

RETHINKING MILITARY PERSONNEL EVALUATIONS

BY

LIEUTENANT COLONEL BRIAN A. YATES
United States Air Force

DISTRIBUTION STATEMENT A:

Approved for Public Release.
Distribution is Unlimited.

USAWC CLASS OF 2011

This SRP is submitted in partial fulfillment of the requirements of the Master of Strategic Studies Degree. The views expressed in this student academic research paper are those of the author and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. Government.



U.S. Army War College, Carlisle Barracks, PA 17013-5050

The U.S. Army War College is accredited by the Commission on Higher Education of the Middle State Association of Colleges and Schools, 3624 Market Street, Philadelphia, PA 19104, (215) 662-5606. The Commission on Higher Education is an institutional accrediting agency recognized by the U.S. Secretary of Education and the Council for Higher Education Accreditation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 03-23-2011		2. REPORT TYPE Strategy Research Project		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Rethinking Military Personnel Evaluations				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Lieutenant Colonel Brian A. Yates, USAF				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Captain James Heffernan Department of Command, Leadership and Management				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army War College 122 Forbes Avenue Carlisle, PA 17013				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution A: Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Military personnel evaluation systems are effective at identifying top performers, but struggle to differentiate between average records. Each service has a different approach to evaluation, appropriately reflecting the underlying philosophy and culture of the service. The basic purposes of all these evaluation systems break down into two essential themes: to provide information for decisions on promotion, retention, and assignment, and to provide feedback to the individual. The problem common to all systems is inflation of ratings, either directly manifested through inaccurately high scores or indirectly through exaggerated language and maladapted processes. While the problem of inflation has been well studied in the fields of organizational behavior and industrial psychology, a more simple approach using economic incentives for behavior can aid understanding. This paper offers three scenarios to explore the dynamics of these behavioral incentives and to manage inflation: a quota system, and pass/fail system, and a rater "cost" system. Ultimately, the cost approach offers the greatest potential for improving personnel evaluations and making them more useful for the military services.					
15. SUBJECT TERMS Rating Systems, Economics, Quotas, Performance, Inflation					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UNLIMITED	18. NUMBER OF PAGES 26	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED			19b. TELEPHONE NUMBER (include area code)

USAWC STRATEGY RESEARCH PROJECT

RETHINKING MILITARY PERSONNEL EVALUATIONS

by

Lieutenant Colonel Brian A. Yates
United States Air Force

Captain James Heffernan
Project Adviser

This SRP is submitted in partial fulfillment of the requirements of the Master of Strategic Studies Degree. The U.S. Army War College is accredited by the Commission on Higher Education of the Middle States Association of Colleges and Schools, 3624 Market Street, Philadelphia, PA 19104, (215) 662-5606. The Commission on Higher Education is an institutional accrediting agency recognized by the U.S. Secretary of Education and the Council for Higher Education Accreditation.

The views expressed in this student academic research paper are those of the author and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. Government.

U.S. Army War College
CARLISLE BARRACKS, PENNSYLVANIA 17013

ABSTRACT

AUTHOR: Lieutenant Colonel Brian A. Yates
TITLE: Rethinking Military Personnel Evaluations
FORMAT: Strategy Research Project
DATE: 23 March 2011 **WORD COUNT:** 5,495 **PAGES:** 26
KEY TERMS: Rating Systems, Economics, Quotas, Performance, Inflation
CLASSIFICATION: Unclassified

Military personnel evaluation systems are effective at identifying top performers, but struggle to differentiate between average records. Each service has a different approach to evaluation, appropriately reflecting the underlying philosophy and culture of the service. The basic purposes of all these evaluation systems break down into two essential themes: to provide information for decisions on promotion, retention, and assignment, and to provide feedback to the individual. The problem common to all systems is inflation of ratings, either directly manifested through inaccurately high scores or indirectly through exaggerated language and maladapted processes. While the problem of inflation has been well studied in the fields of organizational behavior and industrial psychology, a more simple approach using economic incentives for behavior can aid understanding. This paper offers three scenarios to explore the dynamics of these behavioral incentives and to manage inflation: a quota system, and pass/fail system, and a rater "cost" system. Ultimately, the cost approach offers the greatest potential for improving personnel evaluations and making them more useful for the military services.

