Theses and Dissertations   1. Thesis and Dissertation Collection, all items

2022-03

PERFORMANCE EVALUATION TRAIT VALIDATION

Luke, Bryan C.

Monterey, CA; Naval Postgraduate School

http://hdl.handle.net/10945/69677

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

PERFORMANCE EVALUATION TRAIT VALIDATION

by

Bryan C. Luke

March 2022

Thesis Advisor:	Marigee Bacolod
Co-Advisor:	Erik Helzer

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

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

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE March 2022	3. REPORT TYPE AND DATES COVERED Master's thesis	
4. TITLE AND SUBTITLE PERFORMANCE EVALUATION TRAIT VALIDATION			**5. FUNDING NUMBERS**
6. AUTHOR(S) Bryan C. Luke			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. This project was funded in part by the NPS Naval Research Program.

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

13. ABSTRACT (maximum 200 words)

The Navy is undergoing several personnel management system changes that impact the daily lives and promotions of all naval personnel. A large part of this change is updating the traits and associated values statements that are graded on individuals' performance evaluations. The Navy has drafted 8 traits, 39 sub-traits, and 82 value statements (TVS) that need to be evaluated to determine their credibility within Navy doctrine.

This thesis synthesizes current civilian and military literature on performance evaluations and promotions to better frame the new TVS. This thesis then compares the new TVS against Navy doctrine and other military services' officer evaluations to evaluate TVS validity.

The results show that the new trait and value statements have much in common with Navy doctrine and other military services' officer evaluations.

This thesis recommends that the Navy develops a single document that defines performance standards; later, surveys should be fielded that further compare these TVS with Navy promotions to determine TVS's predictive validity.

14. SUBJECT TERMS performance evaluation, FITREP, performance appraisal			**15. NUMBER OF PAGES** 95
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

i

THIS PAGE INTENTIONALLY LEFT BLANK

PERFORMANCE EVALUATION TRAIT VALIDATION

Bryan C. Luke
Lieutenant Junior Grade, United States Navy
BS, Marquette University, 2018

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
March 2022**

Approved by: Marigee Bacolod
Advisor

Erik Helzer
Co-Advisor

Marigee Bacolod
Academic Associate, Department of Defense Management

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Navy is undergoing several personnel management system changes that impact the daily lives and promotions of all naval personnel. A large part of this change is updating the traits and associated values statements that are graded on individuals' performance evaluations. The Navy has drafted 8 traits, 39 sub-traits, and 82 value statements (TVS) that need to be evaluated to determine their credibility within Navy doctrine.

This thesis synthesizes current civilian and military literature on performance evaluations and promotions to better frame the new TVS. This thesis then compares the new TVS against Navy doctrine and other military services' officer evaluations to evaluate TVS validity.

The results show that the new trait and value statements have much in common with Navy doctrine and other military services' officer evaluations.

This thesis recommends that the Navy develops a single document that defines performance standards; later, surveys should be fielded that further compare these TVS with Navy promotions to determine TVS's predictive validity.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AZ	Above Zone
BZ	Below Zone
CNA	Center for Naval Analyses
CNO	Chief of Naval Operations
CNP	Chief of Naval Personnel
CO	Commanding Officer
EEO	Equal Employment Opportunities
FITREP	Fitness Report
IRC	Independent Review Commission
IZ	In Zone
NAVPLAN	Naval Plan
NPC	Naval Personnel Command
OSR	Official Service Record
PES	Performance Evaluation System
PET-TM	Performance Evaluation Transformation - Talent Management
RSCA	Reporting Senior Cumulative Average
SECNAV	Secretary of the Navy
SGA	Summary Group Average
SME	Subject Matter Expert
TF1N	Task Force One - Navy
TVS	Trait and Value Statements
USMC	United States Marine Corps

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. MOTIVATION

> To confront the disrupters of our nation's peace and prosperity, crucial attention has been paid to developing the missiles, platforms, and emerging technologies the U.S. Navy will need in the future battlespace. However, these investments are only as lethal as the warfighters who wield them.
>
> —Vice Admiral John Nowell and Lieutenant Daniel Stefanus, 2020

The Navy's performance evaluation system (PES) is undergoing major changes through the initiatives of Sailor 2025 enacted by senior leadership. Sailor 2025 is a set of human resources initiatives endorsed by the U.S. Navy to bring its personnel management system into the modern age. The current Navy PES was developed over a quarter-century ago and has not evolved with technological and social advancements of the last 25 years (Burke, 2017). "Today, national demographics, social norms, and our knowledge economy continue to evolve, while our current systems to identify, groom, and retain talented Americans for national service may again lie locked in the assumptions, policies, and statutes of a previous era" (Moran, 2014, para. 3). The goal of Sailor 2025 is to update the Navy's PES so it can better focus on attracting and retaining the best sailors in a competitive labor market while combatting emerging threats in the world (Weatherspoon, 2016).

Two main priorities within Sailor 2025 are stated: to provide sailors with "immediate performance feedback and a clearer understanding of development opportunities within their career" (Commander of Naval Education Training Command [NETC], n.d.). NETC has the goal of creating a PES that is easily accessible and intuitive for all sailors. Priority is being given to making a PES that is a modern and flexible tool where sailors can conduct all facets of personnel management, including managing their awards, evaluations, re-entry, advancements, and detailing, with simple access on a mobile device. The desired end state is to have this PES accessible to sailors with mobile smartphone access instead of using cumbersome programs and software. Essentially, the

Navy wants to make performance evaluations clear, fair, transparent, constructive, and easier to access for all sailors.

Aside from mechanics and software improvements to the PES, the Navy is turning away from an industrial-age model of performance evaluations to a coaching and developmental model. Vice Admiral Nowell and Lieutenant Stefanus (2020) describe the importance of effectively managing personnel with MyNavy HR's mission: "to recruit, retain, and manage the talent of the sailors who can win those wars and empower them throughout their careers, however long or short" (para. 2). The nature of war and conflict is constantly evolving, as is the nature of recruitment and retainment; the Navy faces the increasingly difficult task of maintaining end strength as it competes for personnel with a "thriving civilian labor market" (Nowell & Stefanus, 2020, para. 3). Considering both the risk of global conflict and the need to retain the most talented sailors, the Navy must develop sailors as better overall assets to the nation with improved quality of life both at work and at home (Witherspoon, 2016). The desired end-state is to reduce PES clutter so sailors can focus on their jobs while at work, and their families while at home.

The Navy uses three separate tools in its PES to evaluate service member performance: Evaluations (EVAL), Chief Evaluations (CHIEFEVAL), and Fitness Reports (FITREP) are used to rate the performance of enlisted sailors, Chief Petty Officers, and officers, respectively. The Commander of Naval Personnel (CNP) has worked alongside military subject matter experts (SMEs) to brainstorm new performance traits to be included on Navy evaluations; this thesis focuses on the Navy officer FITREP and the new proposed trait and value statements (TVS) that it may contain.

B. BACKGROUND

The research team at the Naval Postgraduate School is sponsored by Naval Personnel Command (NPC) to conduct research to ensure the proposed TVS support the organizational goals of the Navy. The new performance traits began as a list of more than 200 before being narrowed down to the current 82 TVS. These statements describe 39 sub-traits and 8 broad character traits; the structure of these trait and associated value statements are described in detail in the methodology section of this thesis. The research team will

work extensively over the next year to gather data on the new TVS as performance measures within the construct of the Navy's evolving PES. The bottom line is to create a narrowed set of unbiased, validated TVS for performance assessment efforts across the Navy.

C. PURPOSE

The goal of this thesis is to establish a framework that grounds the Navy's new PES in the theory of performance evaluations. This thesis synthesizes civilian and military literature to set the scene for ongoing research on the Navy's new TVS. Evaluations are complex, expensive, and subject to many different errors; as the Navy develops new values and goals for the way it manages personnel, leadership needs to be aware of the potential pitfalls of creating a system misaligned with its organizational objectives. This thesis examines the proposed TVS against Navy doctrine and other military officer evaluations to begin the process of narrowing down the TVS to an unbiased and validated set. A valid set of TVS is necessary to accurately evaluate the performance of and develop Naval personnel.

1. Validity Defined

A high-quality performance evaluation tool has at least two properties, construct validity and reliability of the metrics. An evaluation tool is valid when it accurately assesses the dimensions of performance that it purports to measure; it is reliable if the items or traits measured in the performance evaluation provide consistent scores across different raters and timepoints (Campbell & Wiernik, 2015).

There are several different types of validity for any psychometric tool: construct validity, face validity, predictive validity, and convergent validity, among others (E. Helzer, email to author, January 3, 2022). Construct validity describes the degree to which the evaluation tool measures the appropriate theoretical constructs. In this thesis, the TVS are construct valid if they assess characteristics identified as important in Navy doctrine and other service officer evaluations. Face validity asks the question: Does the assessment tool appear to measure what it purports to measure? Face validity, in this thesis, means how well the TVS, as written, appear to represent the traits and values being measured.

Predictive validity reflects how well the evaluation tool predicts outcomes like future job performance or promotion. Convergent validity is the idea that the evaluation tool tracks with similar constructs; there is convergent validity if TVS scores correlate closely with the current FITREP variables. This thesis will focus on construct and face validity since the TVS will be mapped to current Navy doctrine. Predictive and convergent validity require data that is not currently available.

D.   RESEARCH QUESTIONS

The research questions addressed in this thesis are:

1.   How do the evaluation and promotion processes work in the Navy and how do they align with practices among civilian organizations and other military branches? This is the background research question.

2.   How are the newly developed FITREP Trait and Value Statements supported by current Navy doctrine and other service branches? This is the primary research question.

3.   How can the proposed Trait and Value Statements be utilized to effectively assess talent within the Navy? This is the secondary research question.

E.   ORGANIZATION

The rest of this thesis is organized into the following four chapters.

1.   Chapter II is an in-depth literature review of the theories surrounding performance evaluations and of prior research efforts. Topics covered are the purposes for evaluations, PES development, measurement issues, and evaluations in practice. Navy evaluations and promotions are then described in detail and compared against theory.

2.   Chapter III provides a description of the military documents that were used as performance standards in this thesis as well as the methodology

that was employed to cross-reference the TVS with these military documents.

3.   Chapter IV contains the tabled results from the cross-referencing process. This chapter also discusses findings from the analysis of the validity of the TVS.

4.   Chapter V presents conclusions and any recommendations for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

II.   LITERATURE REVIEW

A.   PERFORMANCE EVALUATION SYSTEMS

"A performance evaluation system is a systematic way to examine how well an employee is performing in his or her job" (University of Minnesota Libraries, 2016, para. 1). Performance evaluations are a large part of the overall PES in an organization; this means the PES is typically formalized or codified by management. To better understand the direction that the Navy is pushing for with the new PES and associated evaluations, this thesis digs deep into performance evaluations: how they have developed, what they are, what is contained in them, how they can be structured, and what the desired outcome is, for both military and civilian industry. This thesis focuses specifically on the FITREP, the Navy's performance evaluation instrument for officers, as well as the PES.

1.   Purposes

Joshua Ellison (2014) states that two main purposes for conducting performance appraisals and evaluations in an organization are to "accurately measure individual performance outputs over a period of time" and "to improve individual and organizational productivity" (p. 4). The author notes that many organizations use a rudimentary evaluation tool simply because there "is no better alternative to measure individual performance" (p. 5).

Laszlo Bock, former head of Google's People Operations in the early 2010s, states that "the major problem with performance evaluation systems today is that they have become substitutes for the vital act of actually managing people (Bock, 2015, p. 151). Bock is suggesting that companies are no longer using performance management tools to develop individuals to be better and stronger assets to the company; rather, they are using them to support a quick reward or punishment, potentially in the form of promotions or firing, for the accomplishment or failure to realize goals. These goals are often shared among a whole department or organization, but the evaluation process should still be focused on the individual; a high achieving individual should not receive poor evaluations based on the shortcomings of his or her department (Boice & Kleiner, 1997).

Campbell and Wiernik (2015) lay out, in detail, the main reasons for evaluating performance: "high-stakes appraisal … performance feedback and development … self-managed performance improvement … research … legal support" (p. 59). The first of these, high-stakes appraisal, reflects promotion decisions and bonus opportunities, but is not the only reason to assess performance.

a.   High-Stakes Decisions

The employee and employer must consider "high-stakes" when it comes to employment; high stakes facing individuals are "promotion, dismissal, reassignment, and compensation decisions" (Campbell & Wiernik, 2015, p. 60). Organizations use evaluations to determine whether an employee is a deficient, average, or excellent performer; if the evaluation and PES are rooted in organization goals and employee expectations are clear, then the organization can utilize their PES to supplement the promotion process, support the dismissal process, validate restructuring and reassigning, and endorse pay or benefits changes. The authors discuss that many organizations focus their PES and evaluations on the high-stakes outcomes of promotions and bonuses, and the result is that employees place higher values on getting good performance evaluation ratings than self or organization improvement (Campbell & Wiernik, 2015). This process will not change; I believe that as long as numerical scores are tied to evaluations, individuals will focus on earning a higher numerical rating instead of concentrating on what the rating truly signifies.

b.   Performance Feedback and Development

In Industrial Age thinking, the goal of evaluations was to grade employees on productivity or output against their peers to identify and promote employees with the highest productivity (Seager, 2018). While organizations still use performance evaluations to grade performance and determine viability for promotion, many have turned away from strict productivity ratings in favor of developmental, or coaching, models. Performance development and feedback serve organizations when high-stakes decisions are "not directly involved" (Campbell & Wiernik, 2015, p. 60). The authors note that the goal of feedback and development is to identify "specific, substantive behaviors that need

enhancement or improvement" (p. 60). The process of providing feedback and the use of developmental performance models is discussed in detail later in this thesis.

c. Self-Managed Assessment

Performance assessments can also serve the purpose of self-managed performance improvement: self-managed performance improvement is where the employee sets their own goals based on the organization's values and provides self-feedback based on specific events that can be measured, not general performance over the entire evaluation period (Campbell & Wiernik, 2015).

d. Research

Research can be conducted through performance evaluation studies to determine how well or accurately the performance evaluation tool measures the performance of an employee; this is done by collecting data on employee output and comparing to the standards set for that employee (Campbell & Wiernik, 2015). In this way, the organization can determine whether the tool is valid.

e. Legal Support

If the evaluation tool is valid and reliable, then the organization can use those evaluations as legal support when making changes in personnel: this is demonstrated when an organization fires an employee and uses that employee's poor performance evaluations as grounds for the dismissal (Campbell & Wiernik, 2015). An organization's performance evaluation process must be legally defensible in the case that fired employees file lawsuits out of frustration or spite towards their former employer.

