



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

SUCCESSFUL PRACTICES FOR EMPLOYEE PERFORMANCE EVALUATIONS

by

Laura C. Small

March 2020

Thesis Advisor:

Co-Advisor:

Simona L. Tick

Kathryn J. Aten

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
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 2020	3. REPORT TYPE AND DATES COVERED Master's thesis	
4. TITLE AND SUBTITLE SUCCESSFUL PRACTICES FOR EMPLOYEE PERFORMANCE EVALUATIONS			5. FUNDING NUMBERS	
6. AUTHOR(S) Laura C. Small				
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.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) Stemming from the Chief of Naval Personnel's Sailor 2025 initiatives, the Performance Evaluation Transformation-Talent Management (PET-TM) program is tasked with improving and modernizing the Navy's current performance evaluation system. In support of these initiatives, this thesis conducts a case analysis of the current Navy performance evaluation system to highlight its main shortcomings. In addition, this thesis performs a review of scholarly and industry literature to identify successful practices of employee performance evaluation systems. Finally, based on the findings from the review of scholarly and industry literature, this thesis makes recommendations the Navy might consider to improve its current performance evaluation system.				
14. SUBJECT TERMS personnel management, evaluations, performance, typology, retention, feedback, coaching			15. NUMBER OF PAGES 107	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**SUCCESSFUL PRACTICES FOR EMPLOYEE PERFORMANCE
EVALUATIONS**

Laura C. Small
Lieutenant Commander, United States Navy
BS, U.S. Naval Academy, 2009

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
March 2020**

Approved by: Simona L. Tick
Advisor

Kathryn J. Aten
Co-Advisor

Marigee Bacolod
Academic Associate, Graduate School of Defense Management

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Stemming from the Chief of Naval Personnel's Sailor 2025 initiatives, the Performance Evaluation Transformation-Talent Management (PET-TM) program is tasked with improving and modernizing the Navy's current performance evaluation system. In support of these initiatives, this thesis conducts a case analysis of the current Navy performance evaluation system to highlight its main shortcomings. In addition, this thesis performs a review of scholarly and industry literature to identify successful practices of employee performance evaluation systems. Finally, based on the findings from the review of scholarly and industry literature, this thesis makes recommendations the Navy might consider to improve its current performance evaluation system.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	MOTIVATION	2
B.	BACKGROUND	2
	1. Sailor 2025	3
	2. MyNavy HR Transformation.....	4
C.	PURPOSE	5
D.	RESEARCH QUESTIONS	5
E.	ORGANIZATION	6
II.	RESEARCH APPROACH AND METHODS.....	7
A.	NAVY QUALITATIVE CASE ANALYSIS.....	7
	1. Objective	7
	2. Design	8
	3. Data Collection and Sources	9
	4. Data Analysis	9
B.	NARRATIVE SYSTEMATIC REVIEW AND SYNTHESIS	10
	1. Objective	10
	2. Design	10
III.	ANALYSIS OF PERFORMANCE EVALUATION IN THE NAVY.....	15
A.	THE EVOLUTION OF NAVY PERFORMANCE EVALUATIONS	15
B.	APPLICABLE LAWS AND INSTRUCTIONS.....	18
	1. The 1993 Government Performance and Results Act (GPRA).....	18
	2. U.S. Navy Regulations, 1990, Article 1129.....	18
	3. BUPERSINST 1610.10E	19
C.	NAVY PERSONNEL EVALUATION SYSTEM	19
	1. Applicability	19
	2. Report Types	20
	3. Reporting Seniors and Raters.....	21
	4. Delivery Method.....	21
	5. Purpose.....	21
	6. Composition.....	22
	7. Command Evaluation Process	24
D.	FINDINGS: CURRENT PES SHORTCOMINGS	26
	1. System Age.....	26

2.	Process over Performance.....	27
3.	Past versus Future Focus.....	29
4.	Lack of Transparency.....	30
5.	Inaccurate Measure	31
E.	NAVY PERFORMANCE EVALUATION TRANSFORMATION-TALENT MANAGEMENT (PET-TM)	33
1.	Coaching Tool	33
2.	Performance Evaluation Tool.....	35
IV.	ANALYSIS OF INDUSTRY LITERATURE AND KEY THEMES IN PERFORMANCE EVALUATION SYSTEMS	37
A.	BACKGROUND	37
1.	Define Performance	38
2.	Performance Management Process	39
3.	Summary.....	45
B.	REVIEW OF SCIENTIFIC LITERATURE.....	45
1.	Background	47
2.	Evaluation Purpose.....	49
3.	Employee Involvement	51
4.	Methods of Evaluation.....	56
5.	Rating	59
6.	Evaluation Process	62
C.	INDUSTRY BEST PRACTICES	64
1.	Google.....	64
2.	Deloitte	67
D.	CONCLUSION	70
V.	CHARACTERISTICS OF SUCCESSFUL PERFORMANCE EVALUATIONS	71
1.	Define the Evaluation Purpose	71
2.	Create a Culture of Communication.....	73
3.	Foster a Perception of Fairness	76
4.	There is No “One Size Fits All” Solution	78
VI.	CONCLUSION AND RECOMMENDATIONS.....	79
A.	CONCLUSION	79
B.	RECOMMENDATIONS.....	81
	LIST OF REFERENCES.....	83

INITIAL DISTRIBUTION LIST	89
---------------------------------	----

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	Case study process. Source: Yin (2018).	8
Figure 2.	Flow diagram of literature review process. Adapted from Waddington et al. (2012).	13
Figure 3.	Example performance evaluation process	25
Figure 4.	PET-TM coaching tool. Source: NPC (2019a).	34
Figure 5.	Sample coaching summary report. Source: NPC (2019a).	35
Figure 6.	Performance equation. Source: Rao (2016).	38
Figure 7.	Performance management process. Source: Pulakos (2004).	41
Figure 8.	Guidelines for effective feedback. Source: Pulakos (2004).	42
Figure 9.	Performance evaluation process. Source: Mondy and Martocchio (2016).	44
Figure 10.	Common reasons for performance rating manipulation. Source: Pulakos (2004).	54
Figure 11.	Traditional intermittent feedback system versus on-going feedback system. Source: Corporate Leadership Council (2016).	55
Figure 12.	360-degree feedback data sources. Source: London and Beatty (1993).	58
Figure 13.	Deloitte core performance design. Source: Pulakos and Battista (2020).	68
Figure 14.	Sample scatterplot of employee project performance. Source: Pulakos and Battista (2020).	69

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Sailor 2025 Initiatives. Adapted from NPC (2019b).	3
Table 2.	Periodic Evaluation Cycle. Adapted from CNP (2019).	20
Table 3.	Performance evaluation traits. Adapted from CNP (2019).	24
Table 4.	HR functions supported by performance evaluation. Adapted from Mondy and Martocchio (2016).	43
Table 5.	Key themes and categories in performance evaluation literature	46
Table 6.	Purposes for performance evaluation. Adapted from Cleveland et al. (1989).	49
Table 7.	Common performance evaluation biases. Adapted from Rao (2016).	53
Table 8.	Common traditional evaluation methods. Adapted from Mondy and Martocchio (2016).	57
Table 9.	Recommended methods for improving performance evaluations. Adapted from Adler et al. (2016).	62

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACDU	Active duty
ADE	Authoritative data environment
BuNav	Bureau of Navigation
CHIEFEVAL	Chief petty officer evaluation
CMC	Command master chief
CNP	Chief of Naval Personnel
CO	Commanding officer
CPO	Chief petty officer
CSR	Coaching summary report
DIVO	Division officer
EP	Early promote
EVAL	Evaluation (enlisted)
FITREP	Fitness report
FTS	Full-time support
GPRA	Government Performance and Results Act
HR	Human resources
INACT	Inactive duty reserve
I/O	Industrial and organizational
LCPO	Leading chief petty officer
LPO	Leading petty officer
MP	Must promote
MPT&E	Manpower, Personnel, Training, and Education
MSAF	Multi-source assessment and feedback
NPC	Navy Personnel Command
OIC	Officer in charge
P	Promotable
PES	Performance evaluation system
PET-TM	Performance Evaluation Transformation-Talent Management
PMS	Performance management system
QA	Quality assurance

RSCA	Reporting senior cumulative average
SECNAV	Secretary of the Navy
SGA	Summary group average
SHRM	Society of Human Resources
SIOP	Society for Industrial and Organizational Psychology
SM	Service member
SP	Significant problems
TM	Talent management
UIC	Unit identification code
XO	Executive officer

I. INTRODUCTION

To win future conflicts, we must win today's war for talent and that requires a personnel system rooted in, not straining beneath, the twenty-first century

—Vice Admiral Nowell and LT Stefanus

Today's military operating environment of emerging threats, expanding missions, and modern technologies is more complex than ever and demands effective ways of managing the U.S. Navy's most valuable resource: Its people. The Navy faces a war for talent and could lose valuable and well-trained Sailors from its ranks (Nowell & Stefanus, 2020) if the Navy's human resources (HR) practices remain "locked in the assumptions, policies, and statues of a previous era" (Moran, 2014). Innovating the performance evaluation system (PES) is at the heart of the Navy's talent management (TM) initiatives to improve how the Navy recruits, trains, develops, and manages its people (Burke, 2017).

President Trump's vision for expanding the fleet to a force of 355 ships (Larter, 2017) coupled with fierce competition from a robust civilian labor market, presents a challenge to ensuring the fleet is manned with the most talented and skilled Sailors (Moran, 2014). In his *Proceedings* article, Admiral Moran (2014) explains that the Navy seeks talented, technically skilled recruits who "present creative solutions ... demonstrating the power of autonomy, opportunity, and trust where it matters most" (p. 2). Industry similarly values these talents and skills, further challenging the Navy to compete for the pool of less than one percent of the U.S. population who possess a willingness to serve and meet the military's qualification standards (Burke, 2017).

Proper management of Navy personnel must begin with accurate evaluation and effective coaching (Department of Defense [DoD], 2018). Rooted in a 24-year-old system, the Navy's current Sailor performance evaluation practices fail to take advantage of the tremendous advances in modern technology and evaluation techniques (Burke, 2017). This thesis presents recommendations for an improved personnel evaluation

system, following an analysis of the current Navy PES and successful practices found through a systematic review of industry and scientific literature.

A. MOTIVATION

Our rivals are intent on taking the lead from us—we must pick up the pace and deny them. We cannot be satisfied with achieving minimum standards—we are a Navy focused on being the best we can be, every day. (Richardson, 2016, p. 16)

The margin between our military’s capabilities and those of emerging powers is diminishing (Richardson, 2016). General James Mattis has emphasized that “creativity and talent of the American warfighter is our greatest enduring strength, and one we do not take for granted” (DoD, 2018), suggesting sustainment of a dominating force will rely on our talented personnel. The Navy’s short-term and long-term planning must encompass preparations for a future modern war as the return of great power competition demands the rapid expansion of the naval fleet (Richardson, 2016). A modern Navy PES oriented to identify critical talent, motivate performance, and retain skilled Sailors is crucial to extending and maintain the Navy’s competitive advantage (Burke, 2017).

B. BACKGROUND

In Admiral Richardson’s *A Design for Maintaining Maritime Superiority 1.0* (2016), he explains that his vision for a more capable, lethal fleet hinges on “leaders and teams who learn and adapt to achieve maximum possible performance, and who achieve and maintain high standards to be ready for decisive operations and combat” (p.8). To meet this challenge, Navy HR has implemented programs to enhance personnel management and support “recruiting, developing and retaining the right number of Sailors with the right skills to man our force” (Navy Personnel Command [NPC], 2019b, p.1). Sailor 2025 launched to support fleet readiness through TM initiatives and grew into a large-scale transformation of Navy HR business practices, organizational structure, and IT (Nowell & Stefanus, 2020).

1. Sailor 2025

Sailor 2025 “is a dynamic set of initiatives, process improvements and management tools that address many of the challenges we face in building a more competitive personnel management system” (Moran, 2016). Sailor 2025 began in 2015 under the former Manpower, Personnel, Training, Education (MPT&E) command, with Admiral Moran, former CNP and Vice Chief of Naval Operations (NPC, 2019b). Anticipating future recruiting and retention challenges caused by low unemployment rates, he asked two questions:

1. Should we not design a personnel system that further encourages competition and rewards achievement with more challenge and greater authority, earlier?
2. Should it not enable mobility and the opportunity to serve in more than just a linear path? (Moran, 2014)

Admiral Moran envisioned and designed Sailor 2025 with the focus of updating policy and shifting away from an industrial-age personnel management system (PMS). Recent performance management initiatives within the Navy stem from these efforts to support Sailors and enhance fleet readiness by understanding and addressing Navy TM shortcomings. Sailor 2025 continued to evolve with the succeeding CNPs, Admiral Burke and Vice Admiral Nowell. Shown in Table 1, Sailor 2025 is now built on three pillars: personnel system modernization (PSM), Ready, Relevant Learning (RRL), and career readiness (CR) (Nowell & Stefanus, 2020).

Table 1. Sailor 2025 Initiatives. Adapted from NPC (2019b).

PSM	RRL	CR
Meritorious Advancement Program	Adopt and introduce proved training techniques	Leader Development Framework
Fleet Scholar Education Program	Continuous and tailored training	Expand Child Development Center hours
SECNAV Tours with Industry	Mobile training delivery platforms	Policies improving stability and work-life balance

PSM	RRL	CR
Revamp Pay and Personnel systems	Leverage cloud hosted training	Expanded maternity leave
Rating Modernization		Career Intermission Program
Transform PES		Longer Fitness Center Hours
Detailing marketplace		Improved Physical Fitness Assessment program

Efforts to improve and modernize the Navy’s current PMS stem from these initiatives and the Secretary of the Navy’s (SECNAV) Strategic Guidance. Specific SECNAV guidance focuses on investing in human capital, modernizing business operations, and supporting an integrated naval force structure (Department of the Navy [DoN], 2019).

2. MyNavy HR Transformation

Expanding from Sailor 2025, MPT&E initiated a transformation effort to update the organization, processes, and infrastructure, which support the HR “hire-to-retire” lifecycle of Navy Sailors (Nowell & Stefanus, 2020). In 2019, MPT&E rebranded as “MyNavy HR”, with fleet-wide improvements focused on keeping pace with industry standards and delivering technologically advanced systems (Nowell & Stefanus, 2020). The transformation efforts focus on modernizing the current HR IT system of 55 disparate, outdated systems with plans to convert the data into an authoritative data environment (ADE), where Sailor information is accurate and easily accessible (Burke, 2017). Additionally, the transformation emphasizes overhauling processes within MyNavy HR lines of business to optimize processes and maximize efficiency (Burke, 2017).

As part of the transformation, the Navy’s Performance Evaluation Transformation-TM (PET-TM) initiative started to concentrate on creating a PES that meets the needs of a modern, agile force. The main goals are to provide a more robust and transparent process, produce relevant talent identification data on Sailors, and update

the evaluation methods to more accurately value the potential of each Sailor (Chief of Naval Personnel [CNP], 2018).

C. PURPOSE

The goal of this thesis is to examine the Navy PES and present successful personnel evaluation practices informed by evidence-based research in scientific literature and industry practices. Maintaining the Navy's competitive advantage alongside industry leaders requires investing in Sailors and the management tools of the Navy force. Ensuring the Navy's PMS supports Sailors and builds their potential strengthens the Navy as a fighting force.

This thesis will systematically analyze and organize published research findings from fields including industrial and organizational (I/O) psychology, behavioral science, and neuroscience relevant to personnel performance evaluation practices; it will additionally review industry practices to categorize and identify themes for informing changes in the Navy's future personnel system. The question of the best way to effectively evaluate employee performance prevails in both military and civilian organizations. Employee evaluations impact every aspect of an organization with implications within employee morale, retention, promotion, performance, recruitment, and organizational success.

D. RESEARCH QUESTIONS

The research questions addressed in this thesis are:

- What are the shortcomings of the current Navy performance evaluation system?
- What performance evaluation practices, key traits, and central themes are evident in scholarly and industry literature?
- What performance evaluation practices should the Navy consider in its performance evaluation transformation?

E. ORGANIZATION

This thesis is organized into six chapters.

- Chapter I introduces the research topic and its relevance to the Navy.
- Chapter II covers the qualitative case analysis evaluation methodology for assessing the Navy PES and a synthesis approach to narrative systematic review used to analyze the literature and industry practices of this topic.
- Chapter III provides a case analysis of the Navy's PES, including background information on the history of Navy performance evaluations and the laws, instructions, and institutional rules that govern the Navy performance evaluation system. Chapter III then reviews the current Navy's PES, including the evaluation methods, tools, and processes for all Navy military personnel; it also discusses the Navy's current progress in updating the PES.
- Chapter IV presents findings and recommendations of themes present within industry and scientific literature.
- Chapter V presents practices present in successful performance evaluations and links these themes to Navy PES shortcomings. Further, it makes recommendations for a future Navy performance evaluation system.
- Chapter VI presents the conclusions and recommendations.

II. RESEARCH APPROACH AND METHODS

This thesis utilizes an inductive, qualitative approach, relying on two research methods:

- Qualitative case analysis to organize and evaluate the Navy's PES (Yin, 2018).
- Narrative approach to systemic review to identify synthesize scientific and industry literature, and to identify industry practices and lessons learned (Snilstveit, Oliver, & Vojtkova, 2012).

This thesis integrates the case analysis and systemic review and synthesis. The case analysis provides meaningful background and insight into the development of the PES and the many functions it supports within the Navy and highlights weaknesses in the current system. The systematic review and narrative synthesis of scholarly and industry literature identifies key themes and current practices. A final integration of two analyses results in successful performance evaluation practices and recommendations. This researcher drew on 11 years of personal experience of using the Navy PES at shipboard and shore commands to interpret the analysis. The researcher's use of personal experience is expressly stated and not intended to demonstrate doctrine.

A. NAVY QUALITATIVE CASE ANALYSIS

1. Objective

This thesis presents a qualitative case analysis of the Navy's PES to understand the system's current state and to identify key problems. A critical component to identifying industry contributions for a future PES requires a thoughtful analysis of the Navy's environment, problems, objectives, and requirements. The intended result of this research is to present a baseline understanding of the Navy's PES and identify critical shortcomings within the current system.

2. Design

The performance evaluation is a complex component within the Navy's PMS. The evaluation process reinforces most aspects of the HR functions including the promotion system, non-statutory selection process, and compensation determinations. With such a complex, interlinked system this case analysis largely focuses on the evaluation system components and the link to the other supporting HR functions; however, the inventory of connections is not exhaustive, with this research focusing primarily on chief problems within the PES.