RETHINKING MILITARY PERSONNEL EVALUATIONS

Essentially, all models are wrong, but some are useful.
—George E. P. Box (1987)

The United States military enjoys a reputation for excellence, professionalism, and selfless service. In fact, in recent surveys, 80% of the American public indicated the military is "excellent" (39%) or "good" (41%) at doing their job.¹ To a large extent, this reputation is the result of an intentional organizational approach to select, train, and groom the best candidates for leadership and promotion. This cultural "meritocracy" requires detailed policies and rules tailored to each service in order to enable the selection of the best. Obviously, the system as a whole is doing something right, since it has repeatedly demonstrated the ability to identify outstanding leaders in times of conflict. To some extent, however, the system works *despite* its structure. "Ratings inflation" has been a recurring problem for military personnel evaluation systems,² creating wasted effort by raters and their units and increasing the challenges for promotion and selection boards. This paper will examine the issue of ratings inflation, employ basic economic theory as a framework to explain the phenomenon, and then offer some alternative approaches to improve the usefulness of service personnel evaluation systems.

Why Have a Rating System?

In order to evaluate the effectiveness of current military personnel evaluation systems, one must first determine what the systems are intended to do. Each service has overarching policy objectives expressed in their respective personnel regulations. The Army, for example, states the "[Evaluation Reporting System] is a multifunctional

system with a basic structure that will allow the rater to give shape and direction to the rated Soldier's performance, provide a chain-of-command or supervision evaluation of an individual Soldier's performance and potential, and allow the entire evaluation reporting process to be reviewed."³

Air Force instructions indicate "[t]he Officer and Enlisted Evaluation Systems have varied purposes. The first is to provide meaningful feedback to individuals on what is expected of them, advice on how well they are meeting those expectations, and advice on how to better meet those expectations. The second is to provide a reliable, long-term, cumulative record of performance and potential based on that performance. The third is to provide officer central selection boards, senior NCO evaluation boards, the Weighted Airman Promotion System (WAPS) and other personnel managers sound information to assist in identifying the best qualified officers and enlisted personnel."⁴

Navy instructions state "FITREPs on officers, CHIEFEVALs on chief petty officers (CPOs), and EVALs on other enlisted personnel are used for many career actions, including selection for promotion, advanced training, specialization or sub-specialization, and responsible duty assignments"⁵ and "Planned and scheduled counseling is a major focus of Navy's performance evaluation system."⁶

Finally, the Marine Corps specifies that "Primarily, the [Performance Evaluation System] supports the centralized selection, promotion, and retention of the most qualified Marines of the Active and Reserve Components. Secondly, the PES aids in the assignment of personnel and supports other personnel management decisions as required."⁷

Although each of the services has a slightly different take on the purpose of their evaluation systems, two common goals emerge: (1) provide information for the service to use for promotion, assignment, retention, or other personnel actions, and (2) provide feedback to the member so he or she can improve (with the notable exception of the Marines, who make a point to say that an evaluation is NOT a counseling document).

These goals, however, are insufficient in an examination of the systems. As Dr. Scriven, a notable expert on evaluation systems, states:

The absurdity of [the goal-achievement model of evaluation] would have been obvious if anyone had been thinking about the general logic of evaluation, because everyone does *product* evaluation all the time and everyone knows that you don't evaluate products against the intents of their makers but against the needs of their users or prospective users (see any issue of *Consumer Reports*).⁸

The point is that the effectiveness of the various systems must be measured by their results and effects as well as by the intentions or goals of the policy makers. This is not a question of whether evaluation of performance is generally desirable -- from the stated goals of each service, it clearly is -- but rather whether the methods chosen are effective in achieving the stated purposes. In short, the merits of any system should be judged on whether they produce useful results.

There are several assumptions in the logic behind the evaluation systems that should be understood. First, the systems all assume that past performance is an indicator of future performance. While this is certainly true in many cases, it is just as certainly *not* true in all cases. Second, the systems assume that the performance is being measured using defined standards. While there are indeed standards for conduct and performance throughout the military, these standards are used in practice only as expected minimums, since jobs vary widely across the range of military specialties, as

do the expectations for the level of performance (e.g. new recruit compared to experienced veteran). Third, the context of the rating process is assumed to be homogenous, but this is almost never the case. While a person might produce outstanding results all the time, there is a qualitative difference between doing so under adverse conditions such as combat versus more favorable conditions such as peacetime training, yet the scale for scoring is the same. These assumptions represent a mismatch between the intentions of the evaluation systems and the actual effects that the services achieve. It does not, however, mean that the systems are not functional, only that another approach could provide more useful results.

