit productivity beyond expectations. Inspired high levels of *esprit de corps*, even in difficult situations. Selflessly contributed to team efforts in a way that was truly extraordinary. Established relationships and networks across a broad range of people and groups, raising accomplishments of mutual goals to a remarkable level. Inspired unit cohesion through high-quality relationships, inclusion, and respect of others' contributions.

Q9 Based on the criteria above, how would you rate this Sailor on **teamwork**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

COMMUNICATION

Description: Actively listening to people's ideas and concerns to ensure comprehension, allowing others to speak without unnecessary interruptions, asking for clarification when unsure of what is being said or asked. Quickly elevating barriers, transparently sharing knowledge and skills. Presenting information clearly, concisely, and logically. Addressing sensitive issues in an open, constructive, professional manner allowing for rational and open discussion. Keeping leadership informed about progress and problems; consulting with others as necessary to determine priorities.

1	2	3	4	5
Failed to create understanding when sharing information verbally. Lacked preparation, confidence, or logic when speaking with others or in formal briefs. Used inappropriate language, rambled, or interrupted others. Failed to listen carefully or was too argumentative. Failed to communicate obstacles or barriers quickly, held onto information that should have been shared. Created written material that was frequently unclear, verbose, or poorly organized. Seldom proofread. Often submitted correspondence which was grammatically incorrect, tailored to wrong audience, or delivered by an inappropriate medium.		Listened to others actively, creating alignment and understanding. Elevated barriers and potential issues quickly. Effectively expressed ideas and facts in individual and group situations; non-verbal actions consistent with spoken message. Communicated to people at all levels to ensure understanding, effectively navigated sensitive issues. Provided written material that was clear, concise, and logically organized. Proofread conscientiously. Drafted grammatically correct correspondence that was tailored to an audience and delivered by an appropriate medium.		Demonstrated a rare level of verbal, nonverbal, and written communication excellence. Clearly articulated and promoted ideas before a wide range of audiences in both formal and unplanned situations. Presented complex or sensitive issues carefully and in a way that substantially elevated understanding. Keenly prioritized and escalated issues as appropriate. Fostered a culture of active listening. Clearly and persuasively expressed complex or controversial material through writing in support of Navy objectives.

Q10 Based on the criteria above, how would you rate this Sailor on **communication**?

◯ 1

◯ 2

◯ 3

◯ 4

◯ 5

◯ Not observed

CRITICAL THINKING

Description: Managing time and risk effectively to produce optimal decisions. Remaining flexible and objective in response to changing circumstances. Being honest, humble, and transparent about current performance; supporting others in ongoing, honest assessment of self and situations. Effectively integrating best available information in planning and execution. Knowing one's own capabilities and limitations, challenging own beliefs using data, facts, and diverse input.

1	2	3	4	5
Poorly analyzed situations or issues by ignoring evidence or holding onto biased thinking. Failed to make necessary decisions or jumped to conclusions without considering facts, alternatives, and impact. Did not effectively weigh risk, cost, and time considerations. Ignored or minimized own performance limitations or those of others. Demonstrated arrogance or an inability to change own mind in the face of contrary facts.		Made sound decisions with the best available information in a timely manner, leading to mission accomplishment. Collected and evaluated diverse sets of evidence to determine and recommend courses of action. Demonstrated analytical thought and incorporated a wide range of ideas into thought process. Displayed honesty and transparency in assessment of performance and capabilities, both own and those of others. Effectively prioritized and solved multiple complex problems. Subordinated personal interest and biases in favor of optimal decision-making.		Combined unusually keen analytical thought, situational awareness, and insight to make appropriate decisions. Focused on key issues and the most relevant information. Mastered the balance between desire for perfect knowledge and speed. Included many diverse perspectives when thinking through issues, demonstrated an awareness of how decisions impact others. Took reasonable risks to achieve positive results. Prioritized flexibility and objectivity, consistently used evidence-based analysis to provide and encourage honest, humble, and transparent assessments of performance at all levels.

Q11 Based on the criteria above, how would you rate this Sailor on **critical thinking**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

MISSION ACCOMPLISHMENT and PRODUCTIVITY

Description: Getting the job done at an acceptable level of quality and timeliness. Self-correcting, continually identifying and fixing small problems at the lowest level. Ensuring safe and secure mission execution through knowledge and adherence to policy and procedure. Effectively integrating time management and subject matter expertise to get results. Applying Navy problem solving tools and best practices to shift from more activity to better outcomes.