Sasa Baskarada's (2014) elaboration of Robert Yin's (2018) widely cited process for conducting case research, provides a framework for this study. Figure 1 shows Yin's six-phase process case study process, consisting of unique but interconnected stages. The stages of this case process—plan, design, prepare, collect, analyze, and share—allow for a structured, yet iterative, approach to research (Yin, 2018).

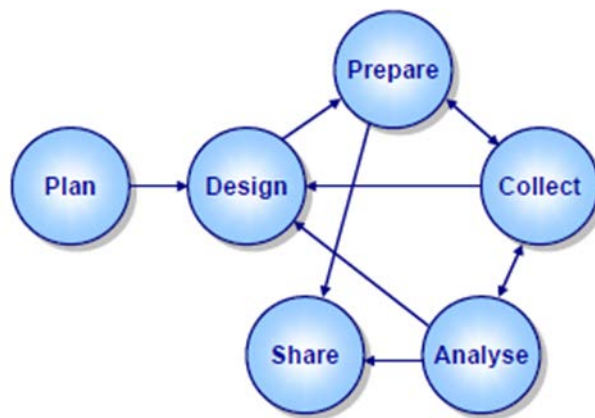


Figure 1. Case study process. Source: Yin (2018).

The planning stage identifies the scope for the case study by defining the research objectives (Baskarada, 2014). This case's objective centers on understanding the historical background, evolution, composition, complications of the Navy's PES. The design, prepare, and collect phases, guide a holistic approach to data collection, designed

to capture a large range of sources (Baskarada, 2014). This approach guided the collection of documents that informed the cases analysis.

3. Data Collection and Sources

Published documents provided the data for the case analysis. The researcher sought documents providing the official Navy description and also the perspective of users of Navy systems.

The researcher searched the Navy reference library, using the search terms *performance*, *evaluation*, *EVAL*, *FITREP*, *CHIEFEVAL*, and *talent management*. This search resulted in 14 Navy instructions and Navy messages related to the PES, including the Navy instruction BUPERSINST 1610.10E. This document is the governing instruction on the evaluation process, composition, method, and procedure for the Navy PES and largely provided the Navy perspective.

PES directly supports Sailor HR functions, thus, the researcher sought accounts providing the perspective of those utilizing the system. *Proceedings*, *U.S. Naval Institute News*, *Navy Times*, *All Hands Magazine*, and *Federal News Network* primarily focus on Navy and military interests and provided this perspective. A search using the terms *Navy*, *performance*, *evaluation*, *appraisal*, *talent management*, *transformation*, and *Sailor 2025* in these sources resulted in 12 documents. Advisors at PERS-3, the Navy's branch responsible for performance evaluations and the PET-TM initiative, provided documents on current efforts and the projected direction of the program.

4. Data Analysis

The researcher analyzed the data by first reviewing all of the documents. The researcher prepared a summary review of foundational statute and Navy publications to understand the regulations and procedures that govern the performance evaluation process. The researcher then reviewed prior research and literature on the development of the Navy's PES to develop an understanding of the background, historical challenges, and unique organizational culture that influences the Navy's current evaluation methods. Recent policy change documents, published articles, interviews with leadership, and

PET-TM informational documents provided an account of present efforts to address changes to the Navy PES.

The researcher noted key events, dates, changes, ideas, recurring themes, and, shortcomings in detailed memos. The researcher then developed a detailed case timeline and narrative from these notes. The researcher paid particular attention to changes in the external and technological environments and mapped these to changes (or lack of changes) in the Navy PES.

B. NARRATIVE SYSTEMATIC REVIEW AND SYNTHESIS

1. Objective

This portion of the thesis synthesizes what is known about industry practices and scientific literature for performance evaluations in order to identify key themes associated in the literature with successful performance evaluation practice. A narrative approach to systematic review and synthesis is particularly useful in analyzing and synthesizing a diverse, complex body of evidence (Snilstveit, et al., 2012).

2. Design

This thesis utilizes the narrative approach to systematic review and synthesis (Snilstveit, et al., 2012). This approach begins with a tentative framework that may be derived from policy priorities and/or practitioner knowledge. This approach connects diverse types of studies, documents, and literature in a systematic and transparent manner to inform an evidence-based conclusion (Snilstveit, et al., 2012). It is principally useful in building the themes seen in multiple sources of research while organizing and analyzing data, particularly qualitative data. Another strength is in the ability to process and analyze data throughout the collection process. The following five-phase process is utilized to conduct this analysis: familiarization, identifying a thematic framework, indexing, charting and mapping, and interpretation (Ritchie & Spencer, 1994).

a. Familiarization

In the first stage, the researcher becomes familiar with the body of knowledge (Ritchie & Spencer, 1994). The initial Navy case analysis provided the researcher familiarity with the Navy case. The PET-TM team members supplied preliminary research articles and reports to frame the scope of the ongoing performance evaluation transformation and provide context through their governing documents. In addition to these, the researcher conducted a preliminary literature search using the term *employee performance evaluation*. The search was restricted to books, book chapters, journal articles, professional conference topics, technical reports, and English language dissertations. The search resulted in over one million records. The researcher then reviewed sources for relevancy to the research questions and selected two titles to provide an initial orientation in performance evaluation literature. Published by the Society of Human Resources Management (SHRM), ***Performance Management: A Roadmap for Developing, Implementing and Evaluating Performance Management Systems*** by Dr. Elaine Pulakos (2004) was particularly useful in providing industry accepted standards for this topic. Additionally, a white paper through the Society for Industrial and Organizational Psychology (SIOP) titled *Encouraging Effective Performance Management Systems* by Dr. Kelsey Kline and Alison Carr, and the scholarly book, ***Human Resource Management*** by R. Wayne Mondy and Joseph Martocchio (2016), further aided in setting the foundation for key ideas and assisted the researcher in identifying a thematic framework.

b. Developing a Thematic Framework

Developing a thematic framework is an iterative and flexible process, which provides for the indexing and classification of data as well as allows for the expansion of connections and themes as the research continues (Ritchie & Spencer, 1994). The researcher develops an initial thematic framework, which guides the research, but may continue to be revised through subsequent iterations of analysis. The thematic framework guides the categorization of the data into themes, which are sorted within a chart or representation of the framework (Ritchie & Spencer, 1994). The framework for this study

consists of a narrative summary of each identified theme, including the evolution of performance evaluation, key supporting functions, and organizational approaches to evaluations.

c. Data Sources

Although a strength of the framework synthesis technique is that it is dynamic and flexible, it is also systematic and transparent (Snilstveit, et al., 2012). Following the preliminary orientation, the researcher drew on the initial thematic framework to identify additional documents for the review, relying on a snowball approach. Snowballing is the process of “reviewing and pursuing references in identified papers, including primary studies and existing reviews, and using these sources to build up...a larger body of evidence” (Waddington, et al., 2012, p. 363). This study used the snowballing technique particularly with Adler et al. (2016), which includes a full body of performance evaluation research to review. Figure 2 shows the researcher’s snowball process.

The researcher conducted an advanced search using the academic databases EBSCOhost, JSTOR, Web of Science, and Google Scholar as well as the Dudley Knox Library catalog for performance evaluation literature using search terms *employee performance evaluation*, *employee appraisal*, *measurement methods*, *talent management*, and *evaluation methods*. This resulted in 2315 documents.

In addition to reviewing scholarly literature, organizations within private industry were reviewed to glean successful performance evaluation factors. Few organizations match the Navy’s large-scale hierarchal organizational structure but most still operate within small teams. Lessons learned and methods to evaluate performance, even in more “flat” organizations, are meaningful. In selecting the organizations for study, the researcher targeted top performing companies on the Fortune 500 list with over 100,000 employees who have undertaken transformational steps to update their employee PES within the past ten years. From this process, Google and Deloitte were selected for review and study. This resulted in an additional 85 documents, and a total of 2359 documents after duplicates were removed.

The researcher then screened the documents, refining results to peer-reviewed journal articles, published books, conference papers, and professional reports, resulting in 111 full texts. The researcher further screened the text by reviewing titles, dates, and abstracts, excluding 77 texts that were either over 50 years old or did not focus on industry study in employee performance evaluations. This resulted in 34 documents, which formed the database for this study. Figure 2 shows a diagram of the literature review process.

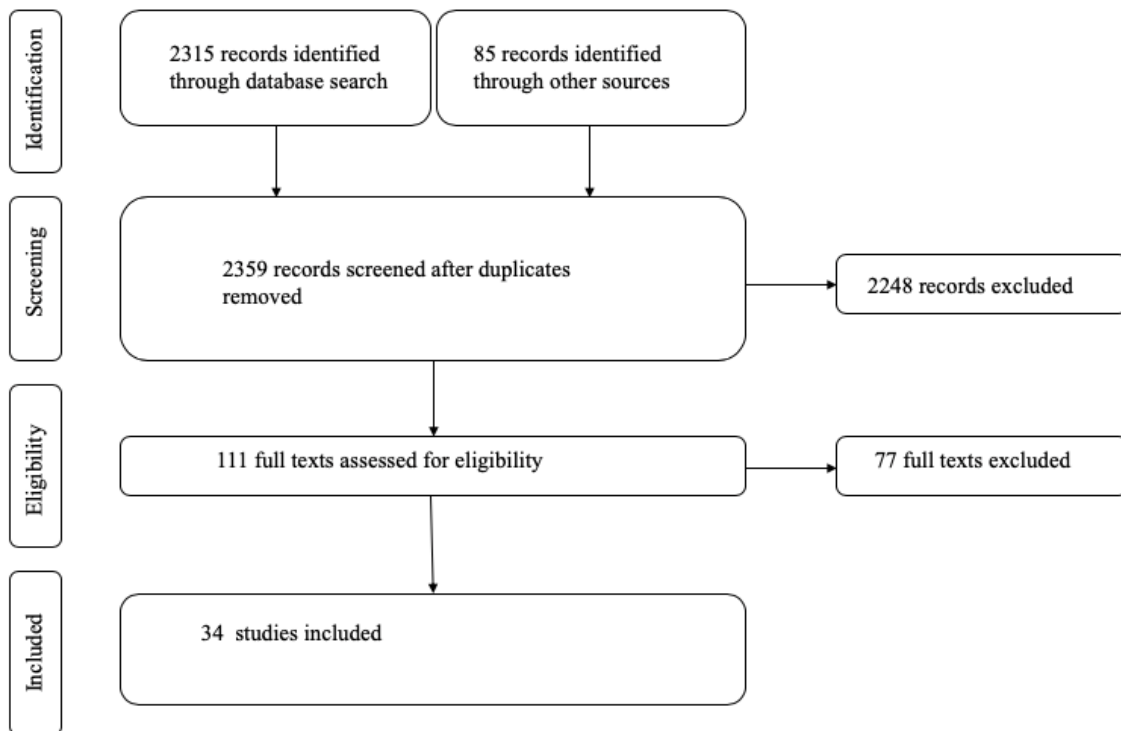


Figure 2. Flow diagram of literature review process. Adapted from Waddington et al. (2012).

d. Charting and Mapping, and Interpretation

In the final mapping and interpretation stage, the researcher demonstrates key concepts and themes in a more concrete manner by “defining concepts, mapping range and nature of phenomena, creating typologies, finding associations, providing explanations, and developing strategies” (Ritchie & Spencer, 1994, p. 186). The

researcher of this study integrated the case analysis during this final mapping and interpretation stage to summarize key findings and develop recommendations for the Navy.

III. ANALYSIS OF PERFORMANCE EVALUATION IN THE NAVY

This scope of this thesis is to explore industry best practices and current academic research regarding effective performance evaluations with the goal of making recommendations for improving Navy performance evaluation processes. This chapter first provides a history of the development of Navy performance evaluations, highlighting the process by which Navy processes advanced and the naval traditions that shaped their composition. This chapter, then, describes the laws and instruction governing the system and current structure and procedures. The chapter concludes by identifying and describing shortcomings in the current system and, finally, by describing MyNavy HR efforts for process improvements.

A. THE EVOLUTION OF NAVY PERFORMANCE EVALUATIONS

For the large part, the personnel management system grew in the same way as the development of the Navy: Through necessity and trial and error. The Navy's performance management system grew as an iterative process, beginning from scratch but growing when personnel problems threatened mission accomplishment and required swift, workable solutions (Chisholm, 2001). Newly independent and no longer protected by European treaties, the origins of the U.S. Navy grew from the necessity to protect its merchant fleet from the Barbary pirates of North Africa (Chisholm, 2001). At the urging of President George Washington, Congress begrudgingly approved the commissioning and temporary manning of six frigates with the Naval Act of 1794; however, no system existed defining recruitment, training, promotions, evaluation methods, separations, disciplinary actions, and other personnel matters (Chisholm, 2001).

Enacted in 1799, *An Act for the Government of the Navy of the United States* first mentioned expected conduct for officers and men within the naval service, with commanding officers charged "to show in themselves a good example of honour and virtue...and to be very vigilant in inspecting the behavior of all such as are under them" (Congress, 1799). Commanders were instructed on proper documentation of the names of

the crew within the ship's log and their duty to ensure discipline and obedience of the crew (Congress, 1799). Thus, the origins of the Navy's method for evaluating officers and enlisted Sailors began with no master plan for a formal system, but developed through constant tweaking and restructuring over the next 200 years—little study of the validity and scientific relevancy of performance evaluations was used to inform changes to either the officer or enlisted evaluation system (Raimondo, 2009).

Ever more recognized as essential to national security and economic stability, the Navy's accomplishments during the War of 1812 lead to increased funding and organizational expansion. In the 1815, *Rules, Regulations, and Instructions for the Naval Service of the United States* continued to progress the idea of marking performance by holding Commanding Officers responsible for reporting on the character of officers serving in their command, "particularly as to his sobriety, diligence, activity and abilities," to the Secretary of the Navy and the Board of Navy Commissioners and to also leave a copy with the ship in the "hands of the commandant of the yard" (Department of the Navy [DoN], 1815, p. 45). Commanders were also charged with the duty of rating his crew—New recruits particularly faced examination for the state of their hygiene so as to not introduce "filth and contagion" (DoN, 1815, p. 51). The ships' company was reviewed by their abilities, and promotions of enlisted Sailors were handled by the commanding officer.

Outside the requirements detailed in Navy Regulations, the operations of the ship, including the standards to which the crew was expected to perform, was dictated by the commanding officer. With no uniform system for evaluating performance, the Navy experienced a lack of standardization and an inability to influence Sailor quality across the ranks. Assuming sole responsibility for personnel management following the reorganization of the Bureau of Navigation (BuNav) in 1889, instructions for evaluating personnel and documenting performance were issued from circular letters by BuNav in 1921 (Hunt, 1966). This began the Navy's standardization for evaluating officer and enlisted sailor performance. To meet the needs and mission requirements of the era, the system was periodically reviewed and fine-tuned, with a revision issued in 1925 and essentially left unchanged until 1956 (Hunt, 1966).

Beginning as a word-of-mouth system between commanding officers, the Navy's early evaluation system was based on a subjective appraisal of work performance. As discussed above, limited instruction existed to formally guide the assessment and differentiation of performance, leading to perceptions of unfairness. According to thesis research conducted by Hunt (1966), the need for enlisted evaluations originated largely from questions on how to accurately classify characterization of service of enlisted personnel following discharge or who to select for special programs. Navy enlisted performance evaluations were initially conducted quarterly and were rated by their division officers (with consultation of senior enlisted) on the junior Sailor's obedience, sobriety, proficiency in rating, and, depending on a man's rating, in seamanship and/or mechanical ability (Hunt, 1966). At the end of a Sailor's enlistment, the marks received in the service record were averaged to determine if the Sailor was eligible for a good-conduct discharge. Until the 1950s, Hunt (1966) notes the commanding officer held authority for petty officer advancement regardless of the man's performance. Major changes to the original formal system included extra emphasis on the formal evaluation method as a necessity to properly identify the top Sailors for promotion, and not just for characterization of service upon their separation.

However, the officer community faced a different challenge that necessitated the need for a functional evaluation system. Adopting many of the traditions and standards from the British Royal Navy, the appointments for officers commissioned within the naval profession were generally treated as a lifetime privilege (Chisholm, 2001). While many fine officers filled the ranks, the officer management system presented little upward mobility and harbored inept performers until they succumbed to either illness or death. Rear Admiral Yates Stirling Jr. highlighted that "promotion by seniority, waiting for dead men's shoes, is a sad blow to efficiency for it stifles initiative and offers no incentive" (Chisholm, 2001, p. 767). However, in 1916, with the institution of "selection up", the officer community began to "shift away from the concept of commission as the right of the individual officer once commissioned to one that made the commission contingent upon meeting standards of performance and the requirements of the navy" (Chisholm, 2001, p. 773). With these shifts in promotion and selection, the Navy

recognized the importance of developing an evaluation system to assess the performance of officers fairly and attempt to make promotion recommendations to the best officers.

The past two centuries have seen many changes and modifications to the Navy's evaluation system. This brief history illustrates the progression of a temporary fleet of six sailing ships to becoming the greatest, most technologically advanced Navy in the world. Especially seen through the mid-20th century, the development of the Navy's current system now encompasses progressive recruiting, training, and detailing methods with algorithms projecting manning and strength requirements. An advanced board system for promotion and special assignments selects based on competence and structured standards. An articulated, well-developed evaluation system undergirds every aspect of these personnel actions—The United States Navy could have the best ships and technology, but without the best people, the fleet could soon be inferior.

B. APPLICABLE LAWS AND INSTRUCTIONS

1. The 1993 Government Performance and Results Act (GPRA)

The GPRA was enacted to hold government and federal agencies accountable and improve the American people's confidence in the federal government. The act required the development of performance measurement strategies, including creating and updating strategic plans, performance plans, and performance reports, as well as the improvement of actual program performance. It mandated that performance measures become institutionalized and part of everyday business practices within federal agencies. The modernization efforts (2010) focused partly on ensuring evaluations are in machine-readable formats. Both acts point to the requirement of performance evaluations and the necessity to embrace technology.

2. U.S. Navy Regulations, 1990, Article 1129

This regulation requires records to be maintained on officer and enlisted naval personnel “which reflect their fitness for the service and performance of duties” (Department of the Navy [DoN], 1990, p. 105). Promotion and duty assignments are determined by each individual's records, including records of fitness or evaluations.

Additionally, the “preparation of these reports shall be regarded by superiors and commanding officers as one of their most important and responsible duties” (p. 105).

3. BUPERSINST 1610.10E

BUPERSINST 1610.10E, dated 6 December 2019, is the Navy’s latest instruction for providing guidance for performance evaluations for all active duty (ACDU), full-time support (FTS), and inactive duty reserve (INACT) naval personnel. It is the single authority on procedures and requirements for conducting performance evaluations and provides the required forms for conducting performance appraisals.