Service Approaches to Evaluations

To illustrate the challenge of executing an effective evaluation system, consider the different approaches each service takes to the issue. First, Army officers are evaluated on areas of professionalism, Army values and leader attributes, which factor into the final ratings, but are not directly linked to the score. The rater then provides an overall rating of performance and potential, selecting either Outstanding, Satisfactory, or Unsatisfactory, along with a narrative describing the specifics. The senior rater then selects both a rating of promotion potential (Best Qualified, Fully Qualified, or Do Not Promote) and an overall performance rating compared to the ratee's peer group (Above Center of Mass [ACOM], Center of Mass [COM], or Below Center of Mass [BCOM]).⁹ This senior rater's overall rating is among the most influential factor for boards, and the Army closely manages the number of ACOMs allowed by formally tracking statistics for each individual senior rater to ensure the number is less than 50 percent.¹⁰

Army NCO evaluation reports also have an area for Army values, with a more detailed breakdown of duty performance, rated Excellence, Success, or Needs

Improvement. The rater then assesses the potential for promotion as Among the Best, Fully Capable, or Marginal. The senior rater provides scores between 1 and 6 for both overall performance and potential for promotion, with a score of 1 being the best.¹¹

There are no limits imposed on the number of each level of rating allowed, although senior raters are charged with "... using all reasonable means to prepare a fair, correct report that evaluates the NCO's duty performance, values/NCO responsibilities, and potential."¹²

In contrast to the Army's "Center of Mass" approach, Air Force Officer Performance Reports (OPRs) evaluate six performance factors using a simple two block scale, either Meets Standards or Does Not Meet Standards. The rater and additional rater each make bulleted comments, which provide the preponderance of the information for the report. Since there is no score aside from the pass/fail ratings, the importance of the comments is magnified, and raters spend a great deal of time crafting nuanced statements that will allow differentiation. The governing instruction gives specific instructions about both mandatory and prohibited comments (for example, promotion recommendations are not allowed on an Air Force evaluation; these are reserved for a Promotion Recommendation Form accomplished when the member is eligible for a promotion board).¹³

Air Force NCO evaluations use a four-block system to capture performance and leadership qualities, which impact but are not directly linked to the five-block ratings by the rater and additional rater, with 5 being the highest rating (Poor, Needs Improvement, Average, Above Average, and Truly Among the Best).¹⁴ Each rater also provides additional comments and stratification if desired, subject to the same limits as officer

evaluations, except that promotion recommendations are not prohibited, but expected.¹⁵ There is no limit to the quantity of particular ratings assessed (i.e. the rater can give every member all 5's), and no review process for the rater's scores.

Navy reports use the same approach for all ranks, although the reports are named differently administratively. Performance traits are graded on a 1 to 5 scale, with 5 being the best. While there are no limits on the number of high (or low) scores that can be assessed, the graded performance traits are averaged for an overall score, which is compared to the reporting senior's overall trait average for that reporting group. The reporting senior also makes a promotion recommendation using a five box scale (Significant Problems, Progressing, Promotable, Must Promote, and Early Promote). This recommendation is independent of the trait average, but should be consistent. The highest rating is limited to no more than 20 percent of reports, and the next highest rating is also limited for some ranks (for example, the total number of Early Promote and Must Promote ratings cannot exceed 40% of the total for O-5s and O-6s).¹⁶

Finally, the Marine Corps mirrors the Navy approach in using a single process for E-5 through O-6. The rater scores 13 attributes describing mission accomplishment, character, leadership, and intellect, using an A to G scale, with G being the highest. These ratings are translated to a numerical scale of 1 to 7, respectively, and then averaged to provide the overall score for the report. This specific score is compared to the average scores by that rater for all previous rated Marines of the same grade, and converted to a "relative value" score, where 80 is the lowest score ever assessed, 90 is average, and 100 is the highest.¹⁷ The Reviewing Official also provides a comparative assessment of how the Marine compares to other Marines of the same grade, with the

score ranging from 1 (Unsatisfactory) to 8 (The Eminently Qualified Marine). This comparative assessment is also tracked over the reviewing official's career, and the information on how the individual report compares is available for boards and other personnel actions.¹⁸ While these varied approaches have different strengths and weaknesses reflecting underlying service philosophies, they all share a common challenge that must be considered in evaluating whether the individual systems are useful in achieving their stated goals.

"The Problem"

The problem with each of these systems is inflation. To put it simply, the ratings recorded are inaccurately high for the population as a whole. The most extreme examples are the Air Force's Enlisted Performance Reports, which have become notorious for awarding average-caliber performance with the highest scores. This has become so common that reports with maximum ratings are referred to as a "Firewall 5s." Consider from 2000 to 2008, the number of promotion-eligible E-5s earning perfect EPR scores rose from 66 percent to 71 percent; similarly, the numbers rose from 83 to 86 percent for E-6s and from 91 to 95 percent for E-7s.¹⁹ The impact is that the EPR ratings are all but meaningless for the purposes of promotion, leaving differentiation to other factors in practice.

