

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS	2a SECURITY CLASSIFICATION AUTHORITY	2b DECLASSIFICATION/DOWNGRADING SCHEDULE
--	-------------------------	--------------------------------------	--

3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	4 PERFORMING ORGANIZATION REPORT NUMBER(S) NPRDC TR 85-6
	5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION Navy Personnel Research and Development Center	6b OFFICE SYMBOL <i>(if applicable)</i> Code 41
---	--

6c ADDRESS (City, State and ZIP Code) San Diego, California 92152-6800	7a NAME OF MONITORING ORGANIZATION	7b ADDRESS (City, State and ZIP Code)
--	------------------------------------	---------------------------------------

8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research	8b OFFICE SYMBOL <i>(if applicable)</i> Code 200	8c ADDRESS (City, State and ZIP Code) Arlington, VA 22217
---	---	---

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER 62763N	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO RF63.	PROJECT NO 521.804	TASK NO .031	WORK UNIT NO

11 TITLE (Include Security Classification)
OFFICER PERFORMANCE EVALUATION SYSTEMS: LESSONS LEARNED FROM EXPERIENCE

12 PERSONAL AUTHOR(S)
Larson, Gerald E., Rimland, Bernard

13a TYPE OF REPORT Final Report	13b TIME COVERED FROM FY83 TO FY84	14 DATE OF REPORT (Year, Month, Day) 1984 November	15. PAGE COUNT 46
---	---	--	-----------------------------

16 SUPPLEMENTARY NOTATION
Companion report to NPRDC TR 85-7.

17 COSATI CODES	18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Performance appraisal, officer evaluation, personnel management, officer assessment, officer personnel					
<table border="1"> <thead> <tr> <th>FIELD</th> <th>GROUP</th> <th>SUB-GROUP</th> </tr> </thead> <tbody> <tr> <td>05</td> <td>09</td> <td></td> </tr> </tbody> </table>		FIELD	GROUP	SUB-GROUP	05	09
FIELD	GROUP	SUB-GROUP				
05	09					

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Pertinent literature was reviewed and cognizant officers interviewed in an attempt to identify methods to improve the Navy officer performance evaluation system. The system has two major weaknesses: (1) inflation in performance ratings, which diminishes the usefulness of officer evaluations as input to decisions concerning promotion and assignment, and (2) a lack of incentives and procedures for ensuring that senior officers convey timely performance information (advice and feedback) to subordinates. Results of this study and those of a companion study (NPRDC TR 85-7) indicate that (1) problems in military performance appraisal result primarily from attitudinal factors rather than from psychometric issues and (2) two systems are needed--one for assignment counseling and one for performance evaluation. It was recommended that (1) the FITREP form and its procedures be modified to reduce inflation, and (2) an assignment planning conference, scheduled for the beginning of the evaluation year, be used to ensure that the subordinate clearly understands his/her duties and priorities.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
---	--

22a NAME OF RESPONSIBLE INDIVIDUAL Larson, Gerald E.	22b TELEPHONE (Include Area Code) (619) 225-6122	22c OFFICE SYMBOL Code 41
--	--	-------------------------------------

NPRDC TR 85-6

NOVEMBER 1984

**OFFICER PERFORMANCE EVALUATION SYSTEMS:
LESSONS LEARNED FROM EXPERIENCE**

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED



**NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152**



**OFFICER PERFORMANCE EVALUATION SYSTEMS:
LESSONS LEARNED FROM EXPERIENCE**

Gerald E. Larson
Bernard Rimland

Reviewed by
Richard C. Sorenson

Approved by
J. W. Tweeddale

Released by
J. W. Renard
Captain, U.S. Navy
Commanding Officer

FOREWORD

Research at the Navy Personnel Research and Development Center aimed at improving the Navy's officer performance evaluation system was conducted under Exploratory Development task areas Career and Occupational Design (RF63-521-804-031) and Future Technologies for Manpower and Personnel (RF63-521-806).

This report describes results of an intensive review of pertinent literature of the past two decades. A companion report (NPRDC TR 85-7) describes results of an anonymous mail-back survey of over 300 Pacific Fleet officers who were asked to respond to a questionnaire covering various aspects of the performance evaluation system.