C. NAVY PERSONNEL EVALUATION SYSTEM

The Navy’s current tool for conducting performance evaluations was implemented in 1996. It organizes the Navy’s method for documenting performance and ultimately determines a service member’s (SM) suitability for promotion and selection to special duties. Understanding the components of the Navy’s PES is important in addressing how industry and scientific best practices may be valuable to informing improvements in evaluating performance in the Navy. Additionally, a thorough understanding of the evaluation system allows for a critical analysis of the strengths and shortcomings of the current system.

While the Navy’s instruction for the evaluation system covers all aspects and every nuance of the special circumstances that arise in the career events of a Sailor, the section below provides a basic and simplified guide for understanding the framework of the Navy’s PES. In addition to special circumstances, different ranks of SM are subject to different rating forms and standards—flag evaluations are under a separate system and not covered in this thesis. This thesis focuses on the common and most used aspects of the system, while highlighting key differences of significance.

1. Applicability

The Navy PES applies to all ACUDU, FTS, and INACT naval personnel (BUPERS). The system uses different forms and criteria to evaluate the different

categories of naval SMs: Officers (W1-O6) utilize fitness report (FITREP), chief petty officers (CPO) (E7-E9) utilize CHIEFEVAL, and enlisted personnel utilize an EVAL.

2. Report Types

“Regular” reports comprise the majority of the evaluations conducted. These periodic reports are conducted annually on a specific cycle determined by the SM’s rank. Other circumstances such as the detachment of the SM, change of reporting senior, or other special evaluations are still regular reports. These reports focus on the day-for-day performance of the SM over the annual reporting period. Concurrent reports can be utilized in cases when the SM is fulfilling temporary additional duty orders or additional duty such as being at training or school. “Operational Commander” reports are optional and are utilized for active duty commanding officers (CO) or officers in charge (OIC) to provide a supplemental record of performance by their operational commander (Chief of Naval Personnel [CNP], 2019). Seen in Table 2, FITREP ending dates are the last day of the month and CHIEFEVAL and EVAL ending dates are the 15th day of the month.

Table 2. Periodic Evaluation Cycle. Adapted from CNP (2019).

	PERIODIC FITREP/CHIEFEVAL/EVAL	
	Officers (All)	Enlisted (All)
Jan	O3	
Feb	O2	
Mar	W5, W4, W3	E5
Apr	O5	E9
May	O1	
Jun		E4
Jul	O6	E3, E2, E1
Aug		
Sep	W1, W2	E8, E7
Oct	O4	
Nov		E6
Dec		

3. Reporting Seniors and Raters

SMs are rated by their reporting senior, which is generally their CO or OIC, depending on the command. Per BUPERSINST 1610.10E, commanding officers are required to rate and evaluate all members within their command. However, they may delegate their authority of reporting senior to the executive officer (XO) or department heads (O-3 and above) within their command.

Raters and senior raters are also included in the evaluation reports for all enlisted Sailors, E6 and below. Their inclusion is required in order to ensure the inputs of the senior enlisted and division officers are coordinated for the final evaluation.

4. Delivery Method

Electronic, fillable forms (NAVPERS 1610/2 for W-2 to O-6; NAVPERS 1616/27 for E-7 to E-9; NAVPERS 1616/26 for E-1 to E-6) are used to complete all performance evaluations. These forms are used within a Microsoft Windows program called NAVFIT98A. Following final inputs onto the electronic NAVPERS form, it is printed, wet signed, and physically mailed to Navy Personnel Command in Millington, Tennessee. From there, it is scanned and uploaded into the Sailor's service record.

5. Purpose

The NAVPERS form functions to provide both professional counseling and evaluation. First, counseling feedback is used to "enhance professional growth, encourage personal development, and improve communication" (CNP, 2019, p 18-1) and is conducted at the midway point in the evaluation period (mid-term counseling). Counseling is conducted by the reporting senior, senior rater, or rater, depending on the rank of the SM, utilizing the NAVPERS form as a counseling worksheet. The SM's strengths, weaknesses, and areas for improvement are discussed and documented as command file. While instruction dictates that counseling take place, the methods and procedures for conducting counseling vary depending on the command.

According to the periodic schedule, above, the service member is evaluated by their reporting senior (or designee), who receives input from the service member's chain

of command. The evaluation provides a look at the SM's performance, looking back at the past period of performance. These evaluations are meant to inform promotion and special selection boards. Procedures for formal evaluations are heavily dictated by instruction, which covers regular reporting requirements and a myriad of special circumstances.

6. Composition

a. Description of Common Elements

There are many similarities between the different evaluation forms used to evaluate officer and enlisted personnel. All forms contain the following sections:

- Basic administrative information of the SM including name, grade/rate, designator, social security number, type of report, physical readiness score, billet, and reporting senior information.
- Command employment and command achievements include a brief statement of significant command achievements during the reporting period. This aids in providing context for the environment in which the SM's duties are performed.
- A description of the SM's primary, collateral, and watchstanding duties provides the duties performed and the number of months they are performed during the period of report.
- Report of mid-term counseling received during the period of report. This counseling session is held between the SM and their supervisor utilizing the SM's evaluation form as a template to discuss their strengths, weaknesses, and areas for improvement.
- Performance traits encompass the quantitative portion of the evaluation, which includes several traits rated on a 5-point scale (1.0-lowest to 5.0-highest). The final trait average results from the average of individual scores of all rated traits—it is compared to the reporting senior's average

score for that rank. The reporting senior cumulative average (RSCA) is an important component of the Navy evaluation system. This is a historical average of all trait score averages that the specific reporting senior has ever given to individuals in the same paygrade. It allows for a promotion board to compare the rating of the SM to the historical rating of the reporting senior for that SM's rank; it is seen as favorable if the SM is at or above the reporting senior's RSCA.

- Promotion recommendations are based on a 5-point scale to include "Significant Problems" (SP), "Progressing," "Promotable" (P), "Must Promote" (MP), and "Early Promote" (EP). These recommendations are made based on performance regardless of where the SM is in the promotion cycle. Forced distribution determines the number SMs assigned a promotion recommendation in the categories of P, MP, and EP.
- The qualitative portion includes a narrative discussion on the SM's performance and description of qualifications/achievements completed during the reporting period.

b. Discussion of Divergent Elements

While BUPERSINST 1610.10E supplies guidance in the differences seen in specific cases, there are several key distinctions in the performance traits and ratings, depending on the SM's rank. Officers (W2-O6) are all evaluated on six performance traits, with the seventh, "tactical performance", used depending on the officer's warfare qualification. They share several performance traits with enlisted Sailors (E1-E6) including "Command or Organizational Climate/Equal Opportunity", "Military Bearing/Character", "teamwork", and "leadership". Table 3 illustrates the differences and shared elements in performance traits.

Table 3. Performance evaluation traits. Adapted from CNP (2019).

Performance Trait	FITREP	EVAL	CHIEFEVAL
Professional Knowledge		X	
Quality of Work		X	
Deckplate Leadership			X
Institutional and Technical Expertise			X
Professionalism			X
Loyalty			X
Active Communication			X
Sense of Heritage			X
Professional Expertise	X		
Mission Accomplishment and Initiative	X		
Command or Organizational Climate/Equal Opportunity	X	X	
Character			X
Military Bearing/Character	X	X	
Personal Job Accomplishment/Initiative		X	
Teamwork	X	X	
Leadership	X	X	
Tactical Performance	X		

While there are commonalities between the ratings, the responsibilities and roles of officers, junior enlisted, and senior enlisted are very different. The evaluation system is designed to capture the unique expectations of each rank while also acknowledging the foundational principles, which are important to all Sailors, regardless of rank or position.

7. Command Evaluation Process

A reporting senior is assigned to rate each SM stationed at their command via their command authority. The SM who is being rated will have their trait score average evaluated in comparison to the associated RSCA. The summary group average (SGA), when applicable, is included when a Sailor is evaluated against a subset of their peers at their current command with whom they share specific characteristics such as paygrade, promotion status, report type, reporting senior, unit identification code (UIC), and competitive category.

While instruction strongly dictates the procedures for completing the evaluation forms, the process of gathering Sailor performance data to populate the evaluations and the rankings procedures of the Sailors within the command are not standardized. To

display an example of a command's process for completing the regular reports for E-6 evaluations, the below figure is provided based on this researcher's personal experience at several commands. Figure 3 illustrates an example command evaluation process.

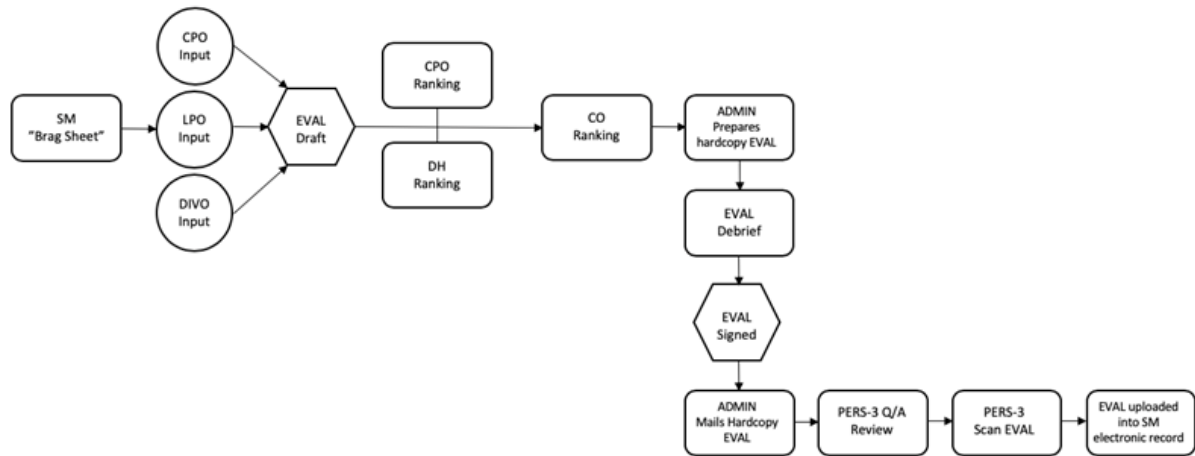


Figure 3. Example performance evaluation process

In this case, the SM being rated provides a “brag sheet” with description of the Sailor’s duties, accomplishments, and contributions to the command throughout the period of evaluation. The divisional leading petty officer (LPO), divisional leading chief petty officer (LCPO), and division officer (DIVO) consider the SM’s contributions and provide their input in the composition of the evaluation to include suggested trait scores and the narrative comments.

Independently, the CPO mess and department heads conduct independent rankings of all E-6's within the command, categorizing each Sailor by the recommended promotion characterization. These rankings and suggested narrative comments are presented to the CO via the XO and the command master chief (CMC). While the CO takes recommendations from the XO and CMC, the CO retains final authority in the SM’s performance evaluations.

Once the evaluations are finalized, the administrative department prepares the summary group evaluations, receiving electronic versions of each department’s E-6

evaluations. They provide quality assurance (QA) reviews and print final copies for signature. Evaluations are debriefed between the reporting senior or their delegate and the SM, where the SM ensures the correctness of the basic evaluation information and receives feedback on their performance over the evaluation period. Once the debrief is conducted, the administrative department receives signed evaluations, makes copies for the command's record, and prepares the summary group's evaluations for mailing to Navy Personnel Command (NPC). Once received, NPC provides another QA on the evaluation batch and either accepts or rejects the evaluations. Finally, the hardcopy evaluations are scanned and uploaded into each SM's electronic personnel record.

D. FINDINGS: CURRENT PES SHORTCOMINGS

The greatest threat to our Navy does not come from the Russians, the Iranians, or the Chinese. It comes from our own ranks. We weaken the world's strongest Navy by promoting the wrong people, a direct result of a misguided evaluation system. (Goforth, 2016)

1. System Age

Implemented in 1996, the current PES supported the Navy in a time of military drawdown and relies on technology no longer compatible with present day advances (Burke, 2017). Further, according to a study by CAPT James Raimondo (2009), the Navy's performance evaluation division (PERS-311) was unaware of any study or analysis conducted to inform the current system's implementation. At 24-years-old, the PES "has not been adapted to the rigors of current and future operating environments, a fully connected world, or unconventional threats" (Raimondo, 2017, p. 86).

Navy performance evaluations use a Microsoft Windows based "form filler" program called NAVFIT98A which "provides help screens, spell-check, calculates individual trait averages, and has validation features that will prevent many common errors" (CNP, 2019, p. 1-1). NAVFIT98A does not connect to the internet, and the evaluations contained in SM records are not searchable (Raimondo, 2017). General Mattis highlights the need to modernize key systems and explains "we cannot expect success fighting tomorrow's conflicts with yesterday's weapons or equipment" (DoD, 2018, p. 6). This extends beyond employing cutting-edge technology and into the

capabilities the Navy uses to evaluate Navy Sailors. The current evaluation system does not support MyNavy HR goals to improve workforce decision-making through the use of predictive analytics and big data (Burke, 2017). Moran expressed that the use of evaluations is a vital TM tool, highlighting that the Navy “still have neither the enterprise-wide metrics to measure nor the tools necessary to nimbly and flexibly shape the naval force in terms of quality rather than quantity” (Moran, 2014).

The National Defense Strategy (2018) warns that the “backlog of deferred readiness, procurement, and modernization requirements has grown in the last decade and a half and can no longer be ignored” (p. 11). Dramatic changes are overdue, and modernization efforts become harder and more cumbersome to achieve as time passes. As the Navy adopts plans to modernize, General Mattis further urges that “software must be designed for routine replacement instead of static configurations that last more than a decade” (DoD, 2018, p. 11). Improving evaluation methods and the supporting technology are ongoing with PET-TM; however, fixing a 24-year-old system is complicated (Burke, 2017).

2. Process over Performance

LT Erik Sand’s article in *Proceedings* (2014) captures the relationship between process and combat readiness in the military, stating that while “the right process increases the likelihood of successful performance, process is only useful to the extent that it actually produces better performance”. While the Navy successfully recruits, trains, and manages its Sailors in today’s demanding operational environment, the HR systems supporting the Navy are “the evil we know” and are complex, outdated, and inefficient (Burke, 2017, p. 8). Management processes must help, not hinder, performance; highlighted by General Mattis, “we must transition to a culture of performance where results and accountability matter. We will put in place a management system where leadership can harness opportunities” (DoD, 2018, p. 10).

However, the current process of performance reports burdens, instead of enhances, the Navy; they “are very manual and time intensive” and lack process standardization (Moran, 2014, p. 3). The command administrative department begins

preparations months in advance of evaluation due dates with rough inputs collected from the SM's chain of command; supervisors must "take time to ensure the best efforts in the beginning of the evaluation cycle to get the best results in the end" (Snyder, 2002, p. 4). The quality of performance evaluations depends on the knowledge and skill level of those preparing the evaluation (Eyer, 2018). Senior officers having the best of intentions still "may have little strategic understanding of the knowledge, skills, and attributes that the Navy desires of its future operational leaders" (Raimondo, 2009, p. 53) and may inadvertently damage the career of the SM.

The administrative burden of evaluations coupled with the requirements of daily routines can result in a diminished and low-quality evaluation process. Performance evaluations constitute the primary source of information regarding a SM's performance and development (DoN, 2019). However, in his article (2016) on the inadequacies of officer fitreps, LT Andrew Poulin notes it is the norm in some commands for evaluations to first be written by the SM themselves, begging the question, "how can your FITREP provide a candid assessment of your performance when *YOU* write your own FITREP" (p. 5).

Noting the need to reduce the burden of process over performance, Admiral Moran stated, "we should work to put the 'bureau' out of business, becoming more of an 'agent' between Sailors and commanders" (Moran, 2014, p. 3). Often, the process takes over in who is deemed a top performing Sailor, with seniority or other reasons taking precedence over performance. Ranking seniority over merit allows for the promotion of lower-qualified Sailors, neither ready nor qualified for the leadership challenges of the next rank. (Hagner, 2003). "Poor or weak performance should not be overlooked in favor of advancing a career at a desirable future job" (Poulin, 2016, p. 4).

Additionally, in his 2017 article, CAPT Raimondo decried that Navy evaluations, "simply do not measure up to what is needed now and will be needed in the future" (p. 1). The "lackadaisical approach" and adherence to process over quality degrade performance evaluations as "exercises in signatures rather than exercises in leadership. They are either congratulatory sessions that ignore significant areas for improvement or disappointing conferences that gloss over tough issues" (Snyder, 2002, p. 1). These behaviors cause the

promotion of enlisted and officers who are bolstered by ineffective evaluations. The PES is a tool, but only as good as the information encapsulated on the evaluation. CAPT Raimondo highlights specific issues with this problem:

In the helicopter community specifically, there have been a number of cases of well-known yet undocumented tragedies and inappropriate actions where the officers involved then went on to positions of greater responsibility and stature, including command. These issues involved the berating of junior personnel and a poisoned command climate. Yet, the behaviors were apparently not documented on fitness reports and unknown to the institutional Navy. In these cases, the promotion boards were diligent in their dictum to only look at documented performance. While the senior officers who did not document the adverse performance on the fitness reports are also at fault, the organization, the Navy, was equally at fault for not identifying these individuals. (Raimondo, 2009, p. 80)

As General Mattis highlights, “current processes are not responsive to need...Delivering performance means we will shed outdated management practices and structures while integrating insights from business innovation (DoD, 2018, p. 10). The Navy’s current PES prioritizes cumbersome, inefficient, and inaccurate processes over enhancing performance and building fleet readiness.

3. Past versus Future Focus

The current PES is designed to inform future promotion, by only assessing the performance of the Sailor over the course of the previous year (DoN, 2019). While it is important for evaluations to address past performance, the PES does not inform Sailor potential for the job they may hold in the future (Poulin, 2016). Additionally, as Poulin (2016) notes, “the problem with the current system is when these sometimes competing objectives of past performance and promotion potential are out of balance with each other” (p. 4).

Utilizing the same form as used on the evaluation, coaching takes place halfway through the evaluation cycle providing inadequate, unhelpful feedback (Poulin, 2016). Counseling should be conducted at the mid-term point and during the evaluation debrief with the purpose “to enhance professional growth, encourage personal development, and

improve communication among all members within the command” (CNP, 2019, p. 18-1). However, “more commonly, these sessions are treated more as a formality...where vague generalities are exchanged, and the debriefing is attempting to cover roughly six months of work where the details of performance recalled from months ago have often faded from memory” (Poulin, 2016, p. 4).

4. Lack of Transparency

The current Navy PES lacks transparency, creating an obstacle for leaders to engage Sailors in an environment of genuine trust (Moran, 2014). The following discusses the issues that arise when the system lacks meaningful feedback and relies on ambiguous signals to inform important personnel decisions. Additionally, the deficiency in transparency diminishes the effectiveness of evaluations by leading to reverse engineering and “gaming the system” (Eyer, 2018).