1	2	3	4	5
Had difficulty accomplishing routine tasks. Produced work that was late or of poor quality. Neglected safety or security requirements. Demonstrated a lack of competence, skill, knowledge, or adherence to policy and procedure. Maintained the status quo despite opportunities to improve. Concentrated on unproductive activities or often overlooked critical demands. Failed to engage other people or resources productively. Did not follow up. Mismanaged information, money or time. Used ineffective tools or left others without means to accomplish tasks. Employed wasteful methods.		Demonstrated consistent competence, skills, and knowledge in designator or rate commensurate with paygrade and experience. Ensured safe and secure mission execution, adhered to policy and procedure. Got the job done in all routine situations and in many unusual ones. Work was timely and of high quality. Results had a positive impact on department or unit. Continuously improved own and organizational effectiveness. Effectively managed time and drove correcting problems at the lowest level. Used and ensured others use resources and best practices to get better.		Demonstrated rare level of competence in designator or rate given paygrade and experience. Maintained optimal balance among quality, quantity, and timeliness of work. Consistently surpassed high expectations. Drove rapid identification and correction of small problems at the lowest levels. Used appropriate tools and best practices to achieve substantially better outcomes and continual improvement. Recognized and exploited new resources, created opportunities. Positively influenced team, unit, or Navy at large through innovation, hard work, and unusual level of expertise.

Q12 Based on the criteria above, how would you rate this Sailor on **mission accomplishment and productivity**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

RESILIENCY and TOUGHNESS

Description: Exhibiting poise and flexibility while executing duties in the face of adversity. Seeking and engaging in positive coping mechanisms when under stress. Courageously aiming high despite risk of failure. Recovering from setbacks with tenacity and renewed purpose, learning from misses and bouncing back. Pushing to find and fix root causes, not just symptoms.

1	2	3	4	5
Was overwhelmed by small obstacles and had difficulty coping appropriately with stressful circumstances. Failed to bounce back and learn from adversity. Set low goals to avoid risk of failure. Failed to take advantage of support from other Sailors or resources. Used unreasonable criteria to set priorities and deadlines. Rarely had plan of action. Failed to focus on relevant information. Overreacted or responded slowly to change in direction or environment.		Performed duties in the face of adversity and pressure with poise, flexibility, discipline, and stability. Proactively sought and used positive coping mechanisms to respond to stress. Recovered from setbacks with tenacity and renewed purpose. Set high goals and worked toward them with perseverance. Pushed to find and fix root causes of problems despite obstacles. Built and maintained high-quality relationships with other team members. Adjusted effort, goals, or methods in response to changing realities. Effectively dealt with volatility, uncertainty, complexity, and ambiguity.		Demonstrated exceptional skill in navigating adversity. Used and encouraged others to use positive coping strategies. Demonstrated seldom-matched presence of mind under the most demanding circumstances. Stabilized difficult situations through the resolute and timely application of direction, focus, leadership, and engagement with others. Maintained a sense of purpose and meaning. Advocated for and led efforts to learn from mistakes and fix tough problems. Rapidly assessed and adjusted to changing conditions. Turned pressure and ambiguity into constructive forces for change.

Q13 Based on the criteria above, how would you rate this Sailor on **resiliency and toughness**?

○ 1

○ 2

○ 3

○ 4

○ 5

○ Not observed

End of Block 2: Trait ratings based on past performance

DIRECTIONS

In this section, please consider this Sailor's *potential to succeed* at the NEXT PAYGRADE or in KEY, PARTICULARLY DEMANDING JOBS. With that in mind, provide your rating of this Sailor's future potential using the dimensions listed below. Remember, you are helping our Navy manage its talent. BE HONEST.

Again, individual responses will NOT be shared with Navy Personnel Command. Individual responses will also be kept completely separate from any Sailor's military records and separate from any other information used to make career decisions about any specific Sailor. Your assessment of this Sailor is entirely for the purpose of helping shape the future of Navy EVALs and FITREPs.

Q14
Learning Mindset and Adaptability
Description: Humility, curiosity, willingness to experiment, and commitment to own development. Sensing and responding rapidly to change.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

51

Q15
Leadership and Teamwork Skills
 Description: Empathy, approachability, adapting to situations, and building collaborative relationships. Inspiring and articulating direction, alignment, and commitment to objectives.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding
job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

Q16
Character Development

Description: Progression toward an increasingly strong, principled ethical orientation. Doing what is right over what is easy, understanding ethical dilemmas and how to promote an ethical climate.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding
job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

Q17
Judgment and Decision-Making
 Description: Critical analysis of situations and information to achieve desired outcomes. Using appropriate mental models and tools to make tough calls.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding
job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

Q18
Experience and Competence
 Description: Expertise and performance in designator or rate; having the breadth and depth of jobs or experiences that are relevant to future success.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding
job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

Q19
Motivation and Drive
 Description: Internal energetic force to take on additional responsibilities. Possessing a genuine desire to do increasingly complex, difficult work.

