

Occupational Bias in Performance Appraisals

Richard Lawrence Taylor

Dissertation submitted to the faculty
of Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree

Doctor of Philosophy
In
Public Administration and Public Affairs

Dissertation Committee:
James F. Wolf, Chair
Larkin S. Dudley
Matthew M. Dull
Colleen A. Woodard

October 20, 2010

Alexandria, Virginia

Keywords: Occupational Bias, Performance Appraisal, Mission Critical Occupations

Copyright 2010, Richard Lawrence Taylor

Occupational Bias in Performance Appraisals

Richard Lawrence Taylor

Abstract

This dissertation investigates the question: are mission critical occupations more favored than other occupations in performance appraisals in pay pools? While many types of bias in performance appraisals occur, such as gender or race, occupational bias--favoring or showing preference for one occupation over another in performance appraisals and subsequent ratings--has not been fully examined. There is a lack of empirical evidence that addresses occupational bias in performance appraisal and ratings in the Federal civil service sector, and more specifically the Department of Defense. The importance of occupational bias in performance appraisals is seen in the cost to organization and taxpayer, the degradation to fairness and trust in the workplace, and the erosion of organizational values.

The methodology used to address this dissertation's hypothesis is a quantitative-qualitative inquiry that investigates performance ratings of the engineering occupational series within three Department of Defense (DOD) engineering agencies. The methodology is comprised of three parts: quantitative analysis of pay pool rating data and qualitative analysis of archived documents and expert interviews. Each part of the methodology is intended to be mutually supportive.

The quantitative analysis yields a null finding of the hypothesis based on two findings. First, indicators of occupational bias were not found using differences of average occupational performance ratings between engineers and other occupations in three DOD engineering organizations. Second, Fiscal Year 2008 engineer occupational series performance ratings in three Department of Defense engineering agencies did not show statistically significant differences when compared to occupations such as personnel management specialists and accountants. This may be due to privacy act limitations in the data set used. Anecdotal evidence of preferences for mission critical occupations in performance appraisals and ratings was found to support the hypothesis.

Acknowledgements

I would like to thank the faculty of the Virginia Tech Center for Public Administration and Policy (CPAP) for a great learning experience. Though their leadership, experience, and counsel, I have expanded my knowledge and understanding of the relationship of the citizen to democratic governance. Most importantly, the Center for Public Administration and Policy fosters continuous and professional growth, particularly for adults, that encourages long term connections to fellow partners in learning, the PhD cohort.

Special appreciation goes to the members of my committee. Professors Larkin Dudley, Colleen Woodard, Matt Dull, and my chair, Jim Wolf provided motivation, encouragement, and enthusiasm for the journey. Particularly, I want to thank Dr. Wolf for his sage guidance in completing this vital phase of the doctoral program.

In pursuing the PhD, many members of my PhD cohort have been critical to my success. Particularly helpful have been Beth Offenbacher and Alejandra Palma, whose encouragement helped me get to the finish line.

Finally, there would most likely not be a Doctorate without the encouragement of my family. It will be rather disappointing not to hear my wife Mary's tongue in cheek cheerleading "are you done yet"? I will also miss my daughters' Kristy and Nikki's admonishments "I can't believe you are doing that!"

I also want to thank the Lord for providing me the gifts of patience, wisdom, and dedication. He answered my prayers more than once.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Chapter 1:.....	1
<i>Backdrop</i>	2
<i>Research Problem</i>	2
<i>Research Question</i>	3
<i>Hypothesis</i>	3
<i>Contribution to the Literature</i>	3
<i>Methodology</i>	4
<i>Assumptions and Limitations of this Dissertation</i>	5
<i>Organization of the Dissertation</i>	6
<i>Distinctions in Terminology</i>	6
<i>Significance Of This Research Topic</i>	8
<i>Costly to the organization.</i>	8
<i>Undermines Fairness and Trust.</i>	10
<i>Degrades civil service and organizational values.</i>	10
<i>Destabilizes the theoretical premises of pay-for-performance.</i>	12
Chapter 2:.....	13
<i>Enactment, Implementation, Abolishment</i>	13
<i>NSPS as a HRM System</i>	14
<i>NSPS: Classification Component</i>	15
<i>NSPS: Compensation Component</i>	17
<i>NSPS: Performance Management Component</i>	21
<i>NSPS Rate Phase: Performance Appraisal Process</i>	22
<i>NSPS Performance Appraisal</i>	22
<i>Step 1: Initiate Appraisal Process.</i>	23
<i>Step 2: Employee Self-Assessment.</i>	25
<i>Step 3: Rating Official Evaluates Performance.</i>	25
<i>Step 4: Higher Level Reviewer Reviews Ratings & Advises.</i>	28
<i>Step 5: Pay Pool Panel Reviews Ratings & Advises.</i>	29
<i>Step 6: Rating Official Communicates the Rating & Employee Acknowledges.</i>	31
Chapter 3:.....	33
<i>Relevance of Theory to the Question</i>	33
<i>Applying Theory to the Process of Performance Appraisal.</i>	36
<i>Expectancy Theory.</i>	36
<i>Implicit Contract Theory.</i>	38
<i>Equity Theory.</i>	39
<i>Empirical Schema Theory.</i>	40
<i>Social Identity Theory (SIT).</i>	41
<i>Framing Occupational Bias in the Literature</i>	43
<i>Human Resources Management (HRM)</i>	44
<i>Human Resource Management (HRM) and Merit Systems.</i>	45
<i>Civil Service Reform Act (CSRA) of 1978.</i>	46
<i>Performance Management</i>	47

<i>Pay for Performance (PfP)</i>	49
<i>Merit Pay</i>	50
<i>Performance Appraisal (PA)</i>	51
<i>Uses of Performance Appraisals</i>	52
<i>Types of Performance Appraisal</i>	52
<i>Process of Performance Appraisal</i>	53
<i>Error in Performance Appraisals</i>	54
<i>Bias in Performance Appraisal</i>	55
<i>Types of Bias in Performance Appraisal</i>	56
<i>Occupational Bias in Performance Appraisal</i>	58
<i>Organizational Influences on Occupational Bias in Performance Appraisal</i>	59
<i>Contribution to the Literature</i>	61
Chapter 4:.....	63
<i>Methodology Part I: Statistical Analysis of Pay Pool Rating Data</i>	64
<i>Data Research</i>	65
<i>Data Collection</i>	68
<i>Data Analysis</i>	69
<i>Methodology Part II: Analysis of Archived Documents</i>	71
<i>Methodology Part III: Expert Interviews Analysis</i>	74
<i>Methodology Part IV: Quantitative-Qualitative Summary Analysis</i>	77
Chapter 5:.....	78
<i>NSPS Pay Pool Performance Rating Analysis</i>	78
<i>Performance Rating Data Test 1</i>	79
<i>Key Finding-1</i>	86
<i>Performance Rating Data Test 2</i>	86
<i>Key Finding-2</i>	91
<i>Analysis of Archived Documents</i>	92
<i>Analysis of Archived Documents: Government Reports</i>	92
<i>Analysis of Archived Documents: Congressional Testimony</i>	97
<i>Analysis of Archived Documents: Books</i>	100
<i>Analysis of Archived Documents: Journals</i>	103
<i>Key Finding-3</i>	107
<i>Analysis of Expert Interviews</i>	108
<i>Conducting Rating Analysis</i>	108
<i>Anomalies in Performance Ratings</i>	109
<i>Occupational Bias in Performance Ratings</i>	109
<i>Key Finding-4</i>	112
Chapter 6:.....	113
<i>Summary of Findings</i>	113
<i>Key Finding-1: No distinctive performance rating distribution anomalies are found in the difference of average occupational performance ratings</i>	113
<i>Key Finding-2. No significant differences of means are found in occupational performance ratings to indicate the potential for occupational bias</i>	114
<i>Key Finding-3. Archived document analysis found anecdotal evidence of occupational</i> .	114
<i>Key Finding-4. Expert interview analysis found anecdotal evidence of occupational bias is present in performance ratings</i>	115

<i>Key Finding- 5: Theoretical basis of basis of occupational bias as applied to performance appraisals is re-validated in this research effort.</i>	116
<i>Conclusions to be drawn based on the findings.</i>	117
<i>Alternative Explanations of the in terms of what was learned.</i>	118
<i>Impact of the study in terms of what was learned.</i>	119
<i>Limitations of this study.</i>	120
<i>Implications for professional practice or decision making</i>	120
<i>Implications for scholarly understanding of the field.</i>	121
<i>Recommendations for further research.</i>	122
<i>Recommendation for changing the research methodology.</i>	123
Bibliography	124
Appendix A: Archived Documents and Interview Summaries	137
Appendix B: NSPS Performance Rating Data Summaries.....	143

List of Tables and Figures

<i>Figure 1.</i> Conceptual Representation of Quantitative-Qualitative Methodology.....	4
<i>Figure 2.</i> Selected Federal Pay-for-Performance Systems	9
<i>Figure 3.</i> Benefits of NSPS Human Resource Management Components.....	15
<i>Table 1.</i> NSPS Pay Schedules and Career Groups	16
<i>Table 2.</i> Sampling of NSPS Occupational Codes	17
<i>Figure 4.</i> Example of NSPS Pay Banding	18
<i>Table 3.</i> NSPS Performance Management Cycle Phase Responsibilities	21
<i>Figure 5.</i> NSPS Performance Appraisal Process.....	23
<i>Figure 6.</i> NSPS Performance Evaluation Process.....	24
<i>Table 4.</i> NSPS Rating Levels.....	26
<i>Figure 7.</i> Theory and Literature Review Framework.....	33
<i>Table 5.</i> Summary of Selected Theories Relating to Performance Appraisal	35
<i>Figure 8.</i> Selected Theories Applied To the Performance Appraisal Process.....	36
<i>Figure 9.</i> Literature Review Framework	44
<i>Table 6.</i> U.S Federal Merit System Principles.....	46
<i>Figure 10.</i> Five Components of Performance Management.....	48
<i>Figure 11.</i> Pay for Performance Approaches.	50
<i>Figure 12.</i> Factors Influencing the Performance Appraisal Process.	54
<i>Table 7.</i> Summary of Rater Errors Found in Performance Appraisal.....	55
<i>Figure 13.</i> Quantitative-Qualitative Methodology	63
<i>Table 8.</i> Sampling of Federal Occupational Series.....	64
<i>Figure 14.</i> The Five Phases of the NSPS Performance Management System.....	65
<i>Figure 15.</i> Six-Step NSPS Performance Appraisal Process	67
<i>Table 9.</i> Example of NSPS Performance Rating Data Set.....	69
<i>Figure 16.</i> Unit of Analysis: Organization-wide Pay Pools	70
<i>Table 10.</i> Sample Framework for Analysis of Archived Documents	73
<i>Table 11.</i> Expert Interview Questions	75
<i>Table 12.</i> Sample Framework for Analysis of Expert Interviews	76
<i>Table 13.</i> Summary of NSPS Organizational Population.....	80
<i>Table 14.</i> Air Force NSPS Average Performance Ratings by Occupational Series	81
<i>Table 15.</i> Navy Average Performance Rating by Occupational Series	82
<i>Figure 17.</i> Weighted Average Performance Ratings-Navy.....	83
<i>Table 16.</i> Army Average Performance Rating by Occupational Series	84
<i>Figure 18.</i> Weighted Average Performance Ratings-Army	85
<i>Table 17.</i> Sample Occupational Series Performance Rating Test of Significance	87
<i>Table 18.</i> Navy Occupational Series Difference of Means Tests of Significance.....	89
<i>Table 19.</i> Army Occupational Series Difference of Means Tests of Significance	91
<i>Table 20.</i> Abbreviated Summary: Occupational Bias Related Archived Reports	96
<i>Table 21.</i> Abbreviated Summary: Occupational Bias Related Archived Testimony	99
<i>Table 22.</i> Abbreviated Summary: Occupational Bias Related Archived Books	102
<i>Table 23.</i> Abbreviated Summary: Occupational Bias Related Archived Journals	106
<i>Table 24.</i> Abbreviated Summary: Occupational Bias in Interviews.....	111

Chapter 1: Introduction

The principles of equity and fairness in Federal human resource management (HRM) systems are at the intellectual center of this dissertation. More specifically, equity and fairness in performance appraisals, a performance management component of human resource management systems, is the focal point of this research effort. A rhetorical question posed by a Federal employee discussing a challenge associated with performance appraisals captures the essence of this dissertation's effort.

There's an institutional bias or preference to certain functions. Once you're in a certain area, you're branded. We were told that really high payouts were skewed to more 'central' functions. How does a system reward people who are deserving and doesn't penalize other people who are doing well in 'less central' functions. (SRA International Inc, 2009, p. 3-7).

Favoring or giving preference to one occupation in performance appraisals runs counter to the principles of equity and fairness. The practice of preferring central or mission critical occupations violates the merit principles that employees should receive fair and equitable treatment in all aspects of personnel management, which includes performance appraisals. Particularly egregious is when the engineering occupations receive the preponderance of high performance ratings in the engineering agency, at the expense of the human resource, resource management or logistical occupations.

The intent of this concentrated effort is to address the basic assumption of the rhetorical question posed earlier: is there occupational bias in performance appraisals—particularly in agencies' mission critical occupations?

Backdrop

Many of the laws, policies, and regulations of today that are designed to insure fairness and equity in Federal personnel management systems began to emerge during the late 19th Century. The Pendleton Act of 1883 is seen as a hallmark legislative action that was initiated to combat the spoils system in the Federal government, and thus create a work environment of equity and fairness. Almost a century later, building upon the foundations set by the Pendleton Act, the Civil Service Reform Act (CSRA) of 1978 institutionalized nine merit system principles into law. These merit principles are a reflection of national foundational values and efforts to combat the spoils system. The theme of equity and fairness is threaded throughout the nine merit system principles codified by the CSRA. Specifically, the second merit principle, which states that all employees should receive fair and equitable treatment in all aspects of personnel management, is at the center of this dissertation's research problem and question.

Research Problem

Favoritism or bias in performance appraisals causes lack of equity and fairness in the workplace, and as highlighted is a violation of Federal law. There are many types of bias in performance appraisals and ratings; some are based on race or gender. A working definition of a performance appraisal is a set of processes or procedures for rating the performance of employees in light of established objectives. Occupational bias, favoring or showing preference for one occupation over another occupation in performance appraisals and ratings, causes lack of equity and fairness in performance management systems and personnel management writ large. Bias or favoritism runs counter to merit systems principles codified in the Civil Service Reform Act (CSRA) of 1978 which in essence states that employees should receive fair and equitable treatment in all aspects of personnel management.

Research Question

The question this dissertation examines is: are mission critical occupations, for example engineers at the U.S. Army Corps of Engineers, more favored than other occupations in performance appraisal pay pools? Mission critical occupations are occupations that management personnel of an agency consider as core to carrying out their missions. Such occupations usually reflect the primary mission of the organization without which mission-critical work cannot be completed (U.S. Office of Personnel Management, 2009). Arbitrary action and personal favoritism in performance appraisals and ratings is a violation of Federal merit principles codified in law.

Hypothesis

This dissertation tests the hypothesis that more favorable performance ratings occur in mission critical occupations in pay pools in DOD agencies. The null hypothesis is that no significant differences in performance ratings are present between mission critical occupations and other occupations in pay pools in DOD Agencies. The Department of Defense (DOD) is a United States (U.S.) Federal department with many subordinate agencies.

Contribution to the Literature

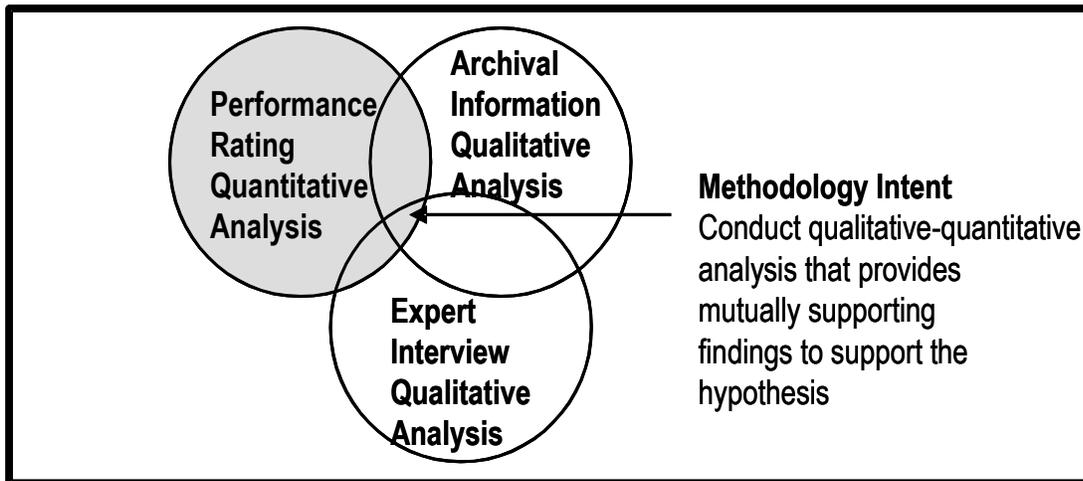
The intent of this dissertation is to fill the gap in information on occupational bias in performance appraisals in three areas. First, from the meta perspective, to enhance the body of knowledge on occupational bias in performance appraisals. Second, provide information on the performance rating component of the Federal sector's most recent experiment in pay for performance, the National Security Personnel System (NSPS). Third, present empirical and anecdotal evidence that supports the proposition that occupational bias in performance appraisals in the Federal sector (ie. DOD agencies) exists and warrants further exploration.

Methodology

The methodology used in this dissertation uses quantitative-qualitative analysis to investigate performance ratings within Department of Defense (DOD) engineering agencies. The rationale for selecting DOD engineering organizations is familiarity with the organizational environment of DOD engineering agencies.

The methodology is comprised of three parts: quantitative analysis of DOD’s National Security Personnel System (NSPS) pay pool performance rating data, and a qualitative analysis of archived documents and expert interviews. Each part of the methodology’s analysis (Figure 1) provides findings that are mutually supportive to test the hypothesis that the engineering occupational series achieves more favorable ratings than other occupational series in DOD engineering agencies. Findings from the analysis are used to form the conclusions of this dissertation.

Figure 1. Conceptual Representation of Quantitative-Qualitative Methodology



Assumptions and Limitations of this Dissertation

Three assumptions are used in this dissertation. First, pay pool data used in this dissertation consist of mixed occupational series; all pay pools are heterogeneous in nature. Some Federal agencies use pay pools that are occupationally homogeneous. The second assumption is that the Department of Defense's National Security Personnel System is abolished and referred to in the past tense. The NSPS system was abolished in late 2009 through the passage of Public Law 111-84. This law mandated that all DOD personnel revert back to the General Schedule system by January 2012; some agencies are still in the process of using NSPS during 2010. Consequently, it could be interpreted that NSPS is an actively used pay for performance system in 2010. The third assumption is that occupational bias is not assumed as a form of systemic bias in this dissertation. The literature review conducted for this dissertation did not find sufficient evidence to classify occupational bias as a form of systemic bias.

Four limitations are associated with this analytical effort. First, the level of analysis used in this dissertation is at the organizational level. Performance rating data on each employee in each pay pool is restricted and not available due to Privacy Act limitations. This impacts the fidelity of the quantitative analysis. Second, the DOD engineering agency data set does not link to demographic data bases; some confounding variables are not available for analysis. The DOD uses two distinct data bases for the storage of performance and personnel data, which are not linked. Thus, analysis of various influences of gender or racial characteristics on performance ratings is not feasible in this research effort. The third limitation of this dissertation is in the scope and time span of the data set examined, which is Fiscal Year 2008. Examination of mission critical occupations in multiple agencies over a number of a fiscal year pay cycles is not performed in this dissertation. The implications are that a longitudinal examination of data over

time over multiple DOD agencies could yield different results than those contained in this endeavor. Finally, the primary focus area in this dissertation is the detection of occupational bias in performance ratings. Simply, is occupational bias occurring? While some discussion of the “why” occupational bias occurs (i.e. influence of pay pool panels) is included in this dissertation, it is not the primary focus.

Organization of the Dissertation

This dissertation is organized into six chapters. This first chapter provides a general introduction to this dissertation. Chapter 2 provides an overview of the National Security Personnel System (NSPS) which was used as the personnel management system in DOD from 2004 until repeal in 2010. Examples of occupational bias are introduced in this chapter. Chapter 3 focuses upon the theoretical foundations of performance appraisals. Five theories that can have impact on the steps in the process of conducting performance appraisals are highlighted. Chapter 3 also provides a literature review on the research topic. Chapter 4 explains the quantitative-qualitative methodology used to test this dissertation’s hypothesis. Analysis of National Security Personnel System (NSPS) performance rating data, archived reports and expert interviews are used. Chapter 5 discusses the findings derived from the quantitative-qualitative analysis and incorporates information from previous chapters that lead to an evaluation of the hypothesis. The final chapter closes this dissertation with a summary of findings, conclusions, and recommendations.

Distinctions in Terminology

Three distinctive terms used in this research effort need to be clarified at the outset. First, the working definition of bias this research effort uses is a partiality or inclination. Closely related is systemic bias. A working definition of systemic bias is tendency of a process or

institution to favor a particular outcome. Devaluing performance ratings of disabled or elderly employees in the workplace can be considered a type of systemic bias. Occupational bias, favoring or showing preference for one occupation over another occupation in performance appraisals and ratings, is not considered as a systemic bias for the purposes of this dissertation. The literature review in this research effort did not find sufficient references to justify the use of the term occupational bias as systemic bias.

Secondly, the terms merit system and merit pay are used throughout this paper and can cause confusion. These two terms are distinct and separate, yet can overlap in application. “Merit pay is a reward that recognizes outstanding past performance. It can be given in the form of lump-sum payments or as increments to the base pay” (Milkovich & Newman, 2005, p. 609) as in the case of NSPS. Merit systems represent a system wide principled approach to personnel management that emphasizes fairness and equity in all aspects of human resources management in the Federal sector.

The third term of note that is highlighted at the outset of this dissertation is mission critical occupation. Mission critical occupations are occupations that management personnel of an agency consider as core to carrying out their missions. The engineering occupational (0800) series is assumed as the DOD engineering organizations’ mission critical occupation. The Office of Personnel Management (OPM) directed that agencies identify mission critical occupations (MCOs). “DOD and Army incorporated input from across the services to form a consolidated list of MCOs for the Department and Army. USACE MCOs are...engineering technician (0802), civil engineering (0810), mechanical engineering (0830), electrical engineering...” (Lovo, 2009, p. 1). There are other interpretations of the definition of mission critical occupations. One such interpretation is that mission critical occupations are shortages of high

demand occupational series, such as contracting specialists that are needed to support mission completion.

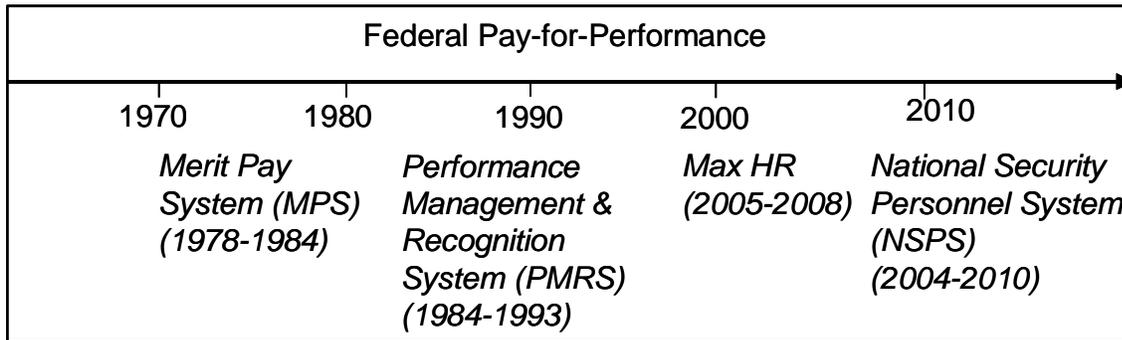
Significance Of This Research Topic

The significance of this research topic can be viewed from two perspectives: public administration and personal. The significance to the field of public administration is that occupational bias in performance appraisals is costly, undermines fairness and trust, degrades the civil service and organizational values, and destabilizes the theoretical foundations of personnel management. The significance of this research topic for me is to determine if performance appraisals are fair and equitable for all employees, no matter the occupation.

Costly to the organization.

Government agencies and the tax payer bear the costs associated with experimentation in public sector pay for performance (PfP) systems such as the recently repealed National Security Personnel System (NSPS) seen in Figure 2. A pay for performance system can be generally viewed as a monetary incentive program that has clear links between performance and pay. Many of the laws that have established pay-for-performance systems have attempted to improve the pay and performance of the Federal civil servant. Pay for performance systems such as the Performance Management and Recognition System (PMRS) and Max HR, a Department of Homeland Security pay for performance initiative, are examples of costly Federal PfP initiatives that have yielded little beyond valuable lessons learned. This dissertation does not address the specific costs of all of the Federal pay for systems found in Figure 2, but instead focuses on some of the cost categories of the NSPS.

Figure 2. Selected Federal Pay-for-Performance Systems



The promise of PfP systems such as NSPS, a Department of Defense Pfp system recently repealed in 2010, offered the potential benefit of a government wide Pfp system. However, the costs of flawed systems, such as those seen in Figure 2, have not achieved the stated benefits.

The cost of implementing and operating a new pay for performance system is often minimized when calculating the total cost of implementing a new Pfp system. Lost opportunity costs were examined by the Congressional Budget Office in their review of the National Security Personnel System (Congressional Budget Office, 2008). Additionally, GAO found that DOD did not determine the full cost of the implementation of NSPS (U.S. Government Accountability Office, 2007b). Significant costs can also be incurred when the organization must respond to performance appraisal bias complaints due to a flawed Pfp system. For example, the Security and Exchange Commission in 2008 settled a Pfp discrimination case for \$2.7 million dollars (Rosenberg, 2008). The case found systemic racial bias in performance ratings. Finally, the loss of organizational resources spent transitioning back and forth between HR systems is a cost not normally assessed. In the case of NSPS, organizational resources must be spent to revert back to the General Schedule (GS) system.

Undermines Fairness and Trust.

Several scholars (Milkovich & Wigdor, 1991; J. L. Perry & Petrakis, 1988; Walker, 2008) have documented trust as a key success factor in implementing an effective performance management system of which performance appraisals and their subsequent ratings are an integral part. “Research has demonstrated that fair treatment can lead to increased trust in supervisors, commitment to the organization, and job satisfaction” (Gilliland & Langdon, 1998, p . 211).

Other researchers (Cleveland & Murphy, 1995; Latham & Wexley, 1994; Shafritz, 2001) suggest that in order for performance appraisals to be fair and to work effectively a number of criteria must be present. Many of these criteria are similar to those found in the National Research Council’s seminal work on pay for performance in the early 1990s (Milkovich & Wigdor, 1991). The National Research Council provided a summation of the key components employees expect in a fair performance appraisal process.

In application to pay for performance, procedural concerns would involve employee perceptions about the fairness of procedures used to design and administer pay. The extent to which employees have the opportunity to participate in the performance appraisal design decisions, the quality and timeliness of information provided them, the degree to which the rules governing performance allocations are consistently followed, the availability of channels for appeal and due process, and the organization’s safeguards against bias and inconsistency are all thought to influence employees perceptions about fair treatment (Milkovich & Wigdor, 1991, p. 93).

Clearly, if a pay for performance system is not trusted by employees due to actual or perceived bias and inconsistency as highlighted by the National Research Council, it will not work effectively, and will contribute to the system’s inability to operate effectively.

Degrades civil service and organizational values.

The third reason the topic of occupational bias in performance appraisal is significant is that preferential treatment or favoritism in civil service systems is an intolerable concept to U.S. citizens. More specifically, favoritism in the administration of the public interests is an affront to

the merit system principles, particularly the value of equity (Denhardt, 1988; Hays & Kearney, 1990; Rohr, 1998). Equity as a value is codified as a common theme throughout many of the nine merit system principles in the Civil Service Reform Act of 1978. Equity, also known as fairness, is explicit in its application to pay for performance systems as seen in three merit principles. Succinctly paraphrased, the second merit principle states that all employees and applicants for employment should receive fair and equitable treatment in all aspects of personnel management. Equal pay should be provided for equal work is seen in the third merit principle. The eighth merit principle highlights that in the interest of fairness, employees should be protected against personal favoritism. The second, third, and eighth merit principles are directly applicable to the conduct of performance appraisals and subsequent ratings in the Federal Civil Service.

The importance of values to an organization also lies in the promise of the ideal organization. For example, an Army organization of employees that lives up to the Army values of loyalty, duty, respect, selfless-service, honor, integrity, and personal courage, can collectively achieve a better organizational team and work environment. Values, are “articulated, publicly announced principles and values that the group claims to be trying to achieve” (Schein, 1992, p. 9).

The Merit Systems Protection Board provides public service values for all Federal departments and agencies to follow. These include: competence, openness, fairness, diversity, neutrality, public interest, effectiveness, and efficiency (U.S. Merit Systems Protections Board, 2002). The DoD and each of the Armed Services has complemented the Federal values set, with their own unique value sets. Plainly, occupational bias in performance appraisals violates the fairness value seen in many organizational value sets.

Destabilizes the theoretical premises of pay-for-performance.

The importance of bias, real or perceived, in performance appraisal and rating systems, and in this case occupational bias, is that it erodes basic theoretical premises upon which performance management systems are based. Performance management systems are used as a way of motivating employees and directing their performance to better achieve organizational goals, objectives, and projected outcomes (Shafritz, 1998). Performance appraisal and ratings are an integral part of performance management systems. Illustratively, if there is gender, racial, or occupational bias in performance ratings, then employees give little credence to the performance goal setting that occurs in the beginning of a performance management cycle. For instance, an employee who perceives that they are the object of bias in the performance appraisal process will give little weight to trying to achieve performance goals, knowing the eventual outcome is biased. Thus the theory of goal-setting is invalidated, but more importantly, performance management techniques are minimized for organizational use. Chapter 3 provides additional insights on the impacts of bias on successful pay-for-performance systems from a theoretical perspective.

Personally, is it really fair?

The motivation for this dissertation began in 2008 when I was discussing my recent performance appraisal of a “3” (i.e. valued employee) with a fellow colleague. We were discussing the fact that it didn’t matter how well you performed, all the high ratings were reserved for the mission critical occupations, in this case the engineers in an engineer organization. While, recognizing that this was most likely an organizational myth, where there is smoke, there is most likely fire. Thus, I set out on an analytic journey to dispel the organizational myth through a determined research effort.

Chapter 2: National Security Personnel System (NSPS)

The quantitative-qualitative inquiry used in this dissertation is comprised of three parts: a quantitative analysis of DOD's National Security Personnel System (NSPS) pay pool performance rating data, and a qualitative analysis of archived documents and expert interviews. The quantitative analysis is focused upon analyzing occupational performance ratings within three DOD engineering agencies: U.S. Army Corps of Engineers (USACE), U.S. Naval Facilities Engineering Command (NAVFACENGCOM), and the U.S. Air Force Civil Engineering Support Agency (AFSESA).

The centrality of the NSPS data set in this research effort makes it necessary to provide sufficient information on NSPS that sets the stage for discussion in latter chapters. Specifically it is necessary to understand: the process steps in the NSPS performance management system where occupational bias can potentially occur, and achieve an appreciation of some anecdotal evidence of biases.

Enactment, Implementation, Abolishment

President George W. Bush signed the National Defense Authorization Act, Public Law (P.L.) 108-136, on 24 November 2003, which enacted the National Security Personnel System. P.L. 108-136, authorized the director of the Office of Personnel Management (OPM), in partnership with the DOD to:

establish and from time to time adjust, a human resources management system for some or all of the organizational or functional units of the Department of Defense. The law protected employees' collective bargaining rights, and required that the system be "fair, credible, and transparent" and provide "effective safeguards to ensure that the management of the system is fair and equitable and based on employee performance (Ginsberg, 2008, p. CRS-4).

The implementation of NSPS occurred over a multiyear period (2004-2009) in a series of phases called spirals. These spirals were designed to cover the roughly 670,000 civilian employees of the DOD. Only “approximately 211,000 employees currently under NSPS” (U.S. Department of Defense Business Board, 2009, p. 3) were covered in 2009 under the newest Federal sector Human Resources Management system.

The goal of the NSPS was to achieve a DOD human resources management (HRM) system more flexible, responsive and adaptable than the General Service (GS) HRM system as measured by five objectives. The performance objectives of NSPS were to establish a: high performance HRM system, responsive workforce, a credible and trusted system, and a HRM system that was fiscally sound (England, 2004). Of particular significance to this dissertation is the key performance objective of establishing a credible and trusted system. Specifically, the Honorable Gordon R. England, Deputy Secretary of Defense, set an objective for NSPS to be a system that adhered to the nine merit principles contained in law (England, 2004).