The cycle of evaluations does little to build trust, with the system built to deliver feedback twice a year (Burke, 2017). While the SM’s evaluation should never come as a surprise (Eyre, 2018), feedback should be given on their performance “throughout their careers, not just when they receive their evaluations (Burke, 2017, p. 1). CAPT Eyre notes in his article (2018) that counseling “should be specific, unemotional, and generated by the leadership team. Then, the individual should be told in exacting detail what he or she needs to do to improve” (p. 2).

Additionally, a transparent system does not require reverse engineering to achieve desired results; the desired results come from allowing evaluations to work as intended, (Burke, 2017). Part of this issue, reporting seniors are burdened by RSCA management and “are incentivized not to objectively grade their subordinates on their performance trait average” (Cordial, 2017, p. 2). Different commanding officers manage how they rate their Sailors differently, and may inadvertently inflate their averages; managing RSCA is important because it allows for the differentiation of superior and poor performance (Cordial, 2017). The metrics must provide meaningful feedback to both the service member and the selection boards, especially when selecting Sailor to service in “special positions of trust” (Moran, 2014, p. 4).

A SECNAV study (2010) examined CO relief for misconduct or poor performance by reviewing officer service records and conducting interviews with personnel within their command. While interviews pointed to known issues with the CO, “without exception, we found no record with documented misconduct or substandard performance” in their service records (SECNAV, 2010, p. 4). However, the report noted that the evaluation system’s design does not support documenting every case of misconduct or poor performance in a SM’s record.

In one case a CO had been relieved for having an affair with the ex-wife of a command member while the member and his ex-wife were reconciling, yet the CO’s post-relief FITREP grades were all 4.0 and the top line of block 42 indicated that he was a “mentor” for his subordinates. We found no cases in which a reporting senior wrote a fitness report contrary to Navy instructions, but some reporting seniors glossed over the substandard performance which led to some COs being relieved. (SECNAV, 2010, p. 4)

Signals directed to a selection or promotion board may not be distinguishable to the Sailor or may even be done inadvertently by the reporting senior (Eyer, 2018). Three-time CO and well-versed expert in the Navy PES, Captain K. Eyer (2018) notes in his article that often the subtleness of the wording and messaging contained in evaluations or fitreps are not fully understood by those writing the evaluations. For example, mentioning the word “potential” in an evaluation may seem benign, but it is often viewed as a “black mark” to the board that the SM may not be ready for promotion (Goforth, 2016). CAPT Eyer discusses similar pitfalls seen in many cases, where good performers miss promotion opportunities, but “the problem can be traced back to a single, unfortunate eval or FitRep, which (and sometimes despite the author’s intent) turned out to be a ‘head shot’” (Eyer, 2018, p. 42). However, cracking the code for writing good evaluations only goes so far when “the process is intrinsically self-inflated, full of hyperbole, and over-embellished” (Poulin, 2016, p. 2).

5. Inaccurate Measure

The Navy’s evaluation system may not be an accurate measure of performance, potentially leading to misalignment in organizational goals and the individuals placed in

certain roles (Moran, 2014). With the Navy piloting several TM initiatives such as “pay for performance”, the validity of the Navy’s evaluation tool is becoming increasingly more important (Burke, 2017). If the system does not provide accurate, defensible feedback, it is certainly not allowing for the proper identification of talent to benefit from these programs. “Today we just don’t have enough fidelity, enough good, objective data on Sailors’ performance that we really could pay based on performance (Burke, 2017, p. 7).

Demotivation and decreases in Sailor performance result from pressure and stress to perform within a system that lacks transparency, is subject to bias, and lacks Sailor confidence (Moran, 2014). Petty Officer Goforth notes his frustrations in seeking evaluation writing advice from senior members of his command, lamenting “how can I be confident in our Navy’s promotion system when its key decision-makers are using irrelevant criteria” (Goforth, 2016).

We want to get rid of some of the artificialities in our current system, one of which is we have a forced distribution system in our current reporting systems that was designed to prevent grade inflation, and today has unintentionally turned into a system that drives the reporting seniors to rate individuals by their seniority rather than talent. We’re looking to drive that behavior out of our reporting seniors so that we really can recognize and report on and reward talent through promotions and then ultimately through pay. (Burke, 2017, p. 8)

If part of the purpose of evaluations is to present an accurate measure of performance, the Navy must innovative the objective measures (Raimondo, 2017, p. 3). One particularly troubling component comes from “Eval inflation”, “when Navy leaders rate all their sailors as superstars because it makes the leaders look good and costs nothing. Not all sailors are superstars, but you would not know it by looking at the average Block 43—Comments on Performance” (Goforth, 2016, p. 1). Burke (2017) further highlights the issues within the Navy’s forced distribution system describing that it has “unintentionally turned into a system that drives the reporting seniors to rate individuals by their seniority rather than their talent” (p. 8).

E. NAVY PERFORMANCE EVALUATION TRANSFORMATION-TALENT MANAGEMENT (PET-TM)

The Navy recognizes that the current system has shortcomings and created the Navy Performance Evaluation Transformation Talent Management (PET-TM) initiative intended to address them. PET-TM is led by PERS-3 at NPC to spearhead the initiative to modernize the Navy PES (NPC, 2019a). Key to the transformation, PET-TM seeks to improve feedback by driving improved individual and unit performance. PET-TM is set to roll out changes in stages, with the current focus on the coaching aspect of the Navy performance evaluation system. Key tenants of the evaluation transformation are: (1) Emphasizing merit over seniority or tenure, (2) rating performance on paygrade-based objective standards focused on professional competence and character, (3) providing meaningful, actionable, and frequent feedback (Chief of Naval Personnel [CNP], 2019a).

1. Coaching Tool

The below figure illustrates the proposed interface of the multi-source assessment and feedback (MSAF) coaching tool, also known as a 360-degree assessment; utilizing modern technology, evaluators can provide candid, real-time feedback to the SM they are rating. The evaluator must respond within a window of time to value statements, utilizing the 9-point ranking scale. Figure 4 shows the proposed coaching tool.

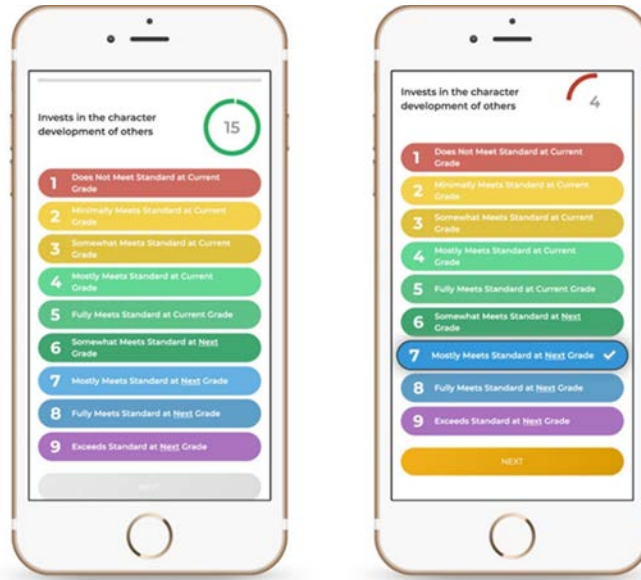


Figure 4. PET-TM coaching tool. Source: NPC (2019a).

The 43 value statements are sorted into eight categories, (character, leadership, initiative and drive, teamwork, communication, critical thinking and decision making, resiliency and toughness, and mission accomplishment) evaluated on a nine-point scale:

1. Does not meet standard at current grade
2. Minimally meets standard at current grade
3. Somewhat meets standard at current grade
4. Mostly meets standard at current grade
5. Fully meets standard at current grade
6. Somewhat meets standard at next grade
7. Mostly meets standard at next grade
8. Fully meets standard at next grade
9. Exceeds standard at next grade

The statements are specific to paygrade, similar to the current system of evaluating different ranks on unique performance traits. The coaching process begins with the SM conducting a self-assessment; pending policy changes will require the

coaching process to take place within six months of checking into a new command. From there, the SM receives input from subordinates, peers, and their direct supervisor. This compilation provides a well-rounded view of the different focus areas from many different perspectives and removes forced distribution through peer comparison. MSAF assessments will be conducted on a cycle dependent upon the Sailor's arrival to the command but can also be requested at any time by the SM or their supervisor. Figure 5 shows an example of the coaching summary report (CSR) utilizing the proposed feedback tool:

EXAMPLE COACHING SUMMARY REPORT*

Value Statement	Member	Evaluator	Peer**	Subordinate**
Exhibits confidence	1.00	6.00	7.00	6.50
Supports others through change	7.00	7.00	4.00	7.50
Has a motivational command presence	7.00	5.00	6.00	6.00
Leads within mess or association	1.00	6.00	6.50	7.50
Fosters a sense of military heritage	9.00	3.00	6.00	4.50
Articulates mission requirements to others	4.00	3.00	5.50	6.00
Actively contributes to an environment of fairness and mutual respect	7.00	6.00	4.00	3.50
Demonstrates proper military courtesies	6.00	5.00	5.00	4.00
Trains Sailors effectively to achieve mission	9.00	4.00	9.00	5.50
Leads in the community	1.00	4.00	7.00	9.00
Pursues challenging duties, responsibilities and collaterals	4.00	9.00	3.00	6.00
Performs tasks with accuracy utilizing current instructions directives and procedures	8.00	8.00	5.50	4.50
Displays a positive attitude towards mission, Sailors, and leadership	9.00	4.00	8.00	4.50
Shares knowledge in a clear and simple manner	8.00	3.00	6.50	5.00
Takes pride in work with no need for recognition	7.00	4.00	5.00	4.00

Does not meet standard

Minimally meets standard

Meets standard

Fully meets standard

Meets standard at next level

*Report details subject to change
**As applicable; peer and subordinate ratings are aggregated

Figure 5. Sample coaching summary report. Source: NPC (2019a).

2. Performance Evaluation Tool

As the plan stands currently, the formal performance evaluation tool is expected to be structured closely to the coaching tool, but will mirror the current EVAL/FITREP signature requirements of including the rater, senior rater, and rated SM. Six months following the initial coaching, the formal evaluation will be conducted. Additionally, it is envisioned that a narrative field will be added and included in the evaluation summary report. The results of the evaluation will be reviewed and debriefed between the senior

rater or reporting senior and the SM. Additionally, while the coaching tool metrics will not be stored, the performance evaluation will be maintained in an ADE to be used for future data analytics and talent management objectives.

The shortcomings of the Navy's PES described in the previous section have not fallen on deaf ears. The CNP and his team have taken great steps in recognizing the need for change and are investing manpower and resources in finding a solution. Scientific and industry research have identified many practices that support effective employee evaluation. Subsequent chapters of this study explore this research and document practices to inform the best trajectory for the Navy moving forward.

IV. ANALYSIS OF INDUSTRY LITERATURE AND KEY THEMES IN PERFORMANCE EVALUATION SYSTEMS

This chapter provides an analysis of performance evaluation practices documented in scientific and industry literature. This chapter is divided into two sections. The first section provides background definitions and describes the framework that guides the analysis. The second section presents an analysis and synthesis of the literature.

The first section begins by defining performance and discussing the organizational benefits of employee engagement. Then, this section defines performance evaluation and performance management and discusses the links and interdependent relationships between them. Next, this section describes Dr. Elaine Pulakos's (2004) schema for understanding the nature of performance management and evaluation within an organization. Dr. Pulakos, a leading expert in I/O Psychology, is president of the Personnel Decisions Research Institute. Pulakos's (2004) study synthesizes research from leading experts within the Society of Human Resources Management and SIOP to create the schema, which forms a conceptual framework that guides this analysis.

The second section of this chapter synthesizes the trends within the performance evaluation literature and industry practices documented in industry literature. This section explores different evaluation methods and models for conducting performance evaluations as well as industry evaluation practices at Google and Deloitte. The insights supplied by this examination inform recommendations of successful practices that the Navy could implement to improve the PES.

In the literature, authors use terms performance evaluation and performance appraisal interchangeably. Both refer to the evaluation of performance. This researcher uses the term *performance evaluation* in analysis and discussion.

A. BACKGROUND

This section defines performance, performance management, and performance evaluation and describes their link to each other within the performance management

process. Additionally, this section discusses the components of performance management based on the conceptual framework presented by Dr. Pulakos. Finally, the components of performance evaluation are described along with their role in the performance management process.

1. Define Performance

In his book, *Performance Management: Toward Organizational Excellence* (2016), T. V. Rao explains that performance depends upon a person's "ability or competence multiplied by motivation or work effort or commitment further multiplied by organizational support or a culture of empowerment and support" as shown in Figure 6 (p. 30).

$$\text{Performance} = \text{Competence} \times \text{Commitment} \times \text{Culture that is supportive and enabling} + \text{Chance factors}$$

Figure 6. Performance equation. Source: Rao (2016).

The management of workforce performance is increasingly recognized as essential to organizational success (Wigert & Harter, 2017). Key to the equation in Figure 6, performance must be defined, as without clear definition, nothing can be accurately measured (Rao, 2016). Additionally, Rao (2016) explains that performance measurement must present exactly "what the performer is expected to do and achieve" (p. 31). Going further, Michael Armstrong (2008), a noted expert in the field of performance management, explored the "mixed model" believing performance to be a combination of results and behavior, with the latter being controllable by the organization.

In the equation seen in Figure 6, an employee's *competence* constitutes a mix of innate intelligence, education, and training, which is recognized and developed through an organization's recruiting and training elements (Pulakos, 2004). *Commitment*, also

known as employee engagement, is seen in employees who “are willing to finish tasks in their own time and see a strong link between the firm’s success and their own career prospects” (Mondy & Martocchio, 2016, p. 234). *Culture* is an important component, which is determined by the organization; to enhance performance, it should be supportive of employee goals and seek to uplift employees (Rao, 2016). Additionally, *chance factors* are outside the foreseeable control of the employees and the organization but could encompass large-scale economic shifts, personal opportunities, or personal conflicts (Rao, 2016).

In a study commissioned by the United Kingdom’s Secretary of State for Business, Mondy and Martocchio, (2016) notes that when a company creates a supportive, trusting culture, the benefits are a more engaged workforce with “less absenteeism...greater employee satisfaction, less mistakes, fewer employees leaving, and naturally higher profits” (p. 234). Additionally, study by Aguinis (2009) noted that trends point to employees following individually guided career paths. While building commitment for the organization is important, the evaluation system an organization employs guides desired employee behaviors and sets individualized opportunities to enhance productivity (Aguinis, 2009).

2. Performance Management Process

This section presents a definition for performance management and uses Dr. Pulakos’s (2004) schema to describe the performance management process, which forms a conceptual framework that guides further analysis. The schema details how a cycle of performance planning, ongoing feedback, employee input, performance evaluation, and performance review all support the organization’s strategy and goals, (Pulakos, 2004).

Performance management is an ongoing, linked system within an organization that encompasses training, evaluation, total rewards, promotion, and other HR functions which contribute to organizational success (Mondy & Martocchio, 2016). The focus of performance management is on “performance improvement in order to achieve organizational, team, and individual effectiveness” and employee development for both

individuals and teams (Armstrong, 2008, p.7). According to Rao (2016), performance management is linked to the performance equation in the following ways:

1. Defining and clarifying what constitutes performance.
2. Providing context for performance, environmental changes, and expectations.
3. Ensuring that the required competencies are identified, available, and developed.
4. Ensuring motivation and commitment—communications and measuring scales.
5. Ensuring organizational support (p. 32).

Having a well-defined PMS ensures these aspects of the performance equation are constantly involved and employed; however, implementation of the system in a careful, organized way is critical to the success of the PMS (Rao, 2016). Understanding that performance management is a shared, two-way process between employees and their managers is an important factor for a PMS (Armstrong, 2008). Armstrong (2008) further highlights that organizations building their PMS “should develop its own version to suit its needs” and “have a conceptual framework within which appropriate processes can be developed and operated” (p. 17). While organizations employ different techniques in performance management, according to Pulakos (2004), most performance management processes typically encompass a variation of the process depicted in Figure 7:

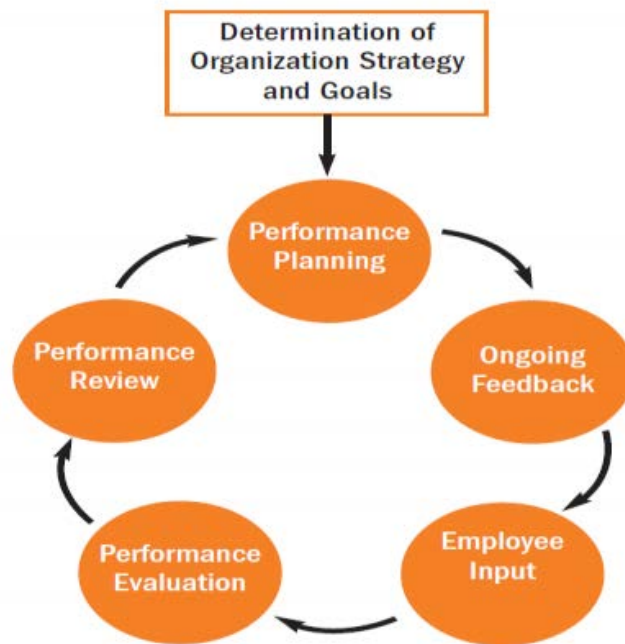


Figure 7. Performance management process. Source: Pulakos (2004).

b. Performance Planning

According to research synthesized by Pulakos (2004), performance planning encompasses a review of expectations for employee behaviors and performance results; it is important to align these expectations to the organization's purpose, values, and strategy. Specificity during this stage is key, as managers should cover objectives and standards of performance, capabilities required for the job, performance measures and indicators, how the capability will be assessed, and the expectation of quality for output (Armstrong, 2008). Rao (2016) asserts this stage to be the more important aspect of a PMS, as it ensures employees understand their role in the organization, and "ramps up the contribution of the individual and enhances his/her self-worth" (p. 33). Pulakos (2004) agrees that this is critical for setting the baseline expectations between the employer and employee, because it fosters transparency and trust in what is expected of the employee in their role.

c. Ongoing Feedback

Ongoing feedback is another critical component to the PMS, as it allows for minor course corrections in employee performance to meet desired goals (Pulakos, 2004). Pulakos (2004) highlights four important components for a successful feedback process: (1) Two-way communication, (2) joint responsibility shared between managers and employees, (3) timely and close to the event, and (4) continuous. Figure 8 shows guidelines for providing feedback based off multiple research studies.