○ Not ready and is unlikely to be ready in the near- to mid-future If "Not ready," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding job: _____

○ Not ready yet, but could be ready in the next 3-5 years with consistent effort If "Not ready yet," briefly describe how this Sailor might become more ready for the next paygrade or for a key, particularly demanding
job: _____

○ Partially ready now and could be fully ready with developmental progress in the next 1-2 years

○ Ready NOW

○ Ready NOW and has been for some time

○ Not observed

End of Block 3: Future Potential

DIRECTIONS

 Please review the figure below and then answer the questions that follow regarding this Sailor's overall performance in the US Navy.

Again, individual responses will NOT be shared with Navy Personnel Command. Individual responses will also be kept completely separate from any Sailor's military records and separate from any other information used to make career decisions about any specific Sailor. Your assessment of this Sailor is entirely for the purpose of helping shape the future of Navy EVALs and FITREPs.

Based on your knowledge of this Sailor's performance, we would like for you to place him or her in one of the following categories. Assume that across all Sailors in the US Navy top- and bottom-tier performance are relatively rare, with most Sailors falling somewhere in the middle of the figure below.

Description	Comparative Assessment
ONE OF THE VERY FEW BEST AND MOST QUALIFIED	5%
ONE OF THE EXCEPTIONALLY QUALIFIED	10%
ONE OF THE HIGHLY QUALIFIED	20%
ONE OF THE MANY QUALIFIED PROFESSIONALS WHO FORM THE MAJORITY OF THIS GRADE	55%
A PROGRESSING SAILOR	10%
AN UNQUALIFIED SAILOR	
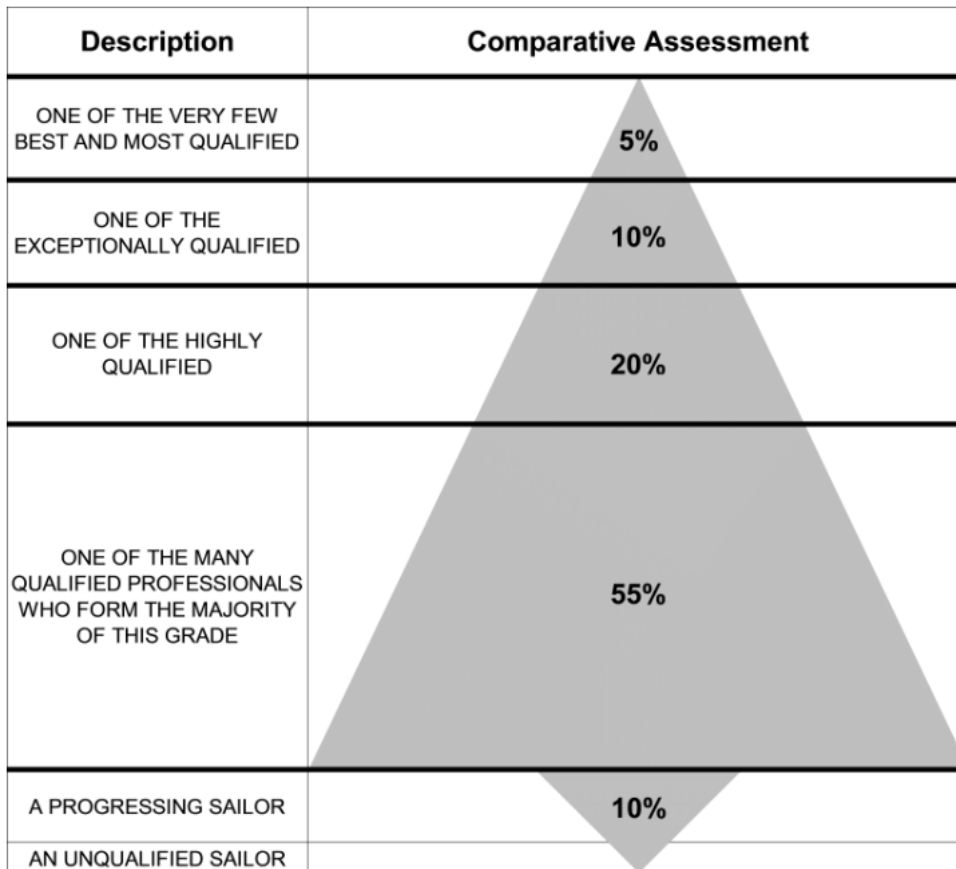
Q20 Compare this Sailor with others of the same paygrade whom you have known in your career. According to the schematic above, in which category would you place the Sailor you are rating? In making the comparison, consider all Sailors of this grade whose professional abilities are known to you personally.