J. W. RENARD
Captain, U.S. Navy
Commanding Officer

J. W. TWEEDDALE
Technical Director

SUMMARY

Problem

The Navy's Report on the Fitness of Officers (FITREP) is the major document used for evaluating naval officer performance. The FITREP serves (1) as a record of the senior officer's evaluation of the performance of his/her subordinates and, hence, as a basis for decisions concerning promotion, retention, assignment, and training, and (2) as a focal point and stimulus for the performance counseling of the subordinate officer by his/her reporting senior. The major problem in using the FITREP for evaluating performance is rating inflation; that is, the nearly overwhelming tendency for ratings to be concentrated at the high end of the scale. Although problems with performance counseling are complex, they appear to be primarily due to the interpersonal discomfort associated with such evaluations and a lack of incentives for candor from both parties.

Purpose

The purposes of this project were to (1) identify, for possible Navy use, innovative strategies, procedures, or rating formats that might be useful in curbing inflation in performance ratings, and (2) identify and propose solutions to the obstacles that hinder effective performance feedback.

Approach

Data were obtained by reviewing the pertinent research literature and interviewing fleet officers and cognizant persons in the Naval Military Personnel Command. A companion report describes data obtained by surveying over 300 Pacific Fleet officers through an anonymous questionnaire.

Findings

1. Performance Evaluation Technology. A major purpose of the research was to identify strategies for controlling inflation in performance ratings that, while they might have failed originally, could be resurrected and made effective by use of computer or other recent technology. However, the literature review indicates that such technological "fixes" are still out of reach, and may forever remain so, largely because the basic problem lies not in the realm of technology but, instead, in the reluctance of the officer corps to accept changes that they perceive as inimical to their interests. The major reasons for inflation are considered to be (a) reluctance to impair the motivation of subordinates, (b) the supposition that overall competency in the officer corps may currently be higher than in the past, (c) the opinion that one's own subordinates are better than average, (d) unwillingness to sacrifice a subordinate to the "up or out" system, (e) concern that leniency on the part of other raters will put one's own subordinates at a disadvantage, (f) desire to enhance group cohesion, and (g) recognition that rewards other than promotions are severely limited in a military environment.

2. Performance Appraisal Interview. The performance appraisal interview, like the inflation problem, is beset with technical and "human" problems that are difficult to surmount, or avoid. Among the various approaches toward improving the performance counseling process is the management by objective (MBO) approach, which, if it worked as advertised, would provide an important improvement to performance evaluation as well. MBO provides a systematized procedure for evaluating performance by comparing it with

established goals. The Marine Corps, Coast Guard, and the Army currently employ MBO type methods as part of their appraisal system.

Although MBO may be too rigid for many applications, the concept of goal-setting is readily acknowledged to be important. If the Navy officer performance system is to be improved, some form of performance counseling/mutual goal-setting seems to be necessary. The survey of fleet officers, described in the companion report, provides support for the performance interview concept and helped clarify the optimal context and procedure for encouraging productive superior-subordinate assignment-setting and performance counseling. Strong support was provided for a midyear assignment counseling interview in which the superior and subordinate can clarify the subordinate's understanding of his or her priorities.