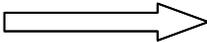
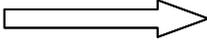
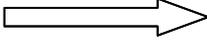
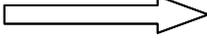
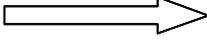
The NSPS system was abolished in late 2009 through the passage of P.L. 111-84. Most DOD organizations began to transition back to the GS system during 2010. It is required by law that all DOD personnel revert back to the GS system by January 1, 2012. Despite the abolishment of NSPS, the system provides several years of performance data from which to research the strengths and weaknesses of Federal pay-for-performance and performance management approaches.

NSPS as a HRM System

NSPS contained five HRM components designed to achieve more flexibility, when compared to similar components in the General Schedule (GS) system. These five HRM components were designed to be more responsive to the new security environment in the 21st

Century. Figure 3 provides the five NSPS human resource management components and the associated benefits to the DOD programs and personnel (U.S. Department of Defense, 2007).

Figure 3. Benefits of NSPS Human Resource Management Components

NSPS HUMAN RESOURCE MANAGEMENT COMPONENT		BENEFITS
Classification		Simple, flexible, streamlined
Compensation		Performance-based, market driven, mission-responsive
Performance Management		Flexible, streamlined, adaptable-right person in the right job
Hiring & Employment		Results-oriented, mission focused
Workforce Shaping		Mission-responsive, less disruptive, performance-focused

NSPS: Classification Component

A distinctive feature of NSPS was the use of banding in both personnel classification and compensation. Occupations in NSPS were classified into four major groups or pay schedules. These four pay schedules were: standard, engineering and scientific, medical, and investigative/protective services. Each schedule seen in Table 1 is further divided into four sub-groups: professional, technical, supervisor, and student (U.S. Department of Defense, 2006). Each sub-group was provided a pay schedule code as seen in the far right column of Table 1.

Table 1. NSPS Pay Schedules and Career Groups

PAY SCHEDULE NAME	PAY SCHEDULE CODE
Standard Career Group-Professional/Analytical Pay Schedule	YA
Standard Career Group-Technical/Support Pay Schedule	YB
Standard Career Group-Supervisor/Manager Pay Schedule	YC
Standard Career Group-Student Educational & Employment Pay Schedule	YP
Engineering & Scientific Career Group-Professional Pay Schedule	YD
Engineering & Scientific Career Group-Technical/Support Pay Schedule	YE
Engineering & Scientific Career Group-Supervisor/Manager Pay Schedule	YF
Medical Career Group-Physical /Dentist Pay Schedule	YG
Medical Career Group-Professional Pay Schedule	YH
Medical Career Group-Technician/Support Pay Schedule	YI
Medical Career Group-Supervisor/Manager Pay Schedule	YJ
Investigative & Protective Services Career Group-Investigative Pay Schedule	YK
Investigative & Protective Services Career Group-Fire Protection Pay Schedule	YL
Investigative & Protective Services Career Group-Police/Security Guard Pay	YM
Investigative & Protective Services Career Group-Supervisor/Manager Pay	YN

Specific occupations were arrayed under career groups. For example, a YA-340 was a Program Manager in the standard career group. Similarly, a YD-0803 was a Safety Engineer in the scientific and engineering career group. Both the YA-340 and the YD-0803 were in the professional pay schedule. A sampling of occupational codes and names by career group and pay schedule (U.S. Department of Defense, 2006) that are used throughout this dissertation are provided in Table 2.

Table 2. Sampling of NSPS Occupational Codes

STANDARD CAREER GROUP	
Occupational Code	Occupational Name
Professional /Analytical Pay Schedule (YA)	
0301	General Analysis
0340	Program Manager
0341	Management Services Specialist
<i>Professional Pay Schedule (YA)</i>	
0501	Finance
0505	Financial Management Specialist
0510	Accountant
SCIENTIFIC AND ENGINEERING CAREER GROUP	
<i>Professional Pay Schedule (YD)</i>	
0801	Engineering
0803	Safety Engineer
0804	Fire Protection and Prevention Engineer
Professional Pay Schedule (YD)	
1301	Physical Science
1306	Health Physicist
1310	Physicist
1370	Cartographer
Note: Represents a sampling of all career groups and occupational codes	

NSPS: Compensation Component

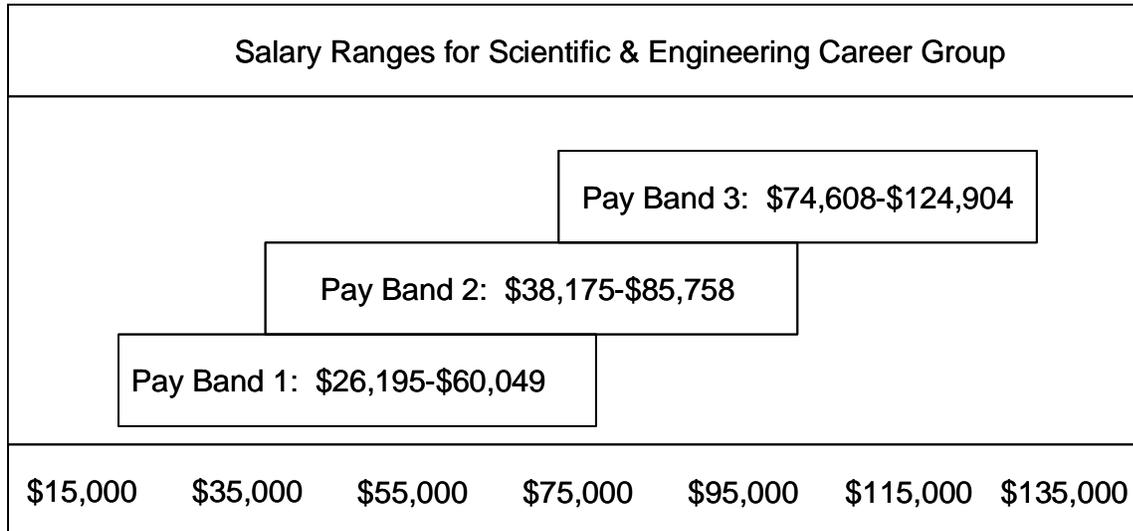
The DOD used a pay-for-performance strategy as the principle compensation component of the NSPS system.

Pay-for-performance is a generic term that applies to a wide variety of monetary incentive programs. Their one unifying theme is the goal of establishing clear and reliable linkages between performance rating and pay, and through the administration of those linkages, to motivate workers and to manage performance (L. G. Nigro, Nigro, & Edward, 2007, p. 282).

NSPS used the concept of pay banding as part of its pay for performance strategy. A working definition of pay banding or broad banding is a pay method that consolidates several pay levels into several broad bands. NSPS used three pay bands across all pay schedules. An early example of pay bands for the YD Career group (scientific and engineering professional) is

seen in Figure 4 (U.S. Department of Defense, 2006). The range in pay bands was varied by career groups to reflect market conditions.

Figure 4. Example of NSPS Pay Banding



NSPS pay bands were designed to provide more flexibility (i.e. market driven) and benefits to supervisors and managers in compensating employees. Comparatively, compensation in the General Schedule personnel system is based primarily on longevity; it uses ten steps or levels of compensation for each grade.

NSPS also used the concept of pay pools as a technique to enable the implementation of DOD’s pay-for-performance (PfP) system. NSPS employees were grouped into pay pools to enable the funding of NSPS pay pools. NSPS pay pools

means the organizational element/units or other categories of employees that are combined for the purpose of determining performance payouts. Each employee is in only one pay pool at a time. Pay pool also refers to the fund designated for performance payouts to employees covered by a pay pool (U.S. Department of Defense, 2008, p. 6).

The composition of an NSPS pay pool was made up of two elements, employees and a pay pool panel. The first element of a pay pool was the employee. In most cases, employees of

an organizational unit, occupations at similar levels or responsibilities, or personnel within a common region were placed into a pay pool of between 35 and 300 employees (U.S. Department of Army, 2009). A minimum and maximum number of employees were required by DOD pay pool policy; this policy guided the formation of pay pools. Pay pools, in most cases, consisted of a variety of occupational specialties.

The second element of a pay pool was the pay pool panel. The pay pool panels were comprised of management officials who were in positions of authority with resource oversight for the groups or categories of employees comprising the pay pool membership. An organization's military commander or deputy commander was the Performance Review Authority (PRA). The roles of the PRA were to oversee the composition of the pay pools, approve the pay pool panel makeup, and appoint and provide guidance to pay pool managers.

The pay pool panel managed the pay pool was comprised of two cells, the pay pool manager (PPM) and the pay pool panelists who were senior management officials chosen to reflect the organizational or functional nature of the pay pool. The primary function of the pay pool panel was to ensure consistency of rating of record. A rating of record was the final numerical rating assigned to a performance appraisal by a PPM. The pay pool deliberations were managed and overseen by the pay pool manager; a pay pool administrator provided personnel management data base support (U.S. Army Corps of Engineers, 2009). Pay pool panels were unique to the NSPS performance compensation component; most GS performance management systems do not use pay pool panels.

Pay pool panels were the focus of criticism that included lack of transparency. One interview conducted for the NSPS 2008 annual evaluation characterized the lack of pay pool transparency to a "black box" (SRA International Inc, 2009, p. 3-17).

A more specific example of the lack of transparency of NSPS panels was the variation of business rules established for the panel. “An issue related to pay pool panel transparency is perceived differences in processes and results across pay pools, raising concerns about unfair advantage and reward” (SRA International Inc, 2009, p. 3-3). One manifestation of the lack of transparency in business rules and guidelines was that:

employees working at Defense agencies—such as the Defense Finance and Accounting Service, Tricare, the Office of the Inspector General, and the Office of the Secretary of Defense—earned higher performance ratings and payouts overall than did their civilian counterparts in the three military service branches. When compared with their counterparts in the Air Force, Navy and Marine Corps, employees at the Defense Department agencies earned more than a full percentage point more on average in their total payout (Losey, 2009, p. 1).

NSPS pay pool panels were also criticized for changing supervisor ratings and not understanding the work of the employees they were evaluating. “A major source of distrust of the pay pool process is linked to the possibility of supervisory ratings being changed by the pay pool manager” (U.S. Department of Defense Business Board, 2009, p. 10). Pay pool panelists and the pay pool manager possessed the ability to adjust ratings ensure consistency. However, this power could be used as a means to influence performance ratings. For example, an influential engineer or engineers on the pay pool panel, when faced with choice of allocating a high performance rating to an engineer or accountant to assure beneficial positioning of the engineer in future administrative decisions (i.e. promotions), may prefer to provide the higher rating to the engineer. The influence of an engineer in power and position in an organizational form such as a pay pool has been discussed by researchers (Friedson, 2001; Mosher, 1968) as a factor in performance appraisal ratings. Significant in their discussions are the means by which professional elite can assert their control over personnel actions. In this case pay pool panels can be viewed as one such means to assert power.

NSPS: Performance Management Component

NSPS as a HRM system used a performance management component that was designed to promote a performance culture in which the contributions of the workforce are recognized and more fully rewarded. Performance management in NSPS was an ongoing process comprised of five phases: plan, monitor, develop, rate, and reward. The responsibilities of the supervisor and employee in each phase of the performance management cycle are outlined in Table 3 (U.S. Department of Defense, 2005b). An automated Performance Appraisal Assessment tool was used to help in the process of conducting a performance appraisal. The PAA was an automated data base tool that recorded the proceedings of the performance management cycle.

Table 3. NSPS Performance Management Cycle Phase Responsibilities

CYCLE PHASE	MANAGER/SUPERVISOR	EMPLOYEE
Plan	<ul style="list-style-type: none"> • Set objectives, expectations, and ways to measure accomplishment • Communicate appropriate work behavior • Determine staff developmental needs 	<ul style="list-style-type: none"> • Set objectives and ways to measure accomplishment • Assess your developmental needs and communicate them to your manager
Monitor	<ul style="list-style-type: none"> • Document observed employee accomplishment • Give appropriate constructive feedback informally and in interim review 	<ul style="list-style-type: none"> • Perform to expectations set in planning stage; continuously monitor yourself • Document your accomplishments • Solicit feed back; advise management of performance obstacles • Actively participate in interim reviews
Develop	<ul style="list-style-type: none"> • Work to improve employee performance through training, mentoring, and coaching • Determine if employees have all the appropriate tools to do their jobs and fix shortfalls • Eliminate unnecessary/low-value tasks 	<ul style="list-style-type: none"> • Participate in developmental activities • Focus on developing your skills and abilities.
Rate	<ul style="list-style-type: none"> • Evaluate employee performance and complete performance appraisal • Submit rating, shares, and distribution recommendations to pay pool manager 	<ul style="list-style-type: none"> • Complete and submit your self-appraisal • Discuss your performance and rating with your manager • Suggest improvement to increase unit performance
Reward	<ul style="list-style-type: none"> • Recognize and reward good performance 	<ul style="list-style-type: none"> • Understand the link between your demonstrated performance, its value to the organization, and compensation

NSPS Rate Phase: Performance Appraisal Process

All phases of the performance management cycle are vital to organizational and individual success, and can be the source of bias. Of particular emphasis in this research is the rate phase of the NSPS performance management cycle. The steps in the rate phase (Table 3) are essential to understanding at what stage in the performance appraisal process occupational bias can occur.

NSPS Performance Appraisal.

The NSPS performance appraisal process determined how well the employee met his or her performance objectives. These objectives were set at the beginning of the rating period. These performance objectives were codified in the employee's performance plan at the beginning of a rating period and agreed to by both supervisor and employee. An employee's performance was monitored and adjusted throughout the rating period. The result of the performance appraisal process was a narrative and numerical rating of an employee's performance. This numerical rating, from 1 to 5, with 5 being the highest rating score, provided the basis for the reward phase. The numerical rating was an output of a six step process and is used as a principle data source for quantitative analysis of performance ratings used in the testing of the hypothesis discussed in the methodology chapter of this dissertation.

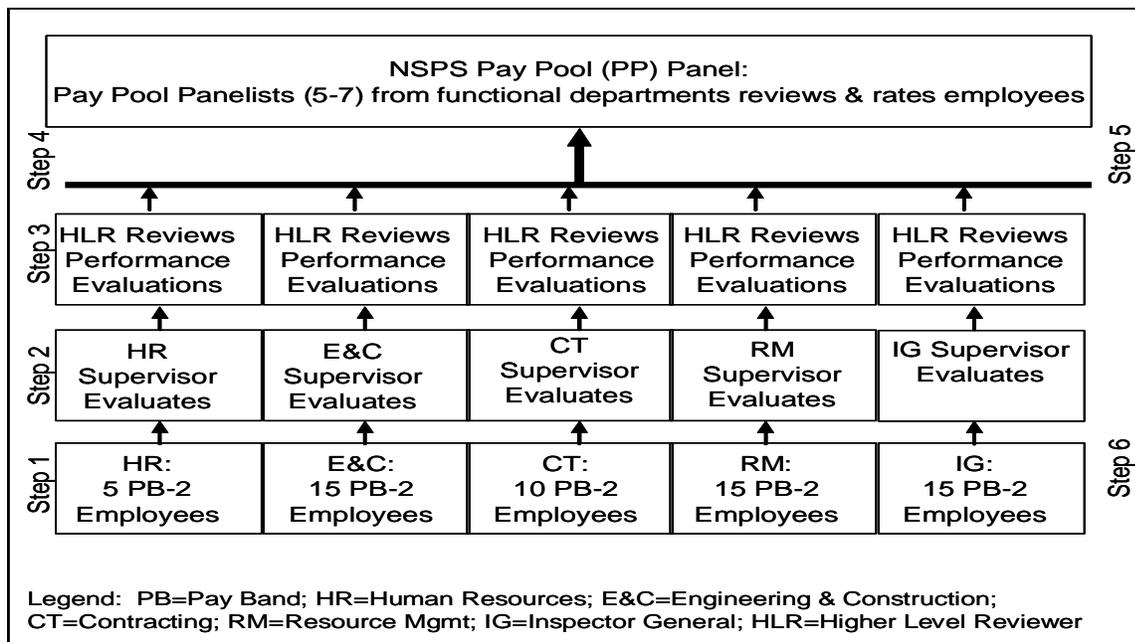
The NSPS performance appraisal process, in contrast to the GS system, contains more steps. The NSPS performance appraisal process introduces the pay pool panel as an additional step as seen in Figure 5. This additional step was the object of criticism for lack of transparency in its business processes.

Step 1: Initiate Appraisal Process.

The end of September of each year signaled the closure of the NSPS performance period and initiated the start of the six step performance appraisal process seen in Figure 5. This six step process resulted in a rating of record for the employee. A rating of record was comprised of narrative justification and the final numerical rating and assigned to a performance appraisal by a pay pool manager.

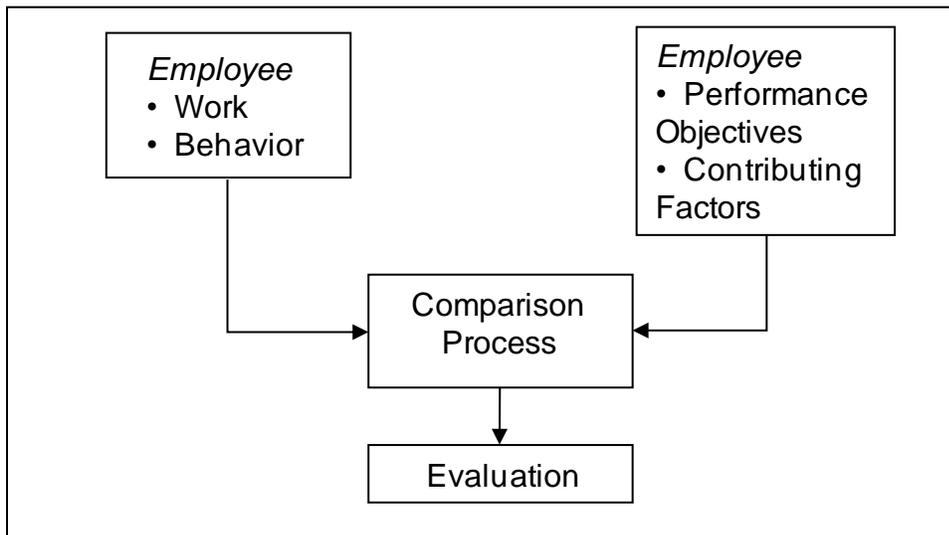
Central to the performance appraisal process was the supervisory evaluation of performance behavior relative to job objectives. Supervisors typically asked themselves when beginning the performance appraisal process: did the employee’s performance behavior meet or exceed the performance objectives specified in the annual performance plan? This rhetorical question is the conceptual foundation of the appraisal process.

Figure 5. NSPS Performance Appraisal Process.



The performance evaluation process seen in Figure 6 is the foundation of a performance appraisal, and is an adaptation from a prominent performance appraisal research team (Cleveland & Murphy, 1995). In its most basic form it is an evaluation. In NSPS, evaluation was a comparison between employee work behavior and the work objective and contributing factors. This comparison process contributed to assuring a rating that was fair and equitable and based on employee performance. The performance evaluation process (Figure 6) is universal in design, yet uniquely applicable to each individual employee. This singular comparison is unique to each employee because of the emphasis on the assignment and completion of individual performance objectives. This comparison, when performed without bias, reflects the theme of equity and fairness that is threaded throughout the nine merit system principles codified in CSRA.

Figure 6. NSPS Performance Evaluation Process.



Step 2: Employee Self-Assessment.

NSPS employees provided a self-assessment of their accomplishments relative to their stated performance objectives. At the end of the NSPS rating period this self-assessment highlighted what the employee achieved, but also how the accomplishment of employee objectives contributed to the organization's mission or strategic direction. This narrative self-assessment provided a basis for the supervisor to perform their assessment of the employee's accomplishment of performance objectives. "Job objectives means an expression of performance expectations in the performance plan that is aligned with the organization's goal(s) and mission(s)" (U.S. Department of Defense, 2005a, p. 4).

Despite emphasis on the basic comparison of their accomplishments relative to their stated performance objectives, NSPS employees perceived that a quota system was in effect which prevented them from achieving a performance rating that was truly based on performance. NSPS employees had a "perception of a quota system or forced distribution to ensure that ratings are mostly 3s" (SRA International Inc, 2009, p. 3-16).

Step 3: Rating Official Evaluates Performance.

The supervisor, also known as the rating official (RO), provided an evaluation of the employee's performance during the rating period. The supervisor compared work behavior to performance objectives and contributing factors. The supervisor used the employee self-assessment as an initial data point in the evaluation process. Building on the employee's self assessment,

supervisors will provide a narrative assessment addressing each job objective describing the employee's accomplishments and contributions to the organization relative to his or her performance expectations, including an assessment of each job objective and associated contributing factors (U.S. Department of Defense, 2008, p. 5).

The supervisor assigned a numerical rating from 1-5 to each job objective contained in the employee’s performance plan. Ratings results seen in Table 4 were based on how well employees completed their job objectives using performance indicators as a guide (U.S. Department of Defense, 2005b).

Table 4. NSPS Rating Levels

RATING LEVEL	LEVEL OF PERFORMANCE	PERFORMANCE DESCRIPTION
Level 5	Role Model	Almost always meets the standards described by the role model benchmarks
Level 4	Exceeds expectations	Almost always meets the standards described by the valued performance benchmarks and typically, but less than almost always meets the standards described by the role model benchmarks.
Level 3	Valued performer	Almost always meets the standards described by the valued performance model benchmarks
Level 2	Fair	Almost always meets the standards described by the valued performance but only as a result of guidance and assistance considerably above that expected at the valued performance level
Level 1	Unsuccessful	Performance below the level 2 rating descriptor or fails the standard performance factor in the performance of a single assignment where such failure has a significant negative impact on accomplishment of mission or where a single failure could result in death, injury, breach of security, or great monetary loss.

Step 3 was an entry point for a rater’s biases to influence an employee’s rating—either upward or downward. Age, gender, past performance, and seniority were potential sources of bias. Illustratively, interviews conducted with NSPS ratees and documented by the Government Accountability Office (GAO) indicated that ratees did not believe that some military supervisors valued the work of employees who perform certain job functions, such as providing child care on

an installation (U.S. Government Accountability Office, 2009). In many cases these employees were classified in the miscellaneous occupational series (0000) or the administrative services occupational series (0300). Congressional testimony by John Gage, President of the American Federation of Government Employees, indicated that senior managers in military agencies explained how the work of employees in the lower grades was systematically undervalued in the NSPS pay system (*Statement of John Gage, president, American Federation of Government Employees, 2008*).

Self bias, also known as “similar to me” error was another potential source of bias in step 3 of the NSPS performance appraisal process. For example, Congressional testimony on NSPS referenced the “similar-to-me” rater error. The American Federation of Government Employees President John Gage employed Professor Jeffrey Pfeffer’s perspectives on rater error noting that “supervisors in charge of judging employees have a natural tendency to favor people like themselves” (*Statement of John Gage, president, American Federation of Government Employees, 2008, p.2*). Implications for occupational bias in performance rating are clear. Engineer supervisors in an engineering organization may favor other engineers in performance appraisals due to self bias.

Once a rating was provided for each objective by the supervisor in NSPS, the rating evaluation process (Figure 6) allowed for adjustment based on contributing factors such as teamwork or technical competency. Performance objectives were weighted to provide a discriminator of the relative importance of objectives to the organization. The weighted scores were added together to obtain the overall score or rating. The supervisor’s recommended rating of an employee was provided to a higher level reviewer (HLR) which is described in the next step.

Step 4: Higher Level Reviewer Reviews Ratings & Advises.

The rating official (RO), in most cases the supervisor, provided the recommended rating to the higher level reviewer (HLR) for his or her review. The HLR was involved in the performance management process throughout the rating period. During the performance evaluation period in October, the HRL conferred with supervisors on recommended ratings. The role of the HLR was to act as a senior rater, someone who ensured a fair and equitable approach to the performance appraisal process, and provided oversight for consistent application of guidance and regulations. The HLR did have the ability to suggest revision of the subordinate rating when deemed necessary. The HLR in collaboration with the supervisor provided a recommended rating of record to the pay pool panel.

This fourth step in the NSPS performance appraisal process is another opportunity for bias and in some cases occupational bias to occur. Organizational pressure to control high performance ratings may have influenced this step. Alternatively, the HLR may have inflated or deflated the recommended rating of the supervisor based on a number of influences. One specific influence may have been the occupation of the ratee, especially if the occupation is a mission critical occupation. A case in point is seen in interviews conducted by GAO on NSPS. GAO found in interviews conducted for their report, that it was difficult for employees in certain positions to receive a rating of five because of the nature of their work or the perceived value their management placed on those positions (U.S. Government Accountability Office, 2009). Earlier efforts by prominent work performance researchers (Landy & Farr, 1983) support the supposition that the characteristics of the position have the potential to influence the ratings that an individual receives. Additionally, a recognized pay for performance consultant in an interview

(Consultant Subject Matter Expert, 2010) conducted for this dissertation also asserted that in general, mission critical job occupations get higher ratings than support jobs.

Step 5: Pay Pool Panel Reviews Ratings & Advises.

Once the pay pool panel received the recommended rating of record from the higher level reviewer, the pay pool panel began to review the recommended ratings of record. The primary function of an NSPS pay pool panel was to:

ensure consistency of ratings of record, share distribution, and pay out allocation decisions within a pay pool. The panel reviews and discusses recommended ratings and establishes a common understanding of performance indicators, benchmarks, and rating levels. The panel ensures that the common understanding is applied to each employee in the pay pool (U.S. Army Corps of Engineers, 2009, p. 9).

The pay pool panel adjusted ratings based on deliberations and discussion of the panel, in coordination with the HLR and the supervisor, when deemed necessary. Particularly, the pay pool panel “promotes equity across the pay pool by neutralizing the effect of “high” and “low” raters”(U.S. Army Corps of Engineers, 2009, p. 10). The panel maintained records of proceedings including changes mandated by the pay pool manager. The panel provided the final rating of record which also included the number of payout shares to the supervisor.

Step 5 of the NSPS Performance Appraisal Process (Figure 5) was the third opportunity for occupational bias to occur. The Pay Pool Panel may have inflated or deflated a recommended rating based on a number of influences. One organizational influence is to control costs by informally advocating forced distribution of performance ratings, a violation of U.S Code. “NSPS performance management guidance may discourage rating officials from making meaningful distinctions in employee performance because this guidance emphasized that most employees should be evaluated as a “3” (or valued performer) on a scale of 1 to 5” (U.S. Government Accountability Office, 2008, p. 6). Performance ratings in a forced distribution

scenario can favor mission critical occupations. An organization that values technical skills more highly than other skills can influence ratings, particularly when competition for ratings among departments occurs at the pay pool level. Differences in status between departments can also influence the way rater's rate employees (Milkovich & Newman, 2005).

One issue at the center of the lack of pay pool panel transparency was the belief that panelists compared one employee job accomplishments to other employees instead of comparing accomplishments to stated objectives. Occupational bias could have occurred in pay pool panel deliberations when for example an engineer was compared to a logistician. "In the pay pool panel, when we go up, we have the impression we're competing with doctors, engineers..." (SRA International Inc, 2009, p. 3-17).

NSPS may have also inadvertently favored employees who work closely or are in direct contact with member of the pay pool panel because those individuals had direct knowledge of their performance (U.S. Government Accountability Office, 2009). Inversely pay pool panel members may have deflated the ratings of employees in situations where they were unfamiliar with the employees work. "Supervisors expressed concern that pay pool members do not understand the work of the employees they are evaluating" (U.S. Department of Defense Business Board, 2009, p. 11). Simply, an employee without occupational series representation on the pay pool panel may have received an inadequate performance review. This is a subtle point when considering the pay pool panelists, normally senior managers in critical mission occupations, were faced sometimes with a hundred or more performance reviews.

The evaluative environment of a pay pool can place a cognitive burden on panelists who then may resort to evaluating on the basis of empirical schema or mental templates. Empirical schema allows raters to make quick decisions or categorize and store information in an efficient

manner based on a cognitive template, such as a hard working, nose to the grindstone engineer. In many cases the information is incomplete, generalized, and prone to error-- particularly when conducting a performance appraisal. The concept originates from Herbert's Simon's work on satisficing which recognizes human limitations on processing large amounts of data needed in decision making.

Step 6: Rating Official Communicates the Rating & Employee Acknowledges.

The NSPS rating official received the rating of record, number of shares and value of shares, and the percentage applied to base salary to the employee from the pay pool panel. A prescribed number of shares were assigned to the employees rating based on a prescribed range and formula. A share was a unit of the total pay pool payout amount awarded to an employee based on performance. A correlation of the number of shares to performance rating was published by the organization early in the process.

The supervisor and employee conducted a performance counseling session in which the employee acknowledged that the appraisal and corresponding share have been communicated. Dissatisfied employees had the option to appeal the rating. NSPS did have a rating appeal process that allowed an employee to challenge the rating of record if the outcome would result in an increase in the overall rating of record.

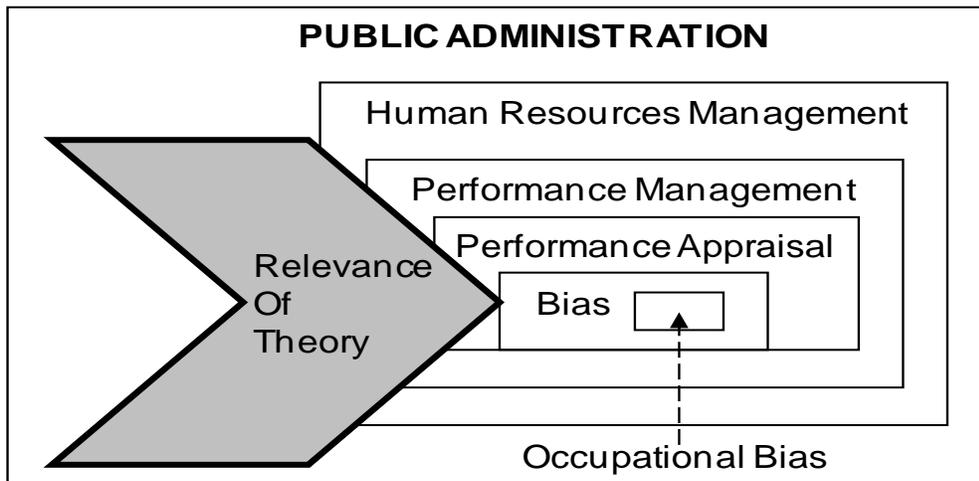
Communicating the results of performance appraisal results to employees and the workforce at large builds trust among the workforce, and has been shown to be a best practice in building a fair and equitable performance management system. Yet, GAO determined in a review of NSPS that "employees at some installations do not have transparency over the final results of the performance management process because DOD does not require commands to publish ratings distributions for employees" (U.S. Government Accountability Office, 2008, p.

6). In the broader field of Federal pay for performance systems, particularly in financial regulatory agencies, a similar lack of communication over final performance management results occurred. These Federal financial regulatory agencies have been encouraged to improve the communication of performance standards and transparency of performance results (U.S. Government Accountability Office, 2007a).

Chapter 3: Literature Review

Two sections comprise this literature review. The first section treats the relevance of theories to performance appraisals (Figure 7). Theories in the academic fields of sociology, psychology, and economics have degrees of application on the practice of performance appraisals. The relevance of several social-psychological theories applied to the practice of performance appraisal, specifically Social Identity Theory are addressed. The second part of this literature review addresses the unique nature of DOD's NSPS and its linkage to the realm of human resources management systems, performance management systems, performance appraisal processes, and bias literature (Figure 7).

Figure 7. Theory and Literature Review Framework



Relevance of Theory to the Question

The field of psychology and associated sub-fields address the practice of performance appraisals, particularly theories on work and motivation. Psychological theories such as Maslow's hierarchy of needs, Herzberg's two factors, and McClelland's needs are covered adequately (Condrey, 2005; Condrey & Brudney, 1992). Other psychological theories such as

reinforcement (Heneman, 1992; Milkovich & Newman, 2005; Smither, 1998) and goal setting (Condrey, 2005; Heneman, 1992; Rabin, 1995), are treated thoroughly in the literature as well. Self-determination (Deci & Ryan, 2000b), intrinsic motivation (Benabou & Jean, 2003; Deci & Ryan, 2000a) and self-esteem (Riccucci & Ban, 1991) are psychological theories that can also have an impact on performance appraisals. Table 5 provides a summary of selected theories relating to the practice of performance appraisal.

The field of economics, particularly theories relating to the exchange of services, addresses the practice of performance appraisals. Marginal productivity (Heneman, 1992), efficiency wage (Condrey, 2005; Heneman, 1992), agency (Eisenhardt, 1989), transaction costs (Riccucci & Ban, 1991), utility (Landy & Farr, 1983), and contingency (Milkovich & Wigdor, 1991), are economic theories that have effects on the process of performance appraisal.