For the 2009 E-7 promotion board, the Air Force selected 1,269 members for promotion to Master Sergeant, all of whom had perfect EPR scores in their records. This could plausibly reflect an intentional disposition to choose those who are "Truly Among The Best," until one considers that an additional 11,502 airmen also had perfect scores, but were not selected.²⁰ The author's experience as a Air Force squadron commander confirms this trend, since the understood (but frustratingly undocumented)

standard within the wing was that any enlisted report which was NOT a "Firewall 5" had to be accompanied with additional justification as to why that individual was not "Truly Among The Best."

The evaluation systems of some of the other services attempt to address this issue by creating limits, which, in execution, become quotas for the highest ratings. The Army's "Above Center of Mass" (ACOM) rating, for example, might on the surface appear to prevent inflation by restricting the number of ACOMs to less than 50 percent, but in fact inflation is still present, only it manifests in different ways. If a senior rater assigns an ACOM rating to a soldier that results in a historical percentage of 50 percent or more of total ACOM ratings by that senior rater, the report is automatically downgraded to "Center of Mass" (COM) by the personnel system. The regulations specifically address this:

"To ensure maximum rating flexibility when rating populations change or to preclude a top box check from inadvertently profiling as a CENTER OF MASS rating, senior raters will need to maintain a "cushion" in their top box rather than simply playing the line at less than 50 percent."²¹

The need to officially address this issue is a recognition of the distortions caused by inflation. These distortions can manifest either through a reactive approach to ratings, where top ratings are simply awarded as they become available, or through an entitlement-based mindset, where perceived need for career progression takes precedence over performance.²² This latter distortion leads to a situation where the number of ACOMs in a performance file becomes one of the most important indicators for promotion, and if these ratings are spread out in search of "fairness," then many records begin to look similar.²³

The Navy's approach is more tightly controlled, limiting the top tier promotion recommendation ratings to a constant 20 percent and the second tier to an additional amount that varies with rank. Even though this may appear to automatically prevent inflation of promotion ratings, in fact the system simply inflates to the maximum it is allowed, just as the Army system does. For the top tier rating, a simple examination will reveal that the specified percentage allowed, 20 percent, is what is routinely allocated by the reporting senior. Inflation then manifests in other ways throughout the system, resulting in the need for informational briefings for raters to learn how to "use" the system properly. A typical example is a personnel briefing that describes the impact of promotion recommendations: "Did you know there are now unlimited MPs ["Must Promotes", the second highest rating] for O-3s? What message is sent if you mark one O-3 as MP?"²⁴ The implication is that raters should rate even average performers as "Must Promote" so as not to injure future promotion chances (unless, of course, that is the intent...). Another example is a personnel management briefing that offers "mitigating techniques" to circumvent the forced distribution: "Would be EP [Early Promote] in any other group, however..." and "Forced distribution unfairly restricts accurate recommendation."²⁵ Given that the promotion chances for Lieutenant Commander are around 90%,²⁶ the impact is that over 70% of the officers competing have the same rating on their most recent report, meaning that the board must use other factors to differentiate between records.

In summary, the ratings under the current systems are not fully effective at identifying different levels of performance, though some systems do better than others. All of the service systems do a credible job of identifying the extremes, i.e. the best and

worst performers, but they struggle with making distinctions between solid, middle-of-the-road records. Since promotion boards have only a few minutes to spend with each record, as board members look for a few key indicators and move on, the ratings for those average records are, in effect, wasted effort. Promotion boards (or other officials) are forced to seek alternative indicators of performance and potential, and these indicators are frequently unofficial, uncontrolled, and subject to "fads" or "word of mouth" changes over time. All of these factors point to the opportunities to improve the ratings systems.

Using Economics to Understand "The Problem"

If inflation is "The Problem," understanding *why* ratings become inflated over time can help solve it. While this issue has been examined and documented in the fields of organizational behavior and industrial psychology, a more intuitive approach using basic economic principles can, perhaps, explain it in a different light. Consider that "a market is an established arrangement by which buyers and sellers come together to exchange particular goods or services."²⁷ Within this idealized market, individual actors take on one of two roles: buyer and seller. In the process of fulfilling his or her wants and desires, the buyer creates demand, which is a collective schedule of "...the amounts of the good or service [buyers] are prepared to buy at different prices during a specified time period."²⁸ Demand is inversely related to price, so for a free good, demand would be theoretically unlimited, but in practice is limited by myriad factors. Supply, contrary to demand, is how much of something a seller is willing to provide at a range of different prices, and the relationship between the two is positive, i.e. at sellers are willing to sell less at lower prices and more at higher prices. The coordinating function for both

supply and demand is the price, and the main function of a price is to "...provide incentives to affect behavior in the use of resources and their resulting products."²⁹

This incentivization of behavior is the core concept that will allow an examination of personnel evaluation approaches. In order for any exchange to occur in a market, the buyer and seller must decide that the *value* of what is delivered at a particular price is preferable to their next best alternative. Note that value is not wholly dependent on cost. For example, a supermarket may offer "loss leaders," or items for sale at a price below what they cost the store to acquire, in order to entice customers to shop at their store. The value is in the increased customer traffic, even though the monetary cost is higher than the price. An automobile buyer might need a vehicle to get to work in the morning, but choose to purchase a premium luxury car because he or she values the derived social status or enjoys the driving experience. The price is greater than the true cost of basic transportation, but the value makes it a desirable choice.