2. PES Development and Performance Standards

Performance evaluation systems are complex and very specific to the organization that they belong to; typically, each organization, including each military branch, has a performance evaluation system that is "tailor-made" for what the organization values (Boice & Kleiner, 1997, p. 197). As time goes on, organizations can change and generate

new values; this can either result in a completely new PES or an old PES adapted to fit the new values.

a.   *PES Development*

Performance evaluations and the associated PES in an organization were, at one point, tailor-made for the organization, even if the PES no longer fits the organization in its current form. Different organizations have distinctive objectives and values, so it is highly unlikely that one organization's values and employee expectations will be generalizable to other organizations. *PES shape* is a term that describes the format and style of questions and responses that an organization provides to managers for rating employees' performance. (Fowler, 2013). Fowler notes that PES shape can impact the results of the evaluation tool. PES shape can be studied and applied to more than one organization if the items within the evaluation are updated to match different organizational values. The goal of performance evaluation is to appraise how well an individual is fulfilling their specific duties and responsibilities while embodying organizational goals (CGT Staffing, 2020). The objective of the PES, under which the performance evaluation falls, is to create an avenue for supervisors to effectively manage and foster employee engagement to achieve high levels of job performance" (Gruman & Saks, 2011).

b.   *Performance Standards*

"A performance standard is a management-approved expression of the performance threshold(s), requirement(s), or expectation(s) that must be met to be appraised at a particular level of performance" (US Office of Personnel Management [OPM], n.d, para. 2). Boice and Kleiner (1997) note that the first step in accurate performance rating is to set performance standards. Once the overarching standards are clearly defined, they can be broken down into departmental, divisional, and/or individual objectives; these sub-groups allow managers to set targets for employees to strive for (Boice & Kleiner, 1997).

An organization's performance standards "outline the expectations of its employees, including their roles and how they should act while in the workplace" (Indeed Editorial Team, 2021, para. 1). Employees can be rated only after the organization's performance standards are set and performance evaluation systems are implemented.

Figure 1 shows the "importance of aligning behaviors" with the organizational mission (Ellison, 2014, p. 7). The mission of the organization must be set before management can build overall strategy. Once strategy is determined, management can set goals and performance standards for employees. These performance standards, which factor into how employee performance is measured, impact employee actions in the workplace.

Additionally, raters need to be educated on the expected performance standards to precisely measure how well an employee performed (Campbell & Wiernik, 2015). Ellison (2014) explains that "rater training is critical to improve the ability of the rater to conduct accurate performance appraisals" (p. 12). The author continues to say that rater training "increases the manager's knowledge" of the PES and makes managers more aware of "biases that unknowingly affect" their decisions (p. 12). Rater training is essential in the conduct of performance evaluations to maintain a high level of trust in the PES.
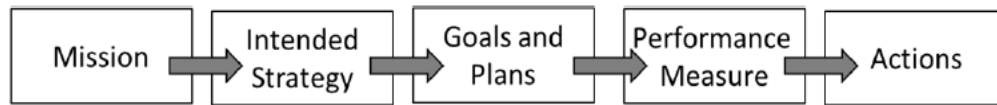
Figure 1.   Relationship Between Individual Actions and Organization
Mission. Source: Ellison (2014).

c.   *Performance Evaluation Elements*

Performance evaluations take various forms, but usually share some similar elements. Shaout and Trivedi (2013) name some "critical elements" or variations of the following graded items: "quality of work … quantity and timeliness of work … reliability and initiative … relationship with others … safety and use of equipment" (p. 405). Each organization determines how those critical elements should manifest at different performance levels. The authors explain that, in quantitative evaluations, there are associated point values for every critical element that gets factored into individuals' overall performance scores. Depending on the style of PES, managers provide feedback to the graded individual on their ratings and/or how they can improve. In my experience, this feedback can be delivered by formal written reports or informal conversations that are

designed to convey the details for why each critical element was graded as such. While these critical elements are broad, senior leaders create specific measures within the mission of their organization based on the critical elements.

B. PERFORMANCE MEASUREMENT ISSUES

1. Subjectivity and Objectivity

Performance evaluations can be either subjective or objective. The differences between subjective and objective are based "on the degree to which human experience, judgement, and feeling (subjective factors) are involved in the derivation of reality. Knowledge is valid (and objective) only when it is not dependent on human processes" (Muckler & Seven, 1992, p.441). Subjective performance measurements rely on an individual's or supervisor's personal judgement of an employee's performance or skill based on observed interactions (Ellison, 2014). Conversely, Ellison (2014) describes objective measurements as "non-judgmental" (p. 9); i.e., those that are independent of the evaluator's experience (Muckler & Seven, 1992). For example, subjective measurements might include perceptions of how well a leader leads or whether an employee is a good team player; objective measures can include directly quantifiable behaviors, such as whether a salesperson is able to achieve a target sales number. Note that the authors differentiate the productivity of an employee from their leadership ability and relationships with others in the workplace in recognition of the fact that some jobs require vastly different metrics based on required output and organizational roles (Campbell & Wiernik, 2015, p.61). The best leader is not always the most productive worker, and the most productive worker is not always the best leader.

While objectivity works well in metric-based goals, subjectivity is unavoidable in interpersonal relations because interpersonal relations are inherently subjective, and it is impossible to separate humans from their personal experiences (Muckler, & Seven, 1992). To rate how well an employee leads, it is imperative to speak to that employee's team to get their personal, subjective input on their leadership, even if this input contains bias. Since subjective evaluations are completely dependent on personal experiences,

supervisors must understand that biases or bitterness can exist amongst employees that need to be accounted for before solidifying any evaluations.

2. Accuracy and Precision

Subjectivity and objectivity are not the only considerations when evaluating performance; accuracy and precision of the evaluation tool are critical as well. Accuracy is how well an evaluation tool effectively reflects the true value of what is being measured; it is seen when a PES can place employees at their exact levels of performance. Precision is seen when a PES can order a list of employees in the true order of lowest to highest performers, but the tool is not perfectly placing individuals at their precise level of performance (Joshi, 2016). An organization can utilize surveys of managers and employees to determine if their PES is accurate and/or precise. An evaluation tool can be valid if it is either accurate or precise; precision will allow the tool to effectively order the performance of all individuals, and management can select the top individuals for promotion. Accuracy can also generate validity because it places the individuals at their exact performance levels, of which management can choose the top levels for promotions.

Studies on performance evaluation systems are expensive ways to obtain information on the precision of a measure, especially in large organizations (Fowler, 2015, p. 2). Fowler (2015) recommends extensive research and a focused, full-scale survey tailored for this "special-purpose" information when looking to study how well a performance evaluation system is working (pg. 2). The data that is expected to be returned when surveying precision should not be available elsewhere; if a researcher can find that information without the survey, then the survey is a waste of time and money.

In addition to subjectivity and objectivity, Campbell and Wiernik (2015) discuss a few challenges with accurately measuring performance on evaluations: "typical vs. maximum performance … performance dynamics … cross-cultural performance assessment … and the distributional properties of performance" (p. 60).

a.   *Typical vs. Maximum Performance*

For performance assessment, the organization must decide if it wants to measure an employee's typical or maximum performance. Typical performance is shown by an employee on an everyday basis. Maximum, or surge, performance is the performance exhibited by "highly focused, increased effort levels, and … higher effort levels for some periods of time" (Campbell & Wiernik, 2015, p 61). Individuals face stressful situations at work and at home that impact their motivation, so the individual may not be able to increase their effort when maximum performance is expected; these conditions must be kept in mind when looking to put metrics on specific performance levels. Deadrick and Gardner (2008) discovered a high correlation between maximum and typical performance. This means that an employee with a higher daily performance is likely to have a higher capability to surge when maximum performance is required.

b.   *Performance Dynamics*

The dynamic nature of performance is the next consideration in accurate performance assessment. The level of performance output or effort, required by an employee is situationally dependent and can change over time based on opportunities, interactions, or trainings provided by the organization and other coworkers (Campbell & Wiernik, 2015). Employees can also experience extenuating circumstances outside of the workspace that can impact their performance levels, for better or worse. The authors discuss that employees facing issues with paying bills on time may be more motivated to work hard and maintain their performance, or they may be so mentally distraught that they cannot focus while at work. Motivation changes in unique ways for every individual even when faced with similar circumstances. Ellison (2014) notes that "productivity varies among individuals based on heterogeneous inputs" because each person has their own "innate ability" which is "seldom observable" (p. 5). The dynamic landscape of performance dictates that an organization understands that employees are inherently different and face circumstances, both in and out of the organization's control, that can impact performance levels at work.

(1)   Recency Bias

Recency bias occurs when managers lean on recent interactions with employees as opposed to performance over the entire evaluation period (Bock, 2015). Chatfield (2016) notes that the "tendency to assume that future events will closely resemble recent experience" is a "universal psychological attribute" (para. 3). Chatfield argues that basic human psychology draws the rater's focus to recent performance, instead of long-term performance, as being more indicative of future performance. Recency bias is not unique to large organizations like Google or the Navy, it occurs everywhere that human subjectivity is present, which is everywhere, according to Muckler (1992). Calibration helps to mitigate the risk associated with recency bias by maintaining accountability across managers.

c.   *Cross-Cultural Assessments*

"Cross-cultural performance assessment" is the idea that practices in one country or culture may not translate accurately to another country or culture, even for employees with different cultural backgrounds within the same organization (Campbell & Wiernik, 2015, p. 61). In 2013, Rotundo and Xie found that performance assessment structures in Western culture share many similarities with those in Chinese industry, but many details and values within each culture are different. For example, in Chinese industry, social aggressions manifested themselves in "indirect and political behaviors," unlike Western cultures where interpersonal aggression is expressed in direct confrontation (Campbell & Wiernik, 2015, p.61). Within this study of the U.S. Navy, a cross-cultural organization, a comparison will be made between the proposed TVS and current Navy doctrine to guard against discovered cultural biases in the PES that would prohibit fair and accurate assessment.

d.   *Distributional Nature of Performance*

Performance has a variable distributional property, which means there is a wide range of performance levels amongst employees, and it is not a normal distribution (Campbell & Wiernik, 2015). By applying a normal distribution, mathematical properties force the average performer to the 50th percentile; the 50th percentile means that the

individual performs better than 50% of others, or is exactly average (Manikandan, 2011) This results in 50% of employees being automatically rated above average and 50% rated below average. If an organization forces a normal distribution on their employees' performances in determining promotion or incentive outcomes, the PES will likely be rewarding average employees for mediocre work (Campbell &Wiernik, 2015). O'Boyle and Aguinis (2012) argue that a Pareto 80/20 relationship better fits performance distribution than a normal distribution: 80% of the organization's success or output is due to 20% of the employees or inputs. Those 20% of employees are the best performers deserving of recognition and rewards, while the other 80% of employees are at or below performance standards. Figure 2 shows the graphical relationship between inputs and outputs when a Pareto distribution is used. The exact mathematical distribution of performance in an organization is disputed, but there is support that performance scores have a distributional value that varies in each organization. Ultimately, the relationship between performance and outputs is not linear; that relationship can take on a multitude of dimensions and forms (Campbell &Wiernik, 2015).
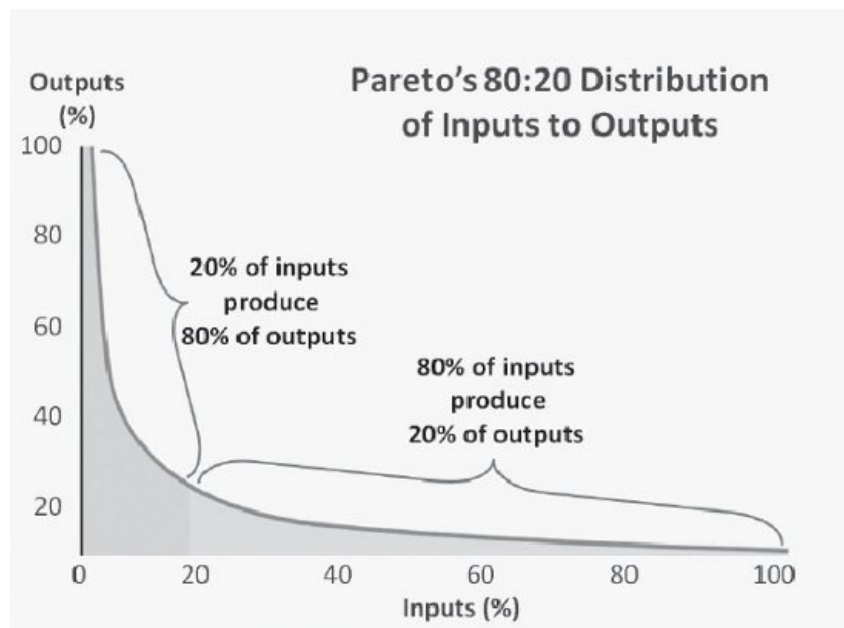
Source: https://www.oreilly.com/library/view/top-market-strategy/9781606493106/ah_id_18.html

Figure 2.   Pareto 80/20 Relationship

e.   ***Calibration***

The process of rating and re-rating employees by a group of seniors is what Laszlo Bock calls "calibration" (Bock, 2015, p. 164). Calibration is a step that Google uses to ensure fairness, trust, and equality in their evaluation process. In the first step of Google's performance evaluations, a manager drafts performance ratings for each employee that he or she is directly responsible for. The second step is the calibration phase; calibration occurs when groups of managers, who oversee similar departments, convene to fine tune scores for all individuals across their departments based on departmental or organizational metrics. This helps instill trust in the system for upholding fairness and objectivity with ratings by relieving "the pressure managers may feel from employees to inflate ratings" (Bock, 2015, p. 164). Because managers are working together to define what performance is, the calibration process also helps create a pseudo-standard for what is expected from employees. Ultimately, calibration forces managers to make tough decisions and rate employees fairly, as well as prevents "discrimination on factors other than merit" (p. 8); this means that employees have a better chance at being rated fairly. The outcome of calibration is some distinction between high, low, and average performers, thus creating a working definition of what performance is for that organization or department.