Theories written in the field of sociology and in the sub-fields of social-cognition and organizational behavior also touch the practice of performance appraisal. Prospect (Cleveland & Murphy, 1991; Daley, 1992), attribution (Locke & Latham, 2004), distributive and social justice (Cohen-Charash & Spector, 2001; Greenberg, 1986), system-justification (Jost, Banaji, & A., 2004), devaluation (Magnusson, 2008), and vertical-dyadic linkage (Cleveland & Murphy, 1991), are theories that can be clustered under the general umbrella of sociology, and have varying degrees of impact on the performance appraisal process.

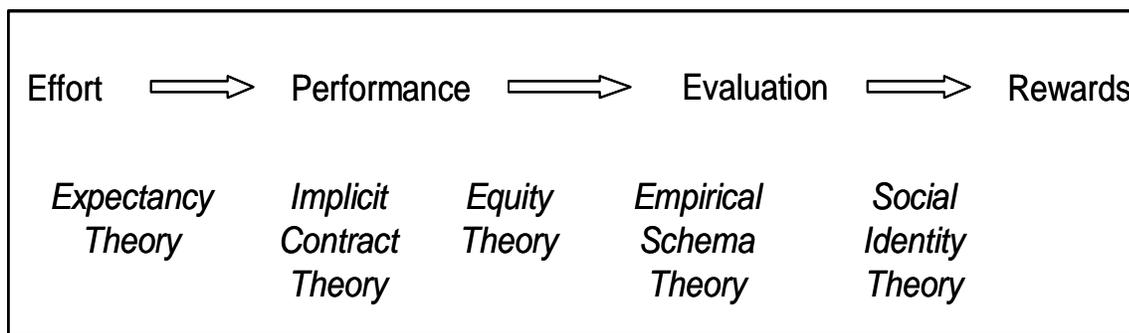
Table 5. Summary of Selected Theories Relating to Performance Appraisal

THEORY	BRIEF SUMMARY
PSYCHOLOGY	
Maslow's Hierarchy of Need	People are motivated by inner needs; needs form a hierarchy from most basic to higher order
Herzberg's Two-factor	Employees are motivated by two type of motivators: hygiene factors and satisfiers
McClelland's	Focuses on three inherent needs; need for achievement (nAch), the need for power (nPwr), the need for affiliation (nAff)
Reinforcement	Rewards reinforce in order to motivate and sustain performance
Goal setting	Challenging performance goals influence greater intensity and duration in employee performance
Self-determination	Distinguish between different types of motivation based on the different reasons or goals that give rise to an action; intrinsic and extrinsic motivation
Intrinsic motivation	Distinction between two types of rewards, extrinsic and intrinsic; extrinsic are those that are controlled by another person; intrinsic are contained in the job itself
Self-esteem	Self-evaluation that individual make and maintain with regards to themselves; threats to self-esteem generated by a pay for performance program lead to rejection of the system
ECONOMIC	
Marginal-productivity	Employers provide higher rates of pay to those employees whose inputs result in higher rates of profit to the organization.
Efficiency wage	Effort is determined by the level of wage; a premium wage is paid to ensure that employee perform up to their maximum levels
Agency	Pseudonym for principal-agent theory; pay directs and motivates employee performance
Transactions costs	Information asymmetries involves a situation in which one of the parties to an exchange possess information unavailable to others in exchange
Utility	The quality or state of being useful; the terms utility and cost-effectiveness are identities in free-enterprise systems; techniques that are profitable are high in utility
Contingency	Certain occupations can command inordinate amounts of power, given their formal hierarchal status. For example, maintenance engineers in tobacco manufacturing plants
SOCIOLOGY	
Prospect	How individuals receive and recall information; there are various limitations humans face in processing large amounts of information
Attribution	Extent to which supervisors perceive or credit external factors with influencing the outcomes of individual initiatives
Distributive & procedural justice	Stressing the important effect on employee satisfaction of the methods or processes by which organizational decisions affecting employees are made
Stakeholder	Attempts to articulate which groups are stakeholders deserving/ requiring management attention or not
System-justification	People are driving the ethnocentric motives to build in-group solidarity, and to defend and justify the interest and identities of fellow in-group members against those of out-group members
Devaluation	Also known as comparable worth discrimination; frequently used sociological explanation of the negative wage effect of the female share occupations; value of labor is gendered
Vertical dyadic linkage (vdl)	Suggests that each supervisor-subordinate dyad involves different style of interaction; theory also suggests that dyads can be classified into in-group and members of the out-group

Applying Theory to the Process of Performance Appraisal.

The relationship of theory, beyond those seen in Table 5, to explicit steps in the process of conducting performance appraisals is addressed by scholars who have contributed work on the topic of performance appraisals (Cleveland & Murphy, 1995; Heneman, 1992; Locke & Latham, 2004). Figure 8 provides a simplified adaptation of a performance appraisal (Heneman, 1992) process that aligns five selected theories to the process steps of a performance appraisal. These five theories are important to consider when evaluating the question of occupational bias in performance appraisals.

Figure 8. Selected Theories Applied To the Performance Appraisal Process



Expectancy Theory.

Expectancy theory (Vroom, 1994) is a recognized theoretical effort that examines how the individual, through a three step approach, satisfies their individual needs. “According to expectancy theory, motivation, or the force to act, results from a conscious, decision-making process undertaken by an individual. The decision to act depends upon three sets of perceptions known as expectancy, instrumentality, and valence” (Heneman, 1992, p. 25). These sets of perceptions speak to three questions considered by the employee. First, can the task be accomplished by the individual? Second, what is in it for the employee? Third, are the rewards sufficiently attractive?

The implications of expectancy theory for performance management and more specifically performance appraisals are fourfold (Heneman, 1992). First, an individual's performance must be accurately measured, thus performance objectives must be clear and specific. Second, increased pay, in whatever form, must be a valued outcome to the employee. Third, the relationship between pay and performance must be clearly defined and linked. Finally, opportunities to improve performance in the future must be achievable.

The relevancy of expectancy theory to occupational bias in performance appraisals is found in the three tenets of Vroom's theory: expectancy, instrumentality, and valence. Of vital importance is the instrumentality perception of expectancy theory in relation to performance appraisal. "Instrumentality is the degree to which the behavior in question is seen to be likely to result in a specific outcome (i.e. getting a pay raise)" (Lloyd G. Nigro & Nigro, 1994, p. 68). There must be a belief on the part of the individual that a certain level of effort (i.e. accomplishing work objectives at a high level), is associated with certain outcomes (i.e. high performance rating). If the outcomes are influenced through bias or favoritism in performance appraisal, this has negative impacts on organizations using expectancy theory as a basis for more effective organizational performance. A case in point is the multi-step performance appraisal process used in NSPS, specifically the pay pool panel process. The NSPS process had the propensity to dilute the instrumentality to which Vroom refers. Congressional testimony on NSPS speaks to Vroom's linkage of effort to outcome. "NSPS is "a "black box" to most employees and dilutes the performance pay linkage" (*Statement of Charles Tiefer, Professor, University of Baltimore School of Law, 2008, p. 7*).

Implicit Contract Theory.

Performance appraisals can be perceived as a contract. The promise of an economic exchange is conducted during the process of performance appraisals. An employee's service or performance is exchanged in return for wage or merit increases as defined through the performance appraisal which is signed by both parties—supervisor and employee. Hence, the performance appraisal can be logically viewed through the lens of contract theory. Implicit contract theory (Gordon, 1974) speaks to the variability in many jobs and tasks, and how to measure and compensate adequately. Many Federal occupations, particularly those in the program or policy management fields are inherently variable in the scope and frequency of tasks. This contrasts significantly with wage or hourly production compensation approaches. Again, this variability in production and measurement is found in many white-collar jobs in the public sector. It is essential to specify performance objectives to minimize variability and cost. "To minimize these potential costs, it is to the employer's advantage to set up contracts with individuals in whom the relationship between performance and pay is clearly specified rather than assumed" (Heneman, 1992, p. 38).

Three propositions have significance when viewing implicit contract theory and performance appraisals. First, the performance appraisal process begins with reviewing performance objectives by both rater and ratee, and ends with an appraisal of performance achievement. This process constitutes an implicit contract for many in the Federal sector. Second, performance must be measured, not assumed (Heneman, 1992). Third, performance ratings should be adjusted for factors outside an employee's control.

The importance of implicit contract theory to the performance appraisal process is found in the exchange relationship between rater and ratee. The relationship is both an economic and

psychological one. Both parties sign the performance management contract” at the beginning of the term or rating period. Midway and at the end of the performance period, an evaluation is made, and a performance rating is provided. The introduction of preference or bias that can inflate or deflate performance ratings constitutes an invalidation of the contract. For example, when a pay pool panel inflates or deflates a rating, the performance plan contract, and subsequent theory between employee and supervisor can be invalidated.

Equity Theory.

Equity theory is a motivational theory which compares individual efforts to those performing similar tasks throughout the organization, and asks if it is fair or equitable (J. S. Adams, 1963). The essence of equity theory is that employees are motivated when perceived outputs, for example pay, is equal to perceived inputs—that is their work effort. A key theme embedded throughout equity theory is the employee’s perception of fairness in the performance appraisal and reward process.

Fairness is judged on the basis of whether the reward is appropriate for the amount of effort needed to receive it. Employees also judge fairness on the basis of how the reward compares with that given to other employees. Thus they look at other employees’ levels of effort and rewards. If they are perceived as fair, the rewards can motivate, but if they are perceived as unfair, they can be de-motivators (J. N. Cayer, 2004, p. 97).

The importance of equity theory to the performance appraisal process in Federal agencies can be seen from two perspectives. First, individual equity can be applied to the disparity in performance ratings and rewards between personnel within a single organizational unit, such as a department, office, or branch. A recent case highlighted by the Congressional Research Service documented one such example where “employees with a lower rating in one office could receive a higher pay bonus than an employee with a higher rating in another”(Ginsberg, 2009, p. 15). The example is not unique. Well documented is the example of an administrative assistant

performing at a high level who receives the same rating or compensation as a less-productive administrative assistant due to some form of rater preferences (Ricucci & Ban, 2002). If the inequity persists, the superior performer may eventually reduce effort.

The second perspective that can be taken when viewing equity theory is organizational in nature. The application of equity theory between agencies is highlighted in the literature as seen below.

Agencies varied considerably in their ratings, which further discredited the system. Managers at the Department of Justice and State were rated the highest (48 and 59 percent, respectively had ratings of 5), while managers at the Departments of Labor and Treasury had tougher ratings (8 percent and 7 percent, respectively, had ratings of 5). Could this mean that one set of agencies had much better managers than another or, is it that one set of performance raters took the rating process more seriously? (Shafritz, 1992, p. 490).

Empirical Schema Theory.

Empirical schema theory is focused on the concept that individuals use frameworks for organizing, interpreting, and retrieving information to make decisions. For example, the use of the phrase, hard working, outstanding performer, as a tool to characterize an individual's performance is an example of a schema. The theory originates from Herbert's Simon's work on satisficing.

The development of Herbert Simon's notion of *satisficing* along the lines later outlined by Robert Axelrod's (1976) search concept, leads to the empirical schema theory. Human beings recognize performance patterns; new events are compared against the set of existing performance patterns stored in an individual memory (Daley, 1992, p. 115).

Empirical schema theory states that while schema allows raters to make quick decisions or categorize and store information in an efficient fashion, in many cases the information is incomplete, generalized, and prone to error. The use of empirical schema theory when applied to the act of conducting performance appraisals leads to the conclusion that managers and

supervisors will evaluate employee behavior not on stated performance objectives or rational decision making, but by comparing and contrasting their mental models of performance.

The importance of empirical schema theory to occupational bias in performance appraisals is highlighted by a team of leading researchers in the field of productivity (Latham & Wexley, 1994). They state that in-group employee schema is formed as result of a categorization in empirical schema theory. More simply, a supervisor gathers and processes information on an employee's performance using schema-based processing. The supervisor compares the employee's performance to an existing mental model of an excellent performer. Further, this schema can be linked to an in-group that is similar to the characteristics of the rater. This can influence the ratings of employees serving in mission critical occupations. "Raters have a well-developed categorization schema in terms of strong expectations about the performance of their in-group members and as a result tend to give them high ratings consistent with their categorization of these people" (Latham & Wexley, 1994, p. 148).

Social Identity Theory (SIT).

Social Identity Theory, developed in the early 1990's by Henri Tajfel (Hogg & Terry, 2001) has an effect on the process of performance appraisal because of the focus on group comparison and discrimination. Social identity theory addresses many phenomena: social comparison and prejudice, inter-group conflict, and organizational behavior.

In a nutshell, social identity theory holds that (a) we derive a great deal of personal value and meaning from our group memberships so that our self-concepts depend in significant ways upon the ways in which our groups are regarded by ourselves and by others, and (b) the only way to assess value and regard in the social world is through processes of comparison, so that the value and worth of one group is always relative to the value and worth of another reference group (Jost & Elsbach, 2002, p. 183).

One driving force behind SIT is the achievement of individual and collective self-esteem, status, and in some cases power. Status and power are the ideal state. In the organizational

setting, according to SIT, people tend to classify themselves and others into various social categories. Typically departments or functional memberships dominate the organizational group classifications. Engineers as a work group, especially as a mission critical occupation in an engineering organization, constitute a distinct identity and a group membership.

A key component of SIT that distinguishes it from other social theories is the role of favoritism in the organization. It is a mechanism to achieve the ideal state of status, prestige and power for the group. Three implications of favoritism are prominent. First, that favoritism as postulated by Tajfel, is part of organizational behavior and is used by high-status groups. “The reality in society and in business organizations is that members of high-status groups are far more likely to engage in discrimination and in-group bias against lower status group members than vice versa” (Jost & Elsbach, 2002, p. 186). From Tajfel’s perspective, occupational bias in performance appraisal would be a highly probable occurrence. The second implication of SIT to the performance appraisal process is inter-group relations and competition. “The tendency towards subunit identity...is exacerbated by competition between subunits for scarce resources and by reward and communication systems that typically focus on subunit function and performance” (Ashford & Mael, 1989, p. 31).

The implications of sub-unit identification and competition in the NSPS performance appraisal process can result in an in-group bias or favoritism of the engineering occupational series over the accountants in heterogeneous pay pools. This can occur in NSPS pay pool panels where high-status engineers influence performance ratings. This sub-unit identification and competition for performance rating is exacerbated when an organization is pressured to comply with a forced distribution of performance ratings in a pay for performance system. For example, pay pool panelists can be pressured to comply with fitting performance ratings into a bell-shaped

curve with only a few in the performance rating competition achieving a higher than average score. High-status engineer pay pool panelists can influence pay pool deliberations to insure that the limited number of high performance ratings in a forced distribution scenario are allocated to mission critical occupations.

The downstream implications of SIT in the workplace regarding pay-for-performance are discussed by researchers and are compelling. Simply, SIT in organizations using pay-for-performance systems will have harmful consequences for members of devalued groups, in this case the non-mission critical occupations.

Chains of command are still very much in place, and evaluation systems such as “pay for performance” constantly stress differences in achievement or ability among individuals and groups, which can have extremely negative social and psychological consequences for organizational members (Jost & Elsbach, 2002, p. 193).

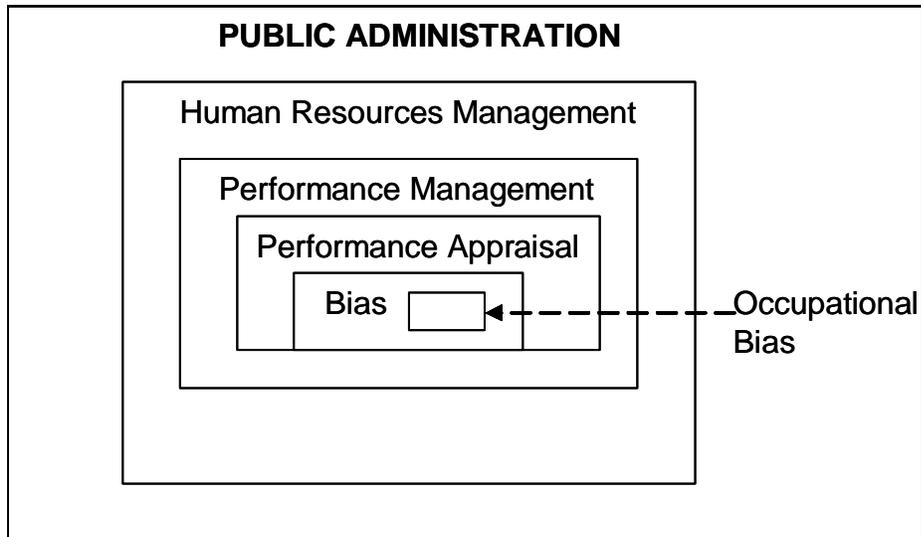
The social and psychological effects of SIT in the workplace regarding occupational bias in performance appraisals are that personnel who do not achieve the top ratings settle for average ratings, and are rewarded as such. Many moderately rated personnel become dissatisfied and leave the organization (Jost & Elsbach, 2002). Another effect in the discussion of winners versus losers in performance ratings leads to work behavior by those who are treated as losers which can lead to further decreases in their performance (Pfeffer, 1993).

Framing Occupational Bias in the Literature

Occupational bias in performance appraisals can be viewed contextually within the framework of various fields of study, such as public administration. The management framework (Ingraham, Joyce, & Donahue, 2003) adapted for the literature review in this dissertation (Figure 9) provides windows from which to view the dissertation question: are mission critical occupations, more favored than other occupations in performance appraisals in pay pools? The literature review framework is nested and includes: human resource

management (HRM), performance management, performance appraisal, bias, and occupational bias. The nested windows serve as roadmap for the literature review in this section, beginning first with a treatment of human resources management and concluding with a discussion of occupational bias in performance appraisals (Figure 9).

Figure 9. Literature Review Framework



Human Resources Management (HRM)

The HRM literature encompasses a broad spectrum of activities that encompass the management of contributions that employees make to an organization. These human resource management activities include the acquiring the right human resources for the job, to the development and motivation of the employee. Federal civil service human resource management literature has been developed by recognized authors (Berman, 2006; Condrey, 2005; Hays & Kearney, 2003). The human resource management literature includes examples of challenges to effective human resource management, and how practitioners and academics have responded to these challenges. Two human resource management challenges central to this dissertation's discussion are: embracing merit in human resource management, and the effects of the Civil Service Reform Act (CSRA) of 1978.

Human Resource Management (HRM) and Merit Systems.

Many of the foundational principles found in Federal sector HRM are based on the concept of merit in civil service. These principles evolved from the Pendleton Act of 1883 (Brook, 2000; B. Rosen, 1975; Rosenbloom & Carroll, 1995). “A merit system is a fair and orderly process for hiring, paying, developing, promoting, retaining, disciplining, and retiring people on the basis of ability and performance” (B. Rosen, 1975, p. 7). Occupational bias in performance appraisals and subsequent performance ratings directly affects the paying and promoting components of merit systems, and negates the objective of establishing a fair and orderly human resources management system.

The concept of merit in Federal civil service, first introduced in the Pendleton Act, has matured through the passage of several subsequent laws. The principles of merit (Table 6) used in today’s Federal civil service were introduced in the Intergovernmental Personnel Act of 1970 and further enhanced in the Civil Service Reform Act (CSRA) of 1978. The central theme of fairness is an inherent part of all merit principles (B. Rosen, 1975). More deliberately, “merit constitutes a value system that is grounded in fairness and equity in personnel decision making, to eliminate other influences” (Woodard, 2000, p. 29). The impact of occupational bias in performance appraisals is that it is a violation of several specific merit principles codified in the CSRA. Explicitly, occupational bias is a direct violation of the eight merit principle that states employees should be protected against arbitrary action and personal favoritism.

Table 6. U.S Federal Merit System Principles

MERIT SYSTEM PRINCIPLES
Federal personnel management should be implemented consistent with the following merit system principles:
(1) Recruitment should be from qualified individuals from appropriate sources in an endeavor to achieve a work forces from all segments of society, and selection and advancement should be determined solely on the basis of relative ability, knowledge, and skills, after fair and open competition which assures that all receive equal opportunity
(2) All employees and applicants for employment should receive fair and equitable treatment in all aspects of personnel management without regard to political affiliation, race, color, religion, national origin, sex, marital status, age, or handicapping conditions, and with proper regard for their privacy and constitutional rights.
(3) Equal pay should be provided for work of equal value, with appropriate consideration of both national and local rates paid by employers in the private sector, and appropriate incentives and recognition should be provided for excellence in performance.
(4) All employees should maintain high standards of integrity, conduct, and concern for the public interest.
(5) The federal work force should be used efficiently and effectively.
(6) Employees should be retained on the basis of the adequacy of their performance, inadequate performance should be corrected, and employees should be separated who cannot or will not improve their performance to meet required standards.
(7) Employees should be provided effective education and training in cases in which such education and training would result in better organizational and individual performance
(8) Employees should be: <ul style="list-style-type: none"> a. protected against arbitrary action, personal favoritism , or coercion for partisan political purposes, and b. prohibited against using their official authority or influence for the purpose of interfering with or affecting the result of an election or a nomination for election
(9) Employees should be protected against reprisal for the lawful disclosure of information which the employees reasonable believe evidences: <ul style="list-style-type: none"> a. a violation of any law, rule, or regulation, or b. mismanagement, a gross waste of funds, an abuse of authority, or a substantial and specific danger to public health or safety.

Civil Service Reform Act (CSRA) of 1978.

The CSRA allowed for experimentation in centralized Federal pay systems particularly in pay-for-performance (PfP) systems. Decades after passage of the CSRA, experimentation in all phases of human resource management continues. An early example of the Federal

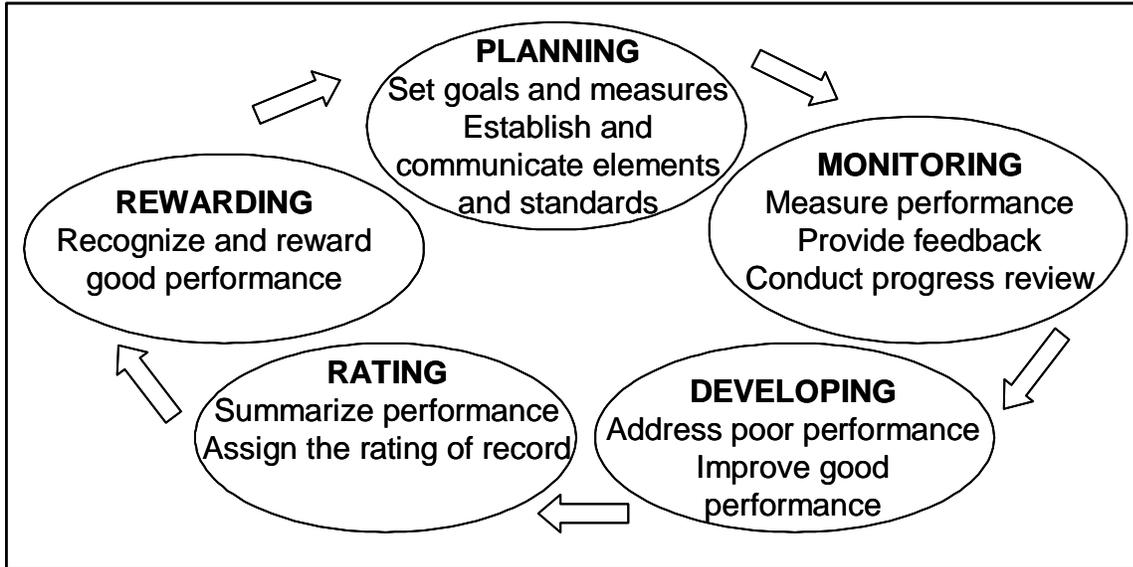
government's experimentation in PfP systems was the Performance Management and Recognition System (PMRS). One of the goals of PMRS was to build the relationship between individual performance and pay (Rainey & Kellough, 2000). PMRS was abolished in the early 1990s. Figure 2 provides a snapshot of recent efforts in establishing Federal Pay-for-Performance systems.

The Department of Defense's National Security Personnel System (NSPS) was a more recent and prominent pay for performance system implemented in the Federal government. The NSPS system was abolished in 2009 by Federal Law. NSPS used the concept of broad-banding, which clustered employees into groups for performance evaluation and pay.

Performance Management

Performance management is a method used to measure and improve the effectiveness of employees in the workplace. The multi-step performance management system approach to enhancing employee effectiveness emerged in the early 1990s (Beer & Ruh, 1990). Further experimentation with performance management systems with respect to Federal pay-for-performance systems continued during the first decade of the second millennium in systems such as NSPS. A multi-step performance management system approach, advocated by the U.S. Office of Personnel Management (OPM) is pictured in Figure 10 (U.S. Office of Personnel Management, 2001).

Figure 10. Five Components of Performance Management



Two key points emerge from the examination of performance management models such as the one seen in Figure 10. First, the organization must establish a performance management system approach that links individual performance to business unit goals. Thus in the planning component of current performance management systems in use, individual objectives are linked to organizational objectives. This approach is designed to improve organizational performance and productivity (Behn, 2006; Berlet & Cravens, 1991; Flannery, Hofrichter, & Platten, 1996; Risher & Fay, 2007). More simply, performance management systems are intended to create a direct azimuth between the organization's mission and performance goals and the performance objectives of the individual employee. If employees perceive bias in the rating component (Figure 10) as reflected in deflated ratings, they are less likely to align and achieve organizational goals. In essence the link between organizational goals and individual goals can be broken.

The second point that emerges from examination of performance management models is the linkage between the ratings and rewarding steps seen in Figure 10. The literature on

performance management systems highlights that in order to enhance performance levels in pay for performance approaches, clear and definitive linkage to pay is necessary (Lawler, 1990; Milkovich & Newman, 2005; Risher, 2004; Rizzotti, 2007). An effect of bias in the performance appraisal process and subsequent performance ratings is that it can erode the linkage between performance and pay.

Pay for Performance (PfP).

A working definition for pay-for-performance is providing a range of monetary incentives for work. PfP is not a new concept. Pay for performance has its roots in Frederick Taylor's Scientific Management movement. The empirical validity of pay for performance has been documented and summarized by scholars in the field (Milkovich & Wigdor, 1991; J. L. Perry, Enghers, Trent, and Jun So Yun, 2009). Organizations at the federal and state levels have experimented with and implemented pay for performance with varying levels of success (Brown & Heywood, 2002; Hyde, 2005; J. Edward Kellough & Nigro, 2002). Pay for performance can consist of a number of types or classes of approaches. Various plans or approaches to pay for performance seen in Figure 11 (Klingner & Nalbandian, 1998) possess associated strengths and weaknesses.

Figure 11. Pay for Performance Approaches.

	Level of Performance	
	Individual	Group
Contribution to Base Salary	Added to base	Merit Pay/ Merit Plans
	Not added to base	Piece rates Commission Bonuses

The pay for performance literature also shows that many PfP systems encounter implementation challenges. One specific challenge is the cost associated with a poorly designed and tested Pfp system. GAO found that DOD needed better internal controls and visibility over costs for implementing its NSPS (U.S. Government Accounting Office, 2007). Another cost associated with implementation of faulty pay for performance systems are the resources expended in settling performance appraisal bias cases (Rosenberg, 2008).

Merit Pay.

Merit pay is a category of pay-for-performance as seen in Figure 11. Merit pay, at the federal level, has its roots in the Performance Rating Act of 1950 and in the CSRA. An early (1980s) collection of literature related to documenting the challenges associated with the implementation of CSRA merit pay systems exists (Pearce & Perry, 1983; J. L. Perry, 1986). Since those works, a series of papers have refreshed the topic of merit pay (Brown & Heywood, 2002; Gabris & Ihrke, 2000; Heneman, 1992) and focused on challenges such as workforce acceptance and satisfaction with merit pay.

The success of merit pay is based in part on the assumption that extrinsic motivation is a governing and effective motivational principle. Several scholars (Benabou & Jean, 2003; Deci & Ryan, 2000a; Meyer, 1975) have discussed and disputed the central point that extrinsic motivation, which is rewards from sources external to the individual, are universally effective. These scholars emphasize the point that intrinsic motivation, such as working to the betterment of society, is an existing and powerful tool.

Deci conducted a series of laboratory studies on the effective of externally mediated rewards such as pay—on subject's intrinsic motivation. He concluded that contingent pay is undesirable because it reduces intrinsic motivation and leads individuals to develop strategies to achieve rewards with minimum effort (J. L. Perry, 2003, p. 125).

The implications of intrinsic motivation on performance appraisals are that employees are not entirely motivated to seek and sustain exceptional performance levels in pursuit of pay.

Combining the implications of intrinsic motivation with other components that degrade the effectiveness of merit pay, such as bias, brings into focus the challenges of establishing an effective Federal merit pay system.

Performance Appraisal (PA)

Performance appraisal is “a process to determine correspondence between worker behavior/task outcomes and employer expectations (performance standards)” (Milkovich & Newman, 2005, p. 612). It is an essential part of the merit pay process. The use of performance appraisals as a tool to measure employee performance has grown steadily from mid 20th century military and governmental organizational use, and is now quite robust. Numerous authors have addressed the strengths and weaknesses of performance appraisals (Bernardin & Beatty, 1984; Cleveland & Murphy, 1989; Latham & Wexley, 1994). Essential to this discussion are the effects of occupational bias on the uses, methods, and processes associated with performance appraisals.

Uses of Performance Appraisals.

The uses of performance appraisals (Cleveland & Murphy, 1995; Daley, 2005; Landy & Farr, 1983) can be divided into two major categories: administrative decision making and development of employees. Within these categories, performance appraisals link to and influence each of the major human resource management components seen in Figure 3.

Performance appraisals provide one basis for deciding who should be promoted, terminated, or given a raise (Cleveland & Murphy, 1995). In essence, performance appraisals have short-term and long-term impact. One such impact of occupational bias in performance appraisals is the promotion of favored occupations to positions of power in the organization by using inflated performance ratings as a tool. This practice is a violation of merit principles.

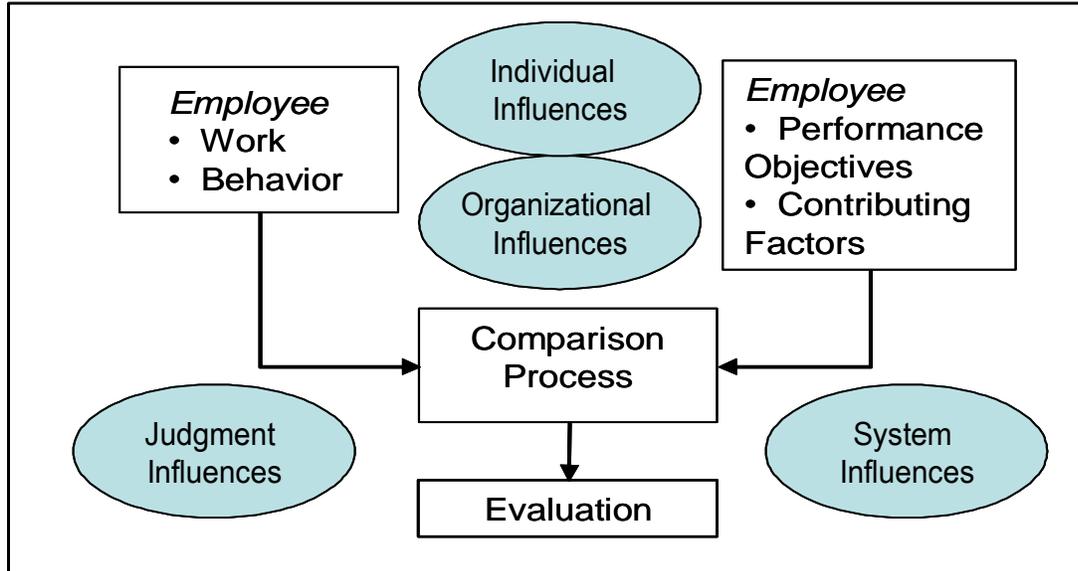
Developmentally, from an organizational management perspective, performance appraisals enable the organization to be more productive by linking strategic goals to individual goals that are stated in performance appraisals. Individual performance goals are linked to and in alignment with organizational goals. Managers and supervisors intent on achieving organizational success must recognize that bias of any type in performance appraisals can break the link to the achievement of organizational goals.

Types of Performance Appraisal. Research has found that there are “18 possible types of appraisal methods” (Bernardin & Beatty, 1984, p. 63). Other scholars (Cardy & Dobbins, 1994; Daley, 2000; Demmke, 2007) found that there are between six to eight common methods of performance appraisal. Among these common methods of appraisal types is management by objective (MBO). The National Security Personnel System (NSPS), used a management by objective performance appraisal method.

MBO programs take a number of forms, but there are several components common to most MBO applications. First, MBO involves participation in goal setting. In particular, MBO programs involve joint supervisor-subordinate determination of specific goals and performance indicators. In particular, supervisors and subordinates jointly set performance goals by agreeing on what is needed to be accomplished and how accomplishment will be measured usually in terms of output or some objective indicator (Cleveland & Murphy, 1995, p. 9).