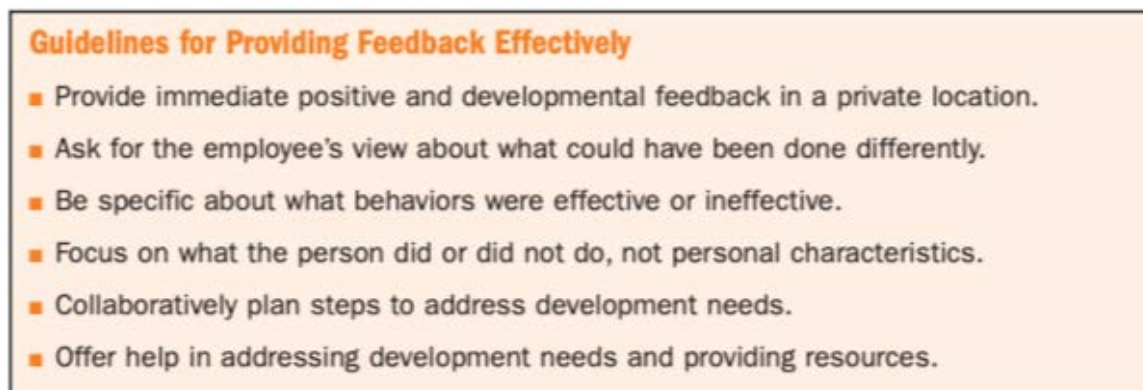


Figure 8. Guidelines for effective feedback. Source: Pulakos (2004).

d. Employee Input

Incorporating employee input in the process creates employee buy-in and a sense of control, which is also linked to higher employee performance (Rao, 2016). Pulakos (2004) states that employee input within the performance management process can be seen through employee performance self-rating and through prepared “statements of their key results or most meritorious accomplishments at the end of the rating period” (p. 8). Further, soliciting employee input within the process has a number of positive effects including: (1) Enhances employee ownership and buy-in to the process, (2) reminds managers on employee achievements and results produced, (3) decreases time spent by managers generating reports from scratch, (4) increases communication between managers and employees, and (5) uses input for HR functions (Pulakos, 2004).

e. Performance Evaluation

Performance evaluation is one of the most important aspects of an organization's PMS because it enables and supports many of the PMS functions (Pulakos, 2004). Performance evaluation is the event of formally reviewing and evaluating the performance of an individual or team (Mondy & Martocchio, 2016). With its primary goal to improve employee and organizational performance, Mondy and Martocchio (2016) additionally highlights that evaluation data is "potentially valuable for virtually every human resource functional area" (p. 237). However, despite its importance, performance evaluation seems to be one of the most dreaded and hated aspects of the PMS across all organizations (Adler et al., 2016). The uses of performance evaluations vary greatly and are shown in Table 4.

Table 4. HR functions supported by performance evaluation. Adapted from Mondy and Martocchio (2016).

HR Function	Description
HR Planning	Uses data from performance evaluation for succession planning and talent identification
Recruitment and Selection	Performance evaluations may predict performance of job applicants by creating benchmarks for desired behaviors of successful employees
Training and Development	Performance evaluations may be used in pointing to employee needs for training and/or development opportunities
Career Planning and Development	Data from performance evaluations may be used to determine employee potential so leadership can ensure proper qualifications and opportunities are implemented into individual career paths
Compensation Programs	Performance evaluations may be used to justify "pay for performance" by linking strong or desired performance to pay increases
Internal Employee Relations	Data from performance evaluations may be used for promotion, termination, or transfer decisions
Assessment of Employee Potential	Performance evaluations may show future potential through an employee's past behavior or performance

Next, the performance evaluation cycle begins with the establishment of what performance traits and goals will be measured. Next, performance criteria are established

by determining which “traits, behaviors, competencies, goals, achievements, and improvement potential” should be evaluated (Mondy & Martocchio, 2016, p. 32). Following the evaluation period, the employee’s performance is reviewed with the rater and a clear understanding of the expected performance criteria is set; from there, the manager and employee can identify any weaknesses and create a plan for better future performance (Mondy & Martocchio, 2016). Figure 9 shows the performance evaluation cycle.

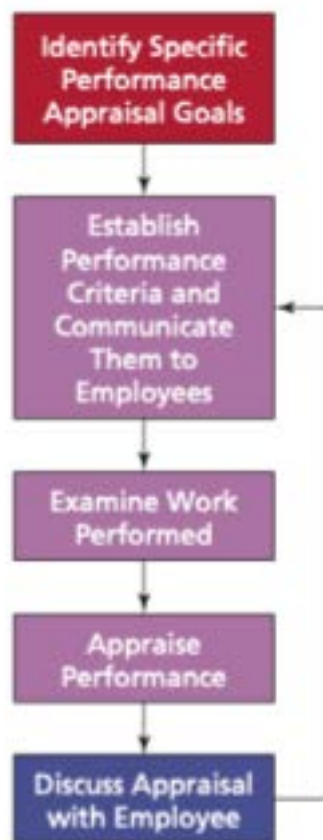


Figure 9. Performance evaluation process. Source: Mondy and Martocchio (2016).

f. Performance Review

The performance review is a discussion between the rater and the employee regarding past performance and areas to improve future performance (Armstrong, 2008). The formal review of performance should never be a surprise, especially if proper attention is paid to providing on-going feedback (Pulakos, 2004). Pulakos (2004) further contends that the formal review should be the time where managers justify “ratings, narratives, and rationale for the evaluation given” (p. 21). Generally, reward and separate administrative decisions should not be discussed during this aspect of the evaluation process (Pulakos, 2004).

3. Summary

This section defines performance, performance evaluation, and performance management and discusses the relationships between these components. In addition, this section describes Dr. Pulakos’s (2004) schema for understanding the performance management process; the components of this applied schema are: Performance planning, ongoing feedback, employee input, performance evaluation, and performance review. This schema forms a conceptual framework that guides further analysis in the next section.

B. REVIEW OF SCIENTIFIC LITERATURE

The following section reviews scientific literature on performance evaluations and analyses the components of the PES. First, this section provides background information on workforce attitudes toward performance evaluations. Next, this section presents literature in categories of (1) evaluation purpose, (2) employee involvement, (3) evaluation methods, (4) rating considerations, (5) evaluation process and (6) selected industry practices. Table 5 categorizes the literature on performance evaluations. The chapter concludes with a review of key themes found through the analysis of the scientific and industry literature.

Table 5. Key themes and categories in performance evaluation literature

Study	Key themes	Category
Adler et al. (2016)	Review of performance evaluation literature	Purpose Rating
Aguinis (2009)	Performance management and impacts on employee performance through feedback	Employee Involvement
Armstrong (2008)	Performance management and evaluation applications	Employee Involvement Rating
Cleveland et al. (1989)	Multiple objectives within evaluations and effect on ratings	Purpose Rating
Colquitt (2001)	Perceptions of fairness	Rating
DeNisi & Pritchard (2006)	- Performance evaluation rating and validity - Measurement issues - Performance improvement through perceptions of fairness	Rating
DeNisi (2016)	- Improving performance through evaluations - Changing behavior: measurement problem vs motivation problem	Purpose
Elicker (2006)	- Impact of supervisor-subordinate exchange relationships on evaluations - Importance of employee voice and impact on perception of justice	Employee Involvement
Greguras et al. (2003)	Measurement	Rating
Haines & St-Onge (2012)	Effective performance management practices	Employee Involvement
Hirsh & Inzlicht (2008)	Study showing human response to uncertainty more negative than actual negative feedback	Employee Involvement
Kavanagh et al. (2007)	Perceptions of fairness and the evaluation process	Process
Latham et al. (2005)	Measurement	Rating
Lawler et al. (2016)	Effective training techniques for conducting evaluations	Process Rating
London & Beatty (1993)	360-degree feedback	Rating

Study	Key themes	Category
Maylett (2009)	• 360-degree feedback implementation as evaluation and developmental tool • How to implement as an evaluation	Rating Process
Murphy (2008)	• Rating/Rater/ Training	Rating
Mondy & Martocchio (2016)	• Performance management and evaluation • Rater error • Methods of evaluation	Rating Employee Involvement
Passmore et al. (2016)	• Training influences within performance management	Rating Employee Involvement
Pulakos (2004)	• Comprehensive review of performance management	All
Pulakos & O’Leary (2011)	• Implementation issues within performance management	Process
Pulakos & Battista (2020)	• Overview of performance evaluation literature • Industry Case Study	Process
Rao (2016)	• Overview of performance management • Methods of evaluation	Rating Purpose
Risher (2013)	• Importance of feedback in rating and coaching	Rating Employee Involvement
Schleicher et al. (2018)	• Performance appraisal and performance management	Purpose
Tyson & Ward (2004)	• Use of 360 Degree Feedback • Evaluation and development uses	Rating
Wigert & Harter (2017)	• Employee dissatisfaction with evaluation system • Feedback	Rating Employee Involvement
Wilkie (2019)	• Feedback	Employee Involvement

1. Background

Research by Wigert and Harter (2017) notes, “traditional performance management systems are broken” and explains that “companies, leaders, managers, and employees have long participated in time-consuming, frustrating performance review that

have not yielded clear improvements in individual or organizational performance” (p. 3). Further research by Adler and colleagues (2016) agree that there are many issues with performance evaluations, but contend new research shows improvements in the field of performance evaluation. There are many challenges faced by employees and managers with traditional methods of performance evaluation.

Analysis by Wigert and Harter (2017) shows that only 29 percent of employees surveyed strongly agreed that they viewed their performance evaluation system as fair and only 26 percent strongly agreed they were accurate. In a separate study, Adler and colleagues (2016) note 90 percent of managers, employees, and HR practitioners felt that their organization’s performance management process failed to deliver a system that was accurate and effective.

A study by the Corporate Executive Council (2016) found that 70 percent of HR executives believe performance reviews do not accurately reflect employee contributions. Additional research shows traditional performance ratings have little correlation with actual business results, leading to demotivation and mistrust in the evaluation systems (Pulakos & O’Leary, 2011). Organizations are concerned with the return on investment on performance evaluations when it is estimated that managers spend about 210 hours and employees spend 40 hours on performance evaluation annually, estimating to cost a company of 10,000 employees \$35 million (Adler et al., 2016).

However, while there are many issues noted with performance evaluations, research also cautions organizations from dismissing performance evaluations completely (Adler et al., 2016). DeNisi (2016) notes the opportunities for organizations to deploy new solutions, highlighting the need to address employee concern to change behavior. DeNisi (2016) further explains performance evaluation is needed for the success of an organization but employees must perceive the system as fair, “transparent, and it is useful if the employee has some voice in the process” (p. 266). Regardless of the existence of formal system, performance is always being evaluated in everyday judgements between employees and managers; promotion, disciplinary, and reward decisions are made and must be based in tangible and defensible criteria (Adler et al., 2016).

2. Evaluation Purpose

This section describes the influence performance evaluation purpose has on the ratings provided by raters and the impact of competing objectives on evaluation quality. The general purpose of performance evaluations depends on the needs of the organization but should “be determined by considering business needs, organization culture and the system’s integration with other human resource management systems” (Pulakos, 2004, p. 3). Overall, the purpose of performance evaluations is to improve work performance (DeNisi, 2016), but additional research conducted by Cleveland, Murphy, and Williams (1989) suggests four sub-categories for evaluation purpose, shown in Table 6.

Table 6. Purposes for performance evaluation. Adapted from Cleveland et al. (1989).

Category	Description
Between person decisions	Promotion and salary decisions
Within person decisions	Feedback and identifying training needs
Systems maintenance	Manpower planning and evaluation of HR Systems
Documentation	Criteria for validity studies and legal

“Within person” (development) systems focus on opportunities for training, mentorship, and feedback for job improvement, while “between person” (administration) systems utilize evaluations for pay, promotion, force shaping, and other administrative actions (Armstrong, 2008).

The purpose of evaluations in organizations influences raters and the effectiveness of the system. “The purpose of evaluations affects both rating processes and rating outcomes” (Cleveland et al., 1989, p. 130). Evaluation purpose determines “what kinds of information a rater searched for and how that information was stored in memory, which affects the actual decisions made” (DeNisi, 2016, p. 263). Complications arise when organizations fail to focus the purpose of their evaluations clearly on their goals. Pulakos (2004) warns, “a performance management system that attempts to achieve too many objectives is likely to die of its own lack of focus and weight” (p. 2).

a. *Influence on Rater*

Evaluation purpose impacts how the rater is influenced in several ways. First, the rating score given by the rater is impacted by the rater's personal beliefs, political agendas, and rating goals (Greguras, 2003). Pulakos (2004) notes that "ratings used for decision-making tend to be more lenient, with most employees receiving ratings on the high end of the scale. Ratings for developmental purposes tend to be more variable, reflecting both employee strengths and development needs" (p. 3). Cleveland and colleagues (1989) further explains that "raters assess the consequences of each purpose and then complete the appraisal form with the single most important purpose in mind, de-emphasizing all others...it is unlikely that the same performance appraisal will provide valid information" (p. 131). Problems with quality and accuracy arise when purpose impacts ratings and it "becomes difficult to get honest and valid ratings when raters have competing goals" (Passmore, 2016, p. 447).

Additionally, Greguras et al. (2003) highlights that "different rating purposes may conflict with the rater's typical role in the organization" (p. 3), which may impact the quality of the evaluation. Cleveland et al. (1989) found that 70 percent of organizations within their study utilized performance evaluations for the dual purpose of "between" (administrative) and "within" (development) comparisons. The purpose of evaluations impacted evaluation quality particularly in cases where the rater was subordinate to the ratee; Greguras et al. (2003) found subordinates were more comfortable and supplied higher quality evaluations when used for developmental purposes. However, peers and subordinates may not feel comfortable providing feedback "impacting administrative decisions that typically are made by supervisors...and violate the typical organizational hierarchy" (Greguras, p. 3).

Furthermore, Greguras et al., (2003) highlights how different purposes of ratings require different levels of cognition and information processed because "raters who provide administrative ratings likely make evaluations of the target's performance relative to other ratees, whereas raters who provide developmental ratings likely make evaluations about the target's absolute level of performance (p. 4). When managers must make judgements that focus on both development and administrative decisions, one

objective must yield to the other; managers require different skills when rating and making decisions based competing purposes, which impacts the quality and accuracy of evaluations (Pulakos, 2004).

b. Objective Overload

Employee dissatisfaction with performance evaluations partially stem from the evaluation targeting too many objectives, (DeNisi, 2016). Research by DeNisi (2017) further highlights that ensuring the employees understand how and why the organization plans to meet objectives through their evaluations aids in transparency and building trust. Research by Cleveland et al. (1989), found “that many organizations require raters to provide ratings that will be used for potentially incompatible purposes, such as developmental feedback and salary administration” (p. 134). Particularly focusing on the “within” and “between” aspects of performance evaluations, Pulakos (2004) further highlights, “while performance management for purposes of decision-making and employee development are certainly related, these two objectives are rarely supported equally well by a single system” (p. 3). According to DeNisi (2016), feedback plays a particularly important role in addressing these issues because communication allows for the employees to pinpoint, with the raters, where the need to change behavior or improve performance is tied to the purpose of the evaluation.

3. Employee Involvement

This section discusses the relationships between the people involved in the performance evaluation and how their interactions impact the PES. This section focuses on the roles evaluation training, bias, and feedback play in the quality of performance evaluations. Armstrong (2008) emphasizes that one of the most important aspects of a PMS is the relationship between managers and employees, noting that they must act as a team ensuring clear goals are set and they jointly act “in monitoring performance and in taking action” (p. 5). Making the connections between work and desirable outcomes increases motivation and performance, even if the goals were not met (Pulakos, 2004).

In traditional systems, raters are generally managers; however, newer models, such as multi-rater feedback solicits input from superiors, peers, subordinates, customers,

and oneself (Armstrong, 2008). Regardless, it is the responsibility of raters to take seriously the appraisal process, as the system is dependent upon capturing the best measure of an individual's performance (Latham et al., 2005). Key components of employee involvement within the PES involve proper training of the system users, a system of feedback, and understanding the rater bias effect on evaluations (Rao, 2016). A study by Colquitt (2001) explored perceptions of fairness within performance evaluations and their influence on retention and performance. Factors including regular feedback and discussions informed system users, and transparent behaviors resulted in more positive attitudes toward the evaluation system (Colquitt, 2001).

a. Training

In a study by Lawler et al. (2016), data supported the importance of training to improve the performance evaluation skills for both the rater and the ratee. Common training areas include: (1) rater error, focusing on teaching raters common rating errors, (2) performance dimension training, focusing on teaching raters what performance aspects will be evaluated, (3) behavioral observation training, focusing on what behaviors are relevant to performance evaluations, and (4) frame of reference, focusing on “which behaviors coincide with each dimension of performance, how to judge behaviors with respect to their effectiveness, and how to combine these judgements into performance ratings” (Passmore et al., 2014, p.443).

Research by Murphy (2008) found that managers in many instances are significantly lacking in training to carry out performance evaluations. Adler et al. (2016) additionally discovered that “rating guidelines are often confusing, poorly understood, and inconsistently applied across raters and, within raters, across ratee” (p. 239). However, evaluation and feedback quality improved when the organization adopted a training program; training aided in raters dealing with uncomfortable discussions, generating more meaningful feedback, and improving the relationship between the manager and employee (Lawler et al. 2016).

b. Bias

A lack of objectivity is a weakness of traditional performance evaluations, as they rely upon the judgements and perceptions of people (Mondy & Martocchio, 2016). Bias is a common shortcoming within performance evaluations, with common biases and errors including: halo error, central tendency, recency, leniency/severity, personal bias (stereotyping), and evaluation manipulation. According to Latham et al., (2005), “appraisals are more often a reflection of the appraiser’s overall biases than they are of the performance of an employee”, which erodes employee confidence in the evaluation system (p. 80). Table 7 describes common biases seen in the evaluation of employees.

Table 7. Common performance evaluation biases.
Adapted from Rao (2016).

Bias	Description
Halo/Horn Error	Raters generalize an employee’s performance (positive or negative) based off a single event or quality
Leniency/Severity	Raters give underserved high ratings (leniency) or overly strict ratings (severity)
Central Tendency	Raters incorrectly rate employees at “average” levels, even if their performance merits lower and higher rating
Recency	Raters utilize employees most recent performance vice considering the entire evaluation period
Personal Bias/Stereotyping	Raters incorrectly apply their own perceptions of ideal employees in their ratings, leading to unfair judgements

In addition to the biases described in Table 7, managers may manipulate employee performance evaluations for a variety of reasons. While performance evaluations are a tool to meet organizational needs, managers may manipulate the evaluations themselves in order to achieve a desired personal objective (Mondy & Martocchio, 2016). Figure 10 lists and describes common reasons for PES manipulation.