The manager calibration phase at Google is comparable to the Navy's deliberation phase, explained in a later section, where members are scored, then "racked and stacked" against members of similar jobs and roles within their unit (Sicard, 2017, para. 7). However, Sicard notes that racking and stacking in the Navy, while designed to simply rank similar sailors, is a process used to incentivize tenure and seniority, not merit (Sicard, 2017). [Say something about how this works against the ultimate goal of calibration, as you described in your last paragraph--- then go on to Burke's recommendation] According to Sicard (2017) the Chief of Naval Personnel (CNP) Vice Admiral Robert Burke recognized this and issued a recommendation for the complete removal of the rack and stack methodology because it undermines the objectivity and transparency that the Navy is pushing for in Sailor 2025 Personnel Management (PM) initiatives.

f.   Costs of Evaluation

The performance evaluation process is one of the most time-consuming processes that organizations, both military and civilian, spend time on. Studies at Google showed some managers spent 24 weeks out of their entire work year "assigning ratings, calibrating ratings …, or communicating ratings" (Bock, 2015, p. 157). Buckingham and Goodall (2015) studied Deloitte, a large audit, tax, and accounting company, and found that it spent nearly "2 million hours a year," on average, on performance management and rating (para. 6). Others found "that the average manager and employee spend 210 and 40 hours, respectively, on PM activities, … costs of 30 million USD annually for a company of 10,000 people (Pulakos et al., 2019, p. 250). Evaluating performance is an expensive and labor-intensive process for organizations to employ, so it must be performed well, or it is a waste of time, funds, and energy. As evaluations happen more frequently in an organization, the cost of the PES will increase concurrently. Periodicity is an important factor in cost evaluation of PES. Typical evaluation periods occur quarterly, semiannually, or yearly (Lattice Team, 2019). An organization must reconcile how important evaluations are to their ethos and how often evaluations need to happen to be effective given the budget for their PES. When Google changed from quarterly to semiannual evaluations it noticed an increase in the quality of evaluations, employees' toleration of the PES, as well as "an instant time savings of 50%" (Bock, 2015, p. 160). He notes that since evaluations are time and energy consuming, many managers were satisfied when Google decreased the number of evaluations performed for each employee in a year. As this example points out, performance evaluations and associated PESs are very expensive systems to get wrong within an organization. Ellison (2014) notes that "if returns from human capital investments exceeds costs, then improved financial productivity should be the result of retaining the most productive individuals" (p. 14). This is to say that if an organization saves money by maintaining its workforce through an effective PES, then the investment in the PES is worth it. The author continues to explain that "financial returns" on PES investments are "generally large" (p. 14). Aside from costs directly related to implementing a PES, there are other costs, called indirect costs, that can be incurred by having an inefficient PES.

g.   *Indirect Costs*

Searching for, hiring, and training new employees are called turnover costs, which are expensive and can cost an employer 33% of an employee's annual salary (Hall, 2019). There are costs, other than hiring and training new employees, called "replacement fees," also associated with turnover: productivity loss, workplace safety issues, and morale down spiral (O'Connell & Kung, 2007, p.1). O'Connell and Kung (2007) discuss that when a new employee is hired, they are often not as productive, fully trained, or qualified as a previous employee, so there is a productivity or performance gap that lasts as long as it takes for a new employee to become proficient in their skills. This same concept applies to learning new safety procedures in the workplace. A new employee would take longer to learn whole "new safety systems" and procedures than it would for an existing employee to be retrained or refreshed (O'Connell & Kung, 2007, p. 3). Morale depletion is the last indirect cost of turnover that O'Connell and Kung discuss in their paper; employees that are fired or decide to quit their job might force other employees to be "overburdened" (p. 2). As morale diminishes, employees will become dissatisfied and look elsewhere for employment, causing a rise in costs for the organization as it recruits and trains new hires.

Most organizations understand how expensive it is to acquire and train new employees, but do not know how to use their evaluations to better supplement their personnel management systems. Ellison (2014) notes that "performance appraisals provide information that managers may use the make critical personnel decisions that affect productivity" (p.6). Effective performance evaluations provide unique opportunities for organizations to avoid expensive turnover costs by equipping the organization's leadership to better develop and manage their current workforce. Ellison (2014) makes the statement that "if returns from human capital investments exceeds costs, then improved financial productivity should be the result of retaining the most productive individuals" (p. 14). This is to say that if an organization saves money by maintaining its workforce through the PES effectively, then the investment in the PES is worth it. The author continues to explain that "financial returns" on PES investments are "generally large" (p. 14).

C.   IMPLEMENTATION OF PERFORMANCE MANAGEMENT

It is evident that having a valid and reliable PES in an organization is important. However, there is no standard way to implement performance evaluations due to differing personnel management styles across different organizations. So, what format does the tangible evaluation take on once the organization has considered its values and define its PES goals? Every organization has unique standards, so there are many ways to look at what performance truly means, but no gold standard for how to do it best. While performance evaluation models are constantly being developed and improved to better reflect what an organization values and expects of its employees, there are a few main styles to which they typically conform.

1.   Performance Assessment Methods

Performance assessments often look different, but most fall under one of the following broad methods: "ratings … samples, simulations, and proxies … or technology-enhanced assessment" (Campbell & Wiernik, 2015, p. 62).

a.   *Ratings*

A rating system is a way for organizations to place a numerical value on an employee's performance to determine their level of skill or achievement in their job or role. Rating systems, either by employers, coworkers, employees, or by self-rating, are most common within organizations because they are relatively easy to employ and help in discriminating between levels of performances (Campbell & Wiernik, 2015). The main issue with applying numbers to rate an employee's performance is that the value associated with the number is not directly related to performance; often, higher scores are due to supervisors trying to reward an employee with bonuses or incentives, not accurately rating their performance. To combat this, what constitutes good or bad performance must be specifically laid out. Raters must be trained on what the organization values and what it expects of specific employees so they can better uphold the precise and unbiased assessment of their employees (Campbell & Wiernik, 2015). Otherwise, as they lay out, raters will rely on their "subjective performance measurements" without fully understanding or representing organizational standards (p. 65).

b.   *Work Samples, Simulations, and Proxies*

An organization can use work samples as another way to evaluate performance in the workplace. A work sample is when an employee is observed doing a real work task "with real job materials;" for a mechanic, this would mean he or she gets rated on how well they fix a real engine (Campbell & Wiernik, 2015, p. 65). A simulation is similar to a work sample but uses fictitious scenarios and replica work materials; an example of a simulation is a driver wearing a video headset and sitting in a fake car seat to practice driving skills. A proxy is simply a situation that is to be used to see how an employee would respond to a task but does not closely resemble the task itself. Proxies are less common, but may use physical or psychological stimuli, like work retreats with employees playing team-building games with or against each other, to see how well an employee will interact with customers back at work; the situations are very dissimilar, but the performance responses therein are similar. Campbell and Wiernik (2015) refer to these three all as "simulations" (p. 65). The argument is that simulations are most advantageous because they truly assess an employee's capability for "performing critical tasks that are otherwise difficult … to assess with any frequency" (Campbell & Wiernik, 2015, p. 65). Examples provided are military and medical simulations where the graded event is only as useful if it feels real to the individual engaged with it. A surgeon may be able to pass a graded exam, but how well will they perform in the operating room under pressure? The main disadvantages to simulations are how costly and time consuming they are to develop and operate, as well as how they lack construct validity since they usually only assess "few (or one) critical job tasks," not the entire job description (Campbell & Wiernik, 2015, p. 65).

c.   *Technology Enhanced*

Technology can be used to enhance assessment by automatically tracking numerical values like sales, deliveries, quotas, or other metric-based outcomes as well as scanning calls and emails to ensure customer satisfaction. Technology must be used with caution because it only picks up measurable outcomes, not necessarily interpersonal performance or individual experiences. Campbell and Wiernik (2015) discuss how big data can augment performance evaluation systems by automatically keeping records on

employees, but the chief complaint is there are perceived "invasions of privacy" (p. 66). For now, technology cannot detect what is directly under the individual's control and what is not; if a delivery truck gets in an accident and a shipment gets delayed, technology may not be able to factor that away from the individual's performance assessment (Campbell &Wiernik, 2015).

2.   Coaching

In a developmental, or coaching, PES model, the focus of performance evaluations is the growth and retention of an employee. When an employee exerts effort to finish a task or achieve a goal and does not attain the objective, they might not "feel motivated to keep performing" (Rodriguez &Walters, 2017, p. 207). An organization that understands this should not employ Industrial Age performance evaluations as a cornerstone of their PES. If an individual is feeling jaded about the relationship between their input effort levels and output results, and receives poor evaluations on top of that, they may see their chances at promotions or pay raises dwindling and [what is the cost to the organization?].

This problem may be especially true in an internal labor market where an employee's option is either to promote or leave, a process called "up or out" in the military (Schirmer et al., 2005, p. 1). Schirmer et al. (2005) note that in the military, an individual must be promoted to specific ranks by certain timelines in their careers; if they are not rated high enough and promoted by that time, they are forced to retire. In an internal labor market like the U.S. Navy, pay and benefits do not change based on individual performance evaluations; military pay and incentives are decided upon in Congress and based on one's official rank in the organization (FederalPay, n.d.). The caveat with the Navy is that evaluations are heavily considered when in contention for a promotion, and promotion to the next rank in the military does include increased base pay and benefits. However, the military cannot grant raises or bonuses based solely on good performance evaluations; raises only occur after promotions.

The performance feedback, or coaching style, model encourages constructive critiques and the development of individual employees to help them reach the next tier or paygrade. When pay and high-stakes decisions are not heavily involved, the organization

can use their performance evaluation to identify behaviors that "need enhancement or improvement" (Campbell & Wiernik, 2015, p. 60). According to Bock (2015), "people love useful information that helps them do their jobs better" (p. 177). To help develop employees with a coaching model, the evaluations need to be specific, not general, when it comes to identifying these behaviors requiring development. The idea of a developmental model is to take perceived negative actions or behaviors and turn them into coachable moments that the employee can improve on. The evaluation should be grounded in specificity and avoid high-level, general performance appraisal of the individual. The developmental model needs to give concrete occasions where the employee can see their actions and improve to better align with the goals and standards set forth by the organization's PES (Campbell & Wiernik, 2015).

3.   Feedback

Feedback is the process of a manager communicating the results of an employee's performance with the employee from a given evaluation period (Elicker et al., 2006). Ellison (2014) points out that feedback is "one of the most cited purposes for conducting performance appraisals" (p. 7). It is an integral part of the performance evaluation process, especially in coaching models, and is often overlooked by many organizations because it is misunderstood and aversive (Steelman et al., 2004). Feedback can be formal or informal. Formal feedback refers to an official process that happens after yearly or semi-annual performance appraisals. Informal feedback is a less structured process that happens more frequently (Katz et al., 2021). Feedback, both formal and informal, is typically given face-to-face.

According to Steelman et al. (2004), a "feedback environment" needs to be created within organizations to encourage employee efforts (p. 166). A feedback environment consists of daily processes and superior-subordinate interactions, not general feedback sessions that occur once per year after performance evaluations are already submitted. Feedback needs to be constant and consistent, and employees or subordinates should be encouraged to ask for feedback as frequently as possible (Bock, 2015). In a feedback environment, employers know the need to develop personnel into stronger assets for their

organization as opposed to simply grading them high or low when compared to peers. Similarly, employees want to know how to improve themselves, not just how they stack against their peers.

Kluger and DeNisi (1996) found that 38% of the time, manager-provided feedback causes employees' performance to worsen. Steelman et al. (2004) argue that worsening performance in response to feedback is related to a misunderstanding of how to provide constructive feedback, not a negative relationship between feedback and performance. Often, managers are reluctant to provide "negative judgements" because it could be received as "de-motivating" to employees (Toppo & Prusty, 2012, p. 2). To remedy this, Google, for example, distributes a productive feedback handout to both evaluators and their subordinates to help map the feedback conversation to maintain its constructiveness (Bock, 2015). Google's method separates the feedback from the superior, who is grading the individual's performance, from the feedback from peers, who are adding input to help better the individual.

The Navy lists feedback as a purpose for counseling and evaluations in the official performance evaluation guidance document, the BUPERSINST1610.10F, but does not have formal processes for providing it. The only feedback recommendation in the 1610.10 is for a supervisor to ask a subordinate to generate feedback on their personal performance before the supervisor provides his or her own feedback on the subordinate (Office of the Chief of Naval Personnel [CNP], 2021). It should be noted that this feedback recommendation falls under formal counseling, which, in my experience, only happens in response to poor or negative behaviors; feedback in the Navy does not fall under the standard performance evaluation process. Essentially, the Navy is not placing developmental value on feedback because the organization does not require the feedback process, it only recommends it.

If provided properly, constructive feedback helps motivate subordinates to improve. Bock (2015) suggests that individuals being evaluated want actionable feedback, or specific instances and ways to improve on their performance, not generalized feedback from the entire observation period. Ellison (2014) states that performance does "vary between individuals; most want those differences to be recognized" (p. 5). Employees at

Google are also encouraged to seek supervisor, subordinate, and peer feedback as often as they want, not just during evaluation periods (Bock, 2015). This is different from the Navy's feedback system where feedback is recommended after evaluations and ratings are complete, or when the individual must be formally counseled for inappropriate behavior (CNP, 2021).

a. Peer and Supervisor Input

Google handles feedback by starting with input from an individual's peers, not just their supervisors (Bock, 2015). Bock notes that as a supervisor or manager, he cannot fully witness an employee's total performance over a given timeframe; he is incapable of providing accurate and precise feedback on all aspects of an individual's work. To address this, Bock decided to gather input from the employee's peers separate from the feedback provided by the supervisor. This helps managers better grasp the extent of employees' work (2015). Google uses input from multiple sources, some senior and some junior to the individual, to help the individual grow and develop.

The Navy has a deliberation and pseudo-calibration period when Chiefs and superiors add input to an individual's performance, but with different methods and end-states. The Navy uses input from different sources to add to the comments on performance section of the officer evaluation but solicits the feedback from superiors in the evaluation phase (CNP, 2021). Per the BUPERSINST1610.10F, input is not returned to the individual until after they are rated; this means the input cannot be used to help coach individuals or positively influence behavior before scores are given (CNP, 2021). The Navy's method closely resembles industrial-aged evaluations where the comments and feedback section are merely used to validate the rating scores, which are often substantiated by superiors looking to influence promotion outcomes (Small, 2020, p. 28).

4. Labor Markets

This thesis surveys existing literature on performance evaluation systems in both the civilian and military sectors, including how those systems feed into promotion decisions. One critical distinction between these sectors is that, unlike many civilian organizations, the military operates within an internal labor market, in which promotions

mainly happen from within the organization (Lazear & Oyer, 2003). Black et al. (2009) define an internal labor market as "the system by which recruitment for senior appointments … is mainly from existing employees in lower-grade jobs. The main merits … are that a firm is likely to know more about the strengths and weaknesses of existing employees than outsiders, and that a reputation for internal promotion as its preferred strategy may assist in recruitment and retention of staff" (no page). This study will use Black et al.'s definition of internal labor market as a foundation for describing the Navy PES environment.

The hierarchy within internal labor markets tends to be more structured and the promotion processes are stricter. Since all employees in an internal labor market organization are in competition with their peers for promotion, the resulting performance evaluations are rigid, bureaucratic, and have direct implications for promotions. Because of steeper barrier to entry in the form of entry-level training and experience, these organizations are less equipped to make "outside hires" from an external market when positions become available (Lazear & Oyer, 2003, p. 4). The specificity of this initial training and education is what prevents internal labor markets from connecting to external labor markets at any point other than the bottom job tiers. Ellison (2014) notes that "under an internal structure, workers have promotion advantages since there is no external competition" (p. 3). However, individuals within the internal labor market are in direct competition with their peers. To be promoted in an internal labor market like the Navy, an individual must be ranked better, or look more promotable, than their immediate peers who are trying to obtain the same promotion. This is where performance evaluations and PES connect to promotion: an individual's promotion chances are directly tied to the organization's current PES and the individual's performance ratings.

5.   Case Studies

Laszlo Bock focused his performance evaluation system on the philosophy of work and behavioral economics at Google; Bock makes the concepts of talent acquisition and human resource management as generalizable as possible (Bock, 2015). Employees at Google, when providing input and feedback on individuals, are required to use a sliding

scale to indicate how well they know the individual's projects and performance; this helps weight their feedback scores based on familiarity with the individual's work (Bock, 2015). Bock was amazed at how similarly his tech company managed personnel compared to Wegmans, a "regional grocery store chain" (p. 9); this shows that organizations can use identical processes for managing personnel despite having different goals. The employees at both companies understand that their leadership, while focused on running a profit-driven business, "will do the right thing for their employees, no matter what it costs" (Bock, 2015, p. 9). This concept is also seen in how they evaluate employee performance: Google is flexible and willing to try new ideas to best evaluate their employees.