Process of Performance Appraisal. The performance appraisal process is covered in works on the measurement of work performance (Landy & Farr, 1983; Vroom, 1994). The post Civil Service Reform Act (CSRA) era literature provides further refinements on the process of performance appraisals in Federal pay-for-performance systems (Cardy & Dobbins, 1994; Daley, 1992; Milkovich & Wigdor, 1991). One of the prominent works (Cleveland & Murphy, 1995) that examines the process of performance appraisal can be used as a common reference point for researchers and practitioners alike. It is adapted for use in this dissertation and seen in Figure 12. Figure 12 incorporates factors or influences that can affect the process of conducting a performance appraisal. This illustration (Figure 12) serves as a platform to view the influences of the individual and organization (represented by the ovals), on the performance appraisal process.

Figure 12. Factors Influencing the Performance Appraisal Process.



Error in Performance Appraisals.

Performance appraisal methods and processes contain error. “Appraisal involves a subjective and retrospective synthesis of information. Obviously, the potential for error and bias due to subjectivity and great memory demand is substantial. This has led many authors to criticize subjective appraisals as being replete with errors” (Cardy & Dobbins, 1994, p. 2). Other scholars have also contributed research and writing efforts on the topic of error in performance appraisal (Cleveland & Murphy, 1995; Daley, 2002; Landy & Farr, 1983; Latham & Wexley, 1994).

Rater errors (Table 7) are a type of error that has a significant influence on the effectiveness of appraisal.

Rating errors are errors in judgment that occur in a systematic manner when an individual observes and evaluates a person, group, or phenomena. The most frequently described rating errors include halo, leniency, severity, and central tendency errors (Milkovich & Newman, 2005, p. 613).

Table 7. Summary of Rater Errors Found in Performance Appraisal

ERROR/EFFECTS	BRIEF SUMMARY
Halo Effect	Some raters allow their ratings on one factor to influence their ratings on other factors
Personal Bias	Interpersonal relations affect rating assigned irrespective of actual performance
Leniency/Strictness	Raters are not uniform in the standards they apply to their employees
Central Tendency	Some raters are reluctant to give either high or low ratings, preferring to view everyone as average
Special Group Feeling	Some raters believe their employees are all better than average because of collective rather than individual effort
Contrast-Effects	Tendency for a rater to evaluate a person relative to other individuals rather than on the requirements of the job
First Impression	Occurs when a manager makes an initial favorable or unfavorable judgment about an employee and then ignores (or perceptually distorts) subsequent information so as to support the initial impression
Similar-to-me	Rater's tendency to judge more favorably those people whom they perceive as similar to themselves
Latest Behavior	Rating influenced by the most recent behavior; failing to recognize the most commonly demonstrated behaviors during the entire appraisal period
Spillover	Allowing past performance appraisal ratings to unjustly influence current ratings. Past performance ratings good or bad result in a similar rating for the current period, although demonstrated behavior does not deserve the rating good or bad

Bias in Performance Appraisal.

Bias, in the context of performance appraisal, is defined as the over or under inflation of performance appraisal ratings because of identification that the rater or organization may have toward a reference group. This definition is consistent with subject matter experts who define bias as systemic distortion which they say can result from limitations in cognitive processing, from individual motivation, or from a combination of motivational and cognitive factors (Smith, DiTomaso, & Farris, 2001).

Other leaders in the field of performance appraisals add that because rater judgment (Figure 12) is such a critical aspect of the appraisal process, some concern must be focused on

the possibility of bias in the exercise of that judgment. “Bias can also obviously be the product of prejudice based on factors such as race, ethnicity, gender, age, or disability” (Ricucci & Ban, 2006, p. 181).

Bias in performance appraisals continues to be found in various forms as seen through the efforts of investigators (Cleveland & Murphy, 1991; Landy & Farr, 1983; Milkovich & Wigdor, 1991). Bias in performance appraisals continues despite repeated efforts to address the problem through training and awareness programs. The presence of bias in performance appraisal may even be considered a wicked problem due to the persistence of the problem.

Controversy continues to rage over the best ways to evaluate present performance. The fundamental weakness of all such evaluation is its subjectivity. Try as we might appraisal of work performance are inevitably influenced by personal considerations and chance. Survey after survey reveals that civil servants do not trust the performance evaluation process. Workers consider their evaluations to be unreliable and prone to favoritism that is both personally and politically motivated (Rabin, 1995, p. 511).

Types of Bias in Performance Appraisal.

Research on bias in performance judgment and ratings emerged in the early 1950s (Wherry & Bartlett, 1982). Subsequent scholars have continued the research effort and focused on types of bias in performance appraisals (Castilla, 2008; Moers, 2004; Viswesvaran & Schmidt, 2005).

Performance appraisal bias based on personal or demographic characteristics such as race, ethnicity, gender, age, disability, pregnancy, has been documented by a number of scholars (Bigoness, 1976; Ferris, Yates, Gilmore, & Rowland, 1985; Pulakos, Oppler, White, & Borman, 1989).

Documented cases of bias in performance appraisals continue to surface during the first decade of the 21st century. For example, racial bias in performance appraisals was documented in the recently abolished NSPS pay for performance system.

White employees received higher average performance ratings, salary increases and bonuses in January than employees of other races and ethnicities. When compared to Asians and Native Hawaiians and Pacific Islanders, whites receive on average a full percentage point more in their total payout of raises and bonuses (Losey, 2009, p. 1).

Bias in performance appraisals can also be attributed to the influence of the organization as illustrated in Figure 12. This influence is recognized by researchers (DiTomaso & Smith, 2007; Jost & Elsbach, 2002; Schminke, Cropanzano, & Rupp, 2002). These researchers have discussed the influence of organizational structure, which includes the authority of vertical or horizontal organizational elements on performance appraisals.

Another influence of the organization on performance appraisal is to favor high graded personnel. This means that high grade personnel receive more favorable performance ratings than middle or lower graded personnel based on their rank or grade. This unique type of bias has been documented at NIST (Eremin, Wolf, & Woodard, 2010). High grade bias was also found in NSPS performance ratings. “Rating and payout analysis suggest that payouts are relatively higher for higher paid employees...preliminary analysis by the NSPS 2008 evaluation report showed a pattern of higher shares being assigned to employees with higher salaries within the 3- and 4-rating ranges” (U.S. Department of Defense Business Board, 2009, p. 17)

The organizational influences on the process of performance appraisal seen in Figure 12 also include factors “such as organizational values, climate or culture within the organization, competition among departments or functions, status differences among functions,” (Cleveland & Murphy, 1991, p. 15) . Competition among departments can lead to organizational politics. The

politics of employee appraisal is alive and well as documented by organizational behavior researchers (Brody & Frank, 2002; Longenecker & Goff, 1992; Prendergast & Topel, 1996). Professor Dennis M. Daley, a scholar in the field of performance appraisal states agrees with these researchers and recognizes that factors external to the appraisal process, such as politics, affect appraisal ratings (Daley, 2005)

Occupational Bias in Performance Appraisal.

The literature and empirical evidence on occupational bias in performance appraisal is less than the treatment of gender and racial bias in performance appraisals. The literature does however provide some implied or anecdotal evidence that can be used as a point of departure for discussion of occupational bias in performance appraisals. Organizational behaviorists and social psychologists performing research into social identity theory and processes have contributed some insights into why certain preferences in performance appraisals can occur.

Empirical evidence on occupational bias in performance appraisals is sparse. Leading contributors on performance appraisal reference anecdotal evidence indicating that personnel whose occupations were closely aligned with the organization's mission were rated higher than personnel whose occupations were in mission support (Cleveland & Murphy, 1995). Similarly, Daley in his publications on performance appraisal sees personnel working in a critical unit central to the success of the organization are favored more than others. "In this case, the importance of the unit to fulfilling the organization's mission is substituted for the job performance of the individual in that unit" (Daley, 2003, p. 160). Others documenting empirical evidence on inflation of ratings noted that highly valued occupational series in the military did demonstrate some very broad occupational group favoritism in performance appraisals (Bjerke, Cleveland, Morrison, & Wilson, 1987).

The phenomena associated with Social Identity Theory (SIT), namely in-group bias or favoritism can manifest itself in similar to me or same as me effect in performance appraisal ratings. The similar to me (Barnes-Farrell, 2001; Higgins, 1983; Pulakos & Wexley, 1983; Pynes, 1997) rater effect states that a higher rating is bestowed upon a ratee who exhibits characteristics similar to that of the rater, or reflects the rater's group. For example, "the more closely an employee resembles the rater in values, attitudes, personality or biographic background, the more likely the rater is to judge that individual favorably" (Latham & Wexley, 1994, p. 141). Placed within the context of occupational bias, when a supervisor who is an accountant, rates accountants more favorably in pay pools where there is a mix of occupations, a similar to me effect can be attributed to the cause of the biased performance rating.

Organizational Influences on Occupational Bias in Performance Appraisal.

Factors influencing the performance appraisal process (Figure 12) include organizational influences. Maintaining an organization's image or high-status departments can influence a manager's decision in performance appraisals (Cleveland & Murphy, 1991). This has been found in the military services, where high-status occupations (i.e. infantry commanders in the Army) are more likely to receive preferential ratings. "High status combat commanders in the navy (e.g. nuclear submarine community) were more likely to suffer from rating inflation than were low-status support communities (e.g. supply and tanker fleets.). In the high-status community, the mere fact that an individual is a member of the community is enough to justify rating of "excellent" on many dimensions" (Cleveland & Murphy, 1991, p. 199). Cleveland and Murphy in their books on performance appraisal close by stating that they know of no research on these phenomena in the private sector. On the whole, according to Cleveland and Murphy, little empirical evidence is found on occupational bias in performance ratings.

Cleveland and colleagues found anecdotal support for the influence of an organizational value system upon ratings. Specifically, employees whose jobs or activities were perceived as more closely linked with the organization's mission or purpose (i.e. what was valued) were rated systematically higher than employees in jobs that were perceived as less central to the organization's mission" (Cleveland & Murphy, 1995, p. 231).

Organizational power and position can also be influencing factors in the performance appraisal process. Illustratively, the power of the professional state within an organization has the potential to influence on behalf of the professional occupations. One method used to elevate the power and prestige of the profession is "the upgrading of rewards (pay) for professionals, and the improvement of their prestige before their associates" (Mosher, 1968, p. 108). Other scholars (Jost & Elsbach, 2002) reinforce Mosher's point by stating that status and power are sustained through the use of evaluations and promotions.

Mosher continues to elaborate on the role of the professional elite, a body of professionals closely associated with each other and the organization. The single occupational group that is closely aligned with the mission of the organization can control key line positions and is a source for key leadership positions. There are sub-groups that maintain prominence within the occupational group because of their close alignment with the primary mission of the organization. This point is illustrated using the examples of engineers in a construction agency.

The most elite of the sub-groups is likely to be one which historically was most closely identified with the end purpose, the basic content of the agency—the officers of the line in the Navy, the pilots in the Air Force, the political officers in the Foreign Service, the civil engineers in a construction agency which also employ electrical and mechanical engineers (Mosher, 1968, p. 114).

An essential point from Mosher's perspective on power in organizations is that mission critical occupations are prone to exert influence on the organization through the upgrading of rewards. One approach is to influence promotions and performance evaluations to

favor mission critical occupations. Complementary to Mosher's perspective, researchers have demonstrated that supervisors consciously distort ratings to reward key subordinates (Cardy & Dobbins, 1994).

Organizational economics can influence ratings in performance appraisals. More specifically, the condition of the economy, the sector that the organization works in, competition from other agencies and industry, and insufficient agency budgets can influence performance appraisals (Cleveland & Murphy, 1991; Daley, 1992). This translates to competition for scarce resources within the organization which can influence ratings towards mission critical occupations. Mission critical occupations are required to sustain the technical competencies needed to continue the legacy of the organization, and may be favored in difficult economic times.

Contribution to the Literature

The literature review in this chapter finds there is a lack of information, data, and research regarding occupational bias in performance appraisal and ratings in the Federal sector, especially in Department of Defense agencies.

There is a lack of empirical evidence that addresses occupational bias in performance appraisals and their associated ratings in Federal sector reports. For example, no government agency reports have been found that analyze cases of mission critical occupation ratings achieving more favorable performance ratings than non-mission critical occupations. There is anecdotal evidence in a series of GAO and DOD Program Executive Office (PEO) reports that point to preferences in performance ratings for mission critical occupations.

The academic literature on occupational bias in performance appraisals can also be characterized as sparse. Journals, books, and briefings containing information on occupational

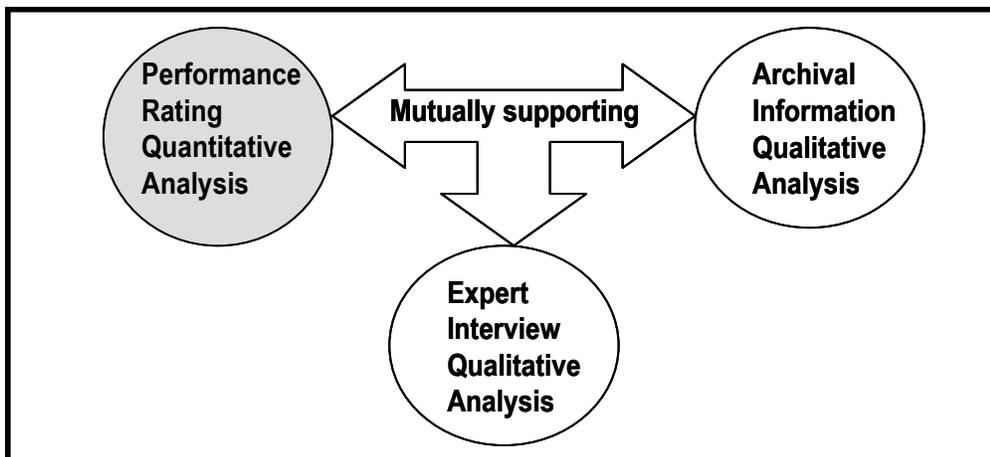
bias in performance appraisals are found to be highly generalized and consist of anecdotal data and dated research. In contrast, articles on gender and racial bias are documented. Noteworthy is the lack of published performance rating analysis and reporting on the National Security Personnel System (NSPS). For example, no articles or reports were found that discussed the topic of more favorable performance ratings occurring in mission critical occupations in pay pools in Federal agencies, and in particular NSPS.

Chapter 4: Methodology

The methodology used in this dissertation addresses the hypothesis that more favorable performance ratings occur in mission critical occupations in pay pools in Federal agencies. Specifically, the engineering occupational series obtains more favorable performance ratings than other occupational series in DOD engineering organizations. Mission critical occupations are occupations that agencies' consider as core to carrying out the agencies' missions. Such occupations usually reflect the primary mission of the organization without which mission-critical work cannot be completed.

The methodology used to address this dissertation's hypothesis is a quantitative-qualitative inquiry that investigates performance ratings within three DOD engineering agencies. The methodology is comprised of three parts (Figure 13): a quantitative analysis of National Security Personnel System performance ratings, and qualitative analysis of archived documents and expert interviews. Each part of the methodology's analysis provides findings that are intended to be mutually supportive of the hypothesis in the concluding chapters. These findings are used to validate the hypothesis.

Figure 13. Quantitative-Qualitative Methodology



Methodology Part I: Quantitative Analysis of Pay Pool Rating Data.

The quantitative analysis of NSPS performance appraisal ratings used in this dissertation is obtained through a Freedom of Information Act (FOIA) request from three DOD engineering agencies. These ratings are used to determine if there are anomalies in the distributions of mission critical occupational performance ratings. An anomaly is a rating or group of ratings that represents a significant statistical departure from the mean. Three NSPS pay pool performance rating data sets are analyzed: the U.S. Army Corps of Engineers (USACE), the U.S. Air Force Civil Engineer Support Agency (AFCESA), and the U.S. Naval Facilities Engineer (NAVFACENGCOR). For the purposes of brevity, these data sets are referred to as the Army, Air Force, and Navy data sets. The performance appraisal ratings are entered into a commercial off-the-shelf spreadsheet analysis software package. This package is used to analyze the organizational pay pool distribution of ratings by occupational code. These occupational codes are numerical identifiers given to federal occupations as seen in Table 8. Analysis of the distribution of performance ratings by occupational code provides insights into the presence of occupational preferences in performance ratings.

Table 8. Sampling of Federal Occupational Series

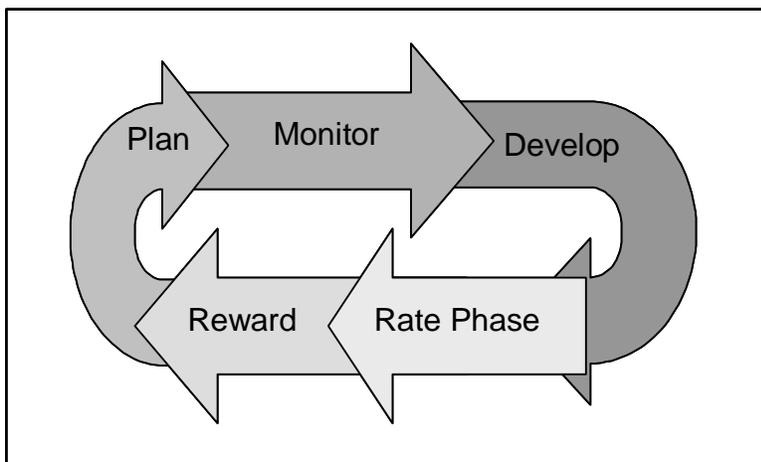
FEDERAL OCCUPATIONS SERIES
GS-000: Miscellaneous
GS-100: Social Science, Psychology, And Welfare Group
GS-200: Personnel Management And Industrial Relations Group
GS-300: General Administrative, Clerical, & Office Services Group
GS-400: Biological Science Group
GS-500: Accounting And Budget Group
GS-600: Medical, Hospital, Dental, And Public Health Group
GS-700: Veterinary Medical Science Group
GS-800: Engineering And Architecture Group
GS-900: Legal And Kindred Group
GS-1000: Information And Arts Group
GS-1100: Business And Industry Group
GS-1200: Copyright And Trade-Mark Group
GS-1300: Physical Science Group

GS-1400: Library And Archives Group
GS-1500: Mathematics And Statistics Group
GS-1600: Equipment, Facilities, And Services Group
GS-1700: Education Group
GS-1800: Investigation Group
GS-1900: Quality Assurance, Inspection, And Grading Group
GS-2000: Supply Group
GS-2100: Transportation Group
Note: List is representative in nature; list is not comprehensive

Data Research.

NSPS performance appraisal ratings were part of the overall NSPS performance management system seen in Figure 14. The NSPS performance management system was in alignment with OPM’s personnel management system discussed in Chapter 3.

Figure 14. The Five Phases of the NSPS Performance Management System



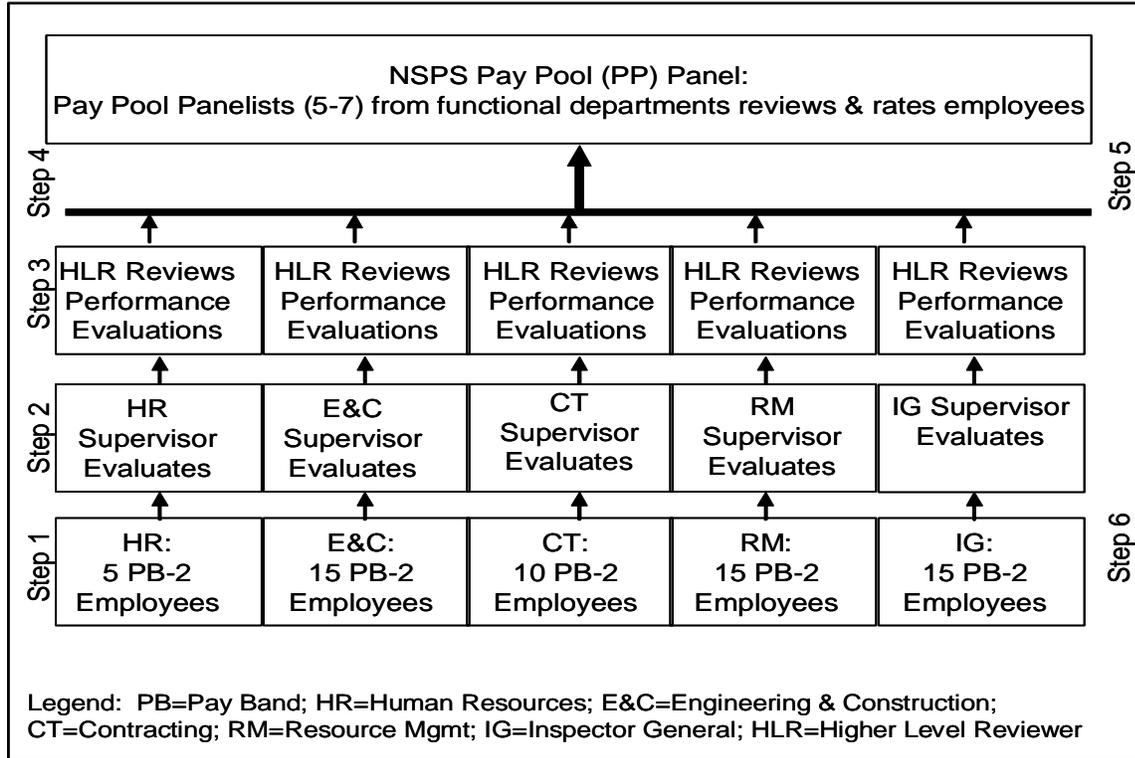
The NSPS tool used in the rate phase seen in Figure 14 is called a performance appraisal. Performance appraisals are a set of “procedures for rating the work performance of individuals according to objectives, standards and criteria applied uniformly across one or several organizations” (Organization for Economic Co-operation and Development, 2005, p. 20). One output of a performance appraisal is a performance rating; some ratings are subject to various

types of bias as discussed in the literature review chapter. One objective of this dissertation's methodology is to determine if performance ratings in DOD engineering agencies are inflated in the 0800 occupational series (i.e. engineers) and deflated for others occupational series (i.e. general administrative (0300), biological (0400), accounting (0500)). A sampling of the occupational codes normally found in Federal agencies, many of which are used herein, are seen in Table 8.

NSPS used a distinct structure and process to conduct the "rate" phase seen in Figure 15. Employees were segregated into pay pools for the purposes of conducting a performance appraisal and rating. Pay pool performance ratings used in this dissertation's quantitative analysis were occupationally homogenous in composition. For example, accounting personnel (i.e. occupational code 0500) are mixed into a pay pool with personnel management employees (i.e. occupational code 0200) and other occupations. Pay pools examined in this dissertation are not occupationally homogeneous. Mixed occupational (human resources managers, engineers, project managers, resource managers, etc.) pay pools were formed in NSPS to insure a sufficient baseline of funds for the reward phase of the performance management system seen in Figure 15.

An example of a typical organizational structure and process associated with a pay pool is seen in Figure 15. This pay pool is comprised of over 50 employees all of whom are in Pay Band 3. Pay band 3 was the highest level of pay pool compensation in NSPS. The pay pool seen in Figure 15 contains a heterogeneous mix of occupational series from resource managers, to engineers, to project managers. The occupational series seen in Figure 15 are associated with functional staff elements located within a Headquarters organization.

Figure 15. Six-Step NSPS Performance Appraisal Process



The NSPS pay pool organizational structures (Figure 15) were hierarchal in nature. At the lowest level of the NSPS performance appraisal structure were the employees who conduct a self-assessment (Step 1). These employees were rated by their supervisors (Steps 2 & 3) and higher level reviewer (HLR) on a scale of 1 (lowest level of performance) to 5 (highest level of performance). The next level of the pay pool structure is a performance panel of between 4-7 senior managers of the organization. Pay pool panel members reviewed each employee's recommended rating (Steps 4 & 5) that was provided by the front line supervisor and HLR, and adjust the recommended ratings to insure that final ratings and awards result in equity and consistency across the performance pool. The pay pool panel included a performance pool manager who represented the highest level of the performance pool structure. The pay pool manager oversaw the operation of the pay pool and was the final arbitrator of pay pool issues.

The rating of record was returned to the employee (Step 6) in a performance counseling session. All ratings are resident in a DOD data base for analysis. These numerical ratings constitute the basis of the quantitative analysis used to evaluate the hypothesis.

Data Collection.

A Freedom of Information Act (FOIA) request was submitted to DOD to obtain the NSPS performance ratings and associated data for three DOD engineering agencies: USACE, NAVFACENGCOM, AFCESA. This data request asked for performance ratings from Fiscal Year (FY) 2007, which runs from October 1, 2006 to September 30, 2007. Additionally, performance ratings for FY 2008 were requested. No personal data or classified information was requested due to Privacy Act restrictions. Privacy Act restrictions also limited the fidelity of performance rating data; many individual ratings were not released and were aggregated. Performance ratings for FY2007 were found to be inadequate for analysis due to the lack of information across the all three DOD engineering agencies.

NSPS performance ratings and the associated data fields in the FOIA request were derived from the NSPS Compensation work bench (CWB) spreadsheet. Three data fields from the NSPS engineering agencies CWB spreadsheets were requested, received, and used in the analysis. They are: pay pool identification code, occupational series, and average rating, as seen in Table 9. Table 9 illustrates a single pay pool consisting of eight occupational series of which only three are releasable due to Privacy Act constraints. A total of 35 personnel in three occupational series (0800, 1100, 0300) comprise the usable data set. Twenty five members of the 0800 (engineering series) comprise the largest occupational population group in the ONHAFCES pay pool. The 0800 occupational series has the highest performance rating average

(3.48) in the Air Force pay pool. The Air Force data set in Table 9 is illustrative of the construct of the larger Navy and Army NSPS data sets.

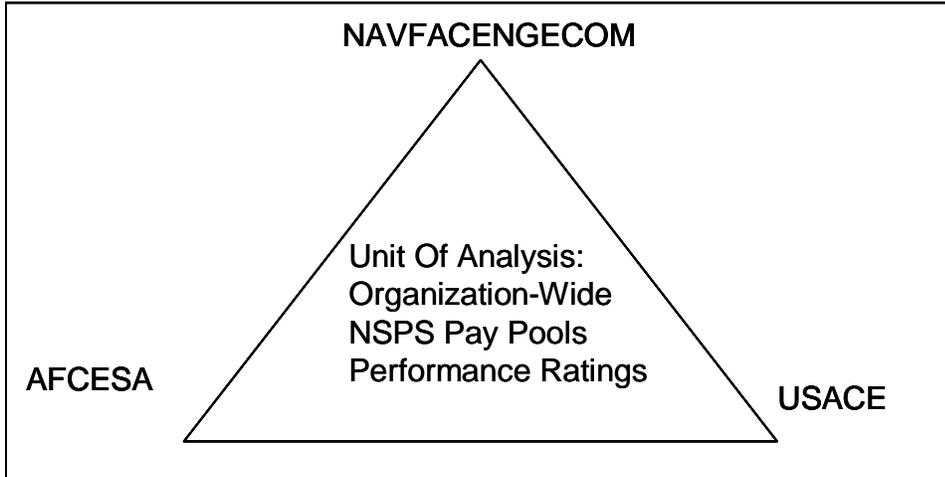
Table 9. Example of NSPS Performance Rating Data Set

AIR FORCE: AVERAGE RATING BY OCCUPATIONAL SERIES BY PAY POOL			
Pay Pool Identification Code	Occupational Series	Total Population in Occupational Series	Average Rating of Total Population by Occupational Series
ONHAFCES	0000	3	Not Releasable
ONHAFCES	0500	2	Not Releasable
ONHAFCES	1000	1	Not Releasable
ONHAFCES	1600	3	Not Releasable
ONHAFCES	2000	1	Not Releasable
ONHAFCES	0800	25	3.48
ONHAFCES	1100	4	3.25
ONHAFCES	0300	6	3.17
Totals	8	45	
Note: non-releaseable indicates restricted data due to Privacy Act.			

Data Analysis.

Quantitative analysis of the distribution of performance ratings by occupational series is used to indicate the presence or absence of normality in the distribution of performance ratings. The analysis examines if there are any significant differences in the distribution of performance ratings by occupational code. For example, can it be shown in pay pool populations that the engineering occupational series (0800), the mission critical occupation in engineering organizations, obtain higher performance ratings compared to other occupations? The unit of analysis is the organizational level as seen in Figure 16. Average occupational ratings from pay pools are combined to achieve an organizational wide unit of analysis.

Figure 16. Unit of Analysis: Organization-wide Pay Pools



Two quantitative tests are performed on the NSPS performance rating data. The first test used for analysis on the performance ratings is a difference of means between all occupational series and the engineer occupational series (0800). It is a test that provides a general point of reference for the occupational series performance rating comparisons. For example, if it is found that the engineering occupations series are found to have consistently higher average performance ratings than all other occupations, this is a notable finding. A second test used for analysis on the performance ratings is the significance of the variation of means. For example is the observed difference in the difference of means between the engineering occupational series and other occupational series significant? These quantitative findings are then used to complement qualitative analysis findings (i.e. expert interviews, archival reports) to assess the hypothesis that more favorable performance ratings occur in mission critical occupations in pay pools.

Methodology Part II: Analysis of Archived Documents.

The qualitative analysis of archived documents is centered on the rate phase of DOD's NSPS (Figure 14) performance management system. The intent of the qualitative analysis of archived documents is to determine the presence and value of Federal agency performance rating information and analysis that contributes to the hypothesis.

A two phased logic in the qualitative analysis of archived documents is used. The first logic inquiry asks: is there information contained in the documents that relates to the performance rating phase of a Federal Performance management system, which contributes to the validation of the hypothesis? Second, if there is information or data, how "valuable" in a qualitative sense is it with regards to the hypothesis? Using the lens of the performance rate phase, three elements of analysis are used to determine the documents' value in relation to validation of the hypothesis. First, is there evidence, in the archived documents that have been examined of a performance rating review framework that is used to analyze performance rating results? In some cases, agencies and departments conduct statistical analysis of performance ratings to detect anomalies in performance rating data. The rationale for using this element of analysis is that it acts as a leading indicator of an agency's ability to detect anomalies in performance ratings—such as occupational bias. Second, have anomalies been detected in performance rating data, as documented in archived documents. For example, has an institution or agency found performance rating anomalies that provide evidence of gender or racial preference in performance ratings? Again, the rationale for this element of analysis is that if an agency is reviewing performance rating data for gender bias, the agency can also be examining for other types of bias (i.e. occupational bias). Third, if anomalies have been detected in rating data, as documented in archived documents, are the anomalies related to occupational bias? For

example, does the analysis of the distribution of performance ratings by occupational codes at the Bureau of the Census, if it were conducted, provide insights into the potential presence of occupational preference for statisticians in the performance rate phase? Selected archived documents are examined and discussed for point of interests relating to the hypothesis.

The identification, collection, review and analysis of archived documents looks for indicators of occupational bias in four categories of documents: U.S. Government reports, U.S. Congressional testimony, books, and journal and news articles. The archived document analysis rates information contained in archived documents on a qualitative scale (1 [lowest value; poor] to 3 [highest value; good]). The rating scale uses three criteria for the qualitative evaluation of archived documents. Is there discussion or analysis on: conducting ratings analysis, anomalies in performance ratings, and evidence of occupational bias in performance ratings which is of high value to the validation of the hypothesis? Table 10 provides a sample framework used to conduct qualitative analysis of archived documents. Column one represents the identification number of the document that is examined and found to contain pertinent performance rating phase information. Documents are numbered sequentially. The second column provides an abbreviated citation. The third, fifth, and seventh column provides a summary of the key point for that specific criterion, which contributes to the qualitative analysis and the hypothesis under examination. The fourth, sixth, and eighth column provides a qualitative assessment or rating of the key point shown in the column. The third row in the table provides a brief explanation of the criteria and rating scale. A blank entry in the table indicates an absence of contributing information for that criterion; it receives the lowest ranking. This framework provides an opportunity to selectively focus on classes of documents, such as GAO reports, or individual documents that contain anecdotal evidence that contribute to the hypothesis.