Overrate

- Get more money for employees.
- Encourage employees.
- Avoid painful outcomes for him/herself by not rating someone low (e.g., the work required by a disciplinary process).
- Avoid hurting employees now or in the future by giving them low ratings.
- Avoid confrontation with an employee by giving him or her a rating lower than he or she expect.
- Get rid of a bad employee so he or she is marketable elsewhere within the company.
- Make themselves look good. Low ratings might imply the supervisor is not doing a good job.
- Get employees access to opportunities that are dependent on higher performance ratings (e.g., development opportunities).

Underrate

- Make an example out of an employee. A supervisor may rate employees lower than they deserve to punish them or teach them a lesson.
 - Get rid of an employee. Supervisors may rate someone lower than he or she deserves simply to create the necessary documentation and evidence to fire someone they don't like.
 - Send a message. A supervisor may rate an employee lower simply to send the message that the employee should leave.
-

Figure 10. Common reasons for performance rating manipulation.
Source: Pulakos (2004).

c. Feedback

On-going feedback and effective communication are the cornerstone to a successful PES (Mondy & Martocchio, 2016) and is designed to cultivate performance improvement and development (Armstrong, 2008). Successful, consistent feedback is linked to increasing engagement between the manager and the employee, which promotes better employee performance and empowers managers (Risher, 2013). However, in their study, Wigert and Harter (2017) found that the majority of organizations maintain the traditional timelines of annual performance feedback; 74 percent of surveyed employees reported receiving performance reviews once a year or less, and 14 percent reported receiving a formal review of performance every six months. Figure 11 illustrates the difference in accuracy seen in comparing intermittent and on-going feedback.



Figure 11. Traditional intermittent feedback system versus on-going feedback system. Source: Corporate Leadership Council (2016).

According to Corporate Leadership Council (CEB) (2016) research, nine out of ten managers are dissatisfied with how their companies conduct annual performance reviews. Part of the issue with annual reviews comes from having to identify past performance over a long period of time (Pulakos, 2004). With a focus on the past, performance evaluations do not foster productivity, performance improvement, and create an “uncomfortable dynamic between managers and employees” (Wilkie, 2019). Additional research shows that the annual evaluation process makes employees defensive and results in a decline of performance (Hirsh & Inzlicht, 2008). In their study, Hirsh and Inzlicht (2008) further highlight that “individuals would rather receive clear but negative information than endure the discomfort of uncertainty” (p. 966). When feedback is on-going and well-managed, Risher (2013) further adds that “year-end reviews would have a very different dynamic. It would significantly reduce the potential for disappointment and surprise. Since weaknesses would have been discussed, the year-end discussion can focus on developmental plans (p. 129).

A study by Elikier et al. (2006) further found that employees responded favorably in “high quality exchange relationships” (p. 547) where employees were given the opportunity to voice their thoughts and managers exhibited active listening while validating employee input when discussing final evaluation decisions. Further study by Aguinis (2009) notes the power of feedback in fostering positive attitudes within the organization and the impact on improved performance. It is noted that that feedback may not be effective in situations where there has been steady conflict and feedback may lead

to further contention; however, “when managers work to create common ground throughout the year, this can be a social foundation on which to build aligned efforts, understanding, and more positive reactions to PA systems” (Eliker, 2006, p. 548).

Additionally, ongoing feedback supports growth and development in employees, allowing them to focus on bettering their work performance (Risher, 2013). Lawler et al. (2016) further argues for the importance of ensuring that developmental and administrative discussions are held at separate times “because of the fact that appraisal discussions are often difficult to hold and can make it difficult for individuals to focus and participate meaningfully in a discussion of their development” (p.193). Technology aids organizations in their ability to conduct real-time feedback sessions, allowing managers to take action in the event that performance deficiencies need to be addressed and also highlight exceptional performance as it happens (Pulakos, 2004).

4. Methods of Evaluation

This section discusses two evaluation methods—traditional systems of evaluation and multisource feedback. A 2017 Gallup study found that traditional methods for evaluating employees failed to meet the goals of motivating employee performance with only 14 percent of surveyed employees strongly agreeing that performance reviews motivated performance improvement (Wigert & Harter, 2017). Research and organizations have identified employee mistrust in the validity of their performance metrics, showing that what has worked in the past may not be the best method for the current employment environment (Adler et al., 2016). 360-degree feedback systems have gained in popularity over the decade and focus on drawing performance feedback from superiors, peers, and subordinates (Adler et al., 2016). The next section discusses the traditional forms of evaluation and the multisource 360-degree feedback method.

a. Traditional Evaluation Methods

Traditional performance evaluations generally evaluate employees once or twice a year on predetermined performance goals or standards (Passmore, 2016). Traditional methods utilize rating scales, critical incidence, essays, work standards, ranking, forced distribution, behaviorally anchored rating scales, and results-based methods (Mondy &

Martocchio, 2016). Table 8 describes the key characteristics of these traditional evaluation methods.

Table 8. Common traditional evaluation methods. Adapted from Mondy and Martocchio (2016).

Method	Description
Rating scales method	• based off defined factors, which are usually contained in five to seven categories; raters rate each category and can create composite score • uses adjectives like “outstanding, meets expectations, needs improvement” • ratings based on job related factors (quality and quantity of work) and personal factors (behaviors as interpersonal skills and traits)
Critical incident method	• written records of key favorable or unfavorable employee work actions • evaluation culmination of critical events throughout evaluation period
Essay method	• rater writes narrative describing employee’s work performance • heavily dependent on rater writing ability
Work standards method	• employee rated on preset standard or level of production • generally seen in production jobs striving for uniformity
Ranking method	• rater ranks all employees within a group based off performance
Forced distribution method	• rater assigns individuals to a limited number of categories based on performance • similar to normal distribution
Behaviorally anchored rating scale method (BARS)	• combines traditional rating scales and critical incident methods • different performance levels shown along scale, describing employee job behavior
Results-based system	• objectives agreed upon by rater and employee • evaluation focuses on accomplishment of the objective

b. Multisource “360-Degree Feedback” Evaluation Method

The 360-degree feedback system (also called multisource assessment or multi-rater feedback) involves evaluation input from multiple numbers and different levels of people from within the organization as well as external sources (Armstrong, 2008). Different from traditional methods of soliciting input from supervisors only, this method includes rankings from “senior managers, the employee himself or herself, supervisors, subordinates, peers, team members, and internal or external customers” (Mondy & Martocchio, 2016, p. 244). Figure 12 illustrates the different sources of feedback in the 360-degree feedback evaluation.

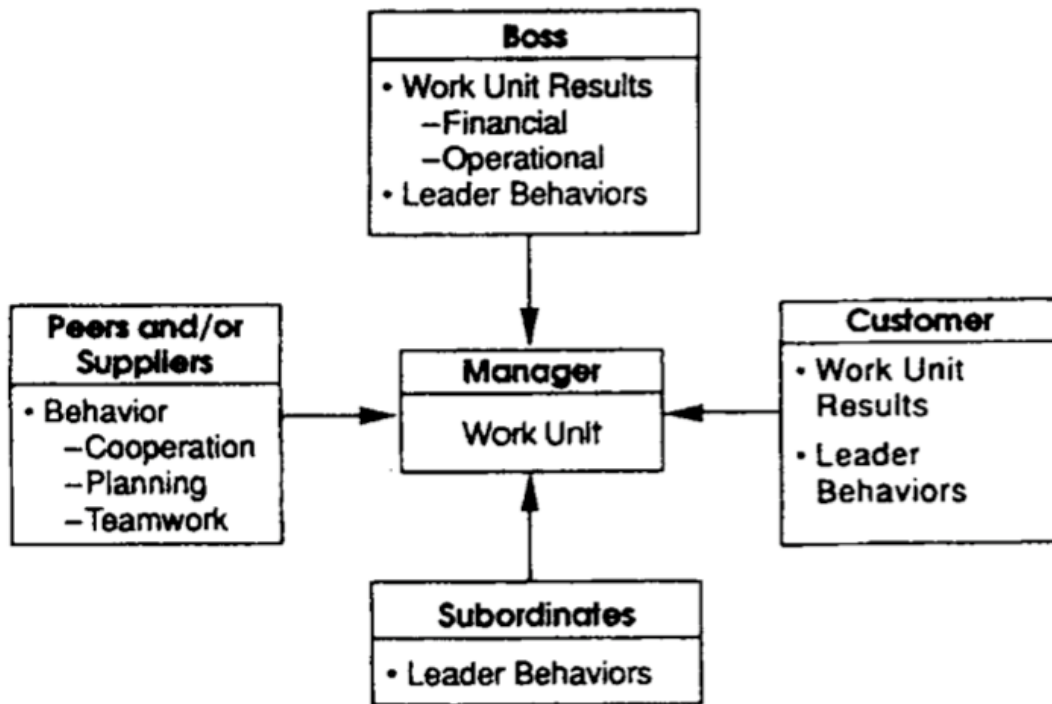


Figure 12. 360-degree feedback data sources. Source: London and Beatty (1993).

The 360-degree feedback system grew from issues with traditional evaluation, seeking to focus on improving workforce performance, enhancing development and feedback, and improving the measurement of performance (London & Beatty, 1993).

Armstrong (2008) notes only 11 percent of organizations surveyed in 1997 used 360-degree feedback; however, this method gained in popularity throughout the 2000s with Mondy & Martocchio, (2016) pointing to nearly 90 percent of Fortune 500 companies employing multisource feedback for evaluation or development. Advantages of the 360-degree feedback system include reducing common errors by having more than one level of input and providing a more objective measure of performance (Mondy & Martocchio, 2016).

Feedback focuses on “skills needed across organizational boundaries” (Mondy & Martocchio, 2016, p. 244). In their study of 360-degree feedback evaluations, London and Beatty (1993) found that by involving “reports of behavior, as well as judgements of performance based on work unit results”, rating scales could be used to allow for comparisons (p. 358). There are several issues with 360-degree feedback seen in “gaming the system” and in peer (when peers are competitors for rewards), they may “intentionally distort the data and sabotage the colleague” (Mondy & Martocchio, 2016 p. 244). However, Latham et al. (2005) notes that “overall, multisource feedback provides a comprehensive way of appraising employees” (p. 82).

A study by Tyson and Ward (2004) found that overall, the 360-degree feedback process was found to be perceived as successful for “improving openness, providing feedback, and involving many different raters” (p. 220). As a tool used for development, employees found benefit when the feedback with integrated with coaching, aiding the mindset of fairness by allowing open, actionable discussion (Tyson & Ward, 2004). However, Tyson and Ward (2004) also warn that it is important to differentiate between the use of 360-degree feedback as an evaluation or development tool; for evaluation purposes, the method requires the addition of more objective, evidence-based outcomes.

5. Rating

Performance rating, the measurement of individual or team performance, can have meaningful and beneficial impacts within an organization; however, rating must “share at least some meaningful true score variance with actual job performance” (Adler et al.,

2016, p. 237). This section describes literature regarding the difficulties in rating accuracy and the challenges ratings present in managing employee performance.

a. *Rating Accuracy*

As discussed in previous sections, accuracy in rating employee performance can never be perfectly objective (Passmore, 2013). Adler et al., (2016) acknowledges methods that are subjective are “the only feasible method for attacking” the problem of performance measurement (p. 238). However, attempting to be as accurate as possible by presenting limitations openly allows for better employee engagement and trust in the system (Colquitt, 2001). Latham et al., (2005) discusses the complicated nature and fundamental flaws of using ratings for comparison among employees:

Measuring the performance between managers and supervisors by using numbers and treating them as the properties of an interval scale (additive, multiplicative, subtractive, and divisive) has serious flaw. When two departmental managers are assessing their juniors, each one of them is using his/her own frame of reference and the circumstances under which each manager or employee functions are not often comparable by virtue of the function and other factors associated with the work. (p. 84)

Addressing the need for evaluations to be transparent and clear, Latham, et al. (2005) notes “hostility toward performance appraisals and the coaching process often occurs when people believe that they are being evaluated on the ‘wrong things’, or on indices over which they have little control” (p.80). Latham et al. (2005) further argues that “no two numbers are comparable in appraisals” because each evaluation contends with unique bias and motivations from the rater (p. 85). Still, even without formal evaluations, employees are still being evaluated, even if informally (Adler et al., 2016).

One remedy to employee perceptions of being evaluated on the ‘wrong things’ is noted by Risher (2013), by exploring evaluations specialized to the type of jobs held by employees. Looking at a “mixed group of craftsmen—carpenters, masons, painters and so on—we would never consider using the same skill profile...the idea of using the same competencies for a mixed-occupation group makes no sense if the goal is to give managers a tool they can effectively use to coach and evaluate employees” (Risher, 2013, p. 129).

b. Managing Performance

Research by DeNisi and Pritchard (2006) highlights the importance of performance evaluations as a tool to inform managers on the best way to improve employee performance. They contend that while accurate appraisal of employee performance is important, the more prudent focus should be on focusing evaluations on “results that are under the control of the employee, apply and communicate clear, consistent standards to the employees, and allow employees voice and participation in the development and/or application of the appraisal system” (DeNisi & Pritchard, 2006, p. 269). This follows the research of Colquitt (2001) in recommending designing performance evaluation strategy that enhances the perception of fairness in the system.

Adler et al., (2016) explores the benefits in quantifying and differentiating performance for reports and promotion functions; research shows that strong performers are attracted to organizations that differentially reward excellent performance, and weak performers will seek employment elsewhere. However, Risher (2013) explains that it is important for managers to continually monitor and update their performance standards to ensure the desired performance is being rewarded. Adler et al. (2016) further explores research regarding improvement of rating scales; in an effort to improve the validity of the rating scales, behaviorally anchored scales are shown to have little impact on performance improvement or rating accuracy. Further, Adler et al. (2016) cites additional research showing “that rating scales probably do not have a large effect on the quality of rating data” (p. 227). In research conducted by Murphy (2008), findings point to rating influence by other sources, including intentional rater distortion, than the employee’s performance, regardless of the type of scale used.

Research by Schleicher et al. (2018) notes that current research is lacking on the best methods for employee observation to ensure feedback is rooted in manager observations. Additionally, research notes the issues with relying fully on one rater, as relationships differ among employees based on their roles; understanding the raters’ role as an observer of performance and defining the best methods for capturing accurate, distinguishable ratings is still an ongoing challenge (Schleicher et al., 2019). Weighing the implications of eliminating performance evaluations from an organization, Adler et

al., (2016) admit that the research is complicated and lacks a clear, definitive path for the best way to evaluation performance. Citing a myriad of research on possible methods for improving performance evaluations, Table 9 summarizes the findings of Adler et al. (2016).

Table 9. Recommended methods for improving performance evaluations.
Adapted from Adler et al. (2016).

Recommendation	Description
Calibration	Managerial meetings to justify and compare desired performance ratings to ensure accuracy and applicability across the different organizational units
Competency modeling	Establishes defined behavioral expectations at each rating level for the job competencies
Multi-rater 360 feedback	Can be delivered formally or informally, with advantages of increasing rater reliability (due to multiple input), reduced bias, and more accurate
Reduce administrative burden	Reduces time and cognitive capital spent on form filling by making the evaluation process more simple and less demotivating
Performance review panels	May be time consuming but may encourage standardization and better talent selection

6. Evaluation Process

Pulakos (2004) describes the importance of using performance evaluation as a meaningful tool within a PMS, but more important is how “effectively it is used and how seriously managers and employees take it” (p. 21). This section discusses the importance of successful system implementation, system audit, and system transparency in maintaining a successfully functioning PES.

a. *System Implementation*

Implementation and integration of performance evaluation within an organization's PMS is often where failures occur (Pulakos & O'Leary 2011). Often, good ideas and quick fixes are well-meaning and aimed to achieve successes; however, recent research by Pulakos and Battista (2020), demonstrates several implementation issues resulting from lack of focus on the bigger picture. In many organizations, upper-level management becomes fixated on the evaluation form, overtaking the goals of the PMS and conversations about "goal clarity, alignment, or feedback...lost in the debates concerning features of the form (Pulakos & Battista, 2020, p. 11). Additional research by Lawler (2016) discusses the importance of integration:

What organizations need to do is create performance management systems that are integrated with the other human resources management systems they have and the overall talent management strategy of the organization. Indeed, they need to go beyond just integrating it with the talent management practices of the organization; they need to make sure it is integrated with the strategy of the organization. There has always been, and our data say there continues to be, a strong correlation between the effectiveness of performance management systems and the degree to which they are driven by the business strategy of an organization. (p. 88)

b. *System Audit*

In addition to thoughtful integration, performance appraisal processes must be reviewed and audited to ensure it is meeting the requirements of the organization (Lawler, 2016). Lawler (2016) further states "this audit should include survey data as well as an analysis of the quality of the materials produced by the appraisal process. Are they complete? Do they cover the right materials? Are they produced in a timely manner? All of these questions need to be answered to have an effective performance management process in an organization" (p. 89). An audit allows for evolutionary change to maintain a relevant system instead of revolutionary change of a system that is entrenched in outdated methods.

c. System Transparency

A study by Kavanagh et al. (2007) sought to review the impact of performance evaluation process on perceptions of fairness in the system. Their findings strongly pointed to enhanced perceptions of fairness from employee “participation in the appraisal process, a belief in the neutrality of the supervisor, and knowledge of the appraisal system” (Kavanagh, 2007, p. 147). Further, Kavanagh et al. (2007) summarizes two consequences of this finding: (1) If employees believe the evaluation system is fair, they are more likely to accept decisions based on the process, and (2) employees will be more satisfied and more likely to be committed to the organization. This research shows that investing in a system that is transparent and understood by the employees yields potential talent management benefits to the organization (Kavanagh, 2007). Maylett (2009) contends that the 360-degree feedback tool can be used for both processes, but highlights the need for organizations to understand that development and evaluation are not interchangeable purposes. Further, “it is important to note...that 360s for performance and 360s for development are likely to produce completely different outcomes” (Maylett, 2009, p. 58).

C. INDUSTRY BEST PRACTICES

The following sections discuss the performance evaluation practices at Google and Deloitte. These companies were selected for study because they both recently overhauled their performance management systems and redesigned their performance evaluation. This section focuses on their lessons identified through the transformation process and reviews their methods for conducting performance evaluation.

1. Google

Founded in 1998 by Stanford University Ph.D. students Larry Page and Sergey Brin, Google began as an internet search engine and has since expanded into the technology giant it is today, (Brant, 2011). The company boasts the most visited website in the world and offers services for work productivity, email, cloud storage, video chat services, maps and navigation, fiber optics, and smartphones (Tran, 2017, p. 4). Google

has built a very strong, collaborative culture and works hard to “attract young, ambitious self-starters and original thinkers” (Garvin, 2013, p. 2).