○ One of the very few best and most qualified

○ One of the exceptionally qualified

○ One of the highly qualified

○ One of the many qualified professionals who form the majority of this grade

○ A marginally qualified sailor

○ An unqualified sailor

Q21 Amplify your comparative assessment mark; evaluate potential for continued professional development to include promotion/advancement, leadership positions/command, assignments, education, and retention.

_____

Q22 Please select ONE of the following promotion recommendations from the list below.

○ Already selected to next paygrade

○ Recently promoted (< 12 months)

○ Promote now

○ Promote with top 20% of peers

○ Promote with peers

○ Promotion potential

○ Retain at current paygrade

○ Do not retain

End of Block 4: Comparative Ranking

Start of Block 5: Workplace Behaviors

Q23 Based on your experiences with the Sailor, how frequently does he or she engage in the following behaviors?

	Never	Once or twice	Once or twice per month	Once or twice per week	Every day
Creates disruptive conflict with others at work	○	○	○	○	○
Musters late without reason or prior notification	○	○	○	○	○
Insults or makes fun of others at work	○	○	○	○	○
Takes time to advise, coach, or mentor fellow Sailors	○	○	○	○	○
Lends a compassionate ear when someone has a work problem	○	○	○	○	○
Helps Sailors who have too much to do	○	○	○	○	○

End of Block 5: Workplace Behaviors

Start of Block 6: Final Comments

Q24 Thank you so much for taking the time to complete this survey and supporting the Navy's performance evaluation transformation process.

If you have any additional comments about the questions we asked or how we asked them, please add them below.

_____

Q25 Would you be willing, once in a while, to participate in additional surveys to help improve how the Navy does performance evaluation? We would also like to be able to contact you with updates as we improve the system so you can see these ideas come to life.

○ Yes

○ No

Q26 If you said "yes" to the question above, please provide an e-mail address where you can be contacted for occasional additional survey requests.

_____

Q27 This is the end of the survey. By clicking on the "Next page" button below, you will submit your survey responses as is and will not be able to revise your responses. If you are satisfied with your responses, please submit them now by clicking the "Next page" button.

If you would like to review your responses you may use the "back" arrow button to do so.

End of Block 6: Final Comments

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B

Figure A-1. Distribution of Current Performance Trait Ratings
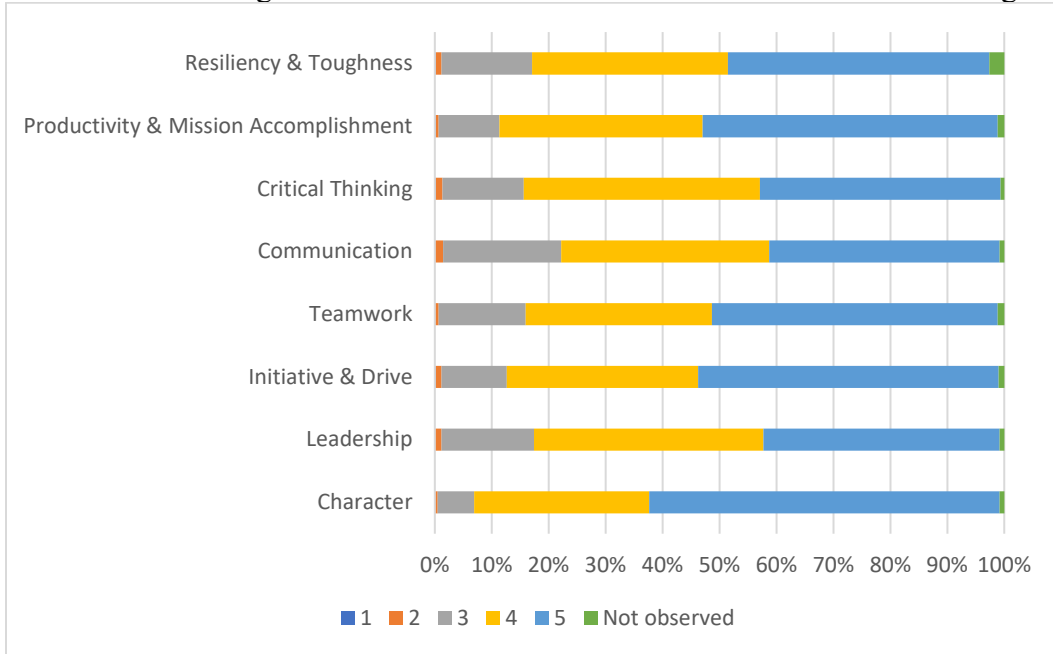