In order to be more effective at evaluating its personnel, the Navy needs to learn from Google and be flexible with its PES. Bock (2015) restructured the Google evaluation tool and PES multiple times over the course of a few years to study and ensure that the evaluation methods best fit the organizational goals. This stands in direct contrast to the rigid Navy PES that has not changed in decades. The Navy must use research and data to create an evaluation tool that can meet its evolving operational and developmental needs. Evaluations are expensive processes to develop and hard to change once implemented, but Google was willing to change that evaluation system at all costs to get the best results. Google is more interested in the purpose of evaluations than maintaining bureaucratic processes (Bock, 2015).

Deloitte, an audit and accounting firm of 65,000 employees, also showed flexibility in its approach to evaluating employees. Prior to a major PES overhaul, Deloitte historically used an industrial-aged rating system (Buckingham & Goodall, 2015). Performance evaluations began with goal setting; performance objectives and goals were set for every individual in the company at the beginning of each year.

> After a project is finished, each person's manager rates him or her on how well those objectives were met. The manager also comments on where the person did or didn't excel. These evaluations are factored into a single year-end rating, arrived at in lengthy "consensus meetings" at which groups of "counselors" discuss hundreds of people in light of their peers. (Buckingham & Goodall, 2015, para. 4)

Seniors at Deloitte understood the need to remain objective when grading whether employees reached their goals. Managers provided subjective feedback in a way that led to a conversation about how employees could improve in a specific area of their work. Deloitte performed evaluations only once a year and had no coaching system in place. They did, however, use a calibration-styled process to arrive at a consensus of all employee ratings at year's end, costing millions of workhours (Buckingham & Goodall, 2015). Deloitte had created a PES that employees did not trust because 62% of the variance in scores were explained by individual managers' "peculiarities of perception;" only 21% of the variance was explained by differences in "actual performance" (Buckingham & Goodall, 2015). This means the PES employed by Deloitte lacked construct validity since their performance measures were adversely impacted by the subjectivity of managers.

Based on research and introspection, Deloitte decided to renovate their entire PES. They started by clearly defining what their goals were for managing performance. Deloitte restructured around three new objectives. The first objective was recognizing performance through variable compensation. Buckingham and Goodall (2015) note that many organizations offer variable pay and incentives based on performance outcomes. However, Deloitte split from other PES by the way it solicited information on employee performance by rephrasing and restructuring the conversation around performance. Deloitte realized it only needed to talk to the direct team leader, not peers, and the questions asked were strictly subjective, not objective (Buckingham & Goodall, 2015). They made this switch because individuals are more consistent at rating their feelings than other peoples' skills (Buckingham & Goodall, 2015). Below is the list of 4 questions that Deloitte asks the immediate team leader of an employee being evaluated:

1. Given what I know of this person's performance, and if it were my money, I would award this person the highest possible compensation increase and bonus [*measures overall performance and unique value to the organization on a five-point scale from "strongly agree" to "strongly disagree"*].
2. Given what I know of this person's performance, I would always want him or her on my team [*measures ability to work well with others on the same five-point scale*].
3. This person is at risk for low performance [identifies problems that might harm the customer or the team on a yes-or-no basis].

4. This person is ready for promotion today [*measures potential on a yes-or-no basis*]. (Buckingham & Goodall, 2015, para 16–19)

These questions dig deeper into an employee's current abilities and future potential, instead of just their rating scores. Deloitte uses this new system to focus the evaluation process on actually developing employees. This style of evaluation makes the critical shift from the difficult process of objectively rating output and performance to thinking in a much more analytical way about the impact an employee has on others.

The last objective of Deloitte's new PES is to "fuel performance" (Buckingham & Goodall, 2015, para. 24). Deloitte's team leaders are tasked with checking in with all the employees that they are responsible for at least once per week; the constant process creates the "feedback environment" that Steelman et al. (2004, p. 166) recommend. The subsequent feedback environment put coaching at the center of the Deloitte PES (Buckingham & Goodall, 2015). The last major note from Buckingham and Goodall (2015) is that some team leaders will be busy and push these team check-ins to a low priority. To address this, they pushed the responsibility onto both the employee and the team leader. Subordinates, rather than team leaders, are pushed to initiate the weekly check-in conversation, in the same way Google requests subordinates to ask for frequent feedback from supervisors (Bock, 2015). Deloitte calls this "radically frequent check-ins" and has viewed an increase in employee engagement because the individual employee, "who more often than not is eager for guidance" wants to have developmental conversations with their team leader (Buckingham &Goodall, 2015, para. 27).

The Navy PES and evaluations are very structured systems that do not foster the feedback environment or the development of individuals. If the Navy's goal is to create a PES centered on coaching and development, the Navy can learn from Deloitte and employ subjective ratings that nurture conversations between superiors and subordinates that drive the growth of the individuals involved in the PES.

D.   NAVY EVALUATION AND PROMOTIONS

1.   Evaluations in the Navy

The goal of this section is to further investigate the Navy's performance evaluation and talent management system initiatives and compare them to stated Department of the Navy (DON) goals.

The Commander of Naval Personnel (CNP) published a 178-page document, the BUPERSINST1610.10F, that governs how performance evaluations should be conducted from the individual to their rater, to their supervisor and groups of supervisors, and to their overall reporting senior (RS) (CNP, 2021). I will focus on the Navy officer evaluation, called a fitness report, or FITREP. The most common FITREP type in the Navy is a "regular" report, which happens annually based on the individual's rank; each rank performs evaluations at different times based on this cycle (Small, 2020, p. 20). FITREPs also take place when an individual detaches from their current command, or their RS detaches from the command (CNP, 2021).

The evaluation process begins with the individual providing input for what they have done over the course of the evaluation period. This self-reported input gets factored into an additional comments section where the RS uses "input from the service member and the member's immediate supervisor(s), as well as the raters' and reporting senior's personal observations" (CNP, 2021). The language that the instruction uses insinuates that the RS will converse with the individual about their input and performance before the comments get edited and approved; in practice, individuals write their own comment section, called a "brag sheet" of what they did in the command since their last evaluation, and submit them for verification (Small, 2020, p. 25). Small (2020) notes that "while instruction strongly dictates the procedures for completing the evaluation forms, the process of gathering Sailor performance data … is not standardized" (p. 24). The sailor is still giving their input, but in a manner that is not aligned with the instruction. This Navy FITREP culture contrasts with the conversation that the instruction uses; as opposed to providing input for and discussing their performance with their supervisor and the RS, sailors are writing their own comments entirely.

Once a member's evaluation comments are submitted, they are considered by senior leadership before the comments are rewritten and/or approved (Small, 2020, p. 25). Often, the comment section looks nothing like the brag sheet provided by the individual.

As the comments are reviewed, seniors assign scores for each observed trait based on that member's performance and interactions over the period of report of the FITREP. Traits evaluated on the current FITREP are displayed in Table 1. Scores are deliberated on by seniors in the command to not only reflect member performance, but to manage the commander's Reporting Senior's Cumulative Average (RSCA). RSCA is a "historical average of all trait score averages that the specific reporting senior has ever given to individuals in the same paygrade" (Small, 2020, p. 23). In the Navy, the individual member's score over the 7 evaluated FITREP traits is averaged and compared to peers within their unit, called a summary group average (SGA), and to their commander's RSCA (Small, 2020, p. 24). The reporting senior is the individual who signs the member's FITREP as the supervisor and senior rater. This is typically the commanding officer (CO) of the unit.

Table 1.   FITREP Traits. Adapted from Chief of Naval Personnel
[CNP] (2021)

Professional Expertise
Command or Organizational Climate/Equal Opportunity
Military Bearing/Character
Teamwork
Mission Accomplishment and Initiative
Leadership
Tactical Performance

Per the 1610.10F, COs also provide what is defined as a "promotion recommendation" for each individual being rated at that time period; from most favorable to least favorable, they are: "early promote" (EP), "must promote" (MP), "promotable" (P), "progressing" (PR), or "significant problems" (SP) (CNP, 2021). Navy RSs have a forced distribution in their promotion recommendation that disallows more than 20% of the

summary group from receiving an EP; for Lieutenant Commander (O-4, the first officer paygrade at where promotion boards occur) distributions, EP and MPs combined cannot exceed 50% of the summary group being evaluated (CNP, 2021). As Campbell and Wiernik (2015) note in their paper, using forced distributions may reward average or below average performers for their unexceptional work or neglect to reward high performers. Those high performers, such as officers who deserve an EP on their FITREP, are unable to receive one due to RS incentives to reward seniority over talent and performance (Small, 2020). Members who have scores above SGA and RSCA, and either a consistent EP or MP in their service record are in advantageous positions to earn promotions at promotion boards when compared to members with below SGA and RSCA or historic promotion recommendations of P, PR, or SP.

2. Naval PES Literature

The Navy is shifting to modernize the PES, and there is much research on the FITREP and PES that identifies key themes and issues with the way the Navy assess talent.

a. *Joshua Ellison, 2014*

Joshua Ellison (2014) compared the Navy PES with the Marine Corps PES to "help ensure optimal use of the performance appraisal to include rating accuracy and differentiation of performance" (p. 2). Ellison explains that "Navy human capital is expensive to acquire, grow, incentivize, and retain" (p. 45). He further describes that when the Navy PES fails to identify and retain the best and "most productive" officers, the Navy is wasting fiscal resources (p.45). Ellison finds that the Navy makes "seven absolute comparisons and one relative comparison" in performance evaluations (2014, p. 25). The seven absolute comparisons are the individual's scores on the seven different traits on the FITREP; the one relative comparison is the individual's average being compared to their SGA. The author describes the difference between relative and absolute comparisons as such: "The absolute method answers questions concerning the overall level of worker productivity while the relative method defines who is the most productive" (Ellison, 2014, p. 9-10). The absolute method refers to a more objective approach to see the exact level of performance an individual achieved, mirroring the discussion about accuracy in the

literature review in this thesis. The relative method is more subjective and parallels the description of precision in this thesis: how correctly can this evaluation tool order a list of individuals in order from highest to lowest performers, regardless of placing individuals at their precise level of performance.

Ellis presents some conclusions that are very relevant when it comes to the Navy PES updates. The first relevant conclusion is that the rater in the Navy PES should be given proper, detailed instruction on expectations so their "behavior is aligned with Navy manpower and budgetary strategy" (p. 2). This concept is clearly supported by Campbell and Wiernik's (2015) argument that raters must be educated on organizational goals and standards before they can precisely measure performance. Ellison's second conclusion is that comparison in the Navy evaluations "should be weighted toward relative comparisons" (p. 2). Relative comparisons rely more on precision, as discussed earlier in this thesis; an individual's exact performance level is not as important in the evaluation process as their relative levels when compared to peers.

Ellison (2014) details that the accuracy of the rater should be emphasized more in the Navy PES. Due to the ambiguity in the PES instruction, there can be a misalignment in the "RS's behavior and Navy manpower strategy" (p. 41). Ellison goes on to describe that the Navy does not have specific training that calibrates rater attitude, which plays a huge role in performance evaluations. Ellison (2014) explains that there is a lack of trust in the Navy PES grounded in a lack of transparency, a concept also supported by Small (2020). Ellison (2014) argues that the Navy should require officers "one level up the chain of command from the RS" to supervise and "monitor for grade inflation and compliance with … the instruction" (p. 41). Ellison continues with, "a decline in an individual's relative ranking should not be interpreted as a decline in performance" (p. 41). He states that if a decline in relative ranking is used as an indicator for declining performance, then RSs will be inclined to continually inflate ratings; this inflation process decreases the overall accuracy of and trust in the Navy PES (2014).

Ellison (2014) focuses on the concept of using relative comparisons in the Navy's PES; this manifests as a clear way to "provide significant improvements over performance traits to differentiate talent" (p. 42). Ellison notes that the performance traits should still be

included in the FITREP because they help create a reference line upon which feedback can be given, but that their accuracy and precision is augmented by including relative rankings as well.

Ellison closes with comments on the incentive structure seen in the Navy. He notes that simply offering bonuses to all Naval Aviators for continued service may not have an increase on productivity despite the increase in retention. If an aviator is highly motivated to take the bonus payout but not very productive in their job, then giving them a bonus to stay in the Navy is not financially aligned with Navy manpower planning goals (Ellison, 2014). To retain the best officers, Ellison suggests "pay for performance" for aviators that would offer different sized bonuses to every aviator based on their relative ranking (p. 46); this idea is only valid if the Navy implements his recommended relative rankings that can differentiate talent enough to draw decisive lines on incentive boundaries.

b.   *Laura Small, 2020*

One of the most recent studies on the Navy FITREP and PES is from a Naval Postgraduate School (NPS) thesis written by Laura Small (2020). Small analyzes the entire Navy PES to "present successful practices informed by evidence-based research" (p. 5). There are several inefficiencies in the FITREP and PES that Small identifies: the five main issues lie in "system age," "process over performance," "past versus future focus," "lack of transparency," and "inaccurate measure" (p. 26-31).

Small (2020) starts by identifying the "system age" as being problematic for the Navy PES (p. 26). The current PES employed by the Navy is overly "complex, outdated, and inefficient" (p. 27). Similarly, Burke (2017) notes that the Navy PES was developed and implemented in 1996 and uses NAVFIT98A, a system that does not connect to the internet or collect data on evaluations to help with MyNavy HR goals of using predictive analytics (Burke, 2017). General Mattis recognized this and understood the need for an updated system that enables individuals to quickly access and search their records (DOD, 2018, p.10).

The second downfall identified by Small (2020) is the Navy PES's focus on the administrative process instead of on measuring performance. Navy FITREPs are energy

draining and riddled with administrative burdens that begin "months in advance of evaluation due dates" because of how many times the comments must be iterated over (Small, 2020, p. 28). If evaluations and associated PES are to be effective, they need to require less administrative burden on the evaluator and provide instantaneous feedback to the individual (Bock, 2015). Issues with seniority being prioritized over performance are a huge factor in FITREP scores. Poulin (2016, para. 7) says that "weak performance should not be overlooked in favor of advancing a career;" placing more importance on seniority than performance on the FITREP can lead to the promotion of unprepared individuals while high performers, capable of taking on the "challenges of the next rank," must wait until they gain seniority. The management of the talent within the Navy must be centered on performance and potential, not the Industrial Age management practices that are inefficient, slow, and only interested in promotion outcomes.

Past versus Future Focus refers to how the Navy FITREP and PES look solely at past performance without regarding any potential success in future jobs (Poulin, 2016). The Navy PES is set up in a way that makes past poor performance very hard to overcome; there is an imbalance between the objectives of measuring past performance and the potential to be promoted in the future. Low ratings on early FITREPs can dominate an individual's record so much that they will be totally unable to promote later, even if they improved their performance after a poor FITREP (Small, 2020).

Transparency has long been an issue within the Navy FITREP and PES; sailors lack genuine trust in a system that has shrouded the feedback loop (Moran, 2014). Small (2020) notes that key point, and highlights that the lack of trust in the PES stems from the Reporting Senior Cumulative Average (RSCA) management done by administrative department and the reporting senior (Cordial, 2017). RSs cannot objectively rate their subordinates because they must keep their RSCA from swelling; RSCA is key to separate top tier performers from the average and below average performers at a promotion board (Small, 2020). This calls into question the effectiveness of the FITREP evaluation tool because seniors and subordinates alike know that the ratings being given may be influenced by RSCA management, not performance.