Two key points emerge from this short review of economics. First, the true cost of something is not merely the price in money, but the value that it has in alternative uses³⁰, and this value is frequently more than just monetary. This concept of value is useful for thinking about military personnel ratings, since the purpose is not merely to evaluate performance against defined standards, but to transmit the *value* of the performance relative to other service members (the next best alternative, in this case) to those who do not have the opportunity to observe the behavior, e.g. promotion boards, etc. The second point is that an exchange benefits both the buyer and seller, or else the exchange would not occur, since either one or both would choose a different alternative. This applies even in situations where there does not appear to be a

desirable choice. A manufacturer with excess inventory may not *want* to sell the products for a loss, but chooses to do so because it is preferable to get at least some money instead of keeping the inventory. A consumer may not *want* to pay five dollars for a gallon of gasoline, but chooses to do so because it is preferable to less driving, or not driving at all.

So how does this apply to personnel ratings and inflation? The assessment of ratings is, in effect, an exchange between the rater and the ratee. The ratee offers some level of performance with an expectation that this will be rewarded with a certain rating. The rater sets the standards and expectations, provides guidance on what it takes to achieve them, and then awards an appropriate rating for the observed performance. Thinking of assessments as an exchange is useful because it helps explain the incentives the systems create. Within this construct, other exchanges also take place; earning medals, achieving expert ratings or qualifications, personal feedback, visible recognition, etc. -- all are part of the interaction between rater and ratee. Of course, the performance of military members is motivated by much more than just ratings, and there is no intent to cast the noble service of our military in such a mercenary light. In fact, systems which try to create incentives by directly linking specific ratings to specific performance can be problematic, since the tendency is for the ratee to become "finely attuned to issues of fairness."³¹ This can result in mere compliance rather than initiative on the part of the service member, and is a perfect example of an unintended, and undesirable, consequence. Even so, ratings do have an impact on performance, even if they are inflated. One only needs to consider the

converse situation -- where ratings are artificially low as opposed to artificially high -- to understand the negative impacts such a situation creates.

So, thinking of the process of evaluation as an exchange allows some insight into what incentives are created by different systems. Recalling the common objectives of the service approaches, the rater has incentives to provide accurate information for future readers of the report, and to provide feedback to the ratee through the ratings. The problem is that there are other incentives in play as well, and those incentives can be at odds with the "official" ones, resulting in ratings inflation. Inflation can occur for a variety of reasons, including: (1) raters may not think the measurements in the system reflect actual performance, (2) raters who practice favoritism may get caught in an upward spiral, (3) raters may not want to discourage a struggling performer, (4) raters may want to be liked or avoid any conflict situations from low ratings, and (5) raters who do not have time to mentor or counsel sufficiently may not want to penalize a mediocre performer for that lack.³² In addition, for those in the military, the career impact of low or even average ratings to career potential is non-trivial, translating almost directly to promotion potential, assignment choice, training, length of service, retirement opportunity, rank and benefits, and sheer monetary earnings. The stakes are indeed high.

Since the existing systems generally do not impose a "cost" to the raters, the raters have little incentive to avoid inflating scores, and plenty of incentive to do so. In other words, the value raters derive by offering higher ratings overcomes the value of being strictly accurate, and while not every rating will be inflated, ratings on the whole become inaccurately high. Some services have measures in place to counteract this

tendency, with the Army and Navy overt limits to the highest ratings being the most apt examples. Arguably the most effective system at mitigating inflation is the Marine system, where the individual ratings are constantly and visibly compared to the rater's average. This works to a large extent because a high (or low) score affects every other rating for the future, meaning that inflated scores will quickly render high scores meaningless, and this performance by the rater is visible to other senior officials. This Marine approach is one way of imposing a "cost" on the rater, and is probably the most effective among all the services at controlling inflation, but it is not perfect, since the system still struggles to differentiate between average records.

Different Possible Approaches

Given the tendency towards rating inflation is common, and understanding why it occurs, an examination of possible alternative systems is appropriate. This paper offers three distinct approaches with differing philosophies. All are designed with the same objective: making personnel ratings more useful for the services.