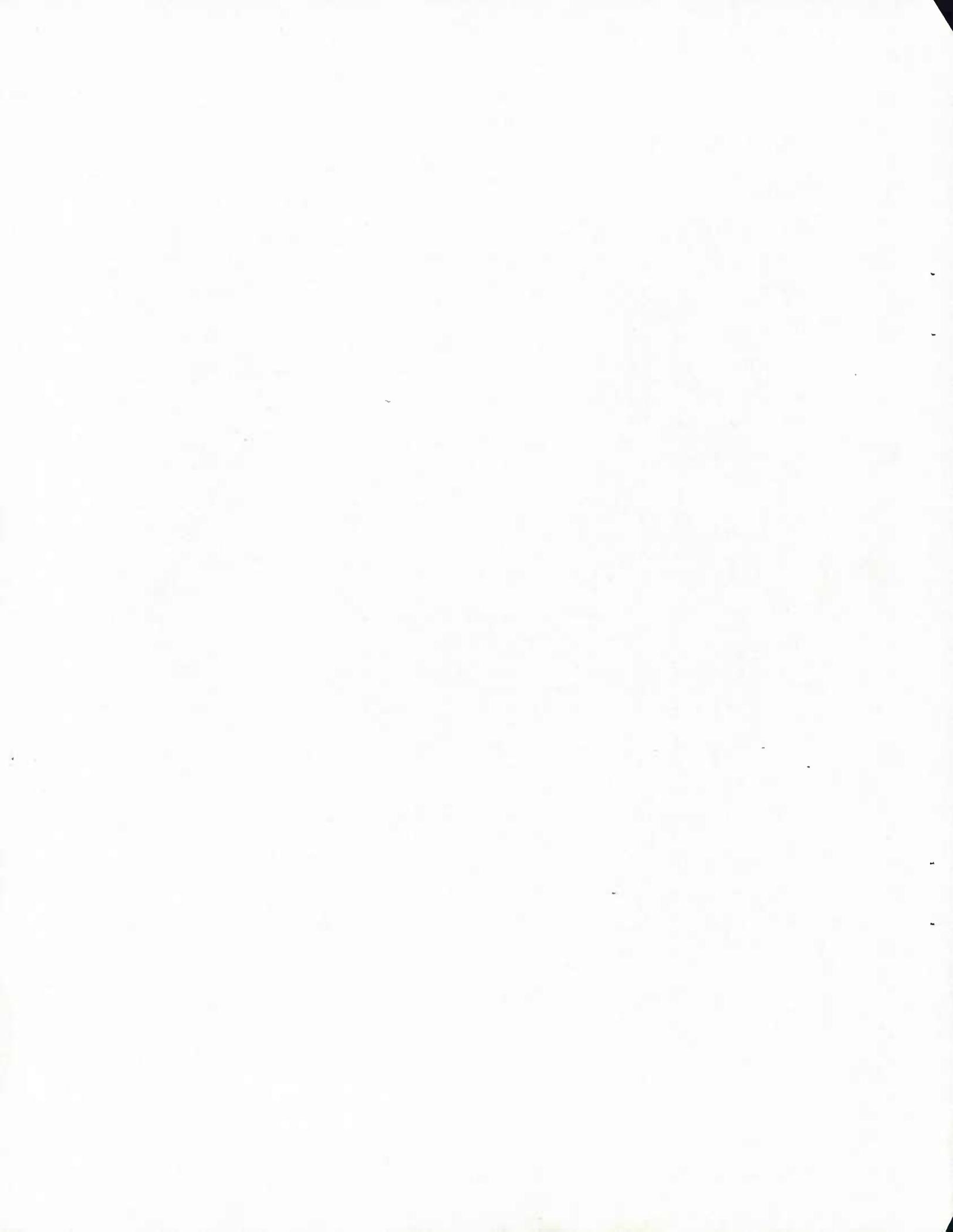
Recommendations

Based on results of the entire project, it is recommended that the Navy's FITREP system be modified as follows:

1. Implement a beginning-of-year assignment conference and midyear assignment and performance review conference between the ratee and the reporting senior, to be held 12 and 6 months prior to the FITREP completion date. These interviews are intended to ensure mutual and clear understanding of the subordinate's duties and priorities. Such circumstances as change of command or reassignment of an officer must be provided for in implementing instructions.
2. Revise the appraisal worksheet by providing expanded definitions of the traits.
3. Revise the current FITREP form by (a) reducing space for the narrative, (b) requiring that the narrative describe specific accomplishments, (c) implementing an "evaluation of potential" section, (d) deleting blocks 53-56 and 77-79 ("trend of performance" and "weaknesses discussed"), and (e) including the "total range of officer value" scale on an experimental basis.
4. Develop rater profiles for the "evaluation of potential" section, with a feedback and enforcement mechanism for dealing with flagrant inflators.
5. Introduce all changes with a significant educational campaign, beginning several months prior to actual system changes.
6. Initiate preliminary research directed toward developing an interactive computer graphic system that would enable selection boards to make on-line inquiries of a data base consisting of all FITREP data for ratees.
7. Make more use of provisions in the recently enacted Defense Officer Personnel Management Act (DOPMA) enabling selective waiver of the "up-or-out" system. These provisions should be broadened to permit a larger range of exceptions to up-or-out. Such policy modifications will become increasingly important as large numbers of officers become involved in narrow but vitally important areas of specialization (e.g., computer technology).