Lastly, Small (2020) acknowledges that accuracy of measure has long been an issue in the Navy PES. The Navy has implemented many talent management initiatives, as seen in Sailor 2025, since the development of the current PES. Since 1996, the Navy has updated its core values and doctrine with documents like Task Force One – Navy (TF1N) and the Signature Behaviors of the 21st Century Sailor. As these organizational values have changed, the values that sailors are being evaluated on have not. Small (2020) argues that "misalignment in organizational goals" has caused the Navy PES to measure performance inaccurately and ineffectively (p. 32). This is consistent with Boice and Kliener (1997) assertion that the FITREP and PES must be updated to reflect the Navy's new attitude towards performance management and organizational goals.

These five issues identified by Small (2020) solidify the need for the Navy to better align its evaluations with PES purposes. An outdated PES that focuses on rigid processes is unable to be flexible for the current needs of the Navy's talent management initiatives. As coaching grows more in popularity within the Navy, more attention will be focused on an individual's potential, or future performance, instead of an Industrial Age focus on the past. The Navy is modernizing its PES to make it accessible and transparent, but it is unlikely that transparency issues with the process of RSCA management will be changed soon. Once clear performance standards are set or updated for the Navy's current values, then Naval personnel can effectively be measured on their performance and ability to attain those standards.

c. *Headquarters Marine Corps (HQMC) CNA Study*

In 2012, HQMC requested a review of the U.S. Marine Corps (USMC) FITREP system from the Center for Naval Analysis (CNA) (Clemens et al., 2012). The resulting review, while focused specifically on the Marine Corps FITREP, reveals some common themes with the literature on the Navy FITREP and PES.

One main commonality between the Navy and USMC FITREP is that RSs are not continually trained or calibrated with the expectations of their ratings. Marines are trained on how to write FITREPs at The Basic School (TBS) "before they have experience with the system" (Clemens et al., 2012, p. 2). Marine officers are provided with "a handout,

three lectures, a case study homework assignment, and a discussion group" for their FITREP training (Clemens et al., 2012, p. 2). The CNA study recommends "expanding education and training on FITREPs" to solidify in officers the long-term implications the ratings they provide (Clemens et al., 2012, p. 4). TBS FITREP training also insinuates that a Marine should have rising scores over time with the same RS, even if their performance is not increasing. The CNA report acknowledges that the FITREP training is "quite appropriate and helpful" and recommends implementing "refresher training" for raters later in their careers (Clemens et al., 2012, p. 52). There is a seemingly unanimous agreement among civilian and military literature that rater training is beneficial for organizations' evaluations.

The CNA review detected a strong correlation between the performance trait ratings given to Marines and their promotion recommendations (Clemens et al., 2012). This discovery provides face and predictive validity for the performance traits used in the USMC FITREP: higher USMC performance evaluation scores are tied to higher promotability. The USMC FITREP marks "are consistent with other indicators of officer quality" (Clemens et al., 2012, p. 61).

d.   *Richard Larger, 2017*

Richard Larger (2017) analyzes the "effectiveness of the U.S. Marine Corps proficiency and conduct marks as measures of job performance for promotion decisions" (p. v). Larger's thesis is based on junior enlisted Marines but has generalizable concepts for performance evaluations.

Larger finds that while proficiency marks on evaluations "appear to be stable year to year" there is not "conclusive evidence to support this form of reliability" (2017, p. 78). The author also discovers that consistency is an issue among raters because the same Marine will receive different scores from other raters without a change in their performance, but notes that further research needs to be done to determine whether those rating differences impact Marines' promotions over time. This again ties closely to the idea that raters must be educated and reeducated on organizational values before they can

precisely measure performance, but are often not (Campbell and Wiernik, 2015). This points to the potential value of adding rater calibration as described by Bock (2015).

Larger discusses that the USMC enlisted proficiency and conduct marks captures the "same type of performance that is recognized under a different performance evaluation system" (2017, p. 78). These marks demonstrate convergent validity based on the observed correlation between the USMC proficiency marks and other performance measures. The measured "person-organization fit, physical fitness, and human capital" factors act as important signals for Marine performance.

Accuracy of graded marks is another concern within the Marine Corps PES. Larger (2017) notes that there is predictive capability between the proficiency and conduct marks on Marines' performance; although there is inflation in those marks, but it does "no apparent harm to the promotion system" (p. 79). With this precise tool, the right people end up in their correct, relative levels of performance, so the tool functions properly. The author describes the possibility that mark inflation on evaluations helps the promotion process by "better differentiation between performance levels" (p. 79). In terms of accuracy and precision, the USMC PES is precise, but not necessarily accurate.

Larger (2017) closes by detailing that the rated marks have "conflicting results" in terms of promotions (p.79). The author discovered that higher proficiency marks are predictive of future performance, but that there is a lack of consistency between USMC raters that may "lead to unfair promotions" (p. 79). Larger concludes that the proficiency and conduct marks are convergent valid since they track with other performance measures and predictive valid since they indicate promotion in the Marine Corps.

e.   *Thesis Contribution*

My thesis continues the prior relevant work evaluating military FITREPs and PES in the following ways.

First, as Ellison (2014) compares the USMC and Navy FITREPs, I use the USMC FITREP as a comparison for the proposed Navy TVS. Since the Navy proposes to change its FITREP using new TVS, my thesis continues the comparative analysis that Ellison

started. Ellison recommends many changes that I find consistent with Navy values and civilian industry literature, such as more rater training and the use of subjective measures, even though rater training and using subjective measures are not the specific focus of my thesis. My thesis uses the USMC FITREP traits as counterweights to the proposed TVS for the Navy's FITREP.

Second, similar to Small (2020), my thesis synthesizes the Navy FITREP and promotion process. Small focuses on the historical context and evolution of the Navy FITREP while my thesis builds off Small's research for the future TVS. Small also finds that there are many inadequacies in the Navy FITREP and PES, inadequacies that are being targeted by Sailor 2025 initiatives and my thesis. Accuracy of measurement on the Navy FITREP is a large issue identified by Small (2020); my thesis looks to authenticate proposed TVS and generate a validated list of these TVS that can used as accurate performance measures on the FITREP. My thesis uses Small's in-depth analysis to better understand how the new TVS can improve on the Navy PES inadequacies identified by Small (2020).

Third, the HQMC CNA study identifies the need for rater training and accountability, which my literature review supports. My thesis will act as the launching point for future research that aims to tie proposed TVS directly to Navy promotion recommendations similar to the CNA study. The goal is to have predictive validity in the proposed Navy TVS the same way as the USMC evaluations.

Finally, Larger (2017) finds that USMC proficiency and conduct marks also track closely with other performance measures in the Marine Corps. In the same way, my thesis compares the proposed Navy TVS, some of which may end up on the Navy evaluation tool, against military service evaluations to determine whether the TVS associate with other performance assessment instruments. Since USMC proficiency and conduct marks correlate well with promotions and other performance measures, my thesis uses this as the conceptual basis for the textual comparison between proposed Navy TVS and expected sailor performance as dictated by military doctrine. If sailors live up to the standards set by Navy doctrine, they will receive higher marks on the TVS excluding adverse impact.

3.   Navy Promotion Boards

The Navy officer promotion process is a bureaucratic and well-structured process that is chiefly centered on the individual who is eligible for promotion and their tracked FITREP averages over time (B. Baran, email to author, December 29, 2021). Essentially, board members see how individuals' FITREP averages and promotion recommendations change throughout their career. The subsequent paragraphs will discuss, in-depth, how promotion boards operate as outlined by the NPC Active Promotion Brief and SECNAVINST 1420.3.

A large portion of the Navy promotion board process hinges on the creation of "zones" (Office of the Secretary of the Navy [SECNAV], 2010). Zones are established by the SECNAV, not the promotion board, and are based on members' eligibility for promotion, whether they have failed to select for promotion to that rank already, and the distribution of officers at the promotion board. In the 1420.3, an officer is "below-zone" (BZ) when they are eligible for the next promotion but still junior to their peers at the board; by not being promoted as a BZ candidate, the member's career will not be hindered. BZ candidates that are not selected for promotion do not incur a "failure of selection" on their record (B. Baran, email to author, December 29, 2021). Being at a promotion board as a BZ candidate is impressive since the member is junior to their peers, but at or above standards for performance measures. In-Zone, or IZ, is when a member is eligible for promotion, is senior to the BZ boundaries and has not failed to select at a promotion before. This is the most common category that members fall in at a promotion board since BZ candidates are allocated fewer promotion selections; only 10% of selections are permitted to be BZ candidates (B. Baran, email to author, December 29, 2021). Above-Zone (AZ) is a more difficult category to get selected for promotion, because the individuals here are eligible for promotion, senior to the senior most IZ officer, but have previously failed to select for promotion to that rank before (SECNAV, 2019). AZ candidates have already failed to select for promotion while being IZ and are typically less likely to select on their second or third try.

The promotion board receives the list of members eligible for promotion so the board can create lists of members in the Navy Selection Board System. Board members

receive a convening order and must swear an oath before they take part in the promotion process (B. Baran, email to author, December 29, 2021). The records for each eligible officer are distributed randomly to board members and they "begin to review and grade all AZ and IZ records." The board members are required to review and analyze member FITREPs, or evaluations, and any other applicable items in that officer's official service record (OSR). It is at this point where board members take notes on promotable members, how their scores compare to their respective RSCA, and their CO's promotion recommendation. The board members compare the individual's RSCAs with other promotable individuals' RSCAs to validate scored averages and to rank individuals for promotion (B. Baran, email to author, December 29, 2021).

The Navy promotion board, "the tank" is where voting occurs (B. Baran, email to author, December 29, 2021). The board members are briefed on individuals and vote on who will get tentatively selected for promotion and who will get dropped from further consideration based on scores from the board members. These "voting motions" are where the board members provide confidence scores, either 100, 75, 50, 25, or 0, with a remote, to grade individuals up for promotion (B. Baran, email to author, December 29, 2021). 100 relates to high confidence in that individual's promotion and 0 represents low confidence. The board members vote based on their analysis of the individual: historic RSCA, FITREP promotion recommendations, and other briefed items in the individual's OSR are all factored into each board member's decision. If an individual's average score from the board members is above a certain score, the individual is tentatively selected for promotion. If an individual is below that certain score, but above the minimum threshold for being not selected, they are recycled and the records will be "crunched" by the board members further (B. Baran, email to author, December 29, 2021). Individuals that fall below the established minimum threshold for non-selects will not be selected for that promotion. AZ and IZ individuals dropped from consideration will receive "failure to select" in their records, which will impact their chances at promotion the next time they are up for promotion. BZ individuals do not receive any "failure to select" marks in their OSR if they do not get selected for promotion.

The board does this in the same room, at the same time, to foster transparency and trust in the system. While each board member provides their own unique scores for every individual, the fact that all the board members do this together closely mirrors Laszlo Bock's calibration concept introduced earlier in this thesis. The board forwards their findings to Office of the Secretary of Defense for final approval; the selected individuals will then promote throughout the next fiscal year as announced by monthly Naval Administrative messages (B. Baran, email to author, December 29, 2021).

The FITREP is the cornerstone of the Navy promotion process. While the promotion board does not crunch each FITREP from an individual up for promotion, the board does compare the individual's RSCA and promotion recommendation histories to determine promotability. In this way, a "single unfortunate FITREP" can lead to "pitfalls... where good performers miss promotion opportunities" (Small, 2020, p.31) The FITREP is an important evaluation instrument that the Navy must study and iterate upon to maintain effectiveness in accurate performance measurement and proper selection for promotions.

E.   RESEARCH OBJECTIVES AND SUMMARY

One purpose of this thesis is to provide an in-depth synthesis of the reasons for, importance of, errors within, and types of performance evaluations to better frame the Navy's current PES and proposed TVS. Performance evaluations are integral portions of managing personnel effectively and creating environments that increase employee productivity and engagement; the Navy operates in an internal labor market and must retain its best and brightest sailors to continue to combat new and emerging threats in the world. The Navy evaluation process is bureaucratic and rigid but must be flexible to adapt to new value systems. As the values of the Navy evolve, so should the tool with which it evaluates its sailors (Small, 2020). This thesis surveys the theoretical constructs for conducting performance evaluations to make a recommendation on further research before the Navy implements any updates to its PES. I want to focus the Navy's attention on updating its PES and evaluation tool to be more accurate and precise, which means better aligning it with new organizational and PMS goals. The Navy's PES leaves room for subjective rating errors; from calibration of individual scores early in the evaluation phase to crunching

records in the promotion board phase, there are many times that biases can negatively impact the integrity of the process. After reviewing PES literature, simply changing the structure and accessibility of the Navy PES is not enough. The Navy must change what it evaluates its sailors on, so those graded items are consistent with Navy doctrine. I hope to encourage a productive conversation about the Navy's new talent management initiatives that can be further developed to create a more effective evaluation tool.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

A. OVERVIEW

In 2019, the Navy assembled personnel and performance evaluation Subject Matter Experts (SMEs) to brainstorm new trait and value statements to be used on the new FITREP. This conference was called the Naval Personnel Command (NPC) PERS-3 Performance Evaluation Transformation-Talent Management (PET-TM), and is referred to as the "TVS Summit" in this thesis.

The SMEs at the summit identified 8 broad character traits with 39 associated sub-traits, which were further broken down into 2–5 value statements per sub-trait (CNP, 2020a). This nested structure for the 8 broad character traits, 39 sub-traits, and 82 value statements is displayed in Table 2. The 8 broad traits that the SMEs determined valuable, which will be further studied in this thesis, are: Character, Leadership, Initiative and Drive, Teamwork, Communication, Critical Thinking and Decision Making, Resiliency and Toughness, and Mission Accomplishment and Productivity. Each of these traits represents some aspect of key behaviors in the Navy, corresponding to Sailor 2025 initiatives or other Navy Doctrine. For the sub-traits, SMEs focused on describing ways that the broad traits could be assessed from "observable behavior" (CNP, 2020a).

SMEs were briefed on the focus of turning the FITREP and PES into personnel coaching tools, and as a result, the language in the TVS was purposely developmental and included many parallels to civilian research on talent management (CNP, 2020b).

My thesis validates the TVS through cross-examination with other Navy and military documentation, described later in this section. As Boice and Kleiner (1997) discuss in their paper on performance evaluations, the first step in creating a PES in an organization is to determine the values of the organization and associated expectations of employees therein. To validate the Navy's long list of proposed TVS, the TVS must be cross-referenced with Navy-specific doctrine, other military service officer evaluations, and future data and research completed by the Naval Postgraduate School research team when it becomes available. The Navy documentation that was analyzed for similarities with the

proposed TVS are as follows: Task Force One – Navy, Hard Truths and the Duty to Change: from the Independent Review Commission on Sexual Assault in the Military, Navigation Plan 2021 (NAVPLAN), Signature Behaviors of the 21st Century Sailor, and the officer evaluations from the Marine Corps, Army, and Air Force.