Quota System. The first option is to institute a straightforward quota system for scores. Assuming that any population of service members has a range of talents, motivation, experience, and output, it should be possible to fit the ratings to some statistical distribution curve that allocates the total scores in specific increments along the curve.³³ The Army restricts ACOM ratings to less than 50 percent, and the Navy restricts Early Promotes to no more than 20 percent, so both approaches have the beginnings of a quota system. To be fully implemented, however, the system needs to allocate the remaining scores as well, not just limit the top tiers. To avoid ambiguity, every person would need to be assigned a position along the distribution curve, and so the system would need enough granularity to differentiate between them (i.e. there must

be sufficient "slots" available). The primary advantage of a quota distribution is clear: everyone would know where they stood relative to others in the rating pool, and that information would be clearly transmitted and useful to anyone who viewed the report.

The disadvantages to a system like this, however, would be challenging to overcome. First is the assumption that the rating group has a predictable distribution, such as a "normal" distribution curve. Other curves might fit the population better, but determining that curve, or even which curve is best across potentially disparate populations within the military, would be difficult at best. A second disadvantage is that it is a zero-sum game. This would mean that extra effort to achieve higher performance would necessarily come at the expense of someone else -- not exactly an environment conducive to unit cohesion. Also, while a quota system would help with rater incentives by forcing the distributed assignment (i.e. the rater would not have to bear that "cost"), ultimately some number of individuals must be rated at the bottom. The "costs" to the rater of presenting that information would be high, and could potentially disincentivize performance and improvement on the part of those ratees. Finally, the nearly inevitable incidents of favoritism would be almost impossible to manage, even if it is only a perception. The rater would have to invest significant effort to demonstrate a transparent link between performance and ratings, forcing time and focus to be spent on task compliance rather than initiative and creativity. In fact, major corporations have been moving away from quota systems in recent years due to lawsuits and morale issues for exactly these reasons.³⁴ Quotas would, by definition, eliminate inflation, but the impact of the negative effects of such a system do not bode well for the success of this approach.

Pass/Fail System. At the other end of the spectrum of options is simply eliminating evaluation scores altogether. Given that promotion boards have a very limited time to evaluate each individual record, much of the energy spent on rating individuals is wasted anyway, at least for that purpose.³⁵ The Air Force's Officer Performance Report (OPR) is an example of this approach. As long as the officer meets or exceeds the minimum levels of performance, the rater selects "Meets Standards," and the vast majority of OPRs are scored this way (reports marked with "Does Not Meet Standards" ratings are considered "referral reports" requiring additional documentation, and are almost always associated with administrative or judicial punishment).³⁶ This means the evaluation is limited to the word picture that the rater can create for the user. The advantage of a pass/fail approach is that it completely bypasses the issues of inflation, at least in terms of quantitative scores. It provides an economy of effort by not forcing organizational and administrative time to be spent on an area a promotion board will not find useful anyway. There are no profiles to track or ratings pictures to project, no overruled ratings due to exceeding limits, and raters are allowed to describe top performers in terms of results rather than forced distributions. Pass/fail also avoids conflict between the rater and average or marginal performers, since the word pictures describing their performance tend to be favorable overall.

The biggest problem with pass/fail, however, is that it simply transfers the inflation to the word picture. Raters resort to exaggerated descriptions about the impacts of performance, often using superlatives to capture the merely mundane. Words that would normally be reserved for rare and exceptional events show up regularly, as raters vie for the limited attention of promotion boards. The presence or

absence of recommendations for command, professional military education, joint service, or special duty become de facto discriminators. The tendency to "stratify," or provide a comparative ranking within a peer group, illustrates that these inflated word pictures are no more effective at differentiating between similar performance than an inflated scoring system would be, especially for the bulk of average records.

"Cost" System. A final approach is to adopt a new system that imposes a cost for the rater. The concept is that every rater would receive a sufficient number of points to rate every ratee as average.³⁷ These points would then be assigned to individual reports based on the rater's assessment, and the scores would fall into "bands" with associated descriptors. The point scale would need to be sufficiently granular to allow differentiation without losing meaning, i.e. a 1 to 5 scale with a score of 3 is inadequate; a 1 to 1,000,000 scale with a score of 561,977 is meaningless. To avoid the stigma associated with perceived low rankings, this paper proposes a 60 to 100 scale. Raters would receive 80 points for every ratee, but could allocate those points as they see fit. This is certainly a zero-sum game, since rating one person with an 85 means five fewer points for others, but with a sufficiently large scale, the negative effects are minimized. The "bands" for performance would be 60 to 69 as marginal, 70 to 89 as successful, with 90 to 94 being outstanding, and 95 or higher truly exceptional. The large middle band normalizes the expectation that the bulk of performance ratings are average, but the scores still provide useful information on where an individual falls within that band.