Figure A-2. Distribution of Potential Performance Trait Ratings
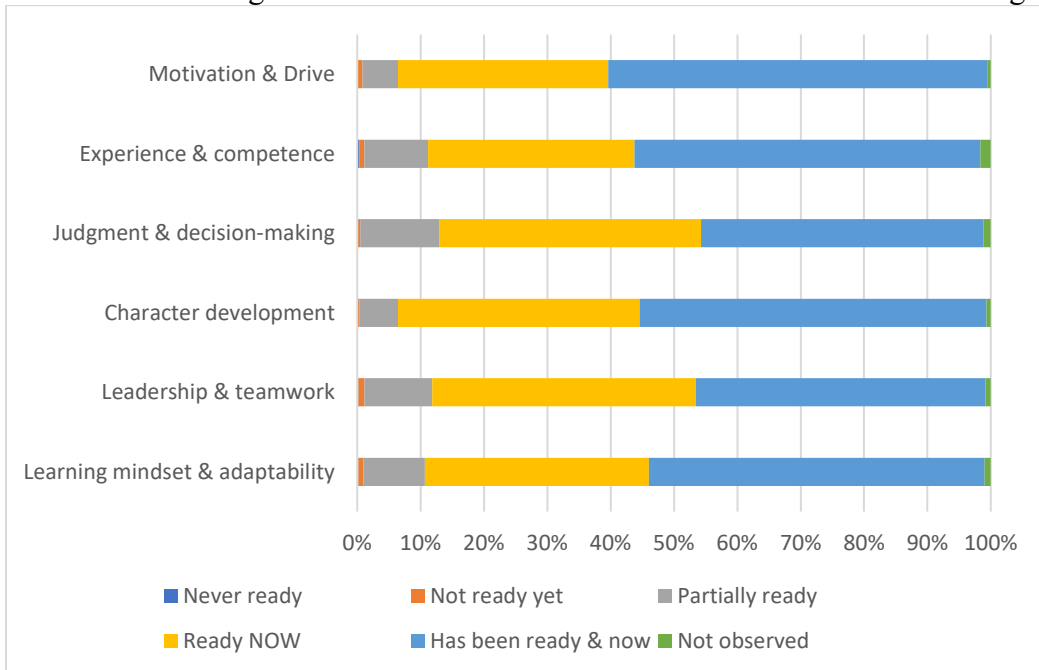
Figure A-3. Counterproductive Workplace and Organizational Citizenship Behaviors
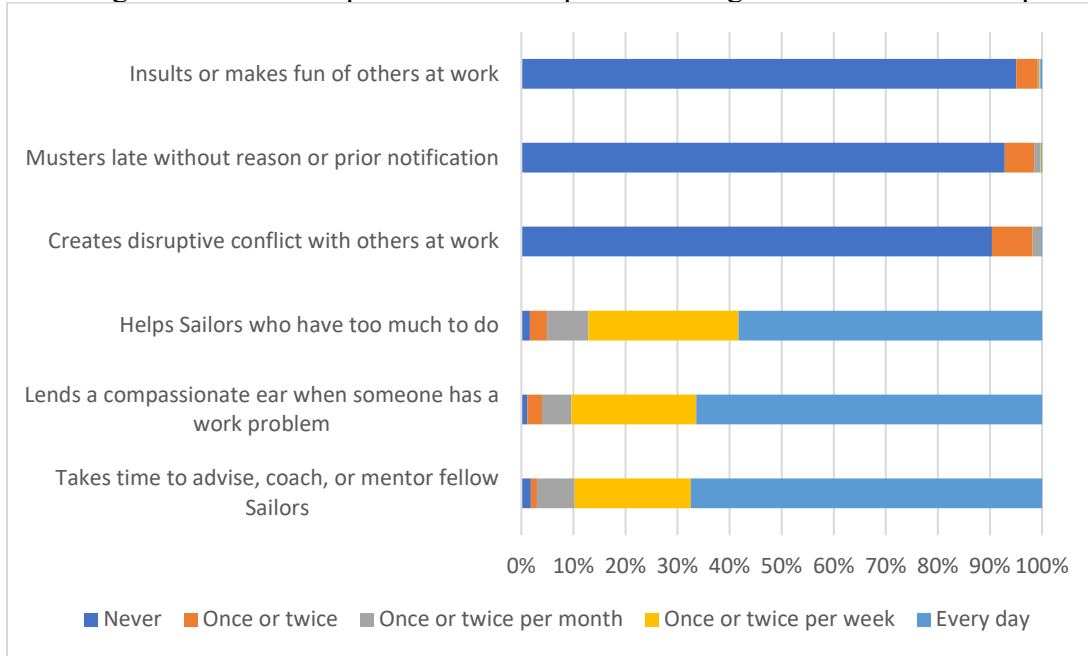
Figure A-4. Promotion Recommendation
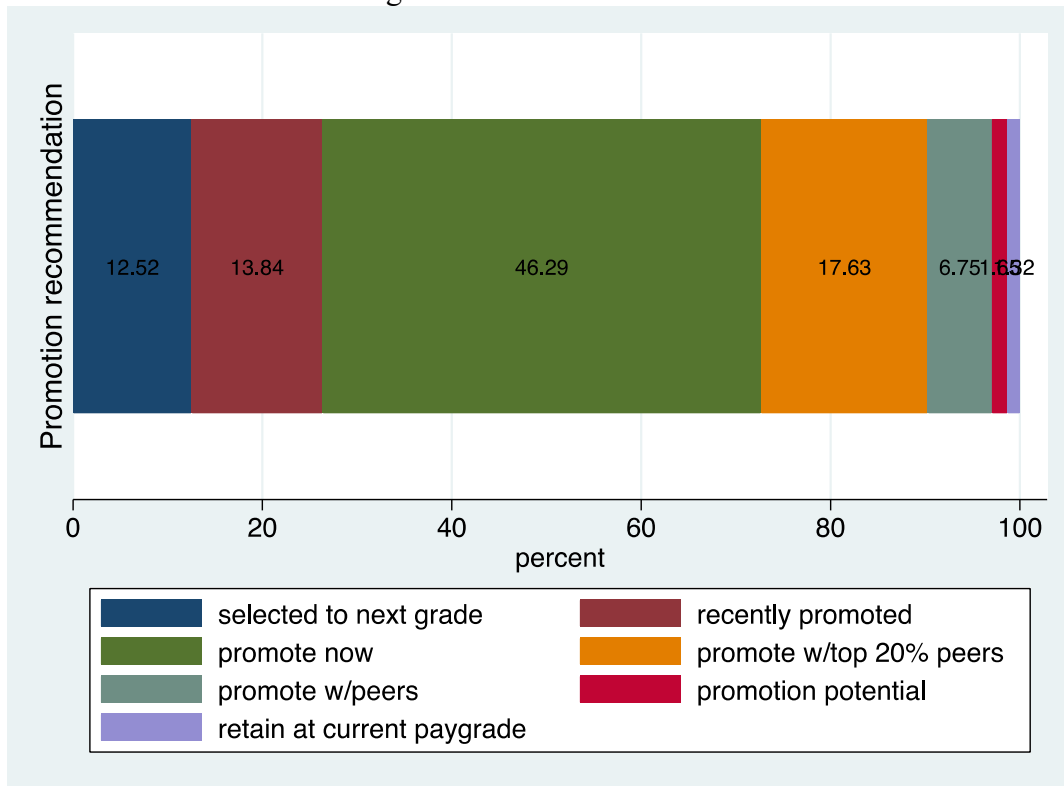
Appendix Table A-1. Raw Correlations