There are many Navy and other military documents that describe what the Navy values; this study uses textual analysis as the methodology to identify similarities across the different references. I describe these documents in detail below. I use these Navy and other military documents to validate the proposed TVS. My textual analysis consists of searching the other documentation for key phrases from the TVS sub-traits. If the sub-trait is not validated by exact word matches from the other literature, textual analysis is then conducted on keywords in the value statements. Since the value statements describe the sub-traits in action, key phrases in the value statements are used as validation tools for that sub-trait. If the sub-trait is explicitly listed in other documentation, cross-referencing the value statements is not necessary. If a sub-trait was validated by a specific Navy document or other service evaluation tool, it received an X in the column associated with that specific document. I use this method to compare every TVS sub-trait with every military document studied in this research.

Table 2.   List of TVSs 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
		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

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
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 accomplishment of shared goals. Includes collaboration with others, inside and outside of the organization.	Team Pride	Demonstrates inclusion and actively supports teamwork and team spirit
		Supports unit cohesion
	Relationships	Develops productive working relationships
		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	Risk Assessment	Assesses risk throughout implementation of a course of action

Trait and Definition (9)	Sub-Trait (39)	Value Statements (82)
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.		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
		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.   **Task Force One–Navy**

Task Force One–Navy (TF1N) is a report that provided 60 recommendations "meant to enhance the Navy's overall diversity and ensure that a culture of inclusivity is evident at every command" (Faram, 2021, para. 2). It is important to note that TF1N took

feedback from hundreds of sailors in focus groups and surveys as well as scanned 6 Navy instructions to decide what initiatives were most valid and necessary to advocate for.

I identified the TF1N report as useful and applicable as a cross-reference with the proposed TVS because the new evaluation items should uphold principles of Equal Employment Opportunities (EEO). Language in the proposed TVS should not "contain offensive, biased, or hampered" phrases as determined by TF1N; the proposed TVS should include values and verbiage that maintain EEO (Faram, 2021). MC1 Faram also noted that TF1N found the Navy enlisted ranks is very diverse, and closely mirrors American societal diversity when looking through the race and ethnicity lens. To validate EEO initiatives within the Navy's proposed TVS, it is imperative to consider the clearest and most recent research done on diversity and inclusion in the TF1N report.

2. Hard Truths and the Duty to Change: from the Independent Review Commission (IRC) on Sexual Assault in the Military

The IRC on Sexual Assault in the military was a 90-day evaluation of sexual assault in the military to find and root out causes of harassment. The IRC makes observations and recommendations for the military to put into action; the recommendations are centered on "accountability, prevention, culture and climate, and victim care and support" (Rosenthal et al., 2021, p. 3). With the focus of these recommendations undoubtedly arranged around the actions of the individual service-member, I deemed it necessary to cross the proposed TVS with the IRC on Sexual Assault. The Department of Defense, at the direction of the President of the United States, emphasizes the need to improve and promote healthy work environments by fostering cultures of dignity and respect; the best way to create these positive environments is by holding individual members to the highest standards of personal conduct, and by evaluating their performance with the results and goals of the IRC on Sexual Assault in mind.

3. NAVPLAN 2021

The Chief of Naval Operations (CNO) released the 2021 NAVPLAN with the focus on 4 main priorities: readiness, capabilities, capacity, and sailors. Readiness refers to fleet and personnel funding, something out of the control of the rank-and-file sailor. The Navy

must not only have ready ships, submarines, and aircraft, it must have a force of sailors that are proficient and ready to dominate all warfare domains (Office of the Chief of Naval Operations [CNO], 2021). Equipment must be ready, and the sailors using that equipment must be trained to exert control over the battlespace. The capabilities priority focuses on the development of lethal and non-lethal weapon systems that can change the course of a conflict, not the capabilities of sailors. Capacity involves the seamless integration of unmanned systems with available talent; without identifying top performers capable of operating these advanced systems, the Navy will not be able to compete in future conflicts. The fourth and final priority is arguably the most important: sailors. The CNO has focused on recognizing and empowering the Navy's sailors to be the best assets they can be. CNO doubles down on upholding equal opportunity and living up to the Culture of Excellence by taking ownership of the Navy's mission and core values (CNO, 2021). The CNO highlights Honor, Courage, and Commitment as our core values while adding the attributes of Integrity, Accountability, Initiative, and Toughness; all of which show up in the proposed TVS. The NAVPLAN is released yearly and is the direction that the CNO is driving the Navy, so any improvements to performance evaluations, which identify top performers in what the Navy values most, should reflect the CNO's NAVPLAN.

4. Signature Behaviors of the 21st Century Sailor ("10 Signature Behaviors")

The Signature Behaviors of the 21st Century Sailor, hereafter referred to as "10 Signature Behaviors," is a document released by the CNO in 2020 with the goal of setting "objectives, programs, and policies" to maximize Sailor performance (Fleet Forces Command, 2020). This document is deeply intertwined with the proposed TVS because the SMEs at the TVS Summit were briefed on the 10 Signature Behaviors before determining what they thought valuable and measurable traits in the Navy (CNP, 2019). The two documents have a circular relationship where the results of the TVS Summit are directly related to the 10 Signature Behaviors since the 10 Signature Behaviors acted as the foundation for SMEs to launch forward with their research. Appendix A shows the 10 Signature Behaviors in their entirety; each signature behavior is elaborated on with its relationship to Navy Core Values, Core Attributes, and paired with several "well will"

statements that act as promises for sailors to uphold (Fleet Forces Command, 2020). Lastly, the signature behaviors have accompanying "in action" statements that show what the specific signature behavior looks like in practice; this portion is a borderline perfect impersonation of the "observable behaviors" value statements that SMEs came up with for in the proposed TVS (CNP, 2020b). The 10 Signature Behaviors is the outline for how the proposed TVS were conceptualized; cross-validation between the 10 Signature Behaviors and proposed TVS yielded a near-perfect match due to the interwoven nature of the two documents.

B.   TVS MEETS OTHER SERVICE EVALUATIONS

The Navy has many similarities with the other warfighting organizations in the United States, including the Army (USA), Air Force (USAF), and Marine Corps (USMC). These branches all operate under the Department of Defense and have similar roles and general responsibilities in the United States of America. I use detailed searches to locate similarities between the Navy's proposed TVS and officer evaluations from the USMC, USAF, and USA. I examine the documents in the same cross-validation effort to find specific words and phrases within these other military branches' officer evaluations that closely resemble the proposed Navy sub-traits. Results of this cross-textual analysis follow in the next Chapter.

IV. RESULTS AND ANALYSIS

Under the definition of construct validity discussed in Chapter I, I validate the proposed TVS developed by Navy SMEs. I use cross-textual analysis to verify if TVS are consistent with the Navy's organizational goals and values and other services' documents. The results of this cross-examination are shown in Table 3, which reports the frequency that each trait is referenced across documents. The sub-traits in Table 3 are organized in order from most valid to least valid, in terms of how frequently they show up in the other Navy doctrine and officer military evaluations. The second column, "10 S.B." refers to the Signature Behaviors of the 21st Century document and is abbreviated to make the table easier to read. Across all columns, sub-traits receive a "1" if they appear in the corresponding document; sub-trait cells are left blank if it is not validated by the other military documents. Frequencies are totaled across the columns and the resulting sums are seen in the last two columns on the right. The column named "Total, Navy only" refers to the sum of all the validated TVS based on Navy doctrine alone. The "Total, ALL" column shows the sum of the TVS frequencies based on Navy doctrine and other service evaluations. While the IRC on Sexual Assault does not belong to any specific service branch, it is included in the "Total, Navy only" column.

A. NAVY VALIDITY

Thirty-three of the 39 TVS sub-traits are validated by at least half of the other Navy doctrine reviewed in this study. However, the sub-traits that are most strongly supported by the Navy doctrine are wellness, ethics, personnel development, innovation, relationships, and feedback. All four "Navy only" doctrines validate these six sub-traits, which are clearly focused on the individual characteristics and interpersonal interactions of the sailor or officer. Wellness, ethics, and relationships all show that the individual must create a healthy and safe work environment and always do the right thing for their subordinates. Personnel development and feedback both contain language that highlights the importance of communicating performance standards and providing effective feedback when subordinates do not achieve these standards. Innovation

describes how an individual adapts to and overcomes situations by taking flexible approaches to solving problems. These six traits are very focused on the individual and their personality, not necessarily their job performance. The least valid traits, as detected in the cross-validation effort, are independence, comprehension, persistence, time management, volunteering, and non-verbal. The first five sub-traits in this list are only supported by a single piece of Navy doctrine; non-verbal [communication] is not supported in any of the Navy literature reviewed in this thesis. The Navy is shifting performance standards from specific work output metrics to coaching and development. This shift is seen clearly in the results of cross-referencing the TVS sub-traits with the Navy doctrine viewed in this study.

B. OTHER SERVICE VALIDITY

Table 4 shows a simplified version of the relationships between the officer performance evaluations of the Army, Air Force, and Marine Corps and the proposed Navy sub-traits. The table in Appendix B shows the full table with detailed references to where the matched words or phrases manifested in the other service evaluations. It is important to note that 18 of the 39 TVS sub-traits have exact matches to all three of the other services' officer evaluations. These matches indicate that there are many similarities between the values of the four service branches, but also significant differences. 18 of the proposed sub-traits were validated by just two of the other branches of service; two of the proposed sub-traits were validated by one other service branches' officer evaluation, and only one proposed Navy sub-trait (volunteering) is completely unique to the Navy's proposed sub-traits. Volunteering does not show up in any of the other military branches' officer evaluations. This is not to say that volunteering is not important across all the services, but it does show that volunteering is not valued highly in officer performance evaluations or promotions.

C. WELLNESS

The results in Table 3 show that "wellness" is unilaterally seen as an integral piece of performance; this is supported by all Navy doctrine and all other service evaluations reviewed in this literature. Wellness falls under the broad trait of

'leadership' and is described by the value statements, "recognizes and addresses signs of stress in others" and "guides others to seek support through available wellness resources" (CNP, 2020a). It is clear, across the board, that the ability to recognize stress and assist others in their struggles is critical to leadership and performance in the military.

D.   CLOSING

As stated before, most of the highly ranked and validated TVS relate to constructs focused on interpersonal interactions. The following 23 sub-traits are validated by 6 or more pieces of other studied doctrine: wellness, ethics, personnel development, innovation, relationships, feedback (from the broad trait leadership), inclusion, integrity, professionalism, feedback (from the broad trait communication), personal development, listening, clarity, problem solving, adaptable, and professional competence. Personnel development, personal development, and professional competence are the only sub-traits that contain language relating to job performance; this reduction shows the shift within the military towards concentrating more attention on personal interactions when it comes to leadership. It is apparent that the entirety of the military is leaning towards evaluating individuals on how well they work with others, not simply how well they perform their job roles and responsibilities.

Based on the results in Table 3, more than 90% (36 of 39) of the developed TVS have support from at least four of the studied Navy doctrine or other service evaluations. This research has also uncovered a few sub-traits deemed most valuable based on the examined doctrine. The TVS Summit and the Post-Summit Validation are used to generate and formalize values considered to be important in fostering a culture of coaching and personnel development in the Navy.

Table 3.   TVS Cross-Validation with Documents

SUBTRAITS	10 S.B.	TF1N	NAV PLAN 2021	IRC	USAF	USMC	USA	Total, Navy only	Total, ALL
Wellness	1	1	1	1	1	1	1	4	7
Ethics	1	1	1	1		1	1	4	6
Personnel Development	1	1	1	1		1	1	4	6
Innovation	1	1	1	1		1	1	4	6
Relationships	1	1	1	1		1	1	4	6
Feedback	1	1	1	1	1		1	4	6
Inclusion	1	`	1	1	1	1	1	3	6
Integrity	1	1		1	1	1	1	3	6
Professionalism	1	1		1	1	1	1	3	6
Feedback	1	1	1		1	1	1	3	6
Personal Development	1	1	1		1	1	1	3	6
Listening	1	1		1	1	1	1	3	6
Clarity	1	1		1	1	1	1	3	6
Problem Solving	1	1	1		1	1	1	3	6
Adaptable	1	1		1	1	1	1	3	6
Professional Competence	1	1		1	1	1	1	3	6
Respect	1	1		1	1		1	3	5
Goals and Expectations	1	1	1			1	1	3	5
Team Pride	1	1		1		1	1	3	5
Contribution	1	1	1			1	1	3	5
Information Sharing	1	1		1		1	1	3	5
Recovery	1		1	1		1	1	3	5
Responsibility	1	1			1	1	1	2	5
Motivation	1	1			1	1	1	2	5
Planning	1		1		1	1	1	2	5
Conflict Management	1	1		1			1	3	4
Moral Courage	1			1		1	1	2	4
Delegation	1	1				1	1	2	4
Change Management	1	1				1	1	2	4
Risk Assessment	1		1			1	1	2	4

SUBTRAITS	10 S.B.	TF1N	NAV PLAN 2021	IRC	USAF	USMC	USA	Total, Navy only	Total, ALL
Evaluates Alternatives	1	1				1	1	2	4
Quality and Attention to Detail	1	1				1	1	2	4
Independence	1				1	1	1	1	4
Comprehension	1				1	1	1	1	4
Persistence	1				1	1	1	1	4
Time Management			1		1	1	1	1	4
Coping	1			1		1		2	3
Non-verbal						1	1	0	2
Volunteering	1							1	1

Adapted from Air Reserve Personnel Center (2015), Chief of Naval Operations (2021), Department of Defense (2021), Department of the Army (2019), Headquarters, United States Marine Corps (n.d.).

Table 4.   Sub-trait Cross-Validation with the USMC, USAF, and USA
Officer Evaluations

PROPOSED SUB-TRAITS	USMC VALID	USAF VALID	USA VALID
Responsibility	X	X	X
Ethics	X		X
Integrity	X	X	X
Respect		X	X
Moral Courage	X		X
Professionalism	X	X	X
Goals and Expectations	X		X
Personnel Development	X		X
Feedback	X	X	X
Inclusion	X	X	X
Wellness	X	X	X
Delegation	X		X
Motivation	X	X	X
Change Management	X		X
Innovation	X		X
Personal Development	X	X	X
Independence	X	X	X
Volunteering			
Team Pride	X		X
Relationships	X		X
Contribution	X		X
Listening	X	X	X
Comprehension	X	X	X
Clarity	X	X	X
Non-verbal	X		X
Feedback		X	X
Conflict Management			X
Information Sharing	X		X
Risk Assessment	X		X
Planning	X	X	X
Evaluates Alternatives	X		X
Problem Solving	X	X	X
Coping	X		
Persistence	X	X	X
Recovery	X		X
Adaptable	X	X	X
Professional Competence	X	X	X
Time Management	X	X	X
Quality and Attention to Detail	X		X

V. SUMMARY AND RECOMMENDATIONS

A. SUMMARY

I pose and answer three questions in this thesis. The first of these is:

(1)   How do the evaluation and promotion processes work in the Navy and how do they best align with practices among civilian organizations and other military branches?