Table 10. Sample Framework for Analysis of Archived Documents

SAMPLE FRAMEWORK ANALYSIS OF ARCHIVED DOCUMENTS							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
4	GAO-09-464T, Apr 2009, Improved Implementation of Safeguards and an Action Plan to Address Employee Concerns	DOD lacks a process to determine whether rating results are non-discriminatory; it does not require a third party to analyze the pre-decisional rating results for anomalies	2	DOD may be unable to fully determine whether ratings are fair and equitable because the department does not have a written policy outlining how it will analyze final ratings	2		1
	Represents the citation of the information being evaluated	This criteria evaluates the presence of some form of analytical system (s) in the document for reviewing or analyzing performance appraisals and their respective ratings. A statistical analysis of ratings represents the highest score	3= a statistical analysis of the agencies performance ratings is found to be present. 2=some form of performance rating analysis or review is present; can be implicitly implied 1=no rating analysis	This criteria evaluates the presence of some form of anomaly in performance ratings, such as racial bias. The highest evaluation score represents empirical evidence of performance rating anomalies	3= empirical evidence of performance rating bias is found to be present. 2= anecdotal or implicit discussion ie. cause of anomaly in performance rating 1=no reference to performance rating anomalies	This criteria evaluates the presence of some form of occupational bias in performance ratings, can be explicit or implicit. The highest evaluation score represents empirical or anecdotal evidence of occupational bias contained in the document	3= empirical or anecdotal evidence of occupational performance rating bias is found to be present. 2=implicit discussion ie. group bias in performance rating 1=no reference to occupational bias

Methodology Part III: Expert Interviews Analysis

The qualitative analysis of expert interviews focuses on Federal pay for performance systems, to include DOD's NSPS Performance Management system. The intent of the qualitative analysis of expert interviews is to determine the presence and or indications of Federal performance rating information or analysis that contributes to the hypothesis under examination. The qualitative analysis of interviews uses a similar logic to the document analysis. If there is information or data revealed during interviews, is it of value to the hypothesis, and can it serve to complement the quantitative analysis or archival document analysis? Using the lens of the performance rate phase, three elements of analysis are used to further examine the research question of occupational bias in performance ratings. First, is there evidence, in interviews that have been conducted with experts, of a performance rating review framework used to analyze performance rating results? In some cases, agencies and departments conduct statistical analysis of performance ratings to detect anomalies in performance rating data. The rationale for this element of analysis is that it acts as a leading indicator for agencies to detect anomalies in performance ratings. Second, have anomalies been detected in performance rating data, as discovered by experts in the field of pay for performance. For example, has a consultant, professor, or researcher found evidence of gender or racial preference in performance ratings in their research or work? Again, the rationale for this element of analysis is that if an agency is reviewing performance rating data for gender bias, the agency could be checking for other types of bias; occupational bias for instance. Third, if anomalies have been detected in rating data, as documented in the interview, are the anomalies related to occupational bias? Illustratively, did an expert, during analysis of the distribution of performance ratings by

occupational codes at the U.S. Bureau of the Census, provide insights into the presence of occupational preference for statisticians in performance ratings?

The collection and analysis of information from expert interviews looks for indicators of occupational bias from four categories of experts: U.S. Government researchers, U.S. Government personnel management experts, personnel management consultants, and pay-for-performance experts in academia. A listing of twenty PfP experts was generated; ten expert interviews were conducted. Fifty percent of the experts declined to be interviewed or did not respond to inquiries for interviews. The identification of the ten experts and their statements is protected under Institutional Review Board (IRB) guidelines.

The conduct of expert interviews uses three questions as seen in Table 11. These questions are designed to solicit information on occupational bias and are linked conceptually to the archived document and quantitative analysis. Hand written notes are recorded during a thirty minute interview with experts. The answers to these questions are recorded on an interview questionnaire data sheet, analyzed, and placed onto a spreadsheet.

Table 11. Expert Interview Questions

ID CODE	DATE
Preface	Some agencies analyze results of annual performance ratings for anomalies. For example, “if the data indicated that there was a statistically significant difference between the ratios of a particular subset of the larger workforce compared to the larger group at the same level, this could constitute an anomaly”. The quote is taken from a recent GAO report. Three questions below relate to this topic.
Question 1	Why do you, your research, or your agency conduct post- performance rating analysis during the annual performance appraisal process?
Question 2	What type of performance rating anomalies have you, your research, or your respective agency found—either by empirical or anecdotal methods?
Question 3	Has your research or analysis found evidence, empirical or anecdotal, of occupational preference in performance ratings?
Closing	Are there any comments that you would like to make at this time? Thank you for your time and insights.

The analysis of expert interviews is scored or rated on a qualitative scale (1 [low value; poor], 2 [moderate value; fair], 3 [high value; good]) for the presence of three key informational elements: conducting ratings analysis, anomalies in performance ratings, and evidence of occupational bias in performance ratings. Table 12 provides a sample of the framework used to conduct qualitative analysis of expert interviews. Column one represents the coded number of the interview that is examined. The second, fourth, and sixth column provides a summary of the key point that contributes to the qualitative analysis and the hypothesis under examination. The third, fifth, and seventh column provides a qualitative assessment or rating of the key point provided. The fourth row in Table 12 provides a brief explanation of the criteria and rating scale. This framework is used for all interviews that have been conducted and analyzed.

This framework, as demonstrated in the qualitative analysis of archived information, provides the opportunity to focus upon specific points in interviews that contain anecdotal evidence that contribute to the hypothesis.

Table 12. Sample Framework for Analysis of Expert Interviews

SAMPLE FRAMEWORK ANALYSIS OF EXPERT INTERVIEWS						
#	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Ratings	Rate
6	Look @ systems for faultiness; look at the mean differences which are of significance in valid systems	3	Empirically found race and gender differences in ratings in operational performance ratings	3	We know the mission critical jobs occs get greater ratings than support jobs in general	2.5
	This criteria evaluates the presence of some form of analytical system (s) in the interview for reviewing or analyzing performance appraisals and their respective ratings. A statistical analysis of ratings represents the highest score	3= a statistical analysis of the agencies performance ratings is found to be present. 2=some form of performance rating analysis or review is present; can be implicitly implied 1=no rating analysis	This criteria evaluates the presence of some form of anomaly in performance ratings, such as racial bias. The highest evaluation score represent empirical evidence of performance rating anomalies	3= empirical evidence of performance rating bias is found to be present. 2=anechdo-tal or implicit discussion ie. cause of anomaly in performance rating 1=no reference to performance rating anomalies	This criteria evaluates the presence of some form of occupational bias in performance ratings, can be explicit or implicit. The highest evaluation score represents empirical or anecdotal evidence of occupational bias contained in the interview	3= empirical or anecdotal evidence of occupational performance rating bias is found to be present. 2=implicit discussion ie. group bias in performance rating 1=no reference to occupational bias

Summary of Quantitative-Qualitative Analysis

Three hypothesis validation points are used. First, did quantitative analysis show a significant difference of means between the engineering occupational series and other occupational series? If significant differences quantitative differences are found, the hypothesis is validated from the quantitative standpoint. The qualitative analysis of documents and interviews are then assessed and provide mutually supportive findings to provide further validation of the hypothesis, or alternative explanations of what was discovered.

Chapter 5: Findings

Bias and more specifically, occupational bias if present in performance appraisals is costly, undermines fairness and trust, and degrades the civil service and organizational values. However, empirical evidence on occupational bias in performance appraisals is lacking, particularly in the Federal sector, and more specifically DOD. This research effort moves to fill this empirical evidence gap. The approach used to fill this gap is to confirm the hypothesis that more favorable performance ratings occur in mission critical occupations in pay pools in DOD agencies. The methodology used to address this dissertation's hypothesis is a quantitative-qualitative inquiry that investigates the status of engineering occupational performance ratings within three DOD engineering agencies. The methodology is comprised of three parts: a quantitative analysis of NSPS pay pool rating data, and analysis of archived documents and expert interview. Each part of the methodology's analysis provides findings to determine the validity of the hypothesis.

NSPS Pay Pool Performance Rating Analysis.

The quantitative analysis of the performance ratings by occupational series is used to determine anomalies in the distribution of performance ratings. The analysis asks if there are statistically significant differences in the distribution of performance ratings by occupational code. Explicitly, do engineering occupational series in engineering organizations consistently obtain higher performance ratings compared to other occupations in pay pools? The unit of analysis is the organizational level (Figure 16). The performance rating data from three DOD engineering organizations is used.

Two quantitative tests are performed on NSPS performance rating data. The first test used for analysis of the performance ratings is a difference of means between the average

performance rating of selected occupational series and the engineer occupational series (0800). A notable finding if found in the difference of means testing would be that engineering occupations series are found to have consistently higher average performance ratings than all other occupations. A second test used for analysis of the performance ratings is the significance of the variation of means. For example, are the difference of means of performance ratings between the engineering occupational series and other selected occupation significant? The quantitative findings are then used in association with archival and interview analysis to assess the hypothesis that more favorable performance ratings occur in mission critical occupations in pay pools.

Performance Rating Data Test 1.

A summary of NSPS organizational population totals by organizational pay pools used for analysis is seen in Table 13. The Fiscal Year (FY) 2008 data set indicates that the Army data set has the largest total organization pay pool population of 12846 in 160 pay pools, while the Air Force data set is the smallest with a population of 45 in one pay pool. Weighted average comparisons are not used with the Air Force data set information due to the limited population. Another point worth highlighting is the number of pay pools across the Navy and Army organizations. The Navy has a relatively smaller number (20) of pay pools in comparison to the Army (160).

Table 13. Summary of NSPS Organizational Population

NSPS SUMMARY: ORGANIZATIONAL POPULATION AND PAY POOLS- FY2008				
Department of Defense	Total Organizational Pay Pool Population	Total Organizational Pay Pool Population Minus Non-Releaseable Population	Usable Pay Pool Population Percentage	Total Number of Organizational Pay Pools
Air Force	45	35	0.78	1
Navy	3266	2998	0.92	20
Army	12684	11900	0.94	160
Note: Selected engineering organizations within the DOD are selected for NSPS performance ratings ie. US Army Corps of Engineers	Note: Total population in all NSPS pay pools by selected organizations, ie. Army=US Naval Facilities Engineering Command	Note: Total population in all NSPS pay pools minus the non-releaseable population	Note: Represents the percentage of the total organizational NSPS population data that is usable (ie. releasable)	Note: Represents the total number of NSPS pay pools by selected DOD engineering organization ie. US Air Force Civil Engineer Support Agency

The first test used in the quantitative analysis is straightforward and intended to determine the distribution of performance ratings among occupations and their relative differences. Specifically, what are the differences between average occupational performance ratings of engineers and other occupations in three DOD engineering organizations?

The difference of average performance ratings by occupational series for the Air Force in Fiscal Year (FY) 2008 is seen in Table 14. As noted, the AF data set is the smallest, and offers few opportunities for occupational performance rating comparisons.

Table 14. Air Force NSPS Average Performance Ratings by Occupational Series

AIR FORCE AVERAGE PERFORMANCE RATINGS BY OCCUPATIONAL SERIES				
Occupational Series	0300	0800	1100	Mean
Occupational Series Ratings Average of All Pay Pools	3.17	3.48	3.25	3.3
Difference of Occupational Series Ratings Averages of All Pay Pools vs 0800	0.31		0.23	0.27

Despite the small Air Force population data set of 35 personnel, several observations are seen. First, the engineering series (0800) performance ratings are higher on average than all other occupational series across the data set. Second, the engineering series is on average .27 greater than other occupational series. The average difference of .27 (0800 occupational series) is greater on average than other occupational series, and is distinctly larger when compared to the Army (.11) and Navy (.067) data sets seen in Tables 15 and 16.

Recognizing the limitations of the Air Force data set, the information in Table 13 can initiate the formation of a comparative rule of thumb or heuristic for comparison among all organizational performance rating data sets. In this instance, a notable indicator of occupational bias begins to crystallize, if for example the Navy and Air Force average occupational difference means are greater than .25 between engineering occupational series performance rating averages and other occupations. A case in point is seen in the Air Force data set. From a layman's perspective, the difference between the between 0300 and 0800 occupational series average performance ratings of .31 is notable. The reason this is notable is because the 0300 occupational series is a general administrative occupational group when compared to the technical engineering series. This point is cogent because it reinforces documentation (Chapter 2) that engineering (0800) and scientist (1300) occupational performance ratings are found to be higher than administrative related jobs such as those in the 0300 series. In essence, technical occupations are more valued than administrative occupations.

The Navy data set extracted from NSPS data records for FY 2008 is more substantial than the Air Force data set, and allows more insights to be gained. Over 2900 employees in 20 pay pools form the basis for determining the difference of average performance ratings by occupational series in the Navy engineering organization (Table 15).

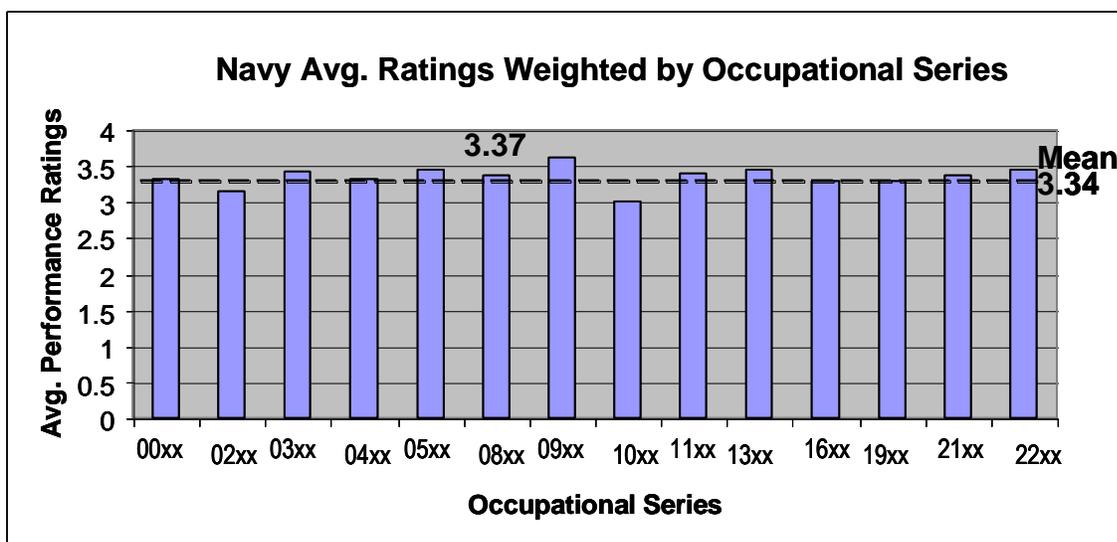
Table 15. Navy Average Performance Rating by Occupational Series

NAVY AVERAGE PERFORMANCE RATINGS BY OCCUPATIONAL SERIES															
Occupational Series	0000	0200	0300	0400	0500	0800	0900	1000	1100	1300	1600	1900	2100	2200	Mean
Occupational Series Ratings Average of All Pay Pools	3.31	3.14	3.46	3.31	3.54	3.42	3.57	3	3.43	3.39	3.29	3.29	3.35	3.51	3.358
Difference of Occupational Series Ratings Averages of All Pay Pools vs 0800	0.11	0.28	-0.04	0.11	-0.12		-0.15	0.42	-0.01	0.03	0.13	0.13	0.07	-0.09	0.067

The Navy performance rating data provides a firmer basis to make observations on the differences between the 0800 series and other occupational series performance ratings. Thirteen occupational performance rating averages are used for comparison (Table 15) to the engineering occupational series (0800). The bottom row of Table 15 indicates that the 0800 series performance rating received higher average ratings in 8 of 13 comparisons. This means that roughly half are above and half are below the mean. In only two comparisons, the human resource management series (0200) and the information management series (1000), did differences in of performance ratings averages exceed .25. The average difference between the engineering series average performance ratings and all others was .067. Interestingly, the 0300 occupational series in comparison to the 0800 occupational series scored higher. This is in contrast to the distinctive difference of .31 seen in the Air Force data set when the 0300 versus 0800 average performance ratings are compared. Holistically, distinctive differences are lacking in the Navy data set using the difference of means approach. Another comparative approach that can be more beneficial to determining aberrations is the use of weighted averages.

Using a weighted average approach is another lens through which to view the comparative differences of Navy occupational performance ratings. A weighted average considers the size of population as a factor when comparing occupational performance rating scores. The weighted averages of the Navy performance ratings are seen in Figure 17. It shows the engineering occupational series average performance rating of 3.37 is above the mean of 3.34 occupational performance rating averages. Yet the difference is minor, and places the engineering series in close proximity to the remainder of the data set. Using the weighted average perspective, the 0800 occupational series ranks 7th of 14; not a distinctive highlight. The legal occupational series (0900) has the highest average occupational performance rating. In a weighted approach the administrative series (0300) ranks higher than the engineering occupational series. Interpreting the results of the Navy comparisons of average performance ratings does not yield notable anomalies that indicate that the 0800 series is favored over other occupations. The differences in the Navy occupational performance rating comparisons are negligible, relatively speaking, when seen in contrast to the Air Force data set in Table 14.

Figure 17. Weighted Average Performance Ratings-Navy



The Army data set is the largest data set and consists of a population 11900 employees in 160 pay pools. This larger data set (Table 16) allows an extensive examination of the difference between performance ratings of the engineering occupational series and the eighteen other occupational series.

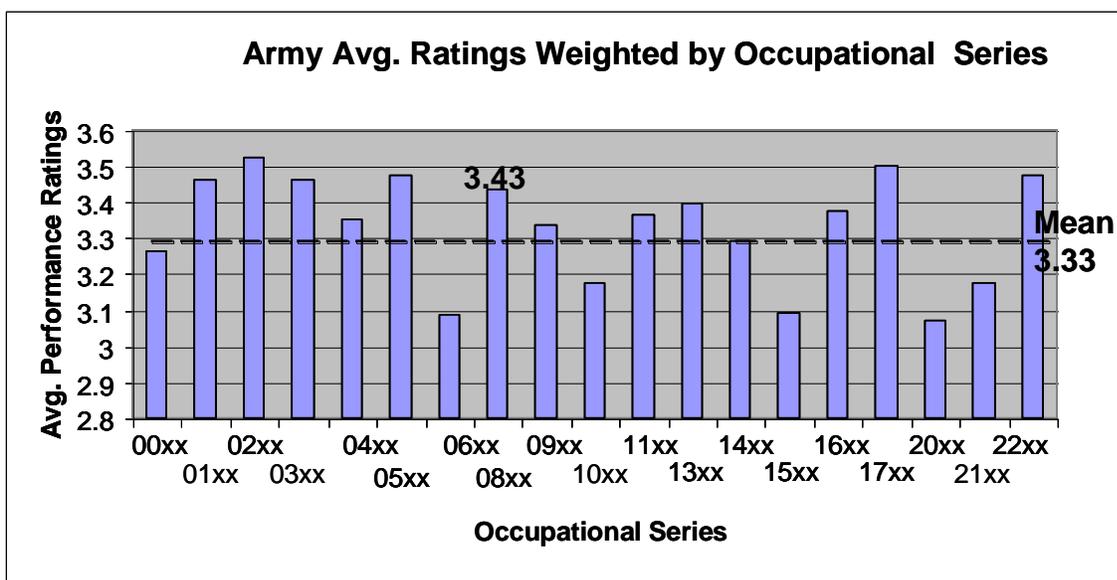
Table 16. Army Average Performance Rating by Occupational Series

ARMY AVERAGE PERFORMANCE RATINGS BY OCCUPATIONAL SERIES																				
Occupational Series	0000	0100	0200	0300	0400	0500	0600	0800	0900	1000	1100	1300	1400	1500	1600	1700	2000	2100	2200	Mean
Occupational Series Ratings Average of All Pay Pools	3.32	3.52	3.56	3.47	3.41	3.52	3.07	3.48	3.35	3.19	3.41	3.41	3.29	3.61	3.35	3.5	3.07	3.19	3.5	3.38
Difference of Occupational Series Ratings Averages of All Pay Pools vs 0800	0.16	-0.04	-0.08	0.01	0.07	-0.04	0.41		0.13	0.29	0.07	0.07	0.19	-0.13	0.13	-0.02	0.41	0.29	-0.02	0.11

Several points are notable from the differences of Army occupational series rating averages featured in Table 16. First, the 0800 occupational series performance ratings scored higher in two thirds (66%) of comparative cases. This contrasts to 60% and 100% in the case of the Navy and Air Force differences of occupational rating averages. Second, the Army average of difference of average occupational performance ratings (.11) is almost twice the level of the Navy (.067). Yet both are distinctively less than the .27 Air Force average of difference of average occupational performance ratings. Third, the margin of difference between occupational series that outscored the 0800 occupational series of the Army data set is minor—no more than .02 with the exception of the 1500 occupational series. The margin of difference was small (.01) in the case of the administrative occupational series (0300) compared to the engineering occupational series. In sum, distinctions of significance are not readily apparent in the Army NSPS performance rating data set for the engineering occupational series in contrast to other occupational series.

The weighted average perspective can also be applied to the average performance ratings of all Army occupational series. A weighted average considers the size of population as a factor when comparing performance rating scores. For example, the Army engineer population is over 5000, while the accountant and resource manager population are close to 500 personnel each. The weighted average of the average performance ratings by occupational series shows the 0800 occupational series average performance rating of 3.43 is above the mean of average of performance ratings (Figure 18), but only by a margin of .10. Furthermore, the 0800 occupational series ranks 7th of 19 occupational average performance ratings. These two observations do not indicate the presence of preferences in performance ratings for the engineering occupational series. In fact, the personnel management occupational series (0200) ranks 1st with a weighted average performance rating of 3.52. Some would characterize the HR function as a support function, not a mission critical function in an engineering organization. This counters observations made by prominent researchers and to paraphrase Mosher, that engineers in the engineering agency are favored.

Figure 18. Weighted Average Performance Ratings-Army



Key Finding-1.

No prominent anomalies are found using comparative differences of average occupational performance ratings between engineers and other occupations in three DOD engineering organizations. The 0800 occupational series did not rank in the top five positions using weighted averages across data sets used. Navy average occupational performance rating scores indicates the 0800 occupational series ranked 7th of 14. This does not indicate a position of preference. Furthermore, the 0800 occupational series ranks 7th of 19 occupational average performance ratings in the Army engineering organization data set.

Performance Rating Data Test 2.

The second performance rating data test asks if there are any significant differences, often referred to as anomalies, in the distribution of performance ratings by occupational code. For instance, did the engineering occupational series in Navy and Army data set during FY 2008 organizations show statistically significant higher performance ratings compared to other occupations in pay pools? The Air Force data set does not provide sufficient data to compare across multiple occupations.

A sample test for significance of the difference of means of the performance ratings of the engineers and biologists in the Navy data set is seen in Table 17. The test begins by subtracting the average performance rating of the 0400 occupational series (3.309) from the 0800 occupational series (3.416). The mean difference between the performance ratings of engineers and biologist is .108. The third column in Table 17 solves for standard error of the sample mean, while the fourth column is the square of the standard error which provides mean variances. By summing the mean variances of the engineer and biologist performance ratings, the variance of the mean difference of .013 is found. The square root of the mean variance (.013) provides the

standard error of the difference between the performance ratings of the engineer occupational series and the biologist series. How does this relate to the hypothesis that that more favorable performance ratings occur in mission critical occupations in occupational pay pools in Federal agencies, such as engineers in the Navy’s engineering organization?

Table 17. Sample Occupational Series Performance Rating Test of Significance

OCCUPATIONAL SERIES	SAMPLE MEAN	STANDARD ERROR OF SAMPLE MEAN	SQUARED STANDARD ERROR (MEAN VARIANCE)
0800 Engineer Series	3.416	.042	.001
0400 Biologist Series	3.309	.105	.011
Mean Difference	.108		
Variance of the mean difference			.013
Standard Error of the mean difference			.113
Test “t” statistic			.095

Using a heuristic similar to that used in *The Essentials of Political Analysis* (Pollock, 2005) aids in understanding the test of significance seen in Table 17. More specifically, “if the sample statistic is at least twice as large as its standard error then the result surpasses the .05 threshold and the null hypothesis can be rejected” (Pollock, 2005, p. 134). In this case the mean difference of .108 between engineers and biologists performance ratings is not twice as large as the standard error of .113, and the null hypothesis should be considered as valid. The null hypothesis in this dissertation is that no significant difference in performance ratings are present between mission critical occupations and other occupations in occupations pay pools in Federal Agencies.

The illustration in Table 17 provides a single test for significance of the difference of means of the performance ratings of the engineers and biologists in Navy NSPS data set from FY2008. Expanding this sample test to an organization wide perspective provides a more robust picture of the potential for determining the case for occupational bias. Thus reiterating, the organizational level is the unit of analysis employed to confirm the hypothesis that more favorable performance ratings occur in mission critical occupations in occupational pay pools. The sample test of significant of the difference of means between the engineering occupation average performance ratings and the biologist occupational series seen in Table 17 is expanded. It is expanded and applied between the engineer series performance ratings and all valid occupational series average performance for the Navy and Army engineering organizations.

The Navy test of significance of the difference of means between the engineering occupational series and other occupations of means is found in Table 18. Repeating the use of the heuristic introduced by Pollock, the sample statistic, in this case the mean difference in Column 8, Table 18 must be least twice as large as its standard error in Column 10 to surpass thresholds to indicate an anomaly.

It is seen in Table 18 that the mean differences found in column 8 are not twice as large as the standard error in column 10. There is one distinctive difference worth noting and that is the differences between the engineering series (0800) and the legal series (0900). The score of -.206 indicates an anomaly for the 0900 occupational series. This may be due to one of the standard effects seen in Table 7. In this case the similar to me effect may be caused by caused by supervisors of lawyers inflating ratings. The cause of the 0900 occupational series may also be attributed to not following local NSPS guidelines that emphasize that supervisors must rate employees against performance objectives.

In sum, there are no significant differences between the engineering occupational series and the other occupational series performance ratings. Therefore, for this part of the analysis, the null hypothesis is validated. In the case of this dissertation, no significant statistical differences are detected between the Navy engineering occupational series performance ratings and other occupations.

Table 18. Navy Occupational Series Difference of Means Tests of Significance

Occupation Series	Average of All Pay Pools (PPs)	Number of PPs	Standard Deviation of Occupational Series Rating Averages	Square Root of Number of PPs	Standard Error of the Mean	Mean Variance	Mean Difference Between 0800 & other occ. series avg. ratings	Variance of mean difference	Standard Error of the Mean Difference	Test Statistics or # of Standard Errors
00xx	3.314	13.000	0.172	3.606	0.048	0.002	0.102	0.004	0.064	1.607
03xx	3.456	20.000	0.193	4.472	0.043	0.002	-0.039	0.004	0.060	-0.653
04xx	3.309	7.000	0.277	2.646	0.105	0.011	0.108	0.013	0.113	0.952
05xx	3.538	9.000	0.241	3.000	0.080	0.006	-0.122	0.008	0.091	-1.341
09xx	3.622	9.000	0.171	3.000	0.057	0.003	-0.206	0.005	0.071	-2.908
11xx	3.432	18.000	0.176	4.243	0.042	0.002	-0.016	0.004	0.059	-0.263
13xx	3.390	7.000	0.218	2.646	0.082	0.007	0.026	0.009	0.093	0.282
16xx	3.285	10.000	0.139	3.162	0.044	0.002	0.131	0.004	0.061	2.152
21xx	3.353	4.000	0.124	2.000	0.062	0.004	0.064	0.006	0.075	0.848
22xx	3.513	10.000	0.269	3.162	0.085	0.007	-0.097	0.009	0.095	-1.019
	Note: This column is the average occupational ratings of all PPs	Note: This column is the number of PPs	Note: This column is the average distance from mean	Note: This column is the square root of all PPs containing occ series	Note: This column is the standard error is equal to the $\text{stddev}/\sqrt{\text{tof}}$ PPs. It is the error associated w/ea_pp	Note: This column is the squared std err from previous column	Note: This column is the mean difference between 08xx and other occ series avg ratings	Note: This column is the sum of the mean variance between the two occ. series performancer ating	Note: This column is the square root of the mean variance, ie the previous column	Note: This column is the mean difference divided by the standard error of the difference; this is the "t" statistic

The Army test of significance of the difference of means is found in Table 19. Table 19 provides the Army average occupational series performance ratings of all pay pools, and all other calculations needed to perform the significance of means test. The test question this calculation seeks to address is: are there significant differences in the engineering occupational performance ratings when compared to other occupational performance ratings in the Army's organizational data set?

The following observations are found in the Army NSPS performance rating data set using the same process (Pollock, 2005) used in the Navy significance of means test. The process

uses a heuristic that states if the sample statistic is at least twice as large as its standard error then the result surpasses the .05 threshold and the null hypothesis can be rejected. Table 19 indicates that in nearly all cases the mean differences found in column 8 when doubled do not exceed the test statistic found in the column 10. The medical (0600), information (1000) and supply (2000) occupational series do show significance. In other words, the engineering occupational ratings in these cases indicate a significant difference that may need to be investigated further for some form of error. However, three of fifteen comparative tests do not signal systemic preferences towards the mission critical occupation (i.e. engineering occupational series). Similar to the previous tests of significance in the Navy data set, the Army analysis leads to a finding that there are no significant differences in the occupational series comparisons.

In sum, there are no significant differences between the engineering occupational series and the other occupational series performance ratings. Therefore, for this part of the analysis, the null hypothesis is validated.

Table 19. Army Occupational Series Difference of Means Tests of Significance

Occupation Series	Average of All Pay Pools (PPs)	Number of PPs	Standard Deviation of Occupational Series Rating Averages	Square Root of Number of PPs	Standard Error of the Mean	Mean Variance	Mean Difference Between 0800 & other occ. series avg. ratings	Variance of mean difference	Standard Error of the Mean Difference	Test Statistics or # of Standard Errors
00xx	3.321	43.000	0.276	6.557	0.042	0.002	0.157	0.002	0.046	3.385
01xx	3.525	22.000	0.317	4.690	0.068	0.005	-0.047	0.005	0.070	-0.667
02xx	3.555	2.000	0.276	1.414	0.195	0.038	-0.077	0.038	0.196	-0.393
03xx	3.466	127.000	0.263	11.269	0.023	0.001	0.012	0.001	0.030	0.401
04xx	3.413	71.000	0.279	8.426	0.033	0.001	0.065	0.001	0.038	1.688
05xx	3.517	48.000	0.319	6.928	0.046	0.002	-0.039	0.003	0.050	-0.773
06xx	3.065	2.000	0.092	1.414	0.065	0.004	0.413	0.005	0.068	6.086
09xx	3.346	43.000	0.056	6.557	0.009	0.000	0.132	0.000	0.021	6.205
10xx	3.188	43.000	0.239	6.557	0.037	0.001	0.290	0.002	0.041	7.021
11xx	3.409	32.000	0.300	5.657	0.053	0.003	0.069	0.003	0.057	1.224
13xx	3.405	43.000	0.286	6.557	0.044	0.002	0.073	0.002	0.048	1.520
16xx	3.352	13.000	0.187	3.606	0.052	0.003	0.126	0.003	0.055	2.266
20xx	3.073	3.000	0.064	1.732	0.037	0.001	0.405	0.002	0.042	9.653
21xx	3.190	3.000	0.259	1.732	0.149	0.022	0.288	0.023	0.151	1.912
22xx	3.498	6.000	0.169	2.449	0.069	0.005	-0.020	0.005	0.072	-0.284
	Note: This column is the average occupational ratings of all PPs	Note: This column is the number of PPs	Note: This column is the average distance from mean	Note: This column is the square root of all PPs containing occ series	Note: This column is the standard error is equal to the stddev/sqrt of PPs. It is the error associated w/ea pp	Note: This column is the squared std err from previous column	Note: This column is the mean difference between 08xx and other occ series avg ratings	Note: This column is the sum of the mean variance between the two occ. series performance rating	Note: This column is the square root of the mean variance, ie the previous column	Note: This column is the mean difference divided by the standard error of the difference; this is the "t" statistic

Key Finding-2.

Navy and Army organization engineer occupational performance ratings and all other valid occupational performance rating difference of means are tested for significance. No systemic (e.g. across all comparisons) anomalies in the difference of means between engineer and other occupations were found. Two significant anomalies were present. In the Navy data set, the score of -.206 indicates an anomaly for the 0900 occupational series when compared to the engineering series (0800). In the Army data set, the medical (0600), information (1000) and supply (2000) occupational series do show significant difference of means when compared to the engineering occupational series. However, three of fifteen comparative tests do not signal systemic preferences towards the mission critical occupation (i.e. engineering occupational series). Recognizing the lack of significant of differences, the quantitative analysis validates the

null hypothesis. The null hypothesis is that no significant differences in performance ratings are present between mission critical occupations and other occupations in occupations pay pools in the Army and Navy organizations.

Analysis of Archived Documents

The methodology used to test the hypothesis is comprised of three parts: a quantitative analysis of NSPS pay pool rating data, analysis of archived documents and expert interview analysis. The qualitative analysis of archived documents determines the presence and value of Federal agency performance rating information and analysis that contributes to the validity of the hypothesis. The analysis of archived documents looks for indicators of occupational bias in four categories of documents: U.S. Government reports, U.S. Congressional testimony, books, and journal articles.

Analysis of Archived Documents: Government Reports.

Eleven U.S. Government reports were analyzed for application to the topic of performance rating preferences in mission critical occupations. The reports originated from the U.S. Government Accountability Office, the U.S. Congressional Research Service (CRS), and other government agencies. An abbreviated summary (Table 20) of pertinent U.S. Government reports provides several contributions to the hypothesis. A completed summary of all government report reviews used in this dissertation can be found in Appendix A.