CONTENTS

	Page
INTRODUCTION	1
Problem and Background	1
Purpose	2
APPROACH	4
FINDINGS	4
Inflation of Performance Ratings	4
Causes of Inflation	4
Strategies to Control Inflation	4
Other Issues in FITREP Design	10
Behaviorally Anchored Rating Scales (BARS)	11
The FITREP Narrative	13
Larger Issues	15
Assessment Centers	15
Performance Appraisal Interview	16
The Need for Separate Systems	16
The Need for Incentives	17
Choosing a Format: Management by Objectives	18
DISCUSSION AND CONCLUSIONS	20
Inflation	20
Other Issues in FITREP Design	20
Performance Appraisal Interviews	21
Conclusions	21
RECOMMENDATIONS	23
REFERENCES	25
APPENDIX--EXAMPLES OF PERFORMANCE RATING FORMS	A-0
DISTRIBUTION LIST	



INTRODUCTION

Problem and Background

The Report on the Fitness of Officers (FITREP) (Appendix A, Figure A-1), the principal document used to manage the career of U.S. Navy officers, has two broad but distinct purposes. First, it serves as a record of the senior officer's evaluation of his/her subordinates and, hence, as a basis for decisions affecting the subordinate's future in the Navy (e.g., those involving retention, promotion, training, assignment, and selection for command). Second, it serves as a performance counseling device. The Appraisal Worksheet (Figure A-2), which is used in preparing the FITREP, is intended for use by the reporting senior during the performance appraisal discussion.¹

* There are many problems that limit the FITREP's effectiveness in filling either role. Inflated evaluations have so greatly reduced the spread of performance ratings that their usefulness to selection and promotion boards may be limited. As a result, decisions affecting officers' careers may be based on factors other than performance--certainly undesirable for both the officers and the Navy. The problems with using the FITREP for performance counseling are due to many factors, including system design, a lack of incentives, and what McGregor (1972) attributes to the supervisors' unwillingness to accept the role of "playing God." [133-138]

Difficulties with performance appraisal are neither new nor unique to the Navy. Vintson (1959), after a review of the literature, confirmed the well-known fact that inflation was the most common problem in all military evaluation systems. No rating method in the history of the military services has proved workable in the long term. Indeed, some systems, particularly the forced-choice system used by the Army (1947-1950) and the controlled-rating format used by the Air Force (1974-1978), have been near disasters (Phillips, 1979; Vintson, 1959). Bayless (1981) described the sequence of unsuccessful evaluation systems in the Air Force:

The Air Force, upon its inception in 1947, adopted the newly developed Army form but dropped it by 1949 due to objections to the "Forced Choice Method" and the inflation of the ratings. In 1949, it established its own "Critical Incident Technique." It was dropped in 1951 due to its complexities and mechanical problems. From 1952 to 1974, the Air Force used the same system but made numerous modifications due to inflation continually reducing its effectiveness. Then, in 1974, the OER (officer effectiveness report) continued its evolution by developing a controlled quota system. In 1977, the quota changed again to control only the top block and, by October 1978, the Chief of Staff of the Air Force discontinued all controls, bringing us to the current system.

While the military's problems with performance appraisal are well known, private industry has certainly not been spared its share of difficulties. After a series of lawsuits, employers became increasingly concerned about the legality of their evaluation systems (Kleiman & Durham, 1981). Cascio (1978) has called performance appraisal "The Achilles Heel" of personnel management.

¹NAVMILPERSCOMINST 1611.1. Subject: Report on the fitness of officers, 12 May 1981.

Because of the magnitude of the problem and the fact that an organization's future is at stake when it chooses its leaders, extensive efforts have been made since World War II to improve the performance appraisal process. In 1964, a Navy review covered over 100 reports on military performance ratings (Shears, 1964). The Navy has been relatively inactive in the area of performance appraisal research during the last two decades; however, the Air Force, Army, and Coast Guard have continued to work on the problem, and to revise their evaluation systems.