	character	leadership	initiative	teamwork	comm	critthink	productivity	resilience	learning	leadteam	character	judgment	experience	motivation	ocb1	ocb2	ocb3	cwb1	cwb2	cwb3
Performance Traits																				
character	1																			
leadership	0.5222	1																		
initiative	0.4433	0.5099	1																	
teamwork	0.4555	0.5882	0.433	1																
communication	0.5021	0.582	0.378	0.5371	1															
critthink	0.4688	0.5576	0.548	0.4852	0.564	1														
productivity	0.4858	0.5613	0.551	0.4998	0.484	0.577	1													
resilience	0.4697	0.5515	0.479	0.4922	0.53	0.523	0.4951	1												
Potential Traits																				
learning	0.4175	0.4824	0.399	0.3934	0.423	0.446	0.399	0.4224	1											
leadteam	0.4043	0.5138	0.341	0.49	0.475	0.468	0.4654	0.4342	0.603	1										
character	0.4116	0.3508	0.322	0.2917	0.351	0.348	0.3505	0.3356	0.461	0.4357	1									
judgment	0.3994	0.48	0.422	0.3451	0.412	0.501	0.4463	0.4507	0.536	0.5432	0.5525	1								
experience	0.3206	0.4376	0.295	0.276	0.317	0.372	0.4144	0.246	0.526	0.5213	0.4668	0.5727	1							
motivation	0.3951	0.4276	0.472	0.2696	0.289	0.371	0.4105	0.3756	0.558	0.4562	0.3887	0.502	0.4617	1						
OCB's and CWB's																				
Advise/coach	0.2831	0.3736	0.278	0.3253	0.325	0.265	0.2872	0.3243	0.357	0.359	0.2662	0.2897	0.2755	0.3266	1					
Listening ear	0.2509	0.3713	0.264	0.3238	0.282	0.285	0.2544	0.3309	0.264	0.3138	0.2394	0.261	0.2154	0.2739	0.624	1				
Helps sailors	0.342	0.3897	0.301	0.4049	0.291	0.31	0.3073	0.3211	0.3	0.3488	0.2399	0.2471	0.2455	0.3008	0.531	0.611	1			
Creates conflict	-0.249	-0.2413	-0.078	-0.2713	-0.26	-0.14	-0.1711	-0.178	-0.12	-0.276	-0.161	-0.138	-0.1193	-0.0863	-0.14	-0.17	-0.23	1		
Musters late	-0.227	-0.1846	-0.117	-0.1996	-0.14	-0.19	-0.1641	-0.162	-0.15	-0.199	-0.162	-0.111	-0.0736	-0.1671	-0.13	-0.17	-0.25	0.39	1	
Insults others	-0.134	-0.101	-0.05	-0.0909	-0.14	-0.05	-0.0215	-0.055	0.007	-0.057	-0.137	-0.088	-0.041	-0.0199	0.007	-0.04	-0.05	0.38	0.38	1

```
. factor $traits $potential
(obs=559)

Factor analysis/correlation                    Number of obs   =        559
    Method: principal factors                  Retained factors =         6
    Rotation: (unrotated)                      Number of params =        69
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	6.37479	5.55920	0.9201	0.9201
Factor2	0.81559	0.51348	0.1177	1.0378
Factor3	0.30211	0.17837	0.0436	1.0814
Factor4	0.12374	0.01154	0.0179	1.0992
Factor5	0.11220	0.06016	0.0162	1.1154
Factor6	0.05204	0.06637	0.0075	1.1229
Factor7	−0.01433	0.00438	−0.0021	1.1209
Factor8	−0.01872	0.04649	−0.0027	1.1182
Factor9	−0.06520	0.05633	−0.0094	1.1088
Factor10	−0.12153	0.00894	−0.0175	1.0912
Factor11	−0.13048	0.02017	−0.0188	1.0724
Factor12	−0.15064	0.02175	−0.0217	1.0506
Factor13	−0.17239	0.00613	−0.0249	1.0258
Factor14	−0.17852	.	−0.0258	1.0000