The advantage of a cost plan is that it makes rating points a scarce commodity which raters would need to allocate carefully. It avoids the problems of a forced distribution, since there is no mandate to score according to a predetermined curve. It

also avoids inflation, since the expected effect would be for the scores of "average performers" to center slightly below 80 in order to provide the opportunity to identify superior performers. Most importantly, by using a more robust rating scale, the cost system maximizes the information available to a promotion board, enabling a finer comparison of performance over time.

One possible difficulty with a cost system is the potential to "run out" of points. If raters do not manage their point totals properly, ratees receiving reports at the end of the rating period could be at a disadvantage. This could be mitigated by rationing the points throughout the year. Another challenge could develop with the expectation that new arrivals or inexperienced personnel would receive lower scores, and have to "earn" the higher scores over time. This is not necessarily a desired result, but it could be an expected response, and it is mitigated by the larger point scale, which will still allow differentiation of performance even if the averages are lower to begin with. A final challenge is how to deal with elite or selectively-manned units, where the expected level of performance is above the average for the service. The system could address this by providing extra points for combat achievements, unit awards, or exceptional service, which could be allocated appropriately by the rater.³⁸

Conclusion

This paper has examined the stated purpose and basic structure of personnel evaluations systems for each of the military service. It also defined the problem of inflation and used basic economic theory as a framework for understanding why inflation occurs. Offering a range of options, as well as a discussion of the advantages and disadvantages of each approach, this paper finds a single method to be preferable to the others. The cost system, while not perfect, would be the best approach. Since

the stated objectives of the services are to provide information for personnel actions and to provide feedback to individuals, the cost system has the biggest potential for success, since it offers the biggest improvement in the ability to distinguish between similar records without jeopardizing the rater/ratee relationship.

The personnel systems of the various military services reflect the values and character of each service, and these systems perform sufficiently well to identify, track, and promote quality leaders. In essence, these systems already work. The compelling need for change, however, could come if the services have to make harder choices in a budget-constrained environment. Senior leaders who must make those hard choices deserve the best information available, and improving the quality and meaning of rating scores could achieve the holy grail of systems: making them more useful.

Endnotes

¹ The Pew Research Center for The People and The Press, "The People and Their Government: Distrust, Discontent, Anger, and Partisan Rancor," April 18, 2010, 57, <http://people-press.org/reports/pdf/606.pdf> (accessed 3 Feb 2011).

² The phenomenon of ratings inflation is a concern for, but not confined only to, military personnel systems. For an extensive review of references describing this effect, see Stephanie L. Wolfgeher, *Inflation of USAF Officer Performance Reports: Analyzing the Organizational Environment*, Master's Thesis (Monterrey, CA: Naval Postgraduate School, December 2009), 1.

³ U.S. Department of the Army, *Evaluation Reporting System*, Army Regulation 623-3 (Washington, DC: U.S. Department of the Army, August 21, 2007), 3-4.

⁴ U.S. Department of the Air Force, *Officer and Enlisted Evaluation Systems*, Air Force Instruction 36-2406 (Washington, DC: U.S. Department of the Air Force, February 17, 2007), 6.

⁵ U.S. Department of the Navy, *Navy Performance Evaluation System*, Bureau of Naval Personnel Instruction 1610.10B (Washington, DC: U.S. Department of the Navy, July 9, 2008), I-1.

⁶ *Ibid.*, 19-1.

⁷ U.S. Department of the Navy, *Performance Evaluation System*, Headquarters United States Marine Corps Order P1610.7F (Washington, DC: U.S. Department of the Navy, May 11, 2006), 1-3.

⁸ Michael Scriven, "The Logic of Evaluation," in *Dissensus and the Search for Common Ground*, eds. Hans V. Hansen, *et. al.*, CD-ROM (Windsor, Ontario: Ontario Society for the Study of Argumentation, June 6-9, 2007), 7.

⁹ U.S. Department of the Army, *Officer Evaluation Report*, Department of the Army Form 67-9 (Washington, DC: U.S. Department of the Army, March 2007), 1-2.

¹⁰ U.S. Department of the Army, *Evaluation Reporting System*, 25.

¹¹ U.S. Department of the Army, *NCO Evaluation Report*, Department of the Army Form 2166-8 (Washington, DC: U.S. Department of the Army, March 2006), 1-2.

¹² U.S. Department of the Army, *Evaluation Reporting System*, 27.

¹³ U.S. Department of the Air Force, *Officer and Enlisted Evaluation Systems*, 27.