Table 1 provides a summary of the current officer evaluation methods of the uniformed services.

Purpose

The purpose of this effort was to address two of the most serious problems with the current Navy FITREP system: (1) inflation of performance ratings, and (2) the FITREP's weaknesses as a performance counseling tool. Efforts by the other branches of the armed services and by private industry to solve similar problems were noted. Current and, in some cases, proposed performance evaluation systems were considered to determine whether they (1) adequately discriminate levels of performance, (2) control for inflation, (3) provide constructive, job-related feedback to the ratee in a manner likely to enhance motivation, and (4) provide valid information to administrative users at low cost and in a reasonably simple manner.

Table 1

Summary of Military Appraisal Systems

Feature	Air Force	Army	Coast Guard	Navy	Marine Corps
Most recent revision	1978	1979	1982	1977	1972
Closed/open system	Closes at colonel	Open	Open	Closes at LCDR	Closes if satisfactory
Frequency of appraisal	O-1 & O-2-- semi-annual Others-- annual	Annual	Semiannual, may go to annual	Annual ^a	Semiannual for all-- BGen and below
Counseling function	Informal (as needed)	Joint support form	Minimum-- start and end of 6-month cycles	Counsel from FITREP worksheet	Separate (check box on FITREP)
Number of forms	One for colonel and below	One for MGen and below	6--Ensign through captain	One	One
Number of signatures	3--Rater, additional rater, and indorser	3--Rater, intermediate rater, and senior rater	3--Supervisor, reporting officer, and reviewing officer	1--Rater	2--Reporting senior and reviewing officer
Appraisal instrument	Graphic scales with behavioral anchoring	Number grades + MBO	BARS + MBO	Peer comparison (letter grade)	Peer comparison (letter grade)
Appraisal of potential	Yes	Yes (by senior rater)	Narrative	Implicit	Implicit
Inflation	Ignored	New system has reduced somewhat	Better with new system (six-box spread)	Inflated	Inflated
Rater profile	No	Yes	Yes	No	No
Feedback to rater	Reviewed at base level and major command level	Annual senior rater profile readout	Yes	Rarely	Yes
Receipt for form	No	No	Copy returned with receipt	No	Copy of OCR scores

^a A change to NAVMILPERSCOMINST 1611.1 that would require semiannual appraisals for CWO2 and LTJG is being considered.

APPROACH

Data were obtained by reviewing the pertinent research literature, interviewing fleet officers and cognizant persons in the Naval Military Personnel Command, and surveying over 300 naval officers by means of an anonymous mail-back questionnaire. The complete results of the survey are being published separately (Hearold, Larson, Rimland, & Lahey, 1984).

FINDINGS

Inflation of Performance Ratings

Causes of Inflation

Inflation is the practice of systematically assigning ratings higher than those deserved. It is a common source of error in performance appraisal and can seriously undermine the usefulness of an evaluation system. As Pappageorge (1974) has pointed out, to give up trying to cure inflation and propose the development and use of other indices of promotability misses the heart of the matter. To be fair, and to ensure an organization's efficiency, promotions must be based on indices of performance.

Many authors have addressed the causes of inflation in officer ratings (e.g., Blakelock, 1976; Grappe, Alvord, & Poland, 1967; Olsen & Oakman, 1979; Tate, 1978). The causes that are most commonly identified, most of which are attitudinal in nature, are listed below:

1. Reluctance to impair the motivation of subordinates.
2. The supposition that overall competency in the officer corps may currently be higher than in the past.
3. The opinion that one's own subordinates are better than average.
4. Unwillingness to sacrifice a subordinate to the "up or out" promotion policy.
5. Concern that leniency on the part of other raters will put one's own subordinates at a disadvantage.
6. Desire to enhance group cohesion.
7. Recognition that rewards are severely limited in a military environment.

Strategies to Control Inflation

A truly impressive array of strategies has been devised in the effort to solve the problem of inflation. In most cases, these solutions have not failed because of technical problems but, rather, because of a tendency to underestimate the "human side" of appraisal (i.e., its motivational aspects, effect on self-esteem, and concern for fairness). Rossi, Pappajohn, Penny, Bassham, Bussey, Delandro, Doctor, Druit, Fountain, Horst, McGraw, Mitchell, Sanders, Sands, Cauglin, & Malone (1974) expanded on this concept in their work with the Army. They viewed inflation as an indication of a lack of confidence in the evaluation system. They noted that (1) most efforts have focused on the appraisal

form, thereby creating an overemphasis on psychometric and quasi-psychometric technology, (2) the consistent failure of these efforts is ironic since, given sufficient trust and confidence, virtually any evaluation form would work, and (3) the Army has never made a concerted, purposeful effort to implement a program designed to build confidence in the officer evaluation system.

Some of the strategies to control inflation, and, where appropriate, the presumed reasons for their failure, are reviewed in the remainder of the section.

Forced-choice System (Army). The Army's experiences with the forced-choice system (1947-1950) are of special interest, for they help illustrate the importance of attitudinal factors. This system, which was briefly adopted by the Air Force (where it was referred to as "the forced-choice confusion system of 1947" (Vintson, 1959)), was designed to combat the inflation that had become an increasingly serious problem since the early 1930s. As Vintson observes:

The procedures required by this form were in direct contrast with previous systems. Rather than indicating how much or how little of each characteristic an officer possessed, it required the rater to choose, from a group of four phrases or single adjectives, one that was most like the officer and one that was most unlike him. It required objective reporting and minimized subjective judgment. The arrangement of the rating elements--sets of four--reduced the rater's ability to produce any predetermined desired outcome by the choice of obviously good or obviously bad traits. In effect, it was designed to eliminate favoritism and personal bias.

The forced-choice form (Form 67-1), which was tested on 50,000 officers, was the first Army efficiency report that was extensively validated and standardized before it was officially adopted. The team that developed it considered it to be superior to any other method.

Although Form 67-1 may have been technically the best form the Army ever had, it was also the most unpopular. Rating officers were unable to determine the rating they were giving and, consequently, felt that they could not make fair and accurate judgments (Taylor, 1952). Also, the raters complained that they were being forced to say things they did not want to say, and that no provision was made for showing the results to the rated officer. Because of strong opposition from the Army officer corps, the forced-choice procedure was discontinued in 1950. The Air Force had abandoned the system in 1949.

Controlled-rating Format (Air Force). The Air Force's controlled-rating format, which was in effect from 30 November 1974 to 10 October 1978, is another example of a technically sophisticated system that was abandoned because of negative perceptions. The format contained a controlled "evaluation of potential" section, such that, on a 6-block scale, only 22 percent of the officers rated could receive a rating of "1" and only 50 percent, a rating of "1" or "2."

Interviews with officer personnel quickly revealed widespread negative reactions to this system (Blakelock, 1976; Neary, 1978; Phillips, 1979). The primary complaint was the requirement that 50 percent of the ratees would receive a rating of "3" or lower. This 50 percent had received two blows: The first blow was the withdrawal of positive reinforcement. As Neary (1978) points out, under the previous, inflated system, more than 90 percent of all officers were receiving the highest possible rating on officer

effectiveness reports (OERs). With the implementation of the controlled-rating format in 1974, 40 percent not only lost their opportunity to be included in the top rating (1) but were excluded from even the second highest rating (2).

Second, officers rated "3" or below perceived that such a rating was a "kiss of death," so far as promotions were concerned. This belief prevailed in spite of the fact that the 22/28/50 percent distribution had been chosen to maintain competitiveness for promotion in the block 3 category (Blakelock, 1976). Brown (1977) examined the evidence and concluded that officers with "3" ratings in their OER index could not possibly be excluded from promotion because there simply would not be sufficient numbers of officers with better ratings to fill promotion quotas. Even though the actual data tended to refute, or at least dilute, apprehensions about the effect of a "3" rating on promotion potential, anxiety about receiving such a rating continued (Phillips, 1979). In August 1977, the Air Force responded to these concerns by eliminating controls from all but the top block. A follow-up survey in August 1978 showed that the controlled OER still had a negative impact on ratees' morale, motivation, career plans, and assignments. Thus, on 10 October 1978, all controls were removed from the OER.

Phillips (1979) has analyzed the Air Force's experiences with the controlled rating concept. He emphasized the need for considering both the self-image aspects of evaluation and the danger that an evaluation system may be perceived, or misperceived, as hostile. He summarized the Air Force's experience as follows:

The accomplishments of the controlled OER in halting inflation, renewing the importance of the OER in the selection program and improving feedback to officers, were not enough to overcome the perceived loss of self-esteem and the high level of anxiety felt by many officers during the life of the controlled rating system. This was the case even though many of these perceptions regarding the system were largely invalid.

Ironically, participants in a 1971 Air Force Human Resources Laboratory (AFHRL) workshop, who initiated the original effort to establish a new performance appraisal system, had pointed out, quite prophetically, that an appraisal system can function only to the extent that raters and ratees accept and support it (Jacobcik, 1976).

The obvious point in all this is that, while clever schemes or exercises in psychometric ingenuity may result in short-term apparent solutions to the inflation problem, they may prove detrimental in the long run. By far the most common mistake made in previous efforts by the services has been the tendency to misjudge the response of the officer corps. Any successful course of action must involve a service-wide attempt to build confidence, understanding, and trust of the system. This is especially likely to be necessary for younger officers, who are a product of a changing society in which authority is frequently challenged and in which new avenues of reward are demanded ("New Breed," 1979; Yankelovich, 1979).

Forced-ranking Procedures (Navy). The Navy currently employs limited forced-ranking procedures by requiring the rank-ordering only of ratees nominated for accelerated promotion. There are several reasons for not applying ranking across the entire officer corps. For one, the use of forced ranking without some method of simultaneously controlling for intercommand differences in officer quality is ambiguous and inequitable. Also, a ranking method by itself provides no information regarding the magnitude of differences between ratees (Codron, 1977). The Army attempted such a system in 1968

but abandoned it a year later when it proved unenforceable. Forty percent of the raters found a reason for not completing the rank order portion of the report.

As Blakelock (1976) points out, a comparative ranking system is also a demoralizing experience for 70 to 80 percent of all personnel. Unless carefully implemented, zero-sum methods such as forced ranking can lead to a variety of motivational problems (Russell, 1977).

Forced Distribution. The appeal of forced-distribution schemes is evident from the fact that, despite the Air Force's highly negative experience with this method, it is still recommended for halting inflation by many researchers (e.g., Bayless, 1981; Neary, 1978; Russell, 1977). Although such a strategy provides clear and concise information in a way that greatly facilitates the job of selection boards, it also entails many negative consequences (e.g., Blakelock, 1976; Codron, 1977; Phillips, 1979), some of which are listed below.

1. Like forced ranking, a forced distribution system is likely to be a demoralizing experience for the majority of ratees.
2. Forced distributions can result in inordinate competitiveness, intentional avoidance of difficult assignments and tasks, and decreased effort on the part of officers whose motivation was diminished by being rated below their expectations.
3. As was the case with a "3" rating under the old Air Force system, raters may perceive, although falsely, that receiving certain ratings makes one unpromotable.
4. Organizations or departments with uniformly outstanding personnel may have difficulty creating an acceptable spread of scores. This may result in potential harm to top performers and to the organization, as other top performers will avoid the assignment.
5. When a supervisor is asked to rate individuals at the bottom part of the distribution, he may begin to think of them as ineffective. If the subordinates sense this attitude, it may, in turn, negatively affect their performance.

Many of the problems with forced-distribution strategies arise as a result of having to single out individuals for the bottom part of the distribution. Several authors have voiced the opinion that this may be an avoidable problem. If one makes the reasonable assumption that the poorest performers are readily identifiable, regardless of the rating system, then controls might be necessary only at the upper level of performance. Bayless (1981), Neary (1978), and Russell (1977) have based their proposals on this idea. In essence, the idea is that only the top five percent or so of officers would be permitted to receive the highest rating. No other rating controls would be mandated. Top performers would therefore be recognized and rewarded, while others would not be stigmatized.

The application of such limited controls would provide useful information to selection boards and would, at least on the surface, appear to be relatively nonthreatening to most ratees. Nevertheless, it may not be well accepted. Experience indicates that rigid controls tend to be a source of much dissatisfaction; it may matter little whether that dissatisfaction is based on accurate or false perceptions.

Use of Endorsers and Additional Raters. Endorsement refers to the review, by individuals at the next higher authority level, of performance ratings assigned by

supervisors. Grappe, Alvord, and Poland (1967), in reviewing data from the Air Force, concluded that "a generally consistent finding across all officer grades and all scale levels, except the two categories at the top and one at the bottom, was that the endorser raised the evaluation more often than he lowered it." Bottenberg (1978), in his study of ratings given Air Force lieutenant colonels from 30 November 1974 to 31 March 1975, found nearly identical means for performance factor ratings given by raters, additional raters, and reviewers. Only rarely were performance factor scores assigned by raters overridden. The differences that did appear on the Evaluation of Potential section were apparently the result of mandatory controls applied to reviewers. All in all, in the absence of provisions for control, the use of endorsers and additional raters does not appear to be effective in dampening inflation. Also, since the use of a single rater is rooted in Navy tradition, any change would probably be resisted.

Rater Training. While some authors (e.g., Spool, 1978) have reported that inflation can be reduced by training raters to minimize errors such as those due to leniency and halo, there is little evidence that the result will be more accurate and valid ratings (Bernardin & Buckley, 1981; Zedeck & Cascio, 1982). In fact, there is evidence that such training can actually decrease accuracy in some cases (Bernardin & Pence, 1980), since a wider spread of scores is not necessarily accompanied by greater observational skills on the part of the rater.

Recently, McIntyre, Smith, and Hassett (1984), recognizing the failure of traditional rater training, investigated a different approach. They sought to develop a common frame of reference among raters by having trainees repeatedly view videotapes of job performance, while critiquing, analyzing, and finally, assigning ratings to the performance. Even though the subjects were specifically trained to observe certain behaviors, and the behaviors themselves were viewed under ideal conditions, training improved rating accuracy only minimally.

Anonymous and Confidential Ratings. Landy and Farr (1980), after reviewing three nonmilitary studies, concluded that ratings given by identified raters are equivalent to those given by anonymous raters. Anonymity, of course, is not the same as confidentiality, but these findings are still of interest. For the single study they reviewed that employed confidential ratings, they found that confidentiality did not affect the mean leniency of ratings but did increase the spread of the ratings.

Research within military settings has tended to provide mixed evidence for the effects of confidentiality on inflation. Robins and Seeley (1956) and Seeley (1954), in studies with Army personnel, found that mandatory showing of ratings is accompanied by increased leniency and by a decreased spread of scores, and that these patterns are maintained over time. The Air Force, however, which employed broad-based restrictions on showing evaluations in the 1960s, found that only temporary improvements resulted from a "no show" policy. In September 1962, AF Form 77 for company grade officers was revised, and the new "no show" policy for all grades was introduced (Grappe, Alvord, & Poland, 1967). Ratings were not to be discussed with or shown to the officer being rated at the time of the rating; however, the system was not truly confidential since the ratee was free to review the document after it had been placed on file, as is the case with all military evaluations. There was an initial dip in the percentage of evaluations in the two highest scale intervals; however, by 1964 the average rating levels had approached those existing in 1962. In addition to the lack of solid evidence for the effectiveness of

confidential ratings in controlling inflation, there are several other arguments against their use:

1. Confidential appraisals have little value as performance counseling tools. An obvious way to circumvent this problem is to restrict the no-show provision to a specific part of the evaluation while sharing the remaining information with the ratee. Dunne (1977), for example, has suggested that the Evaluation of Potential section on Air Force evaluations be temporarily "closed" in most instances. The Air Force's experiences with the controlled OER are of interest in this context. One of the problems with having special controls or procedures applied to only one section of an appraisal is that, if the controls work, that particular measure will differentiate between individuals to a greater extent than will other measures. This is, of course, the rationale for the controls. Consequently, its value as a discriminator gives the controlled measure great weight. According to Neary (1978), the result in the Air Force was that promotion boards began to lose sight of the whole man due to this overwhelming emphasis. The gain may or may not be worth the cost.