Of the government reports analyzed, it was found that that ten of eleven reports addressed the topic of performance rating analysis frameworks, ten of eleven reports addressed performance rating anomalies, while just four of eleven contained some form of notable reference or discussion relating to occupational bias in performance ratings. Each report is assessed by three criteria (conducted rating analysis, anomalies in performance ratings, and

occupational bias in performance rating) and given a qualitative rating that is seen at the bottom of Table 20. The criteria are used to assess the reports on a qualitative rating scale comprised of three levels: 1 (poor), 2 (fair), 3 (good). For example, does the report under review offer a poor (1) level of discussion on the topic of anomalies in performance ratings that contributes to the validity of the hypothesis? This qualitative analysis serves as a gateway for further discussion of the report's information on occupational bias in performance appraisals.

The qualitative rating of 2.15 for the key element of *conduct rating analysis* (Column 3, Row 7, Table 20) indicates that many of the reports provided a fair level of discussion on rating analysis, and more importantly, that some agencies are performing statistical analysis of performance rating data. The qualitative rating of 2.08 (Column 6, Row 7, Table 20) for *anomalies of performance ratings* indicates that there is a fair level of discussion on the topic. Illustratively, a GAO report (U.S. Government Accountability Office, 2005) explicitly discussed documented evidence of gender bias at five Department of Energy locations. The key element of *occupational bias in performance ratings* (Column 7, Row 7, Table 20) has a summary qualitative rating of 1.67 which indicates a poor level of discussion of occupational bias across the U.S Government Reports analyzed for this research effort. Yet, in four of the ten reports examined, there is both explicit and implicit reference to occupational bias in performance ratings. These reports, as discussed below, provide anecdotal evidence that the workforce perceives that occupational bias is occurring in performance appraisals.

The GAO during the period from 2000-2010 focused on the challenges of human capital planning during the new millennium. Between 2005 and 2009 sixteen GAO reports and testimonies documented DOD's effort and challenges in implementing a new HRM system--NSPS. The challenges facing NSPS ranged from the cost of implementation to addressing

employee perceptions about forced distribution of ratings. GAO Report -10-102 (U.S. Government Accountability Office, 2009) documented anecdotal evidence on occupational bias. Discussions in group settings conducted by GAO indicated that “it was difficult for employees in certain positions to receive a rating of “5” because of the nature of their work or the perceived value their management placed on those positions” (U.S. Government Accountability Office, 2009, p. 40). In that same GAO report, NSPS employees told GAO that “they do not believe that some military supervisors value the work of employees who perform certain job functions, such as providing child care on an installation” (U.S. Government Accountability Office, 2009, p. 45).

The Congressional Research Service (CRS) also focused on NSPS issues during the first decade of the 21st Century. One CRS report (Ginsberg, 2008, p. 21) revealed that some civilian employees working at DOD level agencies were provided higher performance ratings than civilian personnel working in the Armed Services. Thus, civilian personnel working higher level jobs at the highest level of DOD, for example positions at the Pentagon, may have received inflated pay-for-performance ratings than those jobs away from the Pentagon.

Congress mandated that annual assessments of NSPS performance be provided to insure compliance with merit principles. More directly, DOD’s NSPS Program Managers were required by P.L. 108-136 to provide safeguards to ensure that the management of the system was fair and equitable. Annual NSPS performance assessments were the safeguard. Three annual assessments were conducted either by or on behalf of the NSPS Program Executive Office (PEO).

The 2008 NSPS annual assessment found that there was a perception that non-mission central jobs and low visibility jobs were likely to receive low ratings. (SRA International Inc,

2009). This point, highlighted in Chapter 1 of this dissertation, was documented through interviews conducted for the annual assessment. One employee stated

There's an institutional bias or preference to certain functions. Once you're in a certain areas, you're branded. We were told that really high payouts were skewed to more 'central' functions. How does a system reward people who are deserving and doesn't penalize other people who are doing well in 'less central' functions. (SRA International Inc, 2009, p. 3-7).

A Merit Systems Protection Board report offered potential for uncovering empirical evidence to support the hypothesis. The report showed that some agencies performance ratings are examined for preferences by geographic location and occupation. This examination is performed at the Office of Currency Control (OCC) at the Department of the Treasury (U.S. Merit Systems Protection Board, 2006). The OCC also examines performance ratings for violations of protected class provisions. However, despite the potential for obtaining empirical evidence, requests for follow-up information at the Office of Currency Control were met with resistance. A GAO report also documented requests for OCC performance ratings analysis by union officials. These union officials met with resistance as well (U.S Government Accountability Office, 2007).

In sum, no empirical on occupational bias was uncovered in government reports, and thus supports the null hypothesis. Yet, anecdotal evidence suggests, in several government reports, that occupational bias is perceived to be present. This suggests that there is potential for obtaining empirical evidence at Federal agencies, such as the OCC, to substantiate occupational bias in performance appraisals.

Table 20. Abbreviated Summary: Occupational Bias Related Archived Reports

ABBREVIATED SUMMARY: OCCUPATIONAL BIAS IN REPORTS							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
1	GAO-10-102, Oct, 2009, Monitoring of Safeguards & Addressing Employee Perceptions Are Key to Implementing a Civilian PMS	Analyzed the rating results for more than by demographic groups, such as race, gender, ethnicity, age, veteran status, and target disability	2	NSPS may inadvertently favor employees who work closely w/ member of the pay pool panel; those individuals have direct knowledge	2	They do not believe that some military supervisors value the work of employees who performance certain job functions, such as providing child care on an installations	3
8	CRS RL 34673, Sep 2008, Pay-for-Performance: The NSPS	Provide effective safeguards to ensure that the management of the system is fair and equitable and based on employee performance	2	NSPS was implemented inconsistently & infused with bias	3	Employees with a lower rating in one office could receive a higher pay bonus than an employee with a higher rating in another	2
10	US Merit System Protection Board: The Practice of Merit: A Symposium Report	Appropriate independent reviews; reasonableness reviews, internal grievance procedures, internal assessments	2		1	We don't look at it just for race, but we look at it by grade, by occupation, and by location	3
11	PEO NSPS 2008 Evaluation Report, May 2009	Data from the PAA and CWB supplemented by other personnel data from DCPDS support the analysis of performance ratings	3	Rating and payout analyses suggest that payouts are higher for higher-paid employees	3	Som employees were under the impression that more mission-central or higher level jobs are more likely to receive true pay-for-performance	3
↓	11 reports are reviewed; full review table can be found in Annex. This abbreviated table shows a sample of entries; final qualitative rating average of all reports seen below. Qualitative ranking is on a 1(low value) to 3 (high value) scale.						
11	Average of Qualitative Ratings		2.15		2.08		1.67
Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography Note: Textual description in column represents a paraphrased summary of a key point							

Analysis of Archived Documents: Congressional Testimony.

Several sessions of U.S. Congressional Testimony were scanned for relevance to occupational bias in performance appraisals. Testimony from the Committee on Oversight and Government Reform, U.S. House of Representatives on February 12, 2008 was analyzed. Pertinent points are found in Table 21, while a completed summary of all selected testimony can be found in Appendix A to this dissertation.

The review of testimony found eight of ten testimonies addressed the topic of performance rating analysis frameworks, eight of ten testimonies addressed performance rating anomalies, while just two of ten testimonies contained some form of significant reference or discussion of occupational bias in performance ratings. An average of qualitative ratings for each criteria (conducted rating analysis, anomalies in performance ratings, and occupational bias in performance rating) of analysis is seen in bottom of Table 21. The qualitative rating of 2.18 for the key element of *conduct rating analysis* (Column 3, Row 7, Table 21) indicates that many of the testimonies provided a fair level discussion of rating analysis. More importantly, Dr. Stanley Ridley states “there must be a statistically significant pass/success rate of the higher scoring/rated employee group relative to the employee group thought to be experiencing the adverse impact” (*Statement of Dr. Stanley E. Ridley, president and CEO of Ridley & Associates, Inc, 2008, p. 4*). The cogent point is that the organization must have a performance rating analysis framework in place to detect or conduct analysis of performance rating anomalies. GAO (U.S. Government Accountability Office, 2008) reported a similar weakness in the NSPS performance management system in 2008. GAO found that NSPS contains no post rating performance analysis system designed to detect anomalies, such as occupational bias. The qualitative rating of 1.91 (Column 6, Row 7, Table 21) for the criteria of *anomalies of*

performance ratings indicates that there is a poor level of discussion on the topic, with two testimonies explicitly discussing documented evidence of bias, most significantly at the Securities and Exchange Commission. The criteria of *occupational bias in performance ratings* (Column 7, Row 7, Table 21) has a summary qualitative rating of 1.18 which indicates a poor level of discussion of occupational bias in Congressional testimony.

Despite the absence of a significant amount of testimony on occupational bias in performance appraisals, there is a notable exception. John Gage, President of the American Federation of Government Employees, discussed the same as me effect, and the bias towards higher levels of positions in the Pentagon. Mr. Gage also revealed that senior managers explained how the work of employees in the lower grades, is also systematically undervalued in the NSPS pay system (*Statement of John Gage, president, American Federation of Government Employees, 2008*). This testimony aligns with findings in an NSPS evaluation report that stated “ratings and payout analysis suggest that payouts are higher for higher-paid employees (SRA International, 2009, p. ES-4). Similar findings, found in another Federal PfP system, found higher grades achieve higher ratings at the National Institute of Standards and Technology (NIST) (Eremin, Wolf, & Woodard, 2010).

A subtle point emerges that correlates with testimony and remarks provided by expert interviews conducted for this dissertation. De-facto occupational bias may occur when high performance ratings favor highly rank personnel. For example, in some organizations, particularly organizations dominated by engineer and scientific occupations, occupational bias in performance ratings may favor physical scientists (1301 series) over administrative support (0300 series). This is because the 1301 series populates the higher grades in an engineering and scientific agency, while the 0300 series populates the lower grades.

Table 21. Abbreviated Summary: Occupational Bias Related Archived Testimony

ABBREVIATED SUMMARY: OCCUPATIONAL BIAS IN TESTIMONY							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
1	Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Testimony of Charles Tiefer	They receive their pay raises in the same way as they receive personnel actions, like promotion and discipline, known to have statistical patterns reflecting systemic risks	3	At the SEC, as similarly would be seen about NSPS, the criteria for pay-for-performance left room for subjective evaluation;for discriminatory stereotyping	2		1
3	Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Testimony of John Gage		1	Supervisors in charge of judging employees have a natural tendency to favor people like themselves	2	A profound bias in favor of employees who work higher up the chain of command...closer to Pentagon; there also may be hierarchies of bias; it is best to work in the Pentagon	2
7	Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Testimony of Max Stier	The GAO took the bold step of publicizing the agency's promotions and performance ratings by race; data revealed a gap between performance appraisals for African-Americans and whites	3	Simply put, a PFP system is doomed to fail if the affected employees do not perceive the system as fair for all, regardless of gender, race, religion, political affiliation, marital status, age, veteran status or disability	2		1
↓	10 testimonies are reviewed; full review table can be found in Annex. This abbreviated table shows a sample of entries; final qualitative rating average of all reports is seen below. Qualitative ranking is on a 1(low value) to 3 (high value) scale.						
10	Average of Qualitative Ratings		2.18		1.91		1.18
Note: Citations are abbreviated for presentation purposes; full citation are found in bibliography Note: Textual description in column represents a paraphrased summary of a key point							

Analysis of Archived Documents: Books.

A number of books were reviewed for anecdotal and empirical materials pertaining to PFP and performance management. The books covered a forty year period from 1968-2008. Twelve selected books are examined and analyzed that addressed the topic of occupational bias in performance appraisal, either explicitly or implicitly. The completed summary of all reviews can be found in Appendix A, with the pertinent selections seen in Table 22.

Compared to information analyzed in government reports and testimonies, the criteria of *occupational bias in performance ratings* (Column 7, Row 7, Table 22) has a higher summary qualitative rating of 2.12 which indicates a fair level of discussion of occupational bias in the books analyzed for this research effort. Three books provide valuable insights into occupational bias in performance appraisals that support the hypothesis.

Frederick Mosher, as early as 1968, recognized that a single occupational group identified with the mission of the organization can influence organizational operations. Mosher highlights classes of occupations, such as the civil engineers in a construction agency, commanding prestige and influence (Mosher, 1968). The influence of the professional elite affects the upgrading of rewards, such as pay, for professionals of the organization. The upgrading of awards is linked to favorable performance appraisal ratings. Mosher recognizes the drive of every profession to improve their prestige among associates and the public. The friction occurs, when this drive runs counter to merit principles. More specifically, friction occurs when the drive to improve the professional elite of an organization aligned with the mission of the agency, runs counter to the merit principles of fair and equitable treatment in performance appraisals. Mosher explains that “the means whereby the professionals assert their control over

personnel policies and actions are many and diverse. Some are specified and required by law and/or regulation; others grow out of gentlemen's agreements within" (Mosher, 1968, p. 125).

Jeffrey Pfeffer supports Mosher's perspectives in his works on examining the sources of power in organizations. Pfeffer attributes one source of power in organizations to those units responsible for accomplishing more important tasks (Pfeffer, 1981). Pfeffer uses the example of the retailing organization and the power of the retailing function over other functional groups in the store organization (Pfeffer, 1981). The sustainment of power by the central function over other functional groups is achieved by a sustained effort to assure that they remain central to organization success.

Murphy and Cleveland, in their seminal work from 1991 explicitly discuss the status difference among functions and their effect on performance appraisals. Similar to the occupational bias anecdotal evidence in government reports, Murphy and Cleveland found anecdotal support for the influence of an agency upon ratings. Previously discussed in this dissertation, their anecdotal evidence found that "employees whose jobs or activities were perceived as more closely linked with the organization's mission or purpose (i.e., what was valued) were rated systematically higher than employees in jobs that were perceived as less central to the organization's mission" (Cleveland & Murphy, 1991, p. 16). Murphy and Cleveland's anecdotal evidence is supported by Bjerke (1987) who indicated that high-status occupational communities were more likely to be recipients of rating inflation than low-status occupational communities.

In sum, Pfeffer, Mosher, and Murphy and Cleveland's research found no empirical evidence to support the hypothesis. However, anecdotal assertions in their work leads to the supposition that those units responsible for accomplishing the mission critical tasks and activities

position themselves for power in the organization by influencing performance appraisals and ratings. This supposition supports the hypothesis.

Table 22. Abbreviated Summary: Occupational Bias Related Archived Books

ABBREVIATED SUMMARY: OCCUPATIONAL BIAS IN BOOKS						
Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
Murphy & Cleveland, 1991, Performance Appraisal: An Organizational Perspective	The variance of the mean ratings calculated for each department in an org. should be positively correlated with the status differentiation within that organization	3	In addition, Cleveland and colleagues found anecdotal support for the influence of an organizational value system upon ratings	2	Specifically, employees whose jobs or activities were perceived as more closely linked with the organization's mission or purpose were rated systematically higher	3
Mosher, 1968, Democracy and the Public Service		1	Control over employment policies and individual personnel actions for its own members in the agency and over the employment of employees not in the elite	2	Within the basic content of the agency, the officers of the line in the Navy...the civil engineers in a construction agency are favored	3
Jeffrey Pfeffer, 1981, Power in Organizations	There was no correlation between departmental prestige and the indicators of subunit power	3			As noted before, none of the staff groups has much power, for the power in the firm was held by retailing, the store management function.	3
12 books are reviewed; full summary table can be found in the Annex. The abbreviated table shows a sample of entries with the final qualitative rating average of all entries seen below. Qualitative ratings is on a 1 (low value) to 3 (high value) scale.						
Average of Qualitative Ratings		1.50		2.27		2.12
Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography Textual description in column represents a paraphrased summary of a key point						

Analysis of Archived Documents: Journals.

Selected journal articles related to PfP and Performance Management Systems and were scanned for information on occupational bias in performance appraisals. Twenty-one articles are reviewed that addressed the topic of Pay-for-Performance (PfP) with some form of implicit or explicit reference to occupational bias. A summary of all reviews can be found in Appendix A, while a summary of key entries is seen in Table 23.

It is found that eight of twenty-one (38%) of the articles contained some discussion of performance rating analysis frameworks; an average qualitative rating of 1.62 (Column 3, Row 7, Table 23). Seventeen of twenty-one (80%) selected articles contained discussion of performance anomalies; an average qualitative rating 2.17 (Column 5, Row 7, Table 23). Seven of twenty-one (33%) of the selected articles contained some explicit or implicit discussion of occupational bias; an average qualitative rating of 1.45 (Column 7, Row 7, Table 23). The low qualitative score of 1.45 seen in Table 23 is misleading upon further examination of findings three researchers.

The importance of mission critical occupations to the organization and their subsequent influence on performance appraisals has been examined by Professor Dennis Daley (Daley, 2003). Daley's discussions on the problems of performance appraisals in the 21st century, as mentioned in the literature review, touch upon occupational bias in performance appraisals.

Daley finds that:

Individuals working in a critical unit may benefit from the perceived centrality or significance of their part of the organization. In this case the importance of the unit to fulfilling the organization's mission is substituted for the job performance of the individual in that unit. In neither instance is the individual's job performance actually measured (Daley, 2003, p. 161).

Daley's assertions run parallel to anecdotal evidence found by Murphy and Cleveland.

Jost and Elsbach (Jost, J. T., & Elsbach, K. D. (2002) provide insights into the intersection of the practice of pay-for-performance systems, performance rewards, and Social Identity Theory. These researchers' observations on the role of mission critical business lines and members of the high-status groups associated with those business lines are central to the hypothesis in this dissertation. Jost and Elsbach, as outlined in this dissertation's theoretical review, state the reality of society and business is that "members of high-status groups are far more likely to engage in discrimination and in-group bias against lower status group members" (Jost, J. T., & Elsbach, K. D, 2002, p. 186).

Applying the concept of in-group bias further, Jost and Elsbach's work can be used to illustrate the tenets of Social Identity Theory (SIT) with occupational bias in performance appraisals. The key component of SIT that distinguishes it from other social theories is the role of favoritism in the organization, as previously noted in Chapter 3. In the case of SIT, it "does a very good job of predicting the behavior of established high-status and powerful group members, who exhibit strong levels of in-group bias" (Jost, J. T., & Elsbach, K. D. , 2002, p. 184). It can be assumed accord to SIT, that favoritism is a mechanism to achieve the ideal state of status--prestige and power for the group. Favoring engineering occupations in engineering organizations through inflated performance appraisals ratings can be considered as a personification of SIT.

The low qualitative rating of 1.45 for discussion of occupational bias in performance appraisals in journals also masks the importance of politics in performance appraisals. As early as 1978 (DeCotiis & Petit, 1978) noted that raters from different organizational levels (i.e. Headquarters vs regional) may have difference perceptions of the impact of ratings to achieve organizational outcomes. This perspective aligns with views (Cook, 1995; Longenecker & Goff,

1992) that politics can influence the principle of merit in performance appraisal. “A surprising number of people will bend those rules to satisfy a personal, departmental, or divisional agenda” (Longenecker & Goff, 1992, p. 47).

Longenecker provides additional insights into how executive appraisal is a political tool used to control organizational personnel and resources (Longenecker & Sims, 1994). His work can be viewed as prescient when applied to the Pay Pool Panels role in NSPS. “Because executive raters have a great deal of autonomy, they are in a position of influence. Ratings have an impact on pay, promotions, power, and career paths, so they can be use to enhance the rater and protect his or her self-interests” (Longenecker & Sims, 1994, p. 53). It is a reasonable assumption that Longenecker would suspect that engineers in a engineering organization pay pool panel would be inclined to favor the 0800 occupational series in performance appraisal ratings.

In sum, as illustrated through the works of Daley, Jost and Elsback, and Longenecker, there is no documented empirical evidence that supports the hypothesis. There is some anecdotal evidence and theoretical foundation for preferences of mission critical occupations in performance appraisals.

Table 23. Abbreviated Summary: Occupational Bias Related Archived Journals

ABBREVIATED SUMMARY: OCCUPATIONAL BIAS IN ARTICLES							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
10	Daley, 2003, The Trials and Tribulations of Performance Appraisal		1		1	The importance of the unit to fulfilling the organization's mission is substituted for the job performance of the person in unit	3
11	Cook, 1995, Performance Appraisal and True Performance		1	Performance ratings suffer from many biases such as age, ethnicity, gender, and physical appearance.	2	Political considerations or private agendas nearly always influence performance ratings	2
15	Jost and Elsbach, 2001, How Status and Power Differences Erode Social Identities		1	Social identity theory does a very good job of predicting high-status members who exhibit strong levels of ingroup bias	2	In business organizations members of high-status groups are far more likely to engage in bias against lower status group members	2
18	Smith, DiThomaso, and Farris, 2001, Favoritism, Bias, & Error in Perf. Ratings	Analyze performance ratings for a sampe of 2,445 scientists and engineers from 24 US companies	3	It is found that there is more evidence of in-group favoritism than out group derogation	3	High status, dominant, and majority group members enjoy favoritism expressed as a global prototype of them as competent	2
↓	21 articles are reviewed; full review table can be found in Annex. This abbreviated table shows a sample of entries; final qualitative rating average of all reports seen below. Qualitative ranking is on a 1(low value) to 3 (high value) scale.						
21	Average of Qualitative Ratings		1.62		2.17		1.45
Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography Note: Textual description in column represents a paraphrased summary of a key point							

Key Finding-3

A qualitative analysis of archived documents was conducted to determine the presence of anecdotal and empirical evidence to support the hypothesis that mission critical occupations are favored in performance ratings in pay pools. There is a lack of empirical evidence in archived documents that supports the hypothesis, thus support for the null hypothesis is validated. Three points support this support this finding.

First, there is evidence that several Federal agencies are conducting ratings analysis for the purposes of determining anomalies in performance ratings. In fact, the use of statistical analysis in the analysis of performance ratings is highlighted in Congressional testimony as a best practice, but has yielded no documented empirical evidence supporting the hypothesis. Second, the topic of anomalies in performance ratings, particularly gender and racial bias in performance ratings, is found throughout all classes of archived documents. Yet little information surfaced on occupational bias anomalies. Third, information on occupational bias in performance ratings is found to be lacking in archived documents—but not completely absent. Government documents provide some anecdotal evidence. This lack of information may be due to the unwillingness of Federal agencies to openly proffer this information. There is anecdotal evidence and discussion on occupational bias in performance ratings in published books and journals, but is limited in scope. In sum, statistical analysis frameworks are being employed to review Federal agency performance rating data in some agencies. Information relating to gender and racial bias was found. Occupational bias in the form of explicit empirical evidence in reviewed archived documents is not evident; some anecdotal evidence is present.

Analysis of Expert Interviews.

The third part of the methodology to determine if mission critical occupations are favored in performance appraisals is analysis of expert interviews. The intent of the interviews is to determine the presence of occupational bias in the work or research of PFP subject matter experts. Over twenty experts were contacted, however only ten subject matter expert interviews were conducted and analyzed. An abbreviated summary of four interviews that provide the most pertinent points of interest is presented in Table 24, while a summary of all interviews is found in Appendix A to this document.

In general, the majority of the experts interviewed were not examining performance rating data for occupational bias. Most performance rating analysis discussions in interviews was focused on detecting anomalies in traditional disadvantaged groups (i.e. gender, race, etc.). The majority of interviews stated that they did not find, nor were they looking for occupational bias. However, four experts from their viewpoint, stated that they found some form of occupational bias in the workplace. Analysis of the interview content is centered on responses to three questions. These questions focus on conducting rating analysis, anomalies in performance ratings, and occupational bias in performance ratings.

Conducting Rating Analysis. Nine of ten interviewees used some form of post rating analysis framework or were aware of a framework used for those purposes. Only three used a rigorous statistical analysis of performance ratings. Interviewees familiar with NSPS and its demise stated that their agency did some post rating analysis for race or gender bias. The qualitative summary rating of 2.10 for all interviews in the key element of conduct rating analysis (Column 4, Row 8, Table 24) indicates that many of the interviews provided a discussion that was valued as fair. The takeaway is that some agencies are performing

statistical analysis on performance rating data, which provides a downstream indicator of the potential for the discovery of occupational bias in performance ratings. Thus, the likelihood of detecting anomalies in performance ratings is greater if an agency is using some form of post-rating performance analysis framework. GAO (U.S. Government Accountability Office, 2008) found similar findings in its analysis of NSPS.

Anomalies in Performance Ratings. Eight of ten interviews conducted for this dissertation provided comments stemming from their work on the issue of performance rating anomalies. Two interviewees did not respond with any explicit knowledge or were aware of anomalies in performance ratings in their agency or research. Four interviewees provided insights into finding racial, sexual, and ethnic group bias in performance ratings. These interviews are in alignment with racial bias cases highlighted in Congressional Testimony (*Statement of Colleen Kelly, president, National Treasury Employees Union, 2008*) and U.S. Government reports (U.S. Government Accountability Office, 2008). One interviewee also highlighted that his work discovered that high grades were achieving higher ratings (Federal Subject Matter Expert, 2010a). In summary, a fair qualitative rating of 2.2 (Column 6, Row 8, Table 24) is given for the discussion of anomalies of performance ratings.

Occupational Bias in Performance Ratings. The evaluative criteria of occupational bias in performance ratings (Column 8, Row 8, Table 26) has a summary qualitative rating of 2.2. This summary rating reflects a fair value of the discussion of occupational bias across all interviews. In four of the ten interviews conducted, experts stated that they had encountered explicit and implicit evidence of occupational bias in performance ratings. A Federal government policy analyst (Federal Subject Matter Expert, 2010a) indicated that occupations with higher status in the organization get higher ratings. A personnel management consultant

stated that in general, it is known that the mission critical jobs series get higher ratings than support jobs (Consultant Subject Matter Expert (May, 2010). A Federal agency personnel management manager stated that there is a difference in ratings in classes of occupations (Federal Subject Matter Expert, 2010d). His example used a comparison that engineers and scientists obtain higher performance ratings than administrative personnel. An academic specializing in the field of Pay for Performance (PfP) stated (Academic Subject Matter Expert, 2010) that there is occupational bias in ratings in academic field. For example some professors within a quantitative oriented department or program are more favored those professors with a qualitative orientation, especially when the raters have a quantitative orientation.

Several interviewees also recognized the “whys” behind occupational bias in performance ratings, which reflect discussions in theoretical literature. Some indicated that some occupational bias can be attributed to the “similar to me” effect. One interview stated “I can understand that people of the same cloth, like people of the same cloth” (Federal Subject Matter Expert, 2010b, p. 1). That same interviewee related what he termed the knowledge bias effect. For example, if the training unit had a member on the pay pool panel and the test unit did not, there is a preference towards members of the training unit. More simply, the pay pool panel member has a preference towards an employee performing a similar function to that of the panel member. These interview comments are in alignment with the tenets of Social Identity Theory discussed throughout this dissertation.

Two interviewees recognized the influence of organizational culture on performance ratings—particularly the value of mission critical occupations to the organization. One interviewee voiced the concern that if an employee worked the high profile programs that employee would obtain higher ratings (Federal Subject Matter Expert, 2010a). Similarly,

another interview revealed that the ability to manage the organization and its mission tends to favor people vital to getting the job done (Consultant Subject Matter Expert, 2010). These comments link to archived discussion by scholars (Daley, 1999) who attribute preferences in performance ratings to high profile programs and associated personnel working on those programs.

Table 24. Abbreviated Summary: Occupational Bias in Interviews

ABBREVIATED SUMMARY: OCCUPATIONAL BIAS IN INTERVIEWS						
Interviews	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
Interview #4	We have some looks at patterns in agencies or govt wide; looked at patterns for equity & merit	2	Do some examinations; found some anomalies; ethnic, gender; very clearly organizational differences	3	There are some occupational bias indications that there is differences; occupations w/ higher status=higher ratings	2.5
Interview #6	Look @ systems for faultiness; look at the mean differences which are of significance in valid systems	3	Empirically found race and gender differences in ratings in operational performance ratings	3	We know the mission critical jobs occupations get greater ratings than support jobs in general	2.5
Interview #7	We do it is because if there is any variance or /w labs, we bring it to the managers attn.	3	Higher end grades tend to get more higher grades	3	There is a difference in classes ie. engineers & scientist vs admin.	2.5
Interview #9	University does post performance review of all faculty	2	Considerable amount of general and race bias in the literature	2	Yes; academics have ideological preferences in rating; qualitative vs qualitative oriented	2.5
10 interviews are conducted; full interview table can be found in Annex. This abbreviated table shows a sample of entries; final qualitative rating average of all interviews seen below. Qualitative ranking is on a 1 (low value) to 3 (high value) scale.						
Average of Qualitative Ratings		2.10		2.20		2.20

Key Finding-4

The qualitative analysis of expert interviews found no empirical evidence to support the hypothesis. The qualitative analysis of expert interviews determined the presence of anecdotal evidence to support the hypothesis that mission critical occupations are favored in performance ratings in pay pools. This is based on three sub-findings. First, there is substantial knowledge in the use of performance ratings analysis frameworks by subject matter experts. The knowledge or use of rating analysis is a best practice for detection of performance rating anomalies. Second, the topic of anomalies in performance ratings, particularly gender and racial performance rating anomalies is acknowledged throughout discussions with interviewees. Third, occupational bias in performance ratings is recognized by subject matter experts interviewed for this research effort. Four of the ten interviews conducted mentioned that they had encountered explicit and implicit occupational bias in their reviews and research. These subject matter experts also highlighted the linkages between conducting rating analysis and the detection of bias. However, similar to archival documents, there is a lack of empirical evidence associated with the interviews conducted for this dissertation that supports the hypothesis.

Chapter 6: Conclusions

This dissertation assesses the hypothesis that more favorable performance ratings occur in mission critical occupations in mixed occupational pay pools in Federal agencies. The null hypothesis is that no significant difference in performance ratings is present between mission critical occupations and other occupations in mixed occupational pay pools.

The methodology used to address this dissertation's hypothesis is a quantitative-qualitative inquiry that investigates NSPS performance ratings within three DOD engineering agencies. The methodology is comprised of three parts: a quantitative analysis of pay pool rating data, analysis of archived documents and expert interview analysis. Each part of the methodology's analysis provides findings to determine the validity of the hypothesis that the engineering occupational series achieve more favorable ratings than other occupational series in DOD engineering agencies.

Summary of Findings

Key Finding-1: No distinctive performance rating distribution anomalies are found in the difference of average occupational performance ratings.

Indicators of occupational bias were not found using differences of average occupational performance ratings between engineers and other occupations in three DOD engineering organizations. An essential observation contributes to this finding.

The engineering occupational series did not rank in the top five positions of performance rating averages using a statistical weighting approach. Navy average occupational performance rating scores showed the engineer occupational series ranked 7th of 14 in average performance ratings in all pay pools. Furthermore, the engineer occupational series ranked 7th of 19 average occupational performance ratings in the Army engineering organization pay pool data set. It

would be expected that if favoritism towards the engineering occupational series in performance ratings was shown, that a more commanding position in relative rankings would surface in the statistical weighting approach.

Key Finding-2. No significant differences of means are found in occupational performance ratings to indicate the potential for occupational bias.

Quantitative analysis of the performance rating data using the statistical significance of the difference of means was conducted on the Navy and Army organization performance rating data sets. The engineer average occupational performance ratings were compared to all other valid occupational average performance ratings. No systemic (e.g. across all performance rating comparisons) anomalies in the difference of means between engineer and other occupations were found. In the Army data set, the medical (0600), information (1000) and supply (2000) occupational series do show significant difference of means preference towards the engineering occupational series. However, three of fifteen comparative tests do not signal significant systemic preferences towards the mission critical occupation (i.e. engineering occupational series).

In sum, no significant variation in the difference of means between engineer and other occupations was found. Based on these quantitative tests, the null hypothesis, that no significant difference in performance ratings is present between mission critical occupations and other occupations in mixed occupations pay pools in the Army and Navy organizations is valid.

Key Finding-3. Archived document analysis found anecdotal evidence of occupational bias in performance ratings.

The qualitative analysis of archived documents showed the presence of anecdotal

evidence that supports the hypothesis that mission critical occupations are favored in performance ratings in mixed pay pools. There is however a lack of empirical evidence in archived documents that support the hypothesis. Three points indicate the potential for validating the hypothesis through the discovery of empirical data in a follow-on research effort.

First, there is evidence that several agencies use performance rating analysis frameworks which is or can be used as a leading indicator for detecting occupational bias. Additionally, the use of statistical analysis of performance ratings is highlighted in Congressional testimony and GAO as a best practice. Second, the topic of anomalies in performance ratings, particularly gender and racial preferences in performance rating has been found in Federal performance ratings (U.S. Government Accountability Office, 2008; U.S. Government Accountability Office, 2005; Rosenberg, 2008). Based on this observation, it is reasonable to assume that occupational bias anomalies may still be found in unanalyzed performance rating data bases in the Federal sector. Finally, GAO reports and annual NSPS performance assessments found anecdotal evidence that occupational bias exists in performance appraisals.