The Navy has well-defined and bureaucratic systems for performance evaluation and promotions that are described in the BUPERISNT1610.10F and SECNAVINST1420.3 respectively. The 1610.10F lays strict guidelines for how the Navy expects performance appraisals to take place. The 1420.3 outlines, in detail, the entire conduct of promotion boards. These rigid Navy processes are contrasted with more agile systems seen in the civilian sector. Google and Deloitte perform evaluations at similar intervals as the Navy but employ more subjective measures that solicit coaching conversations; the coaching and developmental style is centered on the individual's performance and impact on others. The Navy is interested in using a coaching model for performance appraisal which can be seen in the shift to the proposed TVS. For the civilian sector and the Navy, promotion outcomes rely heavily on individuals' performance evaluations. While the bureaucracy of Navy evaluations and promotions is not necessarily a negative trait, the Navy must enact some degree of flexibility to maintain accuracy and effectiveness of performance measurement and talent management.

The second research question I pose is:

(2)   How are the newly developed FITREP Trait and Value Statements supported by current Navy doctrine and other service branches?

To answer this question, I examine the following documents for construct validity: 10 Signature Behaviors, NAVPLAN 2021, IRC, TF1N, and the officer evaluations of the Marine Corps, Army, and Air Force. These Navy documents, as well as other service officer evaluations, are compared at length in a cross-textual analysis to determine the validity of the Navy's new TVS sub-traits.

The majority of the Navy's proposed TVS are heavily supported by other Navy doctrine and other military service officer evaluations. Each of the 39 sub-traits is verified by at least one of the documents reviewed in this thesis; 36 of the 39 traits are validated by half of the documents used in the cross-validation effort. This clearly indicates that what the Navy has developed for new TVS is valid amongst other Navy and military literature at a first glance.

The third research question I pose is:

(3)   How can the proposed Trait and Value Statements be evaluated to effectively assess talent within the Navy?

To effectively evaluate and assess talent within an organization, graded items on performance appraisals should mirror what the organization values. Therefore, the FITREP TVS must reflect what the Navy wants to measure, or values within its organization. As the Navy's values shift towards coaching and development of the individual, the graded FITREP TVS should mirror that shift. The proposed TVS are construct valid if Navy documentation supports them; this is visible in the cross-referencing of Navy doctrine with the TVS to find pinpointed matches in wording and phrasing. The TVS are, for the most part, valid within Navy doctrine but must be further studied to determine predictive and convergent validity.

B.   RECOMMENDATIONS FOR FUTURE RESEARCH

The Navy must maintain an effective fighting force to combat new and emerging threats in the world by retaining its most lethal weapon: the sailor. High-quality sailors and officers that bring out the best in others will enable the Navy to compete in an increasingly complex battle environment. Providing for the common defense starts with presenting sailors and officers clear sets of expectations for their duties so they know what it takes to be a high performer. Sailor 2025 has taken large steps to modernize the Navy PES; these efforts should be continued with focused effort and haste to prevent the loss of highly skilled sailors who end their Naval service due to an outdated and ineffective performance evaluation system.

Further performance evaluation studies should consolidate what the Navy expects from officer performance. Once a single document, or set of PES-specific documents, has been established, the Navy can develop, test, and study exhaustive surveys, of which TVS seem to be most important. The TVS Summit is a great launching point for the creation of a consolidated value system that is consistent with Navy and other military services' documentation. Extreme significance should be placed on forming a single document that guides what is expected of officer performance before any changes can take place within the evaluation tool itself.

Currently, the Navy is refurbishing its entire PES to make it easier to access, but it also needs to consider what it is valuing and how those values align with the described theories and purposes for evaluations. Bock (2015) notes that widespread changes to organization PES require "Herculean efforts" (p. 159). The Navy is making large shifts from Industrial Age mentalities to a much more modern mindset that emphasizes development and coaching. People will scream, people will cry, and people will nearly quit just like they did at Google (Bock, 2015). People often resist change in PES because they place such high value on evaluations and they think that a new PES might negatively impact their promotion chances.

The list of proposed TVS, however, must be studied and merged to fit a survey that can gain valuable insight on the future of the Navy PES. This survey must be narrowed down from the 82 TVS so sailors are willing to participate and provide valuable data. Once data is received and properly analyzed, the Navy can better determine which TVS are valid and must be evaluated on the new FITREP.

On creating the perfect evaluation tool, Bock says, "consensus is impossible" and "we had to develop our own [evidence], working with the leaders of each part of Google to help them test their ideas" (Bock, 2015, p.159). Further research should be conducted to discover evidence for specific values within the Navy; this research should be conducted on and with common Naval personnel, not only high-level admirals that focus on the PES model that they are used to. Performance is impossible to measure without performance standards to measure individuals against.

This thesis has only scratched the surface on the background of what traits from the TVS list *could* be included in the Navy's new evaluation tool based on cross validation with Navy doctrine and other services' evaluations. The next step is to take the proposed TVS list and compare it to overlooked (but important) Naval service doctrine to determine if there are any TVS that do not fit in the Navy's evaluation, or if there are any important traits that were missed or omitted from the TVS that should be included in the new evaluation tool. Once the researchers have narrowed the list down to a reasonable number, a survey should be developed and fielded to a representative sample of sailors and officers, not just admirals. This survey should be tested against the current FITREP to ascertain construct validity and reliability in the new traits being evaluated. The outcome of this survey would be data that estimates how much better (if at all) the new traits are at predicting promotion in the Navy.

Other consideration should be paid towards deciding if the FITREP will be used as a coaching tool or remain solely a performance evaluation measure. If the new traits will factor into a performance evaluation tool only, the traits should relate to officer performance, and the document can look similar to the current FITREP format. If the new FITREP will be used as a coaching tool, the Navy will need to revamp the entire PES, from the evaluation forms to the entire structure of the promotion system. There is currently a mismatch of what is expected of officers and the attributes on which officers are rated. Once expectations of officers are laid out more clearly and concisely, the new FITREP traits will be easier to establish. Actionable feedback, however, is one critical, and missing, piece of the Navy's performance evaluation system. When the Navy decides to get serious about evaluating performance and developing sailors, it will take a page out of Laszlo Bock's book and separate the rating and grading portions from the coaching and developmental portions of the PES.

APPENDIX A. SIGNATURE BEHAVIORS OF THE 21ST CENTURY SAILOR

Table 5.   10 Signature Behaviors of the 21st Century Sailor. Adapted from Fleet Forces Command (2020).

10 Signature Behaviors	Core Value	Core Attribute	Core Attribute (cont.)	We will	In Action (1)	In Action (2)	In Action (3)	In Action (4)	In Action (5)	In Action (6)
Treat every person with respect	Honor	Integrity	Our behaviors as individuals and as an organization align with our values as a professional	Conduct ourselves in the highest manner in all relationships with peers, superiors, and subordinates. Be honest and truthful in our dealings with each other and with those outside the Navy. Be willing to make honest recommendations and accept those of junior personnel. Respect others as a fundamental tenet of our character	"I will engage others with honesty and integrity and will proactively intervene or take action when witnessing those not doing so."	"I will listen to people and solicit, welcome, and respect their contributions."	"I will make honest recommendations and respect those of junior personnel."			
Take responsibility for my actions	Honor	Integrity, Initiative, Toughness	We achieve and maintain high standards	Abide by an uncompromising code of integrity. Take responsibility for our actions and keeping our word. Fulfill or exceed our responsibilities in our public and personal lives 24 hours a day.	"I will not tolerate illegal or improper behavior or the appearance of such behavior and I will proactively counter and report this behavior when I witness or learn of it."	"I will do the right thing whether in uniform or not, whether I am on the job or not."	"I will wear my uniform with pride and represent the Navy in my community and at ceremonies with professionalism and respect."			
Hold others accountable for their actions	Courage	Accountability, Toughness	We are a mission-focused force	Make decisions in the best interest of the Navy and the nation, without regard to personal consequences. Meet the demands of our profession and the mission when it is hazardous, demanding, or otherwise difficult. Reduce the likelihood and acceptance of inappropriate behaviors. Hold ourselves to the highest of personal and professional standards of thought and behavior. Be loyal to our nation, ensuring the resources entrusted to us are used in an honest, careful, and efficient way.	"I will hold Shipmates accountable for infractions even if these Shipmates are my buddies."	"I will not use sexist/ racist comments or jokes and will correct others when they use them."	"I will 'take a stand and intervene when Shipmates are acting or treating others inappropriately."			
Intervene when necessary	Courage	Accountability	Trust and confidence is enhanced by our actions	Correct inappropriate actions of peers, even if unpopular. Alert others when dangerous situations arise. Ensure small infractions are corrected so they don't accumulate into complacency or lead to larger problems. Muster the strength to do what is right, even in the face of personal or professional adversity. Respond to situations in a hostile environment in a measured manner	"I will confront and correct discriminatory behavior when it occurs."	"I will stop hazing and bullying and any other inappropriate behavior if I see it."	"I will intervene to protect my Shipmates from situations that may lead to sexual assault."			

10 Signature Behaviors	Core Value	Core Attribute	Core Attribute (cont.)	We will	In Action (1)	In Action (2)	In Action (3)	In Action (4)	In Action (5)	In Action (6)
Be a leader and encourage leadership in others	Commitment	Accountability, Initiative	Our leaders will take ownership and act to the limit of their authorities	Exhibit the highest degree of character, technical excellence, quality, and competence in what we have been trained to do. Demand respect up and down the chain of command. Care for the safety and the professional, personal, and spiritual well-being of our people. Challenge, question, and be open to change. Communicate well and regularly.	"As a leader, I will demonstrate that navigating through tough decisions with integrity — even when it isn't the popular thing to do — it's always the right thing to do. I will be courageous in conflict."	"As a leader, I will admit my mistakes. In this way, I will show those whom I lead the value of personal integrity and responsibility."	"I will exert peer leadership when necessary to ensure my shipmates successfully navigate through tough decisions, even when it isn't the popular thing to do."			
Grow personally and professionally every day	Commitment	Initiative	On their own, everybody strives to be the best they can be - we give 100% when on the job	Make decisions that reflect the Navy's Core Values of Honor, Courage, and Commitment and contribute to a constructive outcome for us, others, and the Navy. Be committed to positive change and constant improvement. Create an environment that nurtures professional growth and confidence. Live day-to-day with a sense of duty that all Sailors should work together as a team to improve the quality of our work, our people, and ourselves. Choose excellence in all professional and personal commitments	"I will personally mentor a peer/ subordinate who has fallen short of their professional and/or personal goals."	"I will challenge others to meet or exceed my performance on the physical readiness test."	"I will seek advancement opportunities and increased responsibilities whenever they are available"	"I will continue to seek education and training opportunities whenever they are available."	"I will represent my Navy with pride in my appearance, character, and manner."	"I will work hard, do my best, and put 100% effort into each task and goal."
Embrace the diversity of ideas, experiences, and backgrounds of individuals	Commitment	Initiative	Our most junior teammate may have the best idea; we must be open to capturing that idea	Practice inclusion and value diversity. Demonstrate fairness in rewards and promotions. Assist others in understanding equal opportunity in the Navy.	"I will encourage cross-cultural understanding and learning."	"I will not allow Shipmates to be treated unfairly because of race, religion, gender, sexual orientation, or gender identity."	"I will support DOD policy to treat all Service members with professionalism, dignity, and respect."	"I will support cultural and gender diversity, diverse backgrounds, and viewpoints to enhance our readiness and mission effectiveness."		
Uphold the highest degree of integrity in professional and personal life	Integrity	Integrity, Accountability	We actively strengthen each other's resolve to act consistently with our values	Reject illegal or destructive behavior and the appearance of such behaviors. Uphold responsibility and accountability for actions. Speak up for what is right. Live Navy Core Values, Ethos, and Core Attributes.	"I will lead by example, striving for personal and professional excellence in all situations in a manner that brings credit upon me, my Shipmates, my command, and my Navy."	"I will remain loyal to my Navy, and when I see illegal or destructive behavior taking place, I will not tolerate it.				
Exercise discipline in conduct and performance	Discipline	Initiative, Toughness	We foster a questioning attitude and look at new ideas with an open mind	Dedicate ourselves to personal and professional growth. Maintain self-control and sound judgment in thoughts, speech, and actions. Maintain a balance of mental, physical, and spiritual readiness	"I will strive for a high level of professional and operational knowledge, physical and mental fitness, and spiritual strength."	"I will strive for improvement and excellence in all that I undertake."	"I will exercise self-control in my thoughts, speech, and actions."			
Contribute to team success through actions and attitudes	Teamwork	Integrity	As individuals, as teams, and as a Navy, our conduct must always be upright and honorable both in public and when nobody's looking	Dedicate ourselves to Shipmates and the Navy, above ourselves. Appreciate diversity and value the contributions of all. Foster an inclusive environment. Be dependable and reliable to those who are counting on us. Invest personal efforts into team outcomes.	"I will uphold the success of my team, my unit, and the Navy above my personal ambitions."	"I will build a culture of inclusion, considering and valuing the unique and diverse contributions of all team members."	"I will be ready for every team event, whether daily work or special evolutions — physically, mentally, and professionally prepared."	"If I need physical, mental, or spiritual assistance to successfully support my team, I will seek help."		

APPENDIX B. OTHER SERVICE SUB-TRAIT VALIDATION (WITH DIRECT REFERENCES)

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
CHARACTER				
Responsibility	X	X	X	USMC FITREP - individual character, courage. USAF Form 707 - Professional Qualities. USA DA Form 67–10-1A - Leads.
Ethics	X		X	USMC FITREP - leadership, setting the example. USA DA Form 67–10-1A - Leads.
Integrity	X	X	X	USMC FITREP - leadership, setting the example. USAF Form 707 - Professional Qualities. USA DA Form 67–10-1A - Character.
Respect		X	X	USAF - AF Form 707, leadership skills. USA DA Form 67–10-1A - Character; Leads; Develops.
Moral Courage	X		X	USMC FITREP - individual character, courage. USA DA Form 67–10-1A - Leads.
Professionalism	X	X	X	USMC FITREP - leadership, setting the example. USAF Form 707 - Professional Qualities. USA DA Form 67–10-1A - Presence; Leads.

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
LEADERSHIP				
Goals and Expectations	X		X	USMC FITREP - individual character, initiative. USA DA Form 67–10-1A - Leads; Develops.
Personnel Development	X		X	USMC FITREP - leadership, developing subordinates. USA DA Form 67–10-1A - Develops.
Feedback	X	X	X	USMC FITREP - leadership, communication skills. USAF Form 707 - Leadership Skills. USA DA Form 67–10-1A - Develops.
Inclusion	X	X	X	USMC FITREP - leadership, ensuring the well-being of subordinates. USAF Form 707 - Leadership Skills. USA DA Form 67–10-1A - Develops.
Wellness	X	X	X	USMC FITREP - leadership, ensuring the well-being of subordinates. USAF Form 707 - leadership skills. USA DA Form 67–10-1A - Develops.
Delegation	X		X	USMC FITREP - leadership, leading subordinates. USA DA Form 67–10-1A - Leads.
Motivation	X	X	X	USMC FITREP - leadership, leading subordinates. USAF Form 707 - Leadership Skills.