```
    LR test: independent vs. saturated:  chi2(91) = 3901.34 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances
```

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
character	0.6560	−0.1317	0.0310	0.0236	0.1518	0.1067	0.5164
leadership	0.7617	−0.1748	−0.0579	−0.0583	−0.0598	0.0457	0.3769
initiative	0.6516	−0.1550	0.3263	−0.0282	−0.0275	−0.0071	0.4433
teamwork	0.6474	−0.2936	−0.1604	−0.0769	−0.0297	0.0472	0.4599
communicat~n	0.6818	−0.2428	−0.2036	0.0616	0.0523	−0.0233	0.4276
critthink	0.7179	−0.1810	0.0461	0.1053	−0.0967	−0.0980	0.4197
productivity	0.7116	−0.1618	0.1312	0.0397	−0.1256	0.0637	0.4289
resilience	0.6756	−0.2305	0.0207	−0.0080	0.1172	−0.0901	0.4681
pot_learning	0.6973	0.2764	−0.0603	−0.1491	0.0468	−0.0508	0.4068
potleadteam	0.7104	0.1747	−0.2148	−0.0956	−0.0455	−0.0241	0.4069
pot_charac~r	0.5830	0.2586	−0.0109	0.1446	0.1304	0.0579	0.5518
pot_judgment	0.7063	0.2940	0.0190	0.1453	0.0033	−0.0637	0.3892
pot_experi~e	0.6012	0.3961	−0.0635	0.0509	−0.1366	0.0583	0.4529
pot_motiva~n	0.6215	0.2633	0.2238	−0.1434	0.0489	−0.0007	0.4713

Panel B. Rotated factor analysis