Google’s vice president for “people operations”, Laszlo Bock, details the company’s HR practices in *Work rules! Insights from inside Google that will transform how you live and lead*. Bock (2015) explains the major issue with “performance management as practiced by more organizations has become a rule-based, bureaucratic process, existing as an end in itself rather than actually shaping performance. Employees hate it. Managers hate it. Even HR departments hate it” (p.150).

Bock’s first steps in managing HR at Google began with hiring an analytics team with the purpose to “approach HR with the same empirical discipline Google applied to its business operations” (Garvin, 2013, p.2). Project Oxygen began from a data driven research approach, seeking to pinpoint the specific management traits in employees and examine the foundational behaviors that yielded success (Brant, 2011). The research focused on concluding if managers within their organization were required for their organization to succeed; through survey data and qualitative inputs, Google proved managers were necessary and “identified, described, and institutionalized their most essential behaviors” (Garvin, 2013, p. 4). Of those behaviors, Project Oxygen found that coaching and connecting with the employee was at the top of the list of successful behaviors:

At a company like Google, where the staff consists almost entirely of “A” players, managers have a complex, demanding role to play. They must go beyond overseeing the day-to-day work and support their employees’ personal needs, development, and career planning. That means providing smart, steady feedback to guide people to greater levels of achievement—but intervening judiciously and with a light touch, since high-performing knowledge workers place a premium on autonomy. It’s a delicate balancing act to keep employees happy and motivated through enthusiastic cheerleading while helping them grow through stretch assignments and carefully modulated feedback. When the process works well, it can yield extraordinary results. (Garvin, 2013, p. 4)

This project impacted evaluations within Google, showing feedback and employee connectedness were essential talent management tools available within the PMS (Brant, 2011). In research on Google’s organizational culture, Tran (2017) found that

“connection is what transforms a dog-eat-dog environment into a sled dog team that pulls together” (p. 11).

Among industry top performers, successful HR practices are examined and often borrowed, as seen in Google utilizing Intel’s concept of Objectives and Key Results (OKRs), to set organizational goals (Bock, 2015). Aggressive, specific, and measurable goals are set quarterly at an organizational level, with subordinate teams required to set their own goals; a transparent, quarterly report is presented to Google employees to discuss OKRs and company performance (Bock, 2015).

Bock (2015) notes the significant time spent on the rating system before 2013, stating “our managers were spending thousands of hours every three months assigning ratings that were ludicrously precise but that weren’t an accurate basis for determining pay” (p. 157). In the course of transforming Google’s 41-category ranking system based on a 1.0-5.0 scale, Bock (2015) found that in designing the new system “consensus was impossible, people took performance management seriously, and experimentation was vital” (p. 159). “Ratings are tools, simplifying devices to help managers make decisions about pay and promotion” but these tools must be fair and well understood (Bock, 2015, p. 167).

In the design process for the updated system, Bock notes his team went through an iterative and intense process in exploring the options that would work best for Google; at the heart of the present system is the feedback the team took in and instituted to design the current evaluation method (Bock, 2015). OKRs are set quarterly, with one on one check-ins held throughout this time to share feedback; formal annual performance reviews are held with employees in addition to a formal intermediary review (Tran, 2017).

Google now approaches evaluations as a way to calibrate performance and uses multisource feedback to present tangible, coachable areas for improvement; this feedback is use for management’s evaluations of employees, which are conducted twice a year (Bock, 2015). Self-evaluation, multisource feedback, and monthly coaching culminate to inform the manager’s decisions in ranking employees. Bock’s (2015) rules for

performance evaluations are: “set goals correctly, gather peer feedback, use a calibration process to finalize ratings, and split rewards conversations from development conversations” (p. 177).

2. Deloitte

Deloitte, a “Big Four” accounting and professional services organization, estimated the time spent on evaluations was “1.8 million hours across the firm that didn’t fit our business needs anymore” (Orlando & Bank, 2016, p. 42). In 2013, Deloitte began redesigning their performance management model and “wanting to push the boundaries of traditional approaches, they removed goal setting, ratings, and calibration meetings and pushed a much less complex rating process that allowed project leaders to capture ‘snapshots’ of each professional’s performance” (Pulakos & Battista, 2020, p. 15).

Already a leader in innovation, Deloitte developed and built its own \$300 million learning facility, looking to invest in delivering the best learning and development opportunities to the top talent at Deloitte (Pulakos & Battista, 2020). Deloitte’s Managing Director and Chief Learning Officer for Leader Development and Performance, Jeff Orlando, highlights “we realized that our then-current process for evaluating the work of our people was increasingly out of step with our objectives” (Pulakos & Battista, 2020, p. 101). The following describes Deloitte’s approach to performance evaluations prior to their updates:

More than likely, the performance management system Deloitte has been using has some characteristics in common with yours. Objectives are set for each of our 65,000-plus people at the beginning of the year; after a project is finished, each person’s manager rates him or her on how well those objectives are 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, p. 2)

To move forward on a new system, Deloitte conducted a study on their evaluation system and PMS based on previous research and the interview conducted with their employees (Orlando & Bank, 2016, p. 43). The foundation of their new system’s design

came from Deloitte’s research showing they needed to give employees “a chance to use their strengths every day, knowledge of what is expected of them at work, belief in the mission or purpose of the work they are doing” (Pulakos & Battista, 2020, p. 103). Figure 13 shows Deloitte’s performance management model.



Figure 13. Deloitte core performance design. Source: Pulakos and Battista (2020).

Deloitte spent two years testing and piloting the future performance system, taking in performance data, interviews, and design methods to build a better system for their company (Buckingham & Goodall, 2015). According to Buckingham and Goodall (2015), three objectives guided the design of the system: (1) Being able to differentiate and recognize performance, (2) accurately measuring performance, and (3) focus not on skills but future actions for the individual being rated.

Deloitte designed a system to be agile, with performance evaluations completed following project completion or at least every quarter (Buckingham & Goodall, 2015). Now, “snapshot questions” are asked of team members who are rating the performance of their team member on a five-point scale, including: (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, (2) given what I know of this person’s

performance, I would always want him or her on my team, (3) this person is at risk for low performance, and (4) this person is ready for promotion today (Buckingham & Goodall, 2015). An aggregate score for performance is created by averaging snapshot scores and weighting them by the project hours; these scores from the team are plotted on a scatterplot (Pulakos & Battista, 2020). Figure 14 is an example scatterplot used by Deloitte, where each dot represents a person who has been rated by team members, but not their manager.

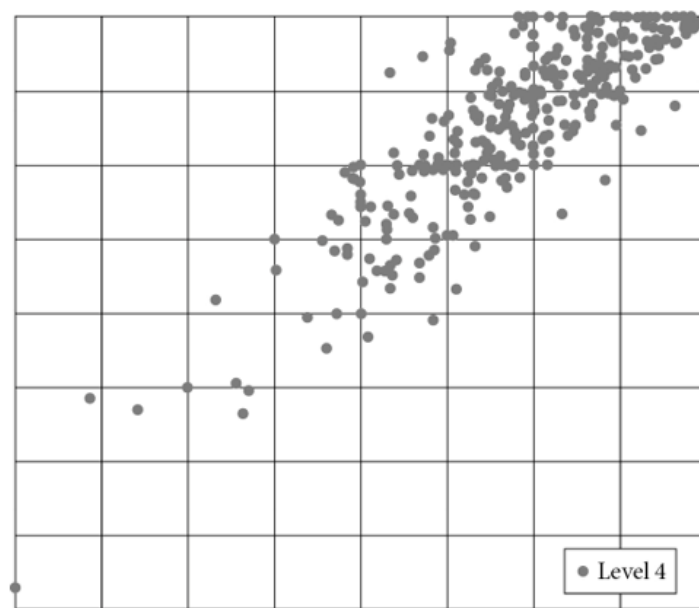


Figure 14. Sample scatterplot of employee project performance.
Source: Pulakos and Battista (2020).

Additional features of Deloitte's evaluation system include weekly check-ins with team members and "pulse surveys", which functions as a developmental only tool to aid and guide feedback outside of performance reviews (Orlando & Bank, 2016). Deloitte reports after a year of implementation, employees report greater levels of engagement, less voluntary turnover, and increased growth and development from the new evaluation system (Pulakos & Battista, 2020).

D. CONCLUSION

This chapter begins with a discussion of performance, performance evaluation, and performance management and presents a conceptual framework for analysis of performance evaluation practices documented in scientific and industry literature. This researcher synthesized research and trends within the performance evaluation literature in the categories of evaluation purpose, employee involvement, evaluation methods, rating considerations, and evaluation process. In addition, performance evaluation practices were analyzed through research into recent performance evaluation transformations at Google and Deloitte.

The researcher's analysis identified four narrative themes that characterize successful performance evaluation practices:

1. Define the Evaluation Purpose
2. Create a Culture of Communication
3. Foster a Perception of Fairness
4. There is No "One Size Fits All" Solution

The insights supplied by this examination inform recommendations of successful practices that the Navy could implement to improve the PES. Chapter VI provides a discussion of these key themes and discusses how these practices could be considered for implementation in the Navy's current PES.

V. CHARACTERISTICS OF SUCCESSFUL PERFORMANCE EVALUATIONS

This researcher's analysis of the scientific and industry literature identified four narrative themes that characterize successful performance evaluation practices. The themes are: (1) Define the Evaluation Purpose, (2) Create a Culture of Communication, (3) Foster a Perception of Fairness, and (4) There is No "One Size Fits All" Solution. Additionally, the findings of this research indicate that there are practices which can be put in place to support a future successful Navy PES. This chapter provides a discussion on the key themes and successful practices that could be useful for implantation in the Navy's PES.

1. Define the Evaluation Purpose

Research conducted by Passmore (2016) highlighted the different purposes of performance evaluations by suggesting that "performance appraisals can be seen to possess the properties of a Janus head, having multiple faces within organizations" (p. 439). While the main goal of any performance management and evaluation system should be to optimize employee performance (Pulakos, 2004), organizations must understand what "face" they will chose for their evaluation purpose. Cleveland et al., (1989) describes the two main purposes of evaluations as development (feedback, job improvement, identification of training needs) and administration (promotion, firing, salary decisions). The research indicated that (1) a single system designed to meet both purposes of development and administration does neither well, and eventually leads to one purpose overtaking the other (Pulakos, 2004; DeNisi, 2016), (2) the quality of feedback diminishes when raters must prioritize administration or development (Cleveland et al., 1989; Greguras, 2003), and (3) rating purpose impacts both rater observations and recorded results (Cleveland et al., Pulakos, 2004). Clear delineation of the type of purpose the performance evaluation seeks to inform particularly impacts several shortcomings within the Navy's PES.

a. *Navy Evaluation Competing Objectives*

The case study into the Navy PES shortcomings revealed the current system faces problems with competing objectives. The performance evaluation focuses on the SM's past performance over the course of the year, but also seeks to address future promotion potential (Poulin, 2016). In addition to these actions, the same NAVPERS form used for formal evaluations is also used for mid-term counseling. Research (Pulakos, 2004; DeNisi, 2016) suggests that using a single system to perform both purposes of development and administration is problematic. The evidence seen in the Navy case study mirror these findings; the Navy PES does not promote open feedback and the evaluation is an inaccurate reflection of performance (Poulin, 2016; Goforth, 2016; Moran, 2014; Burke, 2017).

Clearly defining and differentiating the purpose of the Navy's PES will help make the process more accurate and transparent (DeNisi, 2016). While the structure is in place to conduct counseling, and BUPERSINST 1610.10E highlights the importance of coaching and feedback, the Navy PES, in trying to do meet administrative and developmental goals with one system, has allowed the administrative function to take over (Poulin, 2016; Burke, 2017). Two defined systems, one for administrative and one for development, must be defined and clearly explained to the employees. While the systems may look similar, the conversations, procedure, and criteria used may be adjusted to meet the purpose of either formal evaluation or coaching.

b. *Navy Evaluation Reverse Engineering*

Problems with "reverse engineering" and "gaming the system" are noted within the case study of the Navy's current PES (Eyer, 2018). These problems stem from a lack of transparency within the evaluation system and pressure on the rater from outside influences, other than the SM's performance (Eyer, 2018; Cordial, 2017). Research shows that raters may present inaccurate measures of employee performance when left to weigh the importance of evaluations for either administrative functions or development; when pay or promotion are a factor, evaluation ratings are generally inflated (Cleveland

et al., 1989; Pulakos, 2004). In order to achieve the results that the rater desires, raters may over or underrate employees (Pulakos, 2004; Murphy, 2008).

A particular concern within the Navy PES is seen with RSCA management; rating purpose impacts the recorded results because the SM's performance is not used as the criteria for evaluation (Cleveland et al., 1989; Pulakos, 2004). By defining and fervently communicating the purpose and design of the evaluation system, the competing goals of raters is mitigated (Cleveland et al., 1989; Pulakos, 2004; Greguras, 2003). There are many issues with RSCA management; however, if the system better differentiates from the developmental aspect in counseling and supports administrative functions for the formal evaluation, CO's may be less inclined to treat their ratings as anything other than a strict rating of the SM.

2. Create a Culture of Communication

The topic of feedback is widely identified within performance evaluation literature as a necessary component of a successful PES (Pulakos, 2004; Rao, 2016; Adler et al., 2016; Wigert & Harter, 2017; Elikier et al., 2006; Mondy & Martocchio, 2016). However, simply providing a mechanism for feedback within the PES does not meet the mark. Organizations must create a culture of communication, where feedback is enhanced by training and demonstrated as a priority at all levels within the organization (Pulakos, 2004; Elikier et al., 2006; Aguinis, 2009).

Creating a culture of communication benefits organizational effectiveness in the following ways: (1) Increases employee performance through increased manager-employee engagement (Risher, 2013), (2) Promotes consistent employee performance throughout the reporting period (CEB, 2016; Pulakos, 2004; Hirsh & Inzlicht, 2008), (3) Fosters positive attitudes and increases employee retention (Elikier, 2006; Aguinis, 2009), and (4) Builds trust between the employees and managers (Pulakos & O'Leary, 2011; Colquitt, 2011; DeNisi, 2016; Tyson & Ward, 2004). The review of the scientific and industry literature suggests that many of the problems the analysis of the Navy case identified could be aided by building a culture of communication surrounding the PES

and organization as a whole. These issues include: minimal opportunities for feedback, prioritizing process over performance, and a lack of transparency.

a. Minimal Opportunities for Feedback

The analysis of the Navy case found one of the failing features of the Navy's current PES is the lack of ongoing, effective feedback (Poulin, 2016). The mechanism for feedback is designated in the Navy's instruction for conducting mid-term counseling; however, as pointed out by Poulin (2016), counseling is often overtaken by events and are often not constructive. Burke (2017) further highlights that it is difficult to build trust in a system when it supplies feedback twice (if that) a year. In addition, the quality of feedback is impacted when developmental feedback intermingles within the administrative evaluation; Tyson and Ward (2004) warn that developmental and administrative feedback should not be given in the same meeting. Research by Wigert and Harter (2017) shows that employees desire more engagement from their employers. Additional research by the CEB (2016) shows that frequent, ongoing feedback encourages more steady work performance by the employees.

Eyre (2018) highlights important feedback components detailed by Pulakos (2004), including providing specific feedback and developing goals for future performance. However, research additionally emphasizes the importance of communication as a two-way event, where the voice of the employee is heard and validated by the manager. Both Deloitte and Google focused their employee evaluation transformations on providing meaningful, frequent, developmental feedback throughout the evaluation cycle; both seek to deliver feedback and foster frequent communication about performance at least quarterly, if not more (Bock, 2015; Pulakos & Battista, 2020). Deloitte particularly focused on differentiating the feedback used for administrative purposes from that of developmental purposes (Pulakos & Battista, 2020). With this research in mind, the MSAF coaching tool is an excellent step toward providing actionable, ongoing feedback. While this tool allows for the ability to enhance feedback and drive more frequent conversations, follow-through and specific guidance for minimum periodicity requirements should follow the models of Google and Deloitte.

b. Process over Performance

The case study of Navy PES shortcomings indicate that current Navy systems hinder the execution of creating a culture of communication. First, the Navy's PES is 24-years-old, and does not support advanced workforce decisions making and feedback opportunities (Burke, 2017). With a system burdened by old technology and antiquated processes, the administrative requirements by the SMs who use the system become more focused on completing the required work (Burke, 2014). A cumbersome administrative system takes away from the time and motivation that should be spent in delivering feedback and engaging in meaningful conversations about employee performance (Snyder, 2002). Bock (2015), highlights the importance of simplicity within Google's evaluation system; users must have easy access to their tools and not become burdened by the process of completing evaluations or feedback. Research by Aguinis (2009) shows that when employees understand their evaluation system, they feel supported by its processes and more engaged by the organization.

c. Lack of Transparency

The Navy case study highlights the lack of transparency within the performance evaluation system (Moran, 2014; Burke, 2017; Eyre, 2018; Poulin, 2016). Deloitte highlights a significant success they encountered in the development of their current evaluation system. While the process took nearly two years, the HR leaders at Deloitte piloted a program that heavily depended on user surveys on what improvements could be made; risks were taken and testing was completed throughout to ensure the users were satisfied with the results (Pulakos & Battista, 2020). As the Navy builds its current system, HR leaders should solicit feedback from the test pilot participants, and should seek to include a larger scope of feedback from the fleet.

The use of 360-degree feedback is shown through research to enhance feedback and increase conversations among all employee levels (Tyson & Ward, 2004). This feedback tool involves more members of the organization, building the communication and increasing the frequency in which performance is discussed; in addition, by engaging many more members of the organization, the culture of communication is normalized and

becomes more accepted (Tyson & Ward, 2004; Pulakos, 2004). The research supports the MSAF coaching tool as a way to engage more members of the command and foster more frequent, open communication.

3. Foster a Perception of Fairness

Research by Latham (2005) emphasized “employee acceptance is critical to the implementation and on-going use of appraisals” (p. 80). Research further shows for employees to accept an evaluation system, they must perceive the system as fair (Colquitt, (2001); Pulakos, 2004; Adler et al. (2016); Tyson & Ward, 2004; Kavanagh et al., 2007). Research shows there are many ways in which organizations foster perceptions of fairness within their evaluation system: (1) Encouraging open, two-way communication (Tyson & Ward, 2004; DeNisi & Pritchard, 2006), (2) Establish evaluation training which focuses on evaluation skills and common biases, and (3) Pursue accurate rating techniques to meet organizational goals (DeNisi & Pritchard, 2006; Latham et al., 2005). The analysis of the Navy system shows that the current Navy PES faces challenges from SMs who do not believe the system is an accurate reflection of performance and is composed of unfair processes (Raimondo, 2009; Raimondo, 2017; Poulin, 2016; Hagner, 2003). The case analysis illustrates several shortcomings in the Navy PES which hinder a perception of fairness including inaccurate measures of performance and opaque processes.

a. Inaccurate Measures

Raimondo (2009) discusses the accuracy of performance evaluations and ease with which concerning behaviors may be omitted from a SM’s record; a report found that a CO recently fired for misconduct had no adverse statements or rankings in their fitness report, even after they was taken to Nonjudicial Punishment. If wrongdoing or poor performance are not accurately reflected, then performance evaluations in the current system will likewise be unable to distinguish talent (Moran, 2014). Adler et al. (2016) similarly highlights the issues with the subjectivity of performance evaluations, but encourages the system users to invest in the best methods available to rate employees.