Key Finding-4. Expert interview analysis found anecdotal evidence of occupational bias is present in performance ratings.

The qualitative analysis of expert interviews indicates the presence of anecdotal evidence to support the hypothesis that mission critical occupations are favored in performance ratings in mixed pay pools. Subject matter experts recognized the linkages between conducting performance rating statistical analysis and the detection of bias. Also, subject matter experts in interviews did recognize occupational bias in performance appraisals. However, similar to the findings from analysis of archival documents, there is a lack of empirical evidence associated

with subject matter expert interviews that supports the hypothesis. Three factors contribute to this assessment.

First, interviews indicated that there is substantial awareness in the PfP community of practice on the use of performance ratings analysis frameworks. In particular, statistical testing of performance ratings for anomalies is well recognized. Second, the topic of bias in performance ratings, particularly gender and racial anomalies, is recognized by subject matter experts. Third, occupational bias in performance ratings is acknowledged by subject matter experts. Four of ten interviewees indicated that they had encountered explicit and implicit evidence of occupational bias in performance ratings in their work and research.

Key Finding- 5: Theoretical basis of basis of occupational bias as applied to performance appraisals is re-validated in this research effort.

Demonstrations of real world challenges to implementing NSPS and a bias free performance appraisal environment re-confirm the validity of equity, empirical schema, and social identity theory (SIT) to practice. The literature review, archived documents, and expert interviews illustrated several examples of the application theory to the practice of performance appraisals.

Equity theory (L. S. Adams, 1965), compares individual efforts to those performing similar tasks throughout the organization, and asks if it is fair or equitable. The essence of equity theory is that employees are motivated when perceived outputs, for example pay, is equal to perceived inputs—that is their work effort. Two NSPS examples are cogent. The Congressional Research Service in their examination states that “NSPS was implemented inconsistently and infused with bias. Employees with a lower rating in one office could receive a higher bonus than an employee with a higher rating in another” (Ginsberg, 2008, p. CRS-18). Additionally it was

found in NSPS performance ratings that white employees received higher average performance ratings, salary increase and bonuses, than employees of other races (Losey, 2009).

The use of empirical schema theory when conducting performance appraisals insinuates that managers and supervisors will evaluate employee behavior not on stated objectives or rational decision making, but by comparing and contrasting their mental models of performance. The validation of empirical schema theory is realized through subject matter expert interviews. One subject matter expert stated that he understood that people of the same cloth, like people of the same cloth in performance appraisals (Federal Subject Matter Expert, 2010b). Congressional testimony (*Statement of John Gage, president, American Federation of Government Employees, 2008*) on the topic of NSPS and pay for performance referenced supervisors who have a natural tendency to favor people like themselves.

A central tenet of Social Identity Theory (SIT) is that favoritism, as postulated by Tajfel, is part of organizational behavior and is used by high-status groups in organizations. This tenet was recognized in archival documents and subject matter expert interviews. “The reality in society and in business organizations is that members of high-status groups are far more likely to engage in discrimination and in-group bias against lower status group members than vice versa” (Jost & Elsbach, 2002, p. 186). This social identity theory perspective is borne out by a PfP subject matter expert interview that from his perspective there is a difference in ratings in occupational classes---engineers and scientists versus administrative (Federal Subject Matter Expert, 2010d).

Conclusions to be drawn based on the findings.

The methodology used to address this dissertation’s hypothesis used a quantitative analysis of pay pool rating data, and qualitative analysis of archived documents and expert

interviews. The first four findings presented in the previous paragraphs in this chapter are used to determine the validity of hypothesis that the engineering occupational series achieve more favorable ratings than other occupational series in DOD engineering organizations in mixed pay pools. FY 2008 engineer occupational series (mission critical occupations) performance ratings in DOD engineering agencies did not show statistically significant differences when compared. Difference of means and significant of difference of means quantitative analysis did not reveal a notable number of anomalies in performance ratings. Occupational bias as indicated by performance rating anomalies in performance ratings is not detected resulting in the null hypothesis. The hypothesis that more favorable performance ratings occur in mission critical occupations in mixed occupational pay pools in Federal agencies is found to be invalid. Anecdotal evidence in archived documents and expert interviews did recognize occupational bias in mission critical occupations as a challenge that has not been documented empirically.

Alternative Explanations of the in terms of what was learned.

The hypothesis that more favorable performance ratings occur in mission critical occupations in mixed occupational pay pools in Federal agencies, which has been found to be invalid, offers two alternative explanations.

First, the complex multi-step performance appraisal process used in NSPS helped to provide checks and balances that prevent occupational bias, in this case more favorable ratings for engineers in engineering organizations. Three levels of review, supervisor, higher-level reviewer, and pay pool panels used in NSPS performance appraisal process serve as abating mechanisms to the occurrence of occupational bias. NSPS also uses a structured performance management system that emphasizes adherence to evaluating how well the employee did in accomplishing specific performance objectives.

Second, occupational bias exists, but has not been proven empirically with the correct Federal agency data set. Various forms of bias in performance appraisals exist as documented in the literature. Gender and racial bias in performance ratings in pay for performance systems has been seen in the recent case of the Security and Exchange Commission. Anecdotal evidence from government reports and interviews suggests that occupational bias in performance appraisals exists. The selection of a relatively immature Federal PfP system, such as NSPS, as the data set for examination of the hypothesis may be less than optimal for statistical validation. Alternatively, the selection of a more mature existing Federal PfP performance rating data set, such as that used in the Office of Currency Control, Department of Treasury may prove to be a more beneficial data set to achieve statistical validation of occupational bias in performance appraisals.

Impact of the study in terms of what was learned.

The literature review in this dissertation finds that there is a lack of information, data, and research regarding occupational bias in performance appraisal and rating. In particular, there is a lack of empirical evidence that addresses occupational bias in performance appraisal and ratings in the Federal sector, especially the DOD. Current literature on occupational bias in performance appraisals can be characterized as sparse. Journals and books consist of anecdotal data and dated research; government reports and data are limited. A paucity of current empirical evidence, particularly in Federal performance management systems, presents an opportunity to contribute to the field of personnel and performance management. This dissertation can serve as a baseline from which to further develop the hypothesis.

One reason there is a lack of information, data, and research on occupational bias in performance appraisal and ratings is that it is not a protected class (i.e. race, age, disability, et.)

under the wide range of Equal Opportunity Public Laws in the United States. Data is collected in support of these laws. One example is the collection of performance appraisal data for employee classes (i.e. race, color, national origin, etc) to protect employees under Title VII of the Civil Rights Act of 1964. One benefit of this dissertation's research effort is that it considers, consolidates and advances the knowledge base on bias in performance appraisal and ratings by a non-protected classification of employee ratings: occupations.

Limitations of this study.

There are four limitations associated with this effort. First, the level of analysis is at the organizational level. This impacts the fidelity of the quantitative analysis. Second, the DOD engineering agency data set does not link to demographic data bases. Thus, analysis of various influences of gender or racial characteristics on performance ratings is not feasible in this research effort. The third limitation of this dissertation is in the scope and time span of the data set examined, which is Fiscal Year 2008. Examination of rating data in multiple agencies over a number of a fiscal year pay cycles is not conducted. Finally, the primary focus area in this dissertation is the detection of occupational bias in performance ratings. Some discussion of the why occupational bias occurs is included in this dissertation; it is not the primary focus.

Implications for professional practice or decision making

Bias in performance appraisal can be considered as a wicked problem that is extremely difficult to solve. Occupational bias has not been conclusively shown to be empirically valid. However anecdotal evidence of occupational bias adds to the list of biases found in performance appraisals. The perception that occupational bias exists, whether empirically proven or not, adds to the complexity of establishing and maintaining a fair and equitable performance management

system. A recent case of bias in performance appraisals illustrates how employees can perceive that unbiased performance appraisals are an effort in futility.

Illustrations from governmental reports and interviews indicate that gender and racial bias in performance appraisals continue today. “Women were paid 2 to 4 percent less than men at five of the six laboratories, while minorities were paid about 2 percent less than whites at one laboratory” (U.S. Government Accountability Office, 2005, p. 6). Similarly, in another agency, “the arbitrator found that African-American employees above grade 8 and older employees received significantly fewer pay increases than would be expected given their representation in the pool of eligible employees” (*Statement of Colleen Kelly, president, National Treasury Employees Union*, 2008, p. 4) Given these examples of recent bias in performance appraisals, it can be reasonably assumed, as illustrated in anecdotal evidence in this study that occupational bias occurs. Moreover, the reluctance of some Federal subject matter experts to be interviewed for this study is indicative of the sensitivity of the topic. Many were reluctant to discuss or reveal the extent of published reports of findings of their annual analysis on anomalies in performance appraisals.

Implications for scholarly understanding of the field.

Performance Management and PfP in the Federal government have had a difficult path to success. While some Federal pay for performance demonstration projects show progress, NSPS is the fourth major Federal pay for performance system failure in the past thirty years. These are costly ventures to the taxpayer. Numerous factors contribute to the downfall of Federal PfP systems, one of which can be the absence of fair and equitable performance appraisal processes. Occupational bias, while not empirically proven in this research effort, would be another

challenge to overcome in the implementation of a successful pay for performance system in the Federal government. A scholar captures the implications succinctly.

Performance appraisal has remained an unsolved and perhaps unsolvable problem in human resource management...to be sure, many gimmicky innovations have been devised for reviewing managerial and employee performance, yet they have proven to be short lived. The search for the Holy Grail in performance appraisal goes on (Schay, 1993, p. 649).

Recommendations for further research.

Bias is present in the performance appraisal process. Occupational bias is found to present—as seen in anecdotal comments. The challenge of determining occupational bias empirically still exists. Anecdotal evidence serves as an indicator of potential success. Federal agencies that have demonstrated competence in operating PfP systems offer opportunities to further the occupational bias research—and to validate the hypothesis that mission critical occupations are favored in mixed pay pools through empirical evidence.

For example several Federal agencies, specifically the Farm Credit Agency of the Department of Agriculture, and the Office of Comptroller of the Currency in the Department of Treasury use PfP systems and rating analysis frameworks. These frameworks are used to conduct post-rating analysis of distribution of ratings. The Office of the Comptroller of the Currency indicated that “we don’t look at it just for race, but we look at it by grade, by occupation, and by location” (U.S. Merit Systems Protection Board, 2006, p. 50). However, attempts to obtain information for this research effort were met with resistance. Further determined efforts may prove useful in obtaining empirical evidence of occupational bias in performance appraisals.

Recommendations for changing the research methodology.

The use of a quantitative-qualitative methodology to determine that more favorable performance ratings occur in mission critical occupations in mixed occupational pay pools in DOD agencies has merit for a research project. The probability of meeting with success may be improved by changing the agencies under examination as well using a more mature PfP system.

Availability, type and sensitivity of performance rating data affect the collection of performance rating data. Specifically, the use of NSPS data was problematic because the system was fairly new with incomplete data sets across multiple years. NSPS was implemented in spirals or phases that also limited the number of DOD employees participating in the system. NSPS maintained no direct linkage to demographic data and individual performance rating data. Additionally, Privacy Act limitations on obtaining individual data at the pay pool level may invalidate the use of NSPS as an acceptable data source. A recommendation to overcome the weaknesses of NSPS as a data set is to focus on a Federal agency that has a mature pay for performance system, using post-rating analysis processes.

Bibliography

- Academic Subject Matter Expert (2010, May 2010). [Interview #6].
- Adams, J. S. (1963). Towards and understanding of inequity. *Journal of Abnormal and Social Psychology, 67*(5), 14.
- Adams, L. S. (1965). Injustice in social exchanges. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Orlando, FL: Academic Press.
- Ashford, B., & F. Mael. (1989). Social identity theory and the organization. *Academy of Management Review, 14*(1), 19.
- Barnes-Farrell, J. L. (2001). Performance appraisal: Person perception processes and challenges. In M. London (Ed.), *How people evaluate others in organizations* (1st ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Beer, M., & Ruh, R. A. (1990). Employee growth through performance management. In V. H. Vroom (Ed.), *Manage people, not personnel*. Boston: Harvard Business School Publishing.
- Behn, R. B. (2006). Performance leadership: 11 better practices that can ratchet up performance *Managing for Performance and Results Series* (pp. 30). Washington D.C.: IBM Center for the Business of Government.
- Benabou, R., & Jean, T. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*, 31.
- Berlet, K. R., & Cravens, D. M. (1991). *Performance pay as a competitive weapon*. New York: John Wiley & Sons. Inc.
- Berman, E. M. (2006). *Human resource management in public service : paradoxes, processes, and problems* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, Mass.: Kent Publishing Company.
- Bigoness, W. J. (1976). Effect of applicant's sex, race, and performance on employer's performance ratings: Some additional findings. *Journal of Applied Psychology, 61*(1), 4.
- Bjerke, D. G., Cleveland, J. N., Morrison, R. F., & Wilson, W. C. (1987). *Officer fitness report evaluation study*. (NPRDC TR 88-4). San Diego, CA: Navy Personnel Research and Development Center.
- Booth, W. C., Colomb, G. G., & Williams, J. M. (1995). *The craft of research*. Chicago: The University of Chicago Press.

- Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18(2), 30.
- Brewer, G. A. (2005). In the eye of the storm: Frontline supervisors and federal agency performance. *Journal of Public Administration Research & Theory*, 15, 22.
- Brewer, M. (1996). In-group favoritism: The subtle side of intergroup discrimination. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics*. New York: Russell Sage Foundation.
- Brody, R. G., & Frank, K. E. (2002). Influence of prior involvement on employees' ratings. [Article]. *Psychological Reports*, 90(1), 139-149.
- Brook, D. (2000). Merit and the civil service reform act. In J. Pfeffer & D. A. Brook (Eds.), *The future of merit*. Baltimore: The Johns Hopkins University Press.
- Brown, M., & Heywood, J. S. (Eds.). (2002). *Paying for performance: An international comparison*. Armonk, NY: M. E. Sharpe.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati, Ohio: South-Western Publishing Co.
- Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 6, 48.
- Cayer, J. (2004). Techniques of personnel administration. In J. Cayer (Ed.), *Public personnel administration*. Belmont, CA: Wadsworth/Thomson Learning.
- Cayer, J. N. (2004). Keeping employees motivated and productive In J. Cayer (Ed.), *Public personnel administration* (4th ed.). Belmont, CA: Thomson Publishers.
- Cleveland, J. N., & Murphy, K. R. (1989). Multiple Uses of Performance Appraisal: Prevalence and Correlates. *Journal of Applied Psychology*, 74(1), 6.
- Cleveland, J. N., & Murphy, K. R. (1991). *Performance appraisal: An organizational perspective*. Boston, MA: Allyn and Bacon.
- Cleveland, J. N., & Murphy, K. R. (1995). *Understanding performance appraisal*. Thousand Oaks: Sage Publications.
- Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 32.
- Condrey, S. E. (Ed.). (2005). *Handbook of human resource management in government* (2nd ed.). San Francisco, CA: Jossey-Bass.

- Condrey, S. E., & Brudney, J. L. (1992). Performance-based managerial pay in the federal government: Does agency matter. *Journal of Public Administration Research & Theory*, 2, 19.
- Congressional Budget Office. (2008). A review of the Department of Defense's National Security Personnel System A *CBO Paper*. Washington D.C.: Congress of the United States.
- Consultant Subject Matter Expert (2010, May, 2010). [Interview #7].
- Cook, M. (1995). Performance appraisal and true performance. *Journal of Managerial Psychology*, 10(7), 5.
- Crawford, J., de Vogue, A., & Powell, D. (2009). Court Rules for White and Hispanic Firefighters, Reversing Sotomayor Decision. *ABC News* Retrieved July 17, 2009, 2009, from <http://abcnews.go.com/Politics/SCOTUS/Story?id=7782103&page=1>
- Daley, D. M. (1992). *Performance appraisal in the public sector: Techniques and applications*. Westport, CT: Quorum Books.
- Daley, D. M. (1999). Public sector supervisory performance appraisal: core functions, effectiveness characteristics, and the identification of superstar supervisors. *Review of Public Personnel Administration*, 19(65).
- Daley, D. M. (2000). Performance appraisal techniques and applications. In R. Golembiewski (Ed.), *Handbook of organizational consultation* (2nd ed.): CRC Publishing.
- Daley, D. M. (2002). Compensation. In D. M. Daley (Ed.), *Strategic human resource management*. Upper Saddle River, NJ: Prentice-Hall.
- Daley, D. M. (2003). The trials and tribulations of performance appraisal. In S. W. Hays & R. C. Kearney (Eds.), *Public personnel administration*. Upper Saddle River, NJ: Prentice Hall.
- Daley, D. M. (2005). Designing effective performance appraisal systems. In S. E. Condrey (Ed.), *Handbook of human resource management in government* (2nd ed.). San Francisco: Jossey-Bass.
- Deci, E. L., & Ryan, R. M. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 13.
- Deci, E. L., & Ryan, R. M. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well being. *American Psychologist*, 55(1), 10.
- DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *The Academy of Management Review*, 3(3), 11.

- Demmke, C. (2007). *Performance assessment in the public services of the EU member states: Procedure for performance appraisal, for employee interviews and target agreements*. Maastricht: European Institute of Public Administration.
- Denhardt, K. G. (1988). *The ethics of public service*. New York: Greenwood Press.
- Dewberry, C. (2001). Performance disparities between whites and ethnic minorities: Real differences or assessment bias? *Journal of Occupational and Organizational Psychology*, 74, 12.
- DiTomaso, N., & Smith, D. R. (2007). Effects of structural position on allocation and evaluation decisions for scientists and engineers in industrial R & D. *Administrative Science Quarterly*, 52, 31.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 17.
- Elvira, M., & Town, R. (2001). The effects of race and worker productivity on performance evaluations. *Industrial Relations*, 40(4).
- England, G. (2004). *Requirements document for National Security Personnel System*. Washington, D.C.: U.S. Department of Defense.
- Eremin, D., Wolf, J., & Woodard, C. (2010). Systemic Bias in Federal Performance Evaluations: Does Hierarchy Trump a Performance Management Process? *Public Performance and Management Review*, Forthcoming.
- Federal Subject Matter Expert (2010a, May, 2010). [Interview #4].
- Federal Subject Matter Expert (2010b, May 2010). [Interview #8].
- Federal Subject Matter Expert (2010c, May, 2010). [Interview #9].
- Federal Subject Matter Expert (2010d, May 2010). [Interview #10].
- Ferris, G. R., Yates, V. L., Gilmore, D. C., & Rowland, K. M. (1985). The influence of subordinate age on performance ratings and causal attributions. *Personnel Psychology*, 38, 12.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw Hill.
- Flannery, T. P., Hofrichter, D. A., & Platten, P. E. (1996). *People, performance and pay*. New York, New York: The Free Press.
- Frankfort-Nachmias, C., & Leon-Guerrero, A. (2002). *Social statistics for a diverse society* (3rd ed.). Thousand Oaks, Ca: Sage Publications.
- Friedson, E. (2001). *Professionalism: The Third Logic*. Chicago: University of Chicago Press.

- Gabris, G. T., & Ihrke, D. M. (2000). Improving employee acceptance toward performance appraisal and merit pay systems: the role of leadership credibility. *Review of Public Personnel Administration, 20*(1), 12.
- Gilliland, S. W., & Langdon, J. C. (1998). Creating performance management systems that promote perceptions of fairness. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice*. San Francisco: Jossey-Bass.
- Ginsberg, W. (2008). *Pay-for-performance: The National Security Personnel System*. Washington, D.C.: Congressional Research Service.
- Ginsberg, W. (2009). *Pay-for-performance: Lessons from the National Security Personnel System*. (CRS RL34673). Washington, D.C.: Congressional Research Service.
- Gordon, D. F. (1974). A neo-classical theory of Keynesian unemployment. *Economic Inquiry, 12*(4), 28.
- Grand, C., Szulkin, R., & Tahlin, M. (1994). Organizational structures and job rewards in Sweden. *Acta Sociologica, 37*, 20.
- Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology, 71*, 2.
- Greenhaus, J., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33*(1), 22.
- Guralnik, D. B. (1979). Webster's New World Dictionary of the American Language *Webster's New World Dictionary of the American Language* (Second College Edition ed.). New York: William Collins Publishers, Inc.
- Halpert, J. A., Wilson, M. L., & L., H. J. (1993). Pregnancy as a source of bias in performance appraisals. *Journal of Organizational Behavior, 14*, 14.
- Hays, S. W., & Kearney, R. C. (1995). *Public personnel administration* (3rd ed.). Englewood Cliffs, New Jersey: Prentice Hall.
- Hays, S. W., & Kearney, R. C. (2003). *Public personnel administration: problems and prospects* (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.
- Hays, S. W., & Kearney, R. C. (Eds.). (1990). *Public personnel administration: Problems and prospects* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Heneman, R. L. (1992). *Merit pay: Linking pay increases to performance ratings*. New York, New York: Addison-Wesley Publishing Company, Inc.
- Hickman, G. R., & Lee, D. S. (2001). *Managing human resources in the public sector: a shared responsibility*. New York: Harcourt College Publishers.

- Higgins, T. E. (1983). Effort after meaning in personnel evaluation: The similarity error and other possible sources of bias. In F. J. Landy (Ed.), *Performance measurement and theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hirsch, C. E., & Kornich, S. (2008). The context of discrimination: Workplace conditions, institutional environments, and sex and race discrimination charges. *American Journal of Sociology*, 113(5), 38.
- Hogg, M. A., & Terry, D. J. (2001). *Social Identity Processes in Organizational Contexts*. Ann Arbor: Taylor and Francis.
- Hyde, A. C. (2005). Pay for performance: Has emerged as the new model for federal human resources pay practice and how executives and managers will be challenged to solve its complexities. *The Public Manager*, 34.
- Ingraham, P. W. (1997). *A Laggard's tale: Civil service and administrative reform in the United States*. Paper presented at the Comparative Civil Service Reform, Indiana University, Bloomington, IN.
- Ingraham, P. W., Joyce, P. G., & Donahue, A. K. (2003). *Government performance: Why management matters*. Baltimore, MD: The John Hopkins University Press.
- Ingraham, P. W., & Romzek, B. S. (1994). *New paradigms for government : issues for the changing public service* (1st ed.). San Francisco: Jossey-Bass.
- Jackson, S. E., Schuler, R. S., & Rivero, J. C. (1989). Organizational characteristics as predictors of personnel practices. *Personnel Psychology*, 43, 39.
- Jost, J. T., Banaji, M. R., & A., N. B. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 31.
- Jost, J. T., & Elsbach, K. D. (2002). How status and power differences erode personal and social identities at work. In M. A. Hogg & D. J. Terry (Eds.), *Social Identity Processes in Organizational Contexts*. Philadelphia, PA: Psychology Press.
- Kellough, J. E., & Nigro, L. G. (2002). Pay for performance in Georgia state government: Employee perspectives on GeorgiaGain after 5 years. *Review of Public Personnel Administration*, 22, 146 - 166.
- Kellough, J. E., & Selden, S. C. (1997). Pay-for-performance systems in state government. *Review of Public Personnel Administration*, 17(1), 17.
- Klingner, D. E., & Nalbandian, J. (1998). *Public personnel management: Contexts and strategies*. Upper Saddle River, New Jersey

Prentice-Hall, Inc.

- Klores, M. S. (1966). Rater bias in forced-distribution performance ratings. *Personnel Psychology*, 19(4), 10.
- Kopelman, R. E. (1976). Organizational control system responsiveness, expectancy theory constructs, and work motivation: some interrelations and causal connections. *Personnel Psychology*, 29, 15.
- Kramer, R. M., & Neales, M. A. (Eds.). (1998). *Power and influence in organizations* (1st ed.). Thousand Oaks: Sage Publications.
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance*. London: Academic Press, Inc.
- Latham, G., & Wexley, K. N. (1994). *Increasing productivity through performance appraisal* (2nd ed.). New York, New York: Addison-Wesley Publishing Company.
- Lawler, E. E. (1990). *Strategic pay: Aligning organizational strategies and pay systems*. San Francisco, CA: Jossey-Bass Publishers.
- Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, 73, 18.
- Locke, E. A., & Latham, G. P. (2004). What should we do about motivation theory? Six recommendations for the 21st century? . *Academy of Management*, 29(3), 25.
- Longenecker, C. O., & Goff, Stephen J. (1992). Performance appraisal effectiveness: A matter of perspective. *SAM Advanced Journal*, 14, 6.
- Longenecker, C. O., & Sims, Henry P. (1994). Behing the kask: The Politics of Employee Appraisal. *SThe Academy of Management Executive*, 1,3.
- Losey, S. (2009). Is DoD's new pay system fair? *FederalTimes.com*, 2009(June 29, 2009). Retrieved from <http://www.federaltimes.com/index.php?S=3666513>
- Lovo, J. (2009). *Mission Critical Occupations in the Corps of Engineers Email*. Washington, D.C.: DOD.
- Magnusson, C. (2008). Gender, occupational prestige, and wages: A test of devaluation theory. *European Sociological Review*.
- Meyer, H. H. (1975). The Pay for Performance Dilemma. *Organizational Dynamics*, 3(13), 11.
- Milkovich, G., & Newman, J. M. (2005). *Compensation* (8th ed.). Boston: McGraw-Hill.
- Milkovich, G., & Wigdor, A. (1991). Pay for performance: Evaluating performance appraisal and merit pay (pp. 210). Washington, DC: National Resource Council.

- Miller, D. T., & Ross, M. (1975). Self-serving biases in attribution of causality: Fact or fiction. *Psychological Bulletin*, 82(2), 12.
- Moers, F. (2004). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1), 67-80.
- Mohrman, A. M., Resnick-West, S. M., & Lawler, E. E. (1989). *Designing performance appraisal systems: Aligning appraisals and organizational realities*. San Francisco, CA: Jossey-Bass Publishers.
- Mosher, F. C. (1968). *Democracy and the public service*. New York, New York: Oxford University Press.
- Nigro, L. G. (2007). Human Resources and Organizational Performance. In L. G. Nigro (Ed.), *The New Public Personnel Administration* (6th ed.). Itasca, Ill: F.E. Peacock Publishers.
- Nigro, L. G., & Nigro, F. A. (1994). *The new public personnel administration* (4th ed.). Itasco: F.E. Peacock Publishers, Inc.
- Nigro, L. G., Nigro, F. A., & Edward, K. J. (2007). Performance appraisal and pay for performance. In L. G. Nigro, F. A. Nigro & K. J. Edward (Eds.), *The new public personnel administration* (6th ed.). Belmont, CA: Thomson Publishing.
- NSPS. (2005). Manager supervisor: Focus on performance (pp. 4): NSPS.
- Organization for Economic Co-operation and Development. (2005). *Performance-related pay policies for government employees*. Paris, France: Organization for Economic Co-operation and Development.
- Pazy, A., & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior*, 22, 13.
- Pearce, J. L., & Perry, J. L. (1983). Federal merit pay: A longitudinal analysis. *Public Administration Review*.
- Perry, J. L. (1986). Merit Pay in the Public Sector: The Case for a Failure of Theory. *Review of Public Personnel Administration*, 7, 1257.
- Perry, J. L. (2003). Compensation, merit pay and motivation. In S. W. Hays & R. C. Kearney (Eds.), *Public personnel administration* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Perry, J. L., Enghers, Trent, and Jun So Yun. (2009). Back to the Future? Performance-Related Pay, Empirical Research, and the Perils of Persistence. *Public Administration Review*, 69(1), 32.
- Perry, J. L., & Paarlberg, L. E. (2007). Values management: Aligning employee values and organizational goals. *The American Review of Public Administration*, 37.

- Perry, J. L., & Petrakis, B. A. (1988). Can pay for performance succeed in government? *Public Personnel Management*, 17(4).
- Pfeffer, J. (1981). *Power in organizations*. Cambridge, Mass: Ballinger Publishing Company.
- Pfeffer, J. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, 38, 25.
- Pollock, P. H. (2005). *The essential of political analysis* (1st ed.). Washington D.C.: CQ Press.
- Prendergast, C., & Topel, R. (1996). Favoritism in organizations. *The Journal of Political Economy*, 104(5), 958-978.
- Pulakos, E. D., Oppler, S. H., White, L. A., & Borman, W. C. (1989). Examination of Race and Sex Effects On Performance Ratings. *Journal of Applied Psychology*, 74(5), 770.
- Pulakos, E. D., & Wexley, K. N. (1983). The relationship among perceptual similarity, sex, and performance ratings in manager-subordinate dyads. *Academy of Management Journal*, 26(1), 10.
- Pynes, J. E. (1997). *Human resources for public and non-profit organizations*. San Francisco: Jossey-Bass.
- Rabin, J. (Ed.). (1995). *Handbook of public personnel administration*. New York, NY: Marcel Dekker.
- Rainey, H. G., & Kellough, E. J. (2000). Civil Service reform and incentives in the public service. In J. P. Pfiffner & D. A. Brook (Eds.), *The future of merit : twenty years after the Civil Service Reform Act* (pp. 127-145). Washington, D.C.: Woodrow Wilson Center Press.
- Rainey, H. G., & Thompson, J. R. (2003). Modernizing human resource management in the federal government: The IRS model *Human Capital Series*. Washington D.C.: IBM Endowment for The Business of Government.
- Riccucci, N. (Ed.). (2006). *Public personnel management: Current concerns, future challenges* (4th ed.). New York, NY: Longman Publishing.
- Riccucci, N., & Ban, C. (1991). *Public personnel management: Current concerns, future challenges*. White Plains, N.Y.: Longman Publishers.
- Riccucci, N., & Ban, C. (2006). Employee performance appraisal in the public sector: Uses and limitations. In N. Riccucci & C. Ban (Eds.), *Public personnel management*. New York: Longman Publishers.
- Riccucci, N., & Ban, C. (Eds.). (2002). *Public personnel management: Current concerns, future challenges* (3rd ed.). New York, NY: Longman.

- Riccucci, N., & Naff, K. (2008). Performance appraisal. In J. M. Shafritz (Ed.), *Personnel management in government: Politics and process* (6th ed.). New York: CRC Press.
- Risher, H. (2004). Pay for performance: A guide for federal managers *Human Capital Management Series: IBM Center for The Business of Government*.
- Risher, H., & Fay, C. H. (2007). Managing for better performance: Enhancing federal performance management practices *Human Capital Management Series: IBM*
- Rizzotti, D. (2007). *Incentives, performance measurement systems, and managerial effort allocation*. University of Catania. Catania, Italy.
- Robbins, T. L., & DeNisi, A. S. (1993). Moderators of sex bias in the performance appraisal process: A cognitive analysis. *Journal of Management* 19(1), 10.
- Roberts, G. E. (1998). Perspectives on enduring and emerging issues in performance appraisal. *Public Personnel Management*, 27(3).
- Rohr, J. A. (1998). *Public service, ethics, and constitutional practice*. Lawrence, Kansas: University Press of Kansas.
- Rosen, B. (1975). *The merit system in the United States civil service*. Washington D.C.: U.S. Government Printing Office.
- Rosen, S. (1985). Implicit contracts: A survey. *Journal of Economic Literature*, 23, 31.
- Rosenberg, A. (2008). SEC, Union settle pay for performance case. *Government Executive*, (October 7, 2008). Retrieved from <http://www.govexec.com/dailyfed/1008/100708ar2.htm>
- Rosenbloom, D. H., & Carroll, J. D. (1995). Public personnel administration and law. In J. Rabin, T. Vocino, W. B. Hildreth & G. J. Miller (Eds.), *Handbook of public personnel administration* (pp. 32). New York: Marcel Dekker, Inc.
- Schay, B. (1993). In search of the holy grail: Lessons in performance management. *Public Personnel Management*, 22(4), 19.
- Schein, E. H. (1992). *Organizational culture and leadership* (2nd ed.). San Francisco: Jossey-Bass.
- Schminke, M., Cropanzano, R., & Rupp, D. E. (2002). Organization structure and fairness perceptions: The moderating effects of organizational level. *Organizational Behavior and Human Decision Processes*, 89(1), 24.
- Shafritz, J. M. (Ed.). (1992). *Personnel management in government: Politics and process*. New York, NY: Marcel Dekker.