¹⁴ U.S. Department of the Air Force, *Enlisted Performance Report*, Air Force Form 910 (Washington, DC: U.S. Department of the Air Force, June 18, 2008), 1-2.

¹⁵ U.S. Department of the Air Force, *Officer and Enlisted Evaluation Systems*, 55.

¹⁶ U.S. Department of the Navy, *Navy Performance Evaluation System*, 2 and 1-17.

¹⁷ U.S. Department of the Navy, *Performance Evaluation System*, 8-9.

¹⁸ *Ibid.*, G-3.

¹⁹ Erik Holmes, "Ideas for improving EPRs," *Air Force Times*, July 28, 2009, http://www.airforcetimes.com/news/2009/07/airforce_EPRs_072009w/ (accessed January 15, 2011)

²⁰ Bruce Rolfsen, "Airmen to chief: Fix evals now," *Air Force Times*, March 26, 2010, http://www.airforcetimes.com/news/2010/03/airforce_epr_032610w/ (accessed February 8, 2011)

²¹ U.S. Department of the Army, *Evaluation Reporting System*, 25.

²² Dennis P. Chapman, "It's Time for a New Round of OER Reform," *Military Review* 86, no. 5 (September-October 2006), 100.

²³ LTC Briggs is an Operations Research Analyst who has an interesting, and award-winning, model for predicting promotion rates in the aggregate. David Briggs, "Modeling Army Officer Promotions," *Phalanx* 38, no. 4 (December 2005), 25.

²⁴ Ralph T. Soule, "Fitness Reports and Selection Boards," December 20, 2002, 5, www.navyfitrep.com/files/Navy_Evaluations_NAVSEA.ppt (accessed March 8, 2011).

²⁵ G.W. Korchowski and J. Weingold, "Personnel Career Management," August 27, 2002, 22, www.navyfitrep.com/files/PERS_Carrier_Management.ppt (accessed March 8, 2011).

²⁶ Bureau of Navy Personnel, "FY-09 Active Duty Navy Lieutenant Commander Line Promotion Selection Board," April 16, 2008, <http://www.persnet.navy.mil/NR/rdonlyres/019504DE-6668-490C-AAEC-039AF482F0F5/0/FY09USNO4Linestats.pdf> (accessed March 8, 2011).

²⁷ Roy J. Ruffin and Paul R. Gregory, *Principles of Economics* (Glenview, IL: Scott Foresman and Company, 1986), 67.

²⁸ *Ibid.*, 68.

²⁹ Thomas Sowell, *Basic Economics* (New York: Basic Books, 2007), 15.

³⁰ *Ibid.*, 23.

³¹ Jerald Greenburg, "Differential Intolerance for Inequity from Organizational and Individual Agents," *Journal of Applied Social Psychology* 16, no. 3 (May 1986): 191-196.

³² A.J. Schuler, "Yearly Employee Evaluations and Rating Inflation," <http://www.managerwise.com/article.phtml?id=360> (accessed February 5, 2011).

³³ The quota approach is based on a comparison of relative performance, rather than evaluating performance to a standard. "The principal advantage of the comparison method is preventing raters from placing all employees in one category (for example, all superior). Two disadvantages—especially when very few workers are involved—include assuming (1) employees fall in a normal distribution (there may be four excellent performers in a group of five, or none in a group of three), and (2) there are similar differences in performance between two adjacent employees, for instance, between those ranked 1 and 2 and those ranked 4 and 5." Gregorio Billikpof, "Performance Appraisal," <http://www.cnr.berkeley.edu/ucce50/ag-labor/7labor/06.htm> (accessed March 23, 2011).

³⁴ Mark Truby, "Ford Ends Grading Quotas," *The Detroit News*, July 10, 2001, <http://detnews.com/2001/autos/0107/10/a01-245553.htm> (accessed March 22, 2011).

³⁵ The pass-fail system has a checkered history in federal service, falling into and out of favor over time. "This month, the Merit Systems Protection Board reported that newly popular five-tier employee evaluation systems in federal agencies are not much better than their pass-fail predecessors." Karen Rutzick, "Grading the Grading System," September 28, 2006, <http://www.govexec.com/dailyfed/0906/092806pb.htm> (accessed March 23, 2011).

³⁶ U.S. Department of the Air Force, *Officer and Enlisted Evaluation Systems*, 36.

³⁷ Similar systems have been proposed in the past, although without considering an alternative rating scale. Captain David Blair's proposal is one example. See Rolfsen, "Airmen to chief: Fix evals now."

³⁸ Captain Blair's ideas on this subject are similar, although he allows for individual awards as well. *Ibid.*