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
				USA DA Form 67–10-1A - Intellect, Develops, Leads.
Change Management	X		X	USMC FITREP - leadership, communication skills. USA DA Form 67–10-1A - Develops, Leads.
INITIATIVE AND DRIVE				
Innovation	X		X	USMC FITREP - mission accomplishment, performance and proficiency. Individual character - initiative. USA DA Form 67–10-1A - Intellect, Develops, Leads.
Personal Development	X	X	X	USMC FITREP - individual character, initiative. USAF Form 707 - Job Knowledge. USA DA Form 67–10-1A - Develops
Independence	X	X	X	USMC FITREP - individual character, initiative. USAF Form 707 - Leadership Skills. USA DA Form 67–10-1A - Leads.
Volunteering				
TEAMWORK				

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
Team Pride	X		X	USMC FITREP - leadership, developing subordinates and setting the example. USA DA Form 67–10-1A - Leads; Develops.
Relationships	X		X	USMC FITREP - leadership, leading subordinates. USA DA Form 67–10-1A - Leads; Develops.
Contribution	X		X	USMC FITREP - leadership, leading subordinates and ensuring the well-being of subordinates and communication skills USA DA Form 67–10-1A - Leads.
COMMUNICATION				
Listening	X	X	X	USMC FITREP - leadership, communication skills. USAF Form 707 - communication skills. USA DA Form 67–10-1A - Leads.
Comprehension	X	X	X	USMC FITREP - leadership, communication skills. USAF Form 707 - communication skills. USA DA Form 67–10-1A - Leads.
Clarity	X	X	X	USMC FITREP - leadership, communication skills. USAF Form 707 - communication skills. USA DA Form 67–10-1A - Leads, Develops.
Non-verbal	X		X	

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
				USMC FITREP - leadership, communication skills. USA DA Form 67–10-1A - Leads.
Feedback		X	X	USAF - AF Form 707, Professional Qualities; Rater Instructions. USA DA Form 67–10-1A - Achieves.
Conflict Management			X	USA DA Form 67–10-1A - Develops.
Information Sharing	X		X	USMC FITREP - leadership, communication skills. USA DA Form 67–10-1A - Leads.
CRITICAL THINKING AND DECISION MAKING				
Risk Assessment	X		X	USMC FITREP - individual character, courage. USA DA Form 67–10-1A - Leads.
Planning	X	X	X	USMC FITREP - intellect and wisdom, decision making ability. USAF Form 707 - organizational skills. USA DA Form 67–10-1A - Achieves; Leads.
Evaluates Alternatives	X		X	USMC FITREP - intellect and wisdom, decision making ability. USA DA Form 67–10-1A - Leads.

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
Problem Solving	X	X	X	USMC FITREP - intellect and wisdom, decision making ability. Mission accomplishment, proficiency. USAF Form 707 - organizational skills. USA DA Form 67–10-1A - Leads; Develops; Achieves.
RESILIENCY AND TOUGHNESS				
Coping	X			USMC FITREP - individual character, effectiveness under stress.
Persistence	X	X	X	USMC FITREP - individual character, effectiveness under stress. USAF Form 707 - judgement and decisions. USA DA Form 67–10-1A - Presence; Develops.
Recovery	X		X	USMC FITREP - intellect and wisdom, decision making ability. USA DA Form 67–10-1A - Presence.
Adaptable	X	X	X	USMC FITREP - individual character, effectiveness under stress. USAF Form 707 - judgement and decisions. USA DA Form 67–10-1A - Intellect.
MISSION ACCOMPLISHMENT AND PRODUCTIVITY				

TRAIT / SUBTRAITS	USMC VALID	USAF VALID	USA VALID	OTHER SERVICE REFERENCE
Professional Competence	X	X	X	USMC FITREP - mission accomplishment, performance and proficiency. Individual character - initiative. USAF Form 707 - Job Knowledge. USA DA Form 67–10-1A - Leads; Develops.
Time Management	X	X	X	USMC FITREP - mission accomplishment, performance. USAF Form 707 - Judgement and Decisions. USA DA Form 67–10-1A - Achieves.
Quality and Attention to Detail	X		X	USMC FITREP - mission accomplishment, performance. USA DA Form 67–10-1A - Leads,

Adapted from Air Reserve Personnel Center (2015), Department of the Army (2019), Headquarters, United States Marine Corps (n.d.).

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Air Reserve Personnel Center. (2015, July 31). *Officer performance report* (Lt thru col). https://www.arpc.afrc.af.mil/Portals/4/Documents/AF707%20Checklist-OPR.pdf?ver=2019-02-14-104738-513

Black, J., Hashimzade, N., & Myles, G. (2009). A dictionary of economics. *Oxford Reference.* https://www.oxfordreference.com/view/10.1093/acref/9780199237043.001.0001/acref-9780199237043

Bock, L. (2015). *Work rules!: Insights from inside Google that will transform how you live and lead.* Twelve.

Boice, D. F., & Kleiner, B. H. (1997). Designing effective performance appraisal systems. *Work Study*, *46*(6), 197–201. https://doi.org/10.1108/00438029710367622

Buckingham, M., & Goodall, A. (2015). Reinventing performance management. *Harvard Business Review.* https://hbr.org/2015/04/reinventing-performance-management#:~:text=Deloitte's%20new%20approach%20separates%20compensation,to%20keep%20performance%20on%20course

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 47–74. https://doi.org/10.1146/annurev-orgpsych-032414-111427

CGT Staffing. (2020, July 31). *Performance appraisals and how they improve employee performance*. https://www.cgtstaffing.com/resources/performance-appraisals/

Chatfield, T. (2016, June 14). The trouble with big data? it's called the 'recency bias'. *BBC Future.* https://www.bbc.com/future/article/20160605-the-trouble-with-big-data-its-called-the-recency-bias

Chief of Naval Operations. (2021, January). *CNO navplan*. https://media.defense.gov/2021/Jan/11/2002562551/-1/-1/1/CNO%20NAVPLAN%202021%20-%20FINAL.PDF

Chief of Naval Personnel (n.d.). *Sailor 2025 Glossary Packet*. Navy.mil. (2019, August 22). https://www.netc.navy.mil/Portals/46/RRL/doc/Sailor2025.pdf

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

Cordial, B. (2017). Revamp FITREPs for lethality. *U. S. Naval Institute Proceedings*, *143/12/1,378*. https://www.usni.org/magazines/proceedings/2017/december/revamp-fitreps-lethality

Deadrick, D. L, Gardner D, G. (2008). Maximal and typical measures of job performance: an analysis of performance variability over time. *Human. Resource. Management, Review*, 18(3), 133–45. https://doi.org/10.1016/j.hrmr.2008.07.008

Department of Defense. (2021, January 26). Task Force One Navy. https://media.defense.gov/2021/Jan/26/2002570959/-1/-1/1/TASK%20FORCE%20ONE%20NAVY%20FINAL%20REPORT.PDF

Department of the Army. (2019, March 01). *Officer evaluation report support form.* (DA Form 67-10-1A0. https://armypubs.army.mil/pub/eforms/DR_a/pdf/ARN17086_A67_10_1A_FINAL.pdf

Elicker, J. D., Levy, P. E., & Hall, R. J. (2006). The role of leader-member exchange in the performance appraisal process. *Journal of Management*, *32*(4), 531–551. https://doi.org/10.1177/0149206306286622

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

Faram, M. (2021, February 3). Task Force one Navy completes report to enhance Navy diversity. *United States Navy.* https://www.navy.mil/Press-Office/News-Stories/Article/2490996/task-force-one-navy-completes-report-to-enhance-navy-diversity/

FederalPay. (n.d.). *Military pay raises - 2004 to 2022*. Retrieved February 21, 2022, from https://www.federalpay.org/military/raises

Fleet Forces Command. (2020, February). *Signature behaviors of the 21st century sailor.* https://www.mynavyhr.navy.mil/Portals/55/Support/21stCenturySailor/COE/signature-behaviors.pdf?ver=XTdBCqpOMsUeIFAH39kOVA%3D%3D

Fowler, F. (2013). *Survey research methods (applied social research methods).* Sage Publications.

Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. *Human Resource Management Review*, *21*(2), 123–136. https://doi.org/10.1016/j.hrmr.2010.09.004

Hagner, K. G. (2003). What shall it profit a man? *U. S. Naval Institute Proceedings, 129/6/1,204.* https://www.usni.org/magazines/proceedings/2003/june/comment-and-discussion

Hall, J. (2019, May 9). The Cost of Turnover Can Kill Your Business and Make Things Less Fun. *Forbes*. https://www.forbes.com/sites/johnhall/2019/05/09/the-cost-of-turnover-can-kill-your-business-and-make-things-less-fun/?sh=601f14527943

Hardison, C., Zaydman, M., Oluwatola, T., Saavedra, A., Bush, T., Peterson, H., & Straus, S. (2014). 360-degree assessments: Are they the right tool for the U.S. military? RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR900/RR998/RAND_RR998.pdf

Headquarters, United States Marine Corps. (n.d.). *Commandant's guidance.* (USMC Fitness Report 1610, NAVMC 10835 [rev. 7-11]). https://www.hqmc.marines.mil/Portals/133/Blank%20FITREP.pdf

Indeed Editorial Team. (2021, September 9). *Performance standards: What are they and why they're important*. Indeed Career Guide. Retrieved February 21, 2022, from https://www.indeed.com/career-advice/career-development/performance-standards

Joshi, R. (2016, November 9). *Accuracy, precision, recall & F1 score: interpretation of performance measures*. Exsilio Blog. Retrieved February 21, 2022, from https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/#:~:text=Accuracy%20%2D%20Accuracy%20is%20the%20most,observation%20to%20the%20total%20observations.&text=Precision%20%2D%20Precision%20is%20the%20ratio,the%20total%20predicted%20positive%20observations

Katz, I. M., Rauvola, R. S., & Rudolph, C. W. (2021). Feedback environment: A meta-analysis. *International Journal of Selection and Assessment*. https://doi.org/10.1111/ijsa.12350

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

Lattice Team. (2019, September 19). *How often should your company have performance reviews?* Lattice. Retrieved February 21, 2022, from https://lattice.com/library/how-often-should-company-performance-reviews

Lazear, E. P., & Oyer, P. (2003). Internal and External Labor Markets: A Personnel Economics Approach. *National Bureau of Economic Research.*

Manikandan, S. (2011). Measures of central tendency: Median and mode. *Journal of pharmacology & pharmacotherapeutics.* https://pubmed.ncbi.nlm.nih.gov/21897729/

Moran, 2016: https://www.usni.org/magazines/proceedings/2014/december/once-again-moment-ripe-change

Moran, B. (2014). Once again . . . a moment ripe for change. *U. S. Naval Institute Proceedings, 140/12/1342*. https://www.usni.org/magazines/proceedings/2014/december/once-again-moment-ripe-change

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

Muckler, F., & Seven, S. (1992). Selecting Performance Measures: "Objective" versus "Subjective" Measurement. *Human Factor: The Journal of the Human Factors and Ergonomics Society, 34*(4), 441–455. https://doi.org/10.1177/001872089203400406

Nowell, J., & Stefanus, J. (2020). We must win today's war for talent. *U. S. Naval Institute Proceedings, 146/1/1,403.* https://www.usni.org/magazines/proceedings/2020/january/we-must-win-todays-war-talent

O'Boyle Jr., E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, *65*(1), 79–119. https://doi.org/10.1111/j.1744-6570.2011.01239.x

O'Connell, M., & Kung, M. (Mei-C. (2007). The Cost of Employee Turnover. *Industrial Management*, *49*, 14–19. https://www.researchgate.net/publication/211392097_The_Cost_of_Employee_Turnover

Office of the Chief of Naval Personnel. (2021, December 01). *Navy performance evaluation system.* (BUPERS Instruction 1610.10F). Department of the Navy. https://www.mynavyhr.navy.mil/Portals/55/Reference/Instructions/BUPERS/BUPERSINST%201610.10F%20Signed%201%20Dec%2021.pdf?ver=Rr-9rEY3g1RUBmcuGoJsjQ%3d%3d

Poulin, A. (2016). Do we need to innovate the way we evaluate? *USNI Blog.* https://blog.usni.org/?p=22104

Pulakos, E. D., Mueller-Hanson, R., & Arad, S. (2019). The Evolution of Performance Management: Searching for Value. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*(1), 249–271. https://doi.org/10.1146/annurev-orgpsych-012218-015009

Rosenthal, L. (2021, July 2). *Hard truths and the duty to change: recommendations from the independent review commission on sexual assault in the military*. defense.gov. from https://media.defense.gov/2021/Jul/02/2002755437/-1/-1/0/IRC-FULL-REPORT-FINAL-1923-7-1-21.PDF/IRC-FULL-REPORT-FINAL-1923-7-1-21.PDF?source=GovDelivery

Rotundo, M., & Xie, J. L. (2008). Understanding the domain of counterproductive work behaviour in China. *The International Journal of Human Resource Management*, *19*(5), 856–877. https://doi.org/10.1080/09585190801991400

Schirmer, P., Levy, D. G., Thie, H. J., Moini, J. S., Harrell, M. C., Hall, K. C., Brancato, K., & Abbott, M. (2005). *Creating New Career Options for Officers in the U.S. Military*. RAND Corporation. https://doi.org/10.7249/RB7569

Seager, T. P. (2018). Performance Appraisal in the Information Age. *Medium*. https://medium.com/swlh/performance-appraisal-in-the-information-age-283f6d086b5f

Shaout, A., & Trivedi, J,. (2013) Performance appraisal system - using a multistage fuzzy architecture. *International Journal of Computer and Information Technology, 02*(03), 405–411. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.6432&rep=rep1&type=pdf

Sicard, S. (2017). The Navy's most radical performance overhaul is coming soon. *Task & Purpose*. https://taskandpurpose.com/leadership/navys-radical-performance-overhaul-coming-soon/

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

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The Feedback Environment Scale: Construct Definition, Measurement, and Validation. *Educational and Psychological Measurement*, *64*(1), 165–184. https://doi.org/10.1177/0013164403258440

Toppo, L., & Prusty, T. (2012). From performance appraisal to performance management. *IOSR Journal of Business and Management, 3*(5), 1–6. https://www.iosrjournals.org/iosr-jbm/papers/vol3-issue5/A0350106.pdf

U.S. Office of Personnel Management. *Developing performance standards*. (n.d.). Retrieved February 21, 2022, from https://www.opm.gov/policy-data-oversight/performance-management/performance-management-cycle/planning/developing-performance-standards/

University of Minnesota Libraries (2016, March 22). *Performance evaluation systems*. Human Resource Management. https://open.lib.umn.edu/humanresourcemanagement/chapter/11-1-performance-evaluation-systems/

Walters, K., & Rodriguez, J. (2017). The Importance of Training and Development in Employee Performance and Evaluation. *Research Gate.* https://www.researchgate.net/profile/Kelley_Walters/publication/332537797_The_Importance_of_Training_and_Development_in_Employee_Performance_and_Evaluation/links/5cefe41092851c4dd01ba833/The-Importance-of-Training-and-Development-in-Employee-Performance-and-Evaluation.pdf

Witherspoon, T. (2016, July 10). *Sailor 2025: What you need to know*. All Hands. https://allhands.navy.mil/Stories/Display-Story/Article/1840217/sailor-2025/

INITIAL DISTRIBUTION LIST

1.   Defense Technical Information Center
        Ft. Belvoir, Virginia

2.   Dudley Knox Library
        Naval Postgraduate School
        Monterey, California