- Shafritz, J. M. (Ed.). (1998). *International encyclopedia of public policy and administration* (Vol. 1-4). Boulder, CO: Westview Press.
- Shafritz, J. M. (Ed.). (2001). *Personnel management in government* (5th ed.). New York, NY: Marcel Dekker, Inc.
- Siegel, G. B. (1994). Three federal demonstration projects: Using monetary performance awards. *Public Personnel Management*, 23(1).
- Smith, D. R., DiTomaso, N., & Farris, G. F. (2001). Favoritism, bias, and error in performance ratings of scientists and engineers: The effects of power, status, and numbers. *Sex Roles*, 45(5/6).
- Smither, J. W. (1998). *Performance appraisal: State of the art in practice*. San Francisco: Jossey-Bass Publishers.
- Spitler, R. A. (1999). *Pay for performance: Can it work in the Department of Defense*. Master of Strategic Studies, U.S. Army War College, Carlisle, PA.
- SRA International Inc. (2009). National Security Personnel System (NSPS) 2008 Evaluation Report. Arlington, VA: Program Executive Office, NSPS.
- Statement of Charles Tiefer, professor, University of Baltimore School of Law, U.S. House of Representatives* (2008).
- Statement of Colleen Kelly, president, National Treasury Employees Union, U.S. House of Representatives* (2008).
- Statement of Dr. Stanley E. Ridley, president and CEO of Ridley & Associates, Inc, U.S. House of Representatives* (2008).
- Statement of John Gage, president, American Federation of Government Employees, U.S. House of Representatives* (2008).
- Stauffer, J. M., & Buckley, M. R. (2005). The existence and nature of racial bias in supervisory ratings. *Journal of Applied Psychology*, 90(3), 5.
- Streib, G., & Nigro, L. (1995). Pay-for-performance in local governments: The views of personnel directors. *International Journal of Public Administration*, 18(2), 19.
- U.S. Army Corps of Engineers. (2009). USACE pay pool business rules and performance management policy. In U. S. Department of Army (Ed.), (pp. 43). Washington, D.C.: U.S. Army Corps of Engineers.
- U.S. Department of Army. (2009). Department of Army NSPS policy. In U. S. Department of Army (Ed.). Washington, D.C.: Department of Army.
- National Security Personnel System, Proposed Rule, 5 CFR, chap. XCIX and part 9901 (2005a).

- U.S. Department of Defense. (2005b). NSPS HR elements primer Retrieved Aug 25, 2009, from <http://www.eustis.army.mil/cpac/NSPS/WebHRPrimer1205.pdf>
- U.S. Department of Defense (2006). NSPS handbook Retrieved June 19, 2009, from http://www.wiesbaden.army.mil/sites/installation/img/NSPS_Handbook.pdf
- U.S. Department of Defense. (2007). HR elements for managers, supervisors, and employees. In U. S. Department of Defense (Ed.), (Spiral 1, Version 3 ed.). Washington, D.C.: U.S. Department of Defense.
- U.S. Department of Defense (2008). DOD Civilian Personnel Manual 1400.25-M, Chapter 1910-50. DOD 1400.25-M. Washington, D.C., U.S. Department of Defense. SC 1910. Subchapter 1940: Performance Management: 25.U.S. Department of Defense. (2009). Filling NSPS vacancies during the NSPS transition period. In U. S. Department of Defense (Ed.), (pp. 5). Washington D.C.: National Security Personnel System Program Executive Office.
- U.S. Department of Defense Business Board. (2009). Review of the national security business board. In U. S. Department of Defense (Ed.), (Vol. Report FY09-06). Washington, D.C.: U.S. Department of Defense.
- U.S. Government Accountability Office. (2005). Information on personnel actions, employee concerns, and oversight at six DOE laboratories. Washington D.C.: U.S. Government Accountability Office.
- U.S. Government Accountability Office. (2007a). Agencies have implemented key performance management practices, but opportunities for improvement exist. In U. S. General Accounting Office (Ed.). Washington, D.C.: U.S. Government Accountability Office.
- U.S. Government Accountability Office. (2007b). Key lessons learned to date for strengthening capacity to lead and implement human capital reforms. In U. S. General Accounting Office (Ed.). Washington, D.C.
- U.S. Government Accountability Office. (2008). DoD Needs to Improve Implementation of and Address Employee Concerns about its National Security Personnel System. Washington, D.C.: U.S. Government Accountability Office.
- U.S. Government Accountability Office. (2009). Monitoring of safeguard and addressing employee perceptions are key to implementing a civilian performance management system in DOD *Human Capital*. Washington D.C.: U.S. Government Accountability Office.
- U.S. Government Accounting Office. (2007). DOD needs better internal controls and visibility over costs for implementing its national security personnel system (pp. 25). Washington, D.C.: Government Accounting Office.
- U.S. Merit Systems Protection Board (2006). The practice of merit: A symposium. Washington, D.C., U.S. Merit Systems Protection Board.U.S. Merit Systems Protections Board.

- (2002). Making the public service work: Recommendations for change. Washington D.C.: The Office of Policy and Evaluation.
- U.S. Office of Personnel Management. (2009). Human capital assessment and accountability framework (HCAAF) resource center glossary Retrieved May 19, 2009, from http://www.opm.gov/hcaaf_resource_center/glossary.asp
- U.S. Office of Personnel Management. (2001). A handbook for measuring employee performance: Aligning employee performance plans with organizational goals. Washington, D.C.: U.S. Office of Personnel Management.
- Viswesvaran, C., & Schmidt, F. L. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling subjective and error influences. *Journal of Applied Psychology, 90*(1), 23.
- Vroom, V. (1994). *Work and motivation*. New York: John Wiley and Sons, Inc.
- Walker, R. W. (2008). The most divisive issue: Pay. *Federal Computer Week* Retrieved December 26, 2007, from http://www.fcw.com/print/13_31/management/103611-1.html
- Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35*(3), 521 - 551.
- Woodard, C. (2000). *Merit in principle, merit in practice: An investigation into merit-based human resources management through the lens of title 5-exempt federal organizations*. Doctor of Philosophy, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Appendix A: Archived Documents and Interview Summaries

SUMMARY OF OCCUPATIONAL BIAS RELATED ARCHIVED REPORTS							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
1	GAO-10-102, Oct, 2009	Analyzed the rating results for more by demographic groups, such as race, gender, ethnicity, age, veteran status, and target disability, but limited its analysis to the aggregate data from its pay pools	2	NSPS may inadvertently favor employees who work closely or are in direct contact with member of the pay pool panel because those individuals have direct knowledge of their performance.	2	They do not believe that some military supervisors value the work of employees who perform certain job functions, such as providing child care on an installation	3
		It was found DODs postdecisional analysis following the 2007 and 2008 NSPS performance management cycle lacked consistency and did not generally include demographic information	2	Officials at six locations told us that they believe that there was a forced distribution of the rating or manipulation of the ratings in order to achieve a predetermined quota	2	It was difficult for employees in certain positions to receive a rating of "5" because of the nature of their work or the perceived value their management placed on those positions	2
2	GAO-10-134, Dec 2009	The department does not have a written policy outlining how it will analyze final rating by demographic categories/groups how components will ID potential barriers	2	Several of our discussion group participants expressed concerns about potential for a "forced distribution" of rating, which could erode meaningful distinctions in employee performance	2		1
3	GAO-09-669R, May 2009	DOD does not require a third party to analyze rating results for anomalies prior to finalizing employee ratings	2	Adequate safeguards are needed, including reasonable transparency and appropriate accountability mechanisms, to ensure the fair, effective, and non-discriminatory implementation	2		1
4	GAO-09-464T, Apr 2009	DOD lacks a process to determine whether rating results are non-discriminatory before they are finalized; it does not require a third party to analyze the pre-decisional rating results for anomalies	2	DOD may be unable to fully determine whether ratings are fair and equitable because the department does not have a written policy outlining how it will analyze final ratings	2		1
		This template provides a sample aggregate workforce report for employees, which contains demographic-based reporting categories, including gender, race, etc.	2	How the components will identify potential barriers, if they exist; or what DOD will do, if anything, with the results of that analysis is unknown because of the lack of written policy	2		
5	GAO-07-901T, May 2007	GAO uses a two part approach; we review statistical data on performance ratings by demographic data within each unit, and where appropriate, we conduct assessment of individual ratings	3		1		1
6	GAO-06-0582R, Mar 2006		1	DOD must assure that certain predecisional internal safeguards exist to achieve the consistency, equity, nondiscrimination, and nonpoliticization of the performance management	2		1
7	GAO-05-190: Feb, 2005	GAO found some statistically significant differences in salaries, merit pay increases patterns for professional women and minorities when compared with men and whites	3	Women were paid 2 to 4 percent less than men at five of the six laboratories, while minorities were paid about 2 percent less than Whites at one laboratory	3		1
8	CRS RL 34673, Sep 2008	The law required that the system be fair, credible, and transparent and provide effective safeguards to ensure that the management of the system is fair and equitable and based on employee performance	2	NSPS was implemented inconsistently and infused with bias. Employees with a lower rating in one office could receive a higher bonus than an employee with a higher rating in another.	3	Employees with a lower rating in one office could receive a higher pay bonus than an employee with a higher rating in another	2
9	GAO 03-965 R, Jul 2003	Implementation of effective transparency and accountability measures, including appropriate independent reasonable review, internal assessments, and employee surveys	2	System must include a validated performance appraisal system; adequate safeguards, such as appropriate accountability mechanisms, to ensure the fair, & non-discriminatory implementation	2		1
10	US Merit System Protection Board: The Practice of Merit: A Symposium Report	Transparency and accountability measures, including appropriate independent reviews; reasonableness reviews, internal grievance procedures, internal assessments	2		1	We don't look at it just for race, but we look at it by grade, by occupation, and by location	3
	QUALITATIVE RATING AVG.		2.08		2.00		1.55
Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography; Textual description in column represents a paraphrased summary of a key point from the document ; Qualitative ranking of document is on a 1(lowest value;poor)-3(highest value; good)							

SUMMARY OF OCCUPATIONAL BIAS RELATED ACHIVED TESTIMONY						
Source*	Conducting Rating Analysis**	Rate	Anomalies in Performance Rating**	Rate	Occupational Bias in Performance Rating**	Rate
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Charles Tiefer	They receive their pay raises in the same ways as they receive personnel actions, like promotion and discipline, known to have statistical patterns reflecting systemic discrimination risks	3	At the SEC, as similiarly would be seen about NSPS, the criteria for pay-for-performance left room for subjective evaluation; wide opening for discriminatory stereotyping, diversity-disparaging, patterns of discriminatory evaluation	2		1
			These pay-for-performance had serious costs and impacts. They have systemic discrimination risks because evaluative ratings, unlike objective government-wide pay increases, are vulnerable to supervisor's stereotyping attitudes, etc.	2		
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Dr. Stanley F. Ridley	There must be a statistically significant pass/success rate of the higher scoring/rated employee group relative to the employee group thought to be experiencing the adverse impact	3	Whenever a performance appraisal system even appears not to clearly provide CARE, perceptions (ie. distributive) among personnel are are likely to ensue.	2		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, John Gage		1	Supervisors in charge of judging employees have a natural tendency to favor people like themselves	2	There seem to be a profound bias in favor of employees who work higher up the chain of command or closer to the Pentagon! there also may be hierarchies of bias; we are told that while it is best to work in the Pentagon, second best is a regional command; senior mgrs. explained how the work of lower grades is systematically undervalued	2
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Carol Bonosaro	OPM has made some improvements to the SES pay system such as it requirement to share informaiton about pay adjustments, performance awards, and policies	2	One observations about the SES pay is the increased potential for politicization that resources from a pay system where so much unreviewable discretion is given to those who make salary decisions	2		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Testimony of Charles	A valid, reliable, unbiased performance management system is a necessary condition for an effective pay-for-performance program.	2	Employees and employee representatives alike have remarked on the mischief that can be done in the share process; a "black box" to most employees, and dilutes the performance pay linkage	2		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Colleen Kelly	The Hay Group Report found only 12 percent of FDIC said they found the pay-for-performance system to be a fair program for rewarding employees performance and contributions.	2	The artibrator found that African-American employees above grade 8 and older employees received significantly fewer pay increases than would be expected given their representation in the pool of eligible employees	3		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Max Stier	The GAO took the bold step of publicizing the agency's promotions and performance ratings by race, gender, age, disability, veteran status, location, and payband. The data revealed a gap between performance appraisals for African-Americans and whites	3	Simply put, a PFP system is doomed to fail if the affected employees do not perceive the system as fair for all, regardless of gender, race, religion, political affiliation, marital status, age, veteran status or disability.	2		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Russell George	In response to our findings, the IRS initiated a third-party evaluation of the IRS's pay-for-performance system. This evaluation will determine whether, or how stongly the PFP system supports is organizationa goals to recruit, retain and motivate future leaders.	2		1		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Robert Tobias	The "don't know answer is based on the fact that 73 percent of the respondents indicated their agency failed to share rating information.	2		1		1
Congress of the United States, House of Representatives, Feb 12, 2008, Committee on Oversight and Government Reform, Charles Fav		1	Research has shown there can be serious problems of rating accuracy. Human judgements are sometimes biased,. Studies have been done on the impact of rater bias, the relationship between the rater and ratee, and the demographics of raters and ratees	2		1
QUALITATIVE RATING AVG.		2.10		1.91		1.10

SUMMARY OF OCCUPATIONAL BIAS RELATED ARCHIVED BOOKS							
#	Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
1	Naff, 2001, To Look Like America: Dismantling Barriers		1	Research on factors affecting performance appraisals found that both African-American raters gave significantly higher rating to members of their own race	3		1
2	Montoya and Graham, 2007, Modernizing the Federal Government		1	Raters knowledge of prior performance affects raters information processing; additionally, it has been found that raters expectations introduce bias	2	Raters give scores according to how the employee performs in the area salient to the rater	2
3	Latham and Wexley, 1994, Increasing Productivity through Performance Appraisals		1	These managers made the similar to me error-raters tendency to judge more favorably those people whom they perceive as similar to themselves	2	Raters have a well developed categorization schema in terms of strong expectations about the performance of their in-group members and as a result tend to give high	2
4	Pynes, 1997, Human Resources Management for Public & NPOs		1	Same as me: giving the ratee a rating higher than deserved because the person has qualities or characteristics similar to those of the rater	2		
5	Milkovich & Newman, 2005, Compensation		1		1	Such things as organization values (values technical skills or interpersonal skills more highly) competition among departments, differences in status between departments, economic conditions-all influence the way raters rate employees	2
6	James W. Smither, 1998, Performance Appraisal		1	In addition to discrimination based on gender, race, national origin, age, disability, or other factors, there is concern with issues such as misrepresentation	3		1
7	Milkovich & Wigdor, 1991, Pay for Performance		1	Rating distortion is a behavior that is officially subject to sanction; orgs. rarely admit that ratings should sometimes be distorted.	2		1
8	Murphy & Cleveland, 1991, Performance Appraisal	The variance of the mean ratings calculated for each department in an org. should be positively correlated with the status differentiation within that organization	3	In addition, Cleveland and colleagues found anecdotal support for the influence of an organizational value system upon ratings.	2	Specifically, employees whose jobs or activities were perceived as more closely linked with the organization's mission or purpose were rated systematically higher	3
						Bjerke and colleagues noted that high-status combat communities in the Navy were more likely to suffer from rating inflation than were low-status support communities.	3
9	Landy & Farr, 1983, The Measurement of Work Performance		1	If the rater is white, or male, or older, or younger, or well educated, it is assumed that these characteristics play a major role	2	The characteristics of the position also have the potential to influence the ratings that an individual receives.	2
				Subjects without the occupational category must spend time and cognitive effort on making a categorization decision	2	Cohen has recently reported that subjects explicitly given the occupation of a target person were better able to recall both category consistency and category	2
10	Murphy & Cleveland, 1995, Understanding Perf. Appraisal	Some depts are high in status and others are very low; expect that the average rating w/ the low-status area will be much lower than the average rating in the high-status area.	3	From interviews with 60 upper level executive, Longenecker and colleagues found evidence of deliberate manipulation of formal ratings by executives.	3	Cleveland and colleagues found anecdotal support for the influence of an organization's value system upon rating.	2
				Managers therefore have their own motive and may manipulate rating in order to accomplish their own goals or agendas.	3	Appraisals designed by staff employees for personnel uses or for larger organizational concerns, yet they are completed line mgrs	2
				Longenecker found that executive adjust or manipulate employee rating in two directions - both higher and lower than the actual performance level.	3	It was noted that high-status combat communities in the Navy were more likely to suffer from rating inflation than were low-status support communities.	3
11	Mosher, 1968, Democracy and the Public Service		1	Control over employment policies and individual personnel actions for its own members in the agency and over the employment of employees not in the elite	2	Within the basic content of the agency, the officers of the line in the Navy...the civil engineers in a construction agency are favored	3
				The most elite of the sub-groups is likely to be the one which historically was most closely identified with the end purpose	2	A continuing drive of each is to elevate its stature...the upgrading of rewards (pay) for professionals; and the improvement of their prestige before their associates	2
12	Jeffrey Pfeffer, 1981, Power in Organizations	There was no correlation between departmental prestige and the indicators of subunit power	3			As noted before, none of the staff groups has much power, for the power in the firm was held by retailing, the store management f	3
						Those units responsible for accomplishing the more important tasks and activities come to have more influence in the organization.	2
12	Qualitative Rating Avg		1.50		2.27		2.12

SUMMARY OF OCCUPATIONAL BIAS RELATED ARCHIVED ARTICLES							
#	Source*	Conducting Rating Analysis**	Rate	Anomalies in Performance Rating**	Rate	Occupational Bias in Performance Rating**	Rate
1	Rosenberg, 2007, Arbitrator Rules Against SES Pay-for-Performance System	Statistical analysis performed for the union showed that only 16 percent of African-Americans SEC employees received raises	3		1		1
2	Losey, 2008, Is DODs New Pay System Fair, Federal Times		1	Civilian employees at DOD agencies were assigned overall higher performance ratings than civilian personnel in the AirForce, Army, Navy, and Marine Corps	3		1
				White employees received higher average performance ratings, salary increase and bonuses, than employees of other races	3		
3	Wiley, 2008, Restrictions on Top Employee Rating Violate Law	U.S. Code states that employees are to have their performance formally rated based on objective criteria and without limitations	2	Agency management routinely restricts the number of top ratings that can be given by a supervisor	2		1
4	Longenecker, Sims, and Gioia, 1987, Behind the Mask	The formal appraisal was linked to compensation, career, and advancement in the organization	2	Political action therefore represents a source of bias or inaccuracy in employee appraisal	2		1
5	Decotiis and Petit, 1978, The Performance Appraisal Process		1		1	Raters from different organizational levels may have different perceptions of the implication of the ratee for organizational outcomes	3
6	Klores, 1966, Rater Bias in Force-Distribution Performance Ratings		1	The biases of raters will in large measure determine the future, insofar as their biases are manifested in those promoted	2		1
7	Gioia and Longenecker, 1994, Delving into the Dark Side		1		1	A surprising number of people will bend those rules to satisfy a personal, departmental, or divisional agenda	2
8	Elvira and Town, 2001, The Effects of Race and Worker Productivity on Performance Evaluations	How much of the unexplained variance in performance ratings is accounted for by race?	3	Controlling for worker productivity and other demographic variables, black employees receive lower performance ratings than whites	3		1
9	Stauffer and Buckley, 2005, The Existence and Nature of Racial Bias in Supervisory Ratings	Some researchers who have found a statistically significant interaction effort have dismissed the effect as not being practically significant	3	We believe that others have failed to find this bias; there is a reluctance to call it bias without first understanding its source	2		1
10	Daley, 2003, The Trials and Tribulations of performance Appraisal		1		1	The importance of the unit to fulfilling the organization's mission is substituted for the job performance of the individual in that unit.	3
						Similarly, individuals working in a critical unit may benefit from the perceived centrality or significant of their part of the organization	3
11	Cook, 1995, Performance Appraisal and True Performance		1	Performance ratings suffer from many biases such as age, ethnicity, gender, and physical appearance.	2	Political considerations or private agendas nearly always influence performance ratings	2
12	Mobley, 1982, Supervisor and Employee Race and Sex Effects	Adverse impact can be evaluated in terms of the generally inadequate "80percent rule of thumb"...or by more appropriate statistical tests	3	The general model of the performance appraisal process... provides a framework for helping understand race and sex effects	2		1
13	Martell, 1991, Sex Bias at Work		1	Men were evaluated more favorably than women when raters were faced with an additional task requiring attention and pressures	3		1
14	Varma and Stroh, 2001, The Impact of Same-Sex LMX Dyads on Performance Evaluations		1	Based on data from supervisor-subordinate dyads, both male and female supervisors exhibit a positive bias to same sex	3		1
15	Jost and Elsbach, 2001, How Status and Power Differences Evoke Social Identities		1	Social identity theory does a very good job of predicting high-status members who exhibit strong levels of ingroup bias	2	In business organizations members of high-status groups are far more likely to engage in bias against lower status group members	2

16	DiThomaso, Post, Smith, Farris, and Cordero, 2007, Effects of Structural Position		1	The article provides attention to a framework of stereotype content that is based on an analysis of structural position among groups in workplace	2		1
				The hypothesis challenges assumptions in the management literature that each group will express in group bias and outgroup derogation	2		
				It is found that US born white males, who constitute the normative in-group, receive advantages in allocation and evaluation decisions	2		
17	Varma, Pichler and Srinivas, 2005, The role of interpersonal affect in performance appraisal		1	Results from the US sample indicated that raters are able to separate their liking for a subordinate when assigning ratings	3		1
18	Smith, DiThomaso, and Farris, 2001, Favoritism, Bias, and Error in Performance Ratings	Analyze performance ratings for a sample of 2,445 scientists and engineers from 24 US companies	3	It is found that there is more evidence of in-group favoritism than out group derogation	3	High status, dominant, and majority group members enjoy favoritism expressed as a global prototype of them as competent	2
19	Prendergast and Topel, 1993, Discretion and Bias in Performance Evaluation		1	Evidence of potential bias in performance appraisals come from a variety of sources: personal relations, race, politics	2		1
20	Bol, 2008, The Determinants and Performance Effects of Supervisor Bias		1	The causes of supervisor bias: employee performance, organizational level, financial position of firm, etc.	2		1
21	Borman, Oppler, Pulakos, and White, 1989, Examination of Race & Sex Effects	The primary purpose of this research was to investigate the interaction of rater-ratee race and sex on performance ratings	2	The overwhelming findings was the proportion of rating variance accounted for by race and gender was extremely small	3		1
21	Qualitative Rating Avg		1.62		2.17		1.45
<p>Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography; Textual description in column represents a paraphrased summary of a key point from the document; Qualitative ranking of document is on a 1(lowest value;poor)-3(highest value; good)</p>							

OCCUPATIONAL BIAS INTERVIEW SUMMARY SHEET						
Source	Conducting Rating Analysis	Rate	Anomalies in Performance Rating	Rate	Occupational Bias in Performance Rating	Rate
Interview #1	You want to find the patterns and or problems that have to be fixed	2		1	I can understand that people of the same cloth, like people of the same	2
Interview #2	We do not have a performance rating system in the agency; a narrative summary	2	To prevent anomalies within there are suggested target brackets	2	Concern that if you had the high profile programs has the higher ratings	2
Interview #3	In NSPS or the agencies that use PFP, there are race or gender issues;look @ outliers	2		1	I have not seen this; everybody just gets a 3	2
Interview #4	We have some looks at patterns in agencies or govt wide; looked at patterns for equity & merit	2	Do some examinations; found some anomalies; ethnic, sex, very clearly organizational differences	3	There are some occupational bias indications that there is differences; occupations w/ higher status=higher ratings	2.5
Interview #5	Compliance to determine that there are no egregious violations or any statistically significant differences.	3	Found bias based on race	3	We do not look for occ bias	2
Interview #6	Look @ systems for faultiness; look at the mean differences which are of significance in valid systems	3	Empirically found race and gender differences in ratings in operational performance ratings	3	We know the MC job occs get greater ratings than support jobs in general	2.5
Interview #7	We do it is because if there is any variance or /wi labs, we bring it to the managers attention	3	Higher end grades tend to get more higher grades	3	There is a difference in classes ie. Engineers & Scientist vs Admin.	2.5
Interview #8	Aligning the standards and direction of the organization related to the mission is a key concern	1	There is a tendency to rate your unit higher than others	2	Test unit vs training unit; if panel member is not represented-there is a knowledge bias effect	2
Interview #9	University does post performance review of all faculty	2	Considerable amount of general and race bias in the literature	2	Yes; academics have ideological preferences in rating; qualitative orientation vs qualitative	2.5
Interview #10	We are not very formal	1	Rating creep	2	The ability to manage the organization; tends to bias are towards people vital to getting job done	2
Qualitative Rating Avg		2.1		2.2		2.2
Note: Citations are abbreviated for presentation purposes; full citations are found in bibliography; Textual description in column represents a paraphrased summary of a key point from the document ; Qualitative ranking of document is on a 1(lowest value;poor)-3(highest value; good)						

Appendix B: NSPS Performance Rating Data Summaries

NAVY NSPS PERFORMANCE RATING DATA SHEET: SIGNIFICANCE OF DIFFERENCE OF MEANS

Pay Pool ID	Occupation Series	Occupation Series Total Population	Average Rating Total Population	Average Rating Total Population	Average of All PPs	Number of PPs	Standard Deviation of All Averages	Square Root of Number of PPs	Standard Error (std dev/sqrt of pps; the error asst w/ea pp)	Mean Variance (standard error sqred;)	Mean Diff Between 0800 & oth (dif tween avg o pps)	Variance o mean diff (sum of mean var tween occs)	Std Err of Mean Difference (sqrt of var of mean)	# of Standard Errors (meandif /by std err)
NV2540083	00xx	6	3	3										
NV25400801	00xx	17	3.12	3.12										
NV2562478	00xx	13	3.15	3.15										
NV25694500	00xx	15	3.2	3.2										
NV2562742	00xx	4	3.25	3.25										
NV2562473	00xx	20	3.3	3.3										
NV2533191	00xx	16	3.31	3.31										
NV25442550	00xx	9	3.33	3.33										
NV25624701	00xx	11	3.36	3.36										
NV2500025	00xx	16	3.44	3.44										
NV2540084	00xx	16	3.5	3.5										
NV25400852	00xx	11	3.55	3.55										
NV25624702	00xx	7	3.57	3.57	3.313846154	13	0.171878759	3.605551275	0.047670591	0.002272485	0.102264957	0.0040497	0.063637255	1.606998259
NV2500025	02xx	7	3.14	3.14										
NV25400801	03xx	31	3.13	3.13										
NV2569218	03xx	7	3.29	3.29										
NV25624702	03xx	20	3.3	3.3										
NV2562583	03xx	22	3.32	3.32										
NV25694500	03xx	38	3.32	3.32										
NV25442550	03xx	12	3.33	3.33										
NV25624701	03xx	34	3.35	3.35										
NV25400852	03xx	34	3.38	3.38										
NV2562478	03xx	12	3.42	3.42										
NV2533191	03xx	23	3.43	3.43										
NV2540192	03xx	7	3.43	3.43										
NV2562473	03xx	18	3.44	3.44										
NV2562742	03xx	9	3.44	3.44										
NV2500025	03xx	68	3.49	3.49										
NV2540083	03xx	28	3.5	3.5										
NV2540084	03xx	8	3.5	3.5										
NV2547408	03xx	7	3.57	3.57										
NV253258A	03xx	6	3.67	3.67										
NV25400851	03xx	5	3.8	3.8										
NV25400802	03xx	9	4	4	3.4555	20	0.193213136	4.472135955	0.043203771	0.001866566	-0.03938889	0.003643781	0.060363737	-0.65252569
NV25400801	04xx	16	3	3										
NV2540083	04xx	5	3	3										
NV2500025	04xx	4	3.25	3.25										
NV2562473	04xx	8	3.25	3.25										
NV25624702	04xx	26	3.38	3.38										
NV25694500	04xx	4	3.5	3.5										
NV25400852	04xx	9	3.78	3.78	3.308571429	7	0.277274147	2.645751311	0.104799777	0.010982993	0.107539683	0.012760208	0.112961091	0.952006406
NV25400801	05xx	6	3.17	3.17										
NV25624701	05xx	54	3.39	3.39										
NV2500025	05xx	25	3.4	3.4										
NV2562473	05xx	10	3.5	3.5										
NV2562742	05xx	6	3.5	3.5										
NV25694500	05xx	14	3.5	3.5										
NV25400852	05xx	12	3.58	3.58										
NV2540083	05xx	5	3.8	3.8										
NV2540192	05xx	4	4	4	3.537777778	9	0.240976025	3	0.080325342	0.00645216	-0.12166667	0.008229375	0.090715905	-1.34118341
NV25400801	08xx	194	3.05	3.05										
NV2562478	08xx	37	3.27	3.27										
NV2533191	08xx	58	3.31	3.31										
NV2562473	08xx	86	3.33	3.33										
NV2500025	08xx	66	3.36	3.36										
NV2540084	08xx	45	3.38	3.38										
NV2540192	08xx	16	3.38	3.38										

ARMY NSPS PERFORMANCE RATING DATA SHEET: SIGNIFICANCE OF DIFFERENCE OF MEANS

Pay Pool ID	Occupation Series	Occupation Series Total Population	Average Rating Total Population	Average Rating Total Population	Average of All PPs	Number of PPs	Standard Deviation of All Averages	Square Root of Number of PPs	Standard Error (std dev/sqrt of pps; the error asst w/ea pp)	Mean Variance (standard error sqred;)	Mean Diff Between 0800 & oth (dif tween avg o pps)	Variance o mean diff (sum o mean var tween occs)	Std Err of Mean Diff (sqrt of var o mean)	# of Standard Errors (meandif / by std err)
CE2W07602D	00xx	9	2.33	2.33										
CE2W07602E	00xx	10	2.9	2.9										
CE1W2R906A	00xx	8	3	3										
CE2W2SJ07	00xx	36	3.06	3.06										
CE2W2SM04A	00xx	17	3.06	3.06										
CE2W2SR06E	00xx	15	3.07	3.07										
CE2W2SJ01B	00xx	9	3.11	3.11										
CE2W2SR02C	00xx	8	3.13	3.13										
CE2W07604D	00xx	22	3.14	3.14										
CE2W2SUAA	00xx	7	3.14	3.14										
CE2W2SM01E	00xx	6	3.17	3.17										
CE2W2SM03E	00xx	12	3.17	3.17										
CE2W2SN03Y	00xx	6	3.17	3.17										
CE2W2ST02F	00xx	29	3.17	3.17										
CE2W2R9AAE	00xx	10	3.2	3.2										
CE2W2R9AA	00xx	26	3.23	3.23										
CE1W2SF05A	00xx	8	3.25	3.25										
CE1W2SM01A	00xx	24	3.25	3.25										
CE2W07106E	00xx	4	3.25	3.25										
CE2W0JVAAE	00xx	4	3.25	3.25										
CE3W4EFAAA	00xx	4	3.25	3.25										
CE2W2R9AAI	00xx	14	3.29	3.29										
CE1W2SJ01A	00xx	9	3.33	3.33										
CE2W2SD03A	00xx	11	3.36	3.36										
CE1W07602A	00xx	8	3.38	3.38										
CE1W2SR05E	00xx	8	3.38	3.38										
CE1W2SR04A	00xx	18	3.39	3.39										
CE2W2V6AAI	00xx	23	3.39	3.39										
CE2W2ST04C	00xx	7	3.43	3.43										
CE2W2R9AAI	00xx	9	3.44	3.44										
CE1W2R905A	00xx	4	3.5	3.5										
CE1W2SJ07	00xx	4	3.5	3.5										
CE1W2ST03A	00xx	10	3.5	3.5										
CE1W2ST04A	00xx	6	3.5	3.5										
CE2W2SR02E	00xx	4	3.5	3.5										
CE2W2ST02D	00xx	4	3.5	3.5										
CE3W2SJ02B	00xx	5	3.6	3.6										
CE2W2SM01C	00xx	8	3.63	3.63										
CE1W2SF01A	00xx	10	3.7	3.7										
CE2W2SJ01D	00xx	4	3.75	3.75										
CE1W2ST02A	00xx	10	3.8	3.8										
CE2W2SM03E	00xx	5	3.8	3.8										
CE2W0JVAAE	00xx	6	3.83	3.83	3.320930233	43	0.276188462	6.557438524	0.042118346	0.001773955	0.157069767	0.002153697	0.046407946	3.384544712
CE2W2R9AAI	01xx	6	3	3										
CE2W2SR06C	01xx	6	3	3										
CE2W2R9AAI	01xx	12	3.17	3.17										
CE2W2SJ07	01xx	5	3.2	3.2										
CE2W2R9AAI	01xx	34	3.21	3.21										
CE2W2SJ01D	01xx	4	3.25	3.25										
CE2W07106M	01xx	7	3.29	3.29										
CE2W2SN03Y	01xx	7	3.29	3.29										
CE2W0JVAAE	01xx	6	3.33	3.33										
CE2W2SN04A	01xx	4	3.5	3.5										
CE2W2SUAAI	01xx	16	3.56	3.56										
CE3W4EFAAA	01xx	16	3.56	3.56										
CE2W2SM01C	01xx	7	3.57	3.57										
CE2W2ST04F	01xx	13	3.62	3.62										
CE2W2SR02D	01xx	8	3.63	3.63										