Research by Latham et al. (2005) found that in addition to blatant manipulation of ratings, ratings often tell more about the rater than the performance of the employee.

While research by Schleicher et al. (2018) contends that consensus is lacking on the best ways to evaluate employees, the need remains to foster an environment of perceived fairness within the PES. However, research by Adler et al. (2016) and practice by Deloitte (Pulakos & Battista, 2020) highlight the need for an organization to constantly calibrate its evaluation processes. As the Navy seeks to improve its current PES, the PET-TM should focus on building a perception of fairness, starting from the development of the new evaluation system. Proceeding past the pilot program and into implementation, Navy HR leaders must solicit feedback resolve large-scale conflict.

b. Opaque Processes

Raimondo (2009) and Eyer (2018) both highlight the murky processes of the current Navy PES and its impact on evaluation quality and stakeholder trust in the system. Raimondo (2009) particularly warns of COs with the best of intentions who inadvertently negatively impact a Sailor's career. Further, Eyer highlights that while the PES is complicated and nuanced, training and consultation with the evaluation experts is the remedy. However, research warns against systems that are labor intensive, complicated, and lack clear instruction (Wigert & Harter, 2017). Research shows that the "employee voice", validation, active listening, and consistent touch points, all aid in fostering trust (Tyson & Ward, 2004). DeNisi and Pritchard (2006) further highlight that when the conversation regarding evaluation processes is transparent, it allows for the system to be questioned, and the disparities to be resolved; this leads to a perception of fairness of the system. Furthermore, Kavanagh et al., (2007) found that when employees believe the system is fair, they are more likely to accept the system outcomes. Lack of communication erodes trust and a perception of fairness (DeNisi & Pritchard, 2006). Research highlights keeping lines of communication open and ensuring processes are transparent is key to fostering a perception of fairness. A future Navy PES would benefit from simple, transparent processes that are apparent to all PES users.

4. There is No “One Size Fits All” Solution

As Lawler and colleagues (2016) point out, “performance management systems are complex and involve a number of features. To some degree, the right features are a function of fit with the design of organization and its strategy. Because of this, no size fits all situations” (p. 192). The research contained in this thesis support that there is no easy solution identifying the best practices an organization should use to conduct effective performance evaluations. As seen the performance evaluation transformations at Google and Deloitte, HR leaders cited the need study the problems with the current system, involve system users in solutions and testing, and approach the process as iterative (Bock, 2015; Pulakos & Battista, 2020).

To ensure an effective future PES, research suggests that successful implementation depends on (1) Focus of the integration between the performance evaluation and how it fits into the overall picture of performance management (Pulakos & Battista, 2020), (2) Rigorous testing and application of user feedback through a thoughtful iterative process (Bock, 2015; Pulakos & Battista, 2020), and (3) Periodically audited to ensure it meets the needs of the organization (Lawler, 2016). The Navy should continue its current work on performance evaluation transformation but ensure the decisions made are customized to fit the unique needs of the organization and incorporate the research findings for successful PES implementation.

VI. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSION

The Navy is well on its way to address the war for talent with the MyNavy HR transformation, with Sailor 2025 initiatives focusing on improving performance management within the Navy. One of the most important talent management tools, performance evaluation, is primed for transformation. This thesis conducted a case analysis of the current Navy PES and of the PET-TM initiatives meant to address the current system's shortcomings. In addition, this thesis reviewed the published industry and scientific literature on performance evaluations to identify main themes on successful performance evaluations. Finally, based on the case study analysis and on the findings from the literature, this thesis proposes evaluation system practices the Navy should consider in order to improve the current PES. This thesis addressed three research questions. The main findings for each are summarized as follows:

- **What are the shortcomings of the current Navy performance evaluation system?**

The findings from the analysis of the Navy case presented many shortcomings including: System age, process over performance, past vs future focus, lack of transparency, and inaccurate measure of performance. First, research showed that the 24-year-old Navy PES was not meeting the current and future demands of the Navy's operating environment (Raimondo, 2009; Moran, 2014). Findings also indicated that the Navy PES prioritizes process over performance with the system's adherence to manual administrative processes (Moran, 2014). Additional impacts from this shortcoming include inconsistent quality of performance evaluations, inadequate feedback opportunities, and misrepresented performance results (Snyder, 2002; Raimondo, 2009; Burke, 2017). Third, the Navy case study showed that the evaluation's focus on past performance limited the ability for raters to give feedback on future potential (Poulin, 2016). Next, the case study on the Navy PES found that a lack of transparency with the current system lead to a lack of trust between Sailors and their leaders and did not foster a steady flow of feedback due to the periodicity of evaluations (Moran, 2014).

Further, the system created an environment where evaluations could be biased by CO's forced distribution to achieve desired results for such purposes as RSCA management (Burke, 2017). Additional issues include clandestine signals directed to a selection or promotion board (Eyer, 2018) and the ability to distort rankings to mask misconduct (Raimondo, 2009). Lastly, findings from the Navy case study revealed shortcomings in the accuracy of the evaluation's performance measurement through bias and irrelevant rating criteria (Moran, 2014; Goforth, 2016).

- **What performance evaluation practices, key traits, and central themes are evident in scholarly and industry literature?**

Four narrative themes were found through a review of scientific and industry literature. Research within "Define the Evaluation Purpose" relates to whether the evaluation is filling an administrative or developmental purpose. The findings indicated that a single system designed to meet both purposes of development and administration does neither well (Pulakos, 2004; DeNisi, 2016). Additionally, the quality of feedback diminishes when raters must prioritize administration or development (Cleveland et al., 1989; Greguras, 2003), and rating purpose impacts both rater observations and recorded results (Cleveland et al., 1989; Pulakos, 2004).

The theme "Create a Culture of Communication" relates to the importance of an open, honest flow of feedback within an organization. The research found that employee performance increased and remained consistent with greater manager-employee engagement (Risher, 2013; CEB, 2016; Pulakos, 2004; Hirsh & Inzlicht, 2008). Additionally, findings indicated that a culture of communication builds trust and promotes positive attitudes, which increases employee retention (Eliker, 2006; Aguinis, 2009).

Findings from the theme "Foster a Perception of Fairness" indicate the importance of encouraging open, two-way communication, establishing evaluation training which focuses on evaluation skills and common biases, and pursuing accurate rating techniques to meet organizational goals (Tyson & Ward, 2004; DeNisi & Pritchard, 2006). Lastly, research within "There is No 'One Size Fits All' Solution" point to the differences among organizations and the need to develop a PES that is the right fit to the

organization. The findings indicated the importance the integration between the performance evaluation and performance management process (Pulakos & Battista, 2020). Addition research highlighted the need to conduct rigorous testing before system implementation as well as periodic audits to ensure it meets the needs of the organization (Bock, 2015; Pulakos & Battista, 2020).

- **What performance evaluation practices should the Navy consider in its performance evaluation transformation?**

Following a review of the Navy case shortcomings and the findings from research of scientific and industry literature, the following practices were identified for Navy consideration:

- (1) Establish two defined systems within the PES, one for administrative and one for development purposes,
- (2) Proceed with testing and implementation of the MSAF coaching tool,
- (3) Foster transparent, frequent communication with Sailors throughout the PES development and implementation process,
- (4) Review transformation lessons from Google and Deloitte on structure of evaluation cycle and feedback design, and
- (5) The PET-TM should follow an iterative process of rigorous testing and apply user feedback to ensure best possible integration to the PMS.

B. RECOMMENDATIONS

In addition to the recommendations stated in the previous section, this thesis recommends the PET-TM commission a study similar to the analysis conducted internally by Deloitte. The research shows that there are many complexities to performance evaluation. Therefore, this thesis recommends consulting with subject matter experts within the performance evaluation field throughout the design, implementation, and audit stages of the Navy's performance evaluation transformation.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Olander-Krane, R., & Pulakos, E. D. (2016). Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9(2), 219–252. <https://doi.org/10.1017/iop.2015.106>
- Aguinis, H. (2009). *Performance management* (2nd ed.). Prentice Hall/Pearson Education.
- Armstrong, M. (2000). *Performance management: Key strategies and practical guidelines*. London: Kogan Page.
- Baskarada, S. (2014). Qualitative case study guidelines. *The Qualitative Report*, 19(40), 1-18. <https://nsuworks.nova.edu/tqr/vol19/iss40/3>
- Bock, L. (2015). *Work rules!: Insights from inside Google that will transform how you live and lead* (1st ed.). New York: Twelve.
- Bryant, A. (2011). Google's quest to build a better boss. *The New York Times*. https://www.nytimes.com/2011/03/13/business/13hire.html?pagewanted=all&_r=0
- Burke, R. (2017). Interview by J. Serbu [Transcript]. Sailor 2025 - 40 Initiatives, *Federal News Radio*. Washington, DC. Retrieved from <https://www.navy.mil/navydata/people/cnp/Burke/Interview/CNP-170601-Sailor2025-FedNewsRadio.pdf>
- Carr, A., & Kline, K. (2016). *Encouraging effective performance management systems*. SIOP. <https://www.siop.org/Portals/84/docs/White%20Papers/PerfMgmt.pdf>
- Chisholm, D. (2001). *Waiting for dead men's shoes: origins and development of the U.S. navy's officer personnel system, 1793-1941*. Stanford, CA: Stanford University Press.
- Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, 74, 130-135.
- Corporate Leadership Council. (2016). *Performance management can be fixed: Driving lasting behavior change through experiential learning*. Gartner. <https://www.cebglobal.com/content/dam/cebglobal/us/EN/talent-management/performance-impact-solutions/pdfs/White%20Paper%20-%20Performance%20Management%20Can%20Be%20Fixed.pdf>

- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- 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>
- Department of Defense. (2018). *National Defense Strategy*. <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>
- Department of the Navy. (1815). *Rules, Regulations, and Instructions, for the Naval Service of the United States*. <https://collections.nlm.nih.gov/catalog.nlm.nlmuid-2575068R-bk>
- Department of the Navy. (1990). *United States Navy Regulations, Article 1129*. https://www.public.navy.mil/surfor/Documents/UNITED_STATES_NAVY_REGULATIONS_CHAPTER_11.pdf
- Department of the Navy. (2019). *Department of the Navy Strategic Guidance FY2020*. https://www.navy.mil/strategic/Department_of_the_Navy_Strategic_Guidance_FY2020.pdf
- DeNisi, A. S. (2016). Managing performance to change behavior. *Journal of Organizational Behavior Management*, 31, 262-276. doi: 10.1080/01608061.2011.619414
- DeNisi, A. S., & Pritchard, R. D. (2006). Performance Appraisal, Performance Management and Improving Individual Performance: A Motivational Framework. *Management and Organization Review*, 2(2), 253–277. doi: 10.1111/j.1740-8784.2006.00042.x
- 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>
- Eyer, K. (2018). Charting a course-write better paper. *U. S. Naval Institute Proceedings*, 144/2/1,380, 42-45). <https://www.usni.org/magazines/proceedings/2018/february/charting-course-write-better-paper>
- Garvin, D. A. (2013, December). How Google Sold Its Engineers on Management. Retrieved from <https://hbr.org/2013/12/how-google-sold-its-engineers-on-management>

- Goforth, H. J. (2016). From the deckplates-take back block 43. *U. S. Naval Institute Proceedings*, 142/6/1,360.
<https://www.usni.org/magazines/proceedings/2016/june/deckplates-take-back-block-43>
- Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*, 56, 1-21.
- 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>
- Haines, V. Y., & St-Onge, S. (2012). Performance management effectiveness: practices or context? *The International Journal of Human Resource Management*, 23(6), 1158–1175. doi: 10.1080/09585192.2011.561230
- Hirsh, J. B., & Inzlicht, M. (2008). The Devil You Know: Neuroticism Predicts Neural Response to Uncertainty. *Psychological Science*, 19(10), 962–967. <https://doi.org/10.1111/j.1467-9280.2008.02183.x>
- Hunt, A. G. (1966). *Enlisted performance evaluation in the United States Navy* [Master's thesis, George Washington University]. NPS Archive: Calhoun.
<https://calhoun.nps.edu/handle/10945/9500>
- Kavanagh, P., Benson, J., & Brown, M. (2007). Understanding performance appraisal fairness. *Asia Pacific Journal of Human Resources*, 45(2), 132–150. doi: 10.1177/1038411107079108
- Latham, G. P., Almost, J., Mann, S., & Moore, C. (2005). New developments in performance management. *Organizational dynamics*, 34(1), 77-87. doi: 10.1016/j.orgdyn.2004.11.001
- Larter, D. B. (2017, December 14). Trump just made a 355-ship Navy national policy. *Defense News*. <https://www.defensenews.com/congress/2017/12/14/trump-just-made-355-ships-national-policy/>
- Lawler, E. E., Benson, G. S., & McDermott, M. (2012). What makes performance appraisals effective? *Compensation & Benefits Review*, 44(4), 191–200.
<https://doi.org/10.1177/0886368712462331>
- London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, 32(2-3), 353-372.
<https://doi.org/10.1002/hrm.3930320211>

- Maylett, T. (2009). 360-degree feedback revisited: The transition from development to appraisal. *Compensation & Benefits Review*, 41(5), 52–59.
<https://doi.org/10.1177/0886368709337922>
- Mondy, R. W., & Martocchio, J. J. (2016). *Human resource management* (14th ed.). Pearson.
- 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>
- Murphy, K. (2008). Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance. *Industrial and Organizational Psychology*, 1(2), 148-160. doi:10.1111/j.1754-9434.2008.00030.x
- Navy Personnel Command. (2019a, April 15). Performance Evaluation Transformation.
<https://www.public.navy.mil/bupers-npc/career/performanceevaluation/Pages/transformation.aspx>
- Navy Personnel Command. (2019b, August 22). Sailor 2025. [Fact sheet].
<https://www.public.navy.mil/bupers-npc/career/talentmanagement/Pages/default2.aspx>
- 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>
- Office of the Chief of Naval Personnel. (2018, November 26). *Performance Evaluation Transformation* (NAVADMIN 279/18). Department of the Navy.
<https://www.public.navy.mil/bupers-npc/reference/messages/Documents/NAVADMINS/NAV2018/NAV18279.txt>
- Office of the Chief of Naval Personnel. (2019, December 06). *Navy performance evaluation system* (BUPERS Instruction 1610.10E CH-1). Department of the Navy. <https://www.public.navy.mil/bupers-npc/reference/instructions/BUPERSInstructions/Documents/1610.10.pdf>
- Orlando, J., & Bank, E. (2016). A New Approach to Performance Management at Deloitte. *People & Strategy*, 39(2), 42–44.
- Passmore, J., Peterson, D. B., & Freire, T. (2016). *The Wiley-Blackwell handbook of the psychology of coaching and mentoring*. Chichester, West Sussex: Wiley-Blackwell.

- Pulakos, E. D. (2004). *Performance management: a roadmap for developing, implementing and evaluating performance management systems*. Alexandria, VA: SHRM Foundation.
- Pulakos, E., & O'Leary, R. (2011). Why Is Performance Management Broken? *Industrial and Organizational Psychology*, 4(2), 146-164. doi:10.1111/j.1754-9434.2011.01315.x
- Pulakos, E. D., & Battista, M. (2020). *Performance management transformation: Lessons learned and next steps*. New York, NY: Oxford University Press.
- Raimondo, J. R. (2009). *You are what you measure: U. S. Navy 'reports on the fitness of officer' and their impact on future joint operational leaders* [Research paper, Naval War College] Center for Naval Warfare Studies.
- Raimondo, J. R. (2017). FitReps: We are what we measure. *U. S. Naval Institute Proceedings*, 143/1/1,367. 85-86.
- Rao, T. V. (2016). *Performance management: Towards organizational excellence*. Los Angeles: SAGE.
- Richardson, J. M. (2016). *A Design For Maintaining Maritime Superiority*. Naval War College Review: Vol. 69: No.2, Article 4. Retrieved from <https://digital-commons.usnwc.edu/nwc-review/vol69/iss2/4>
- Ritchie, J., & Spencer, L. (1994). Qualitative Data Analysis for Applied Policy Research. In Huberman, A. M., & Miles, M. B. *The qualitative researcher's companion* (pp. 305-329). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412986274
- Risher, H. (2013). We need to "fix" performance management. *Compensation & Benefits Review*, 45(3), 127-130. Doi: 10.1177/0886368713497543
- Sand, E. A. (2014). Performance over process. *U. S. Naval Institute Proceedings*, 140/10/1,340. <https://www.usni.org/magazines/proceedings/2014/october/performance-over-process>
- Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management*, 44(6). doi: 10.1177/0149206318755303
- Snyder, W. (2002). Make performance evaluations work. *U. S. Naval Institute Proceedings*, 128/2/1,188. <https://www.usni.org/magazines/proceedings/2002/february/make-performance-evaluations-work>

- Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness*, 4:3, 409-429, DOI: 10.1080/19439342.2012.710641
- Tran, S. K. (2017). Google: a reflection of culture, leader, and management. *International Journal of Corporate Social Responsibility*, 2(1). doi: 10.1186/s40991-017-0021-0
- Tyson, S., & Ward, P. (2004). The Use of 360 Degree Feedback Technique in the Evaluation of Management Development. *Management Learning*, 35(2), 205–223. doi: 10.1177/1350507604043025
- U.S. Congress. (1799) *An Act for the Government of the Navy of the United States of 1799*, Congressional Session. Retrieved from <https://www.navycs.com/public-law/naval-gov-1799.html>
- Waddington, H., White, H., Snilstveit, B., Hombrados, J., Vojtkova, M., Davies, P., Bhavsar, A., Eyers, J., Koehlmoos, T., Petticrew, M., Valentine, J. C., & Tugwell, P. (2012). How to do a good systematic review of effects in international development: a tool kit. *Journal of Development Effectiveness*, 4(3), 359-387. doi: 10.1080/19439342.2012.711765
- Wigert, B., & Harter, J. (2017). *Re-engineering performance management*. Gallup.
- Wilkie, D. (2019, August 16). *Is the annual performance review dead?* SHRM. <https://www.shrm.org/resourcesandtools/hr-topics/employee-relations/pages/performance-reviews-are-dead.aspx>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications, Inc.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California