```
. rotate, varimax blanks(.3)

Factor analysis/correlation                 Number of obs    =        559
    Method: principal factors               Retained factors =          6
    Rotation: orthogonal varimax (Kaiser off)   Number of params =        69

    ----------------------------------------------------------------------
        Factor  |    Variance   Difference       Proportion   Cumulative
    ------------+---------------------------------------------------------
        Factor1 |    3.42982     0.48489            0.4950       0.4950
        Factor2 |    2.94493     1.87531            0.4250       0.9201
        Factor3 |    1.06962     0.92206            0.1544       1.0744
        Factor4 |    0.14756     0.04695            0.0213       1.0957
        Factor5 |    0.10061     0.01267            0.0145       1.1102
        Factor6 |    0.08794         .              0.0127       1.1229
    ----------------------------------------------------------------------

    LR test: independent vs. saturated:  chi2(91) = 3901.34 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances
```

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
character	0.5257						0.5164
leadership	0.6603	0.3518					0.3769
initiative	0.4418		0.5516				0.4433
teamwork	0.6907						0.4599
communicat~n	0.6946						0.4276
critthink	0.5993	0.3053	0.3091				0.4197
productivity	0.5552	0.3002	0.3931				0.4289
resilience	0.6059						0.4681
pot_learning	0.3265	0.6559					0.4068
potleadteam	0.4540	0.6147					0.4069
pot_charac~r		0.5516					0.5518
pot_judgment		0.6593					0.3892
pot_experi~e		0.6906					0.4529
pot_motiva~n		0.5503	0.4144				0.4713

```
(blanks represent abs(loading)<.3)
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Bartram, D. (2005). The Great Eight competencies: a criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185-1203.

Boice, D. F., & Kleiner, B. H. (1997). Designing effective performance appraisal systems. *Work Study, 46*, 197-201.

Boswell, W. R., & Boudreau, J. W. (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology, 16*, 391-412.

Campbell, J.P. (2012). Behavior, performance, and effectiveness in the twenty-first century. In SWJ Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology, Vol. 1* (pp. 159-196). New York: Oxford University Press.

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 47-74.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology, 83*, 615–633.

Clemens, A., Malone, L., Phillips, S., Lee, G., Hiatt, C., & Kimble, T. (2012, July). *An evaluation of the fitness report system for Marine officer*. Headquarters Marine Corps. https://www.hqmc.marines.mil/Portals/138/DRM-2012-U-001003-Final.pdf

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, *74*, 130-135.

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? *Journal of Applied Psychology, 102*, 421.

Ellison, J. (2014). *Improving the signal for U.S. Navy officer productivity.* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/44554

Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. *Human Resource Management Review*, *21*, 123-136.

Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation. *The International Journal of Human Resource Management*, *17*, 504-522.

Larger, R. (2017). *Effective of the Marine Corps' junior enlisted performance evaluation system: an evaluation of proficiency and conduct marks* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/53006

Luke, B. C. (2022). *Performance Evaluation Trait Validation*. [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/69677

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures:" Objective" versus" subjective" measurement. *Human factors*, *34*(4), 441-455.

Pettijohn, C. E., Pettijohn, L. S., & d'Amico, M. (2001). Characteristics of performance appraisals and their impact on sales force satisfaction. *Human Resource Development Quarterly*, *12*, 127-146.

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*, 370-395.

Serbu, J. (2017). Navy overhauling its evaluation system in drive toward pay-for-performance. *Federal News Network*, https://federalnewsnetwork.com/defense/2017/06/navy-overhauling-its-evaluation-system-in-drive-toward-pay-for-performance/

Small, L. (2020). *Successful practices for employee performance evaluations* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/64880

Uggerslev, K. L., & Sulsky, L. M. (2008). Using frame-of-reference training to understand the implications of rater idiosyncrasy for rating accuracy. *Journal of Applied Psychology, 93,* 711–719.

Weatherspoon, T. (2016). Sailor 2025: What you need to know. *All Hands*, https://allhands.navy.mil/Stories/Display-Story/Article/1840217/sailor-2025/

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189-205.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
    Ft. Belvoir, Virginia

2. Dudley Knox Library
    Naval Postgraduate School
    Monterey, California

3. Research Sponsored Programs Office, Code 41
    Naval Postgraduate School
    Monterey, CA 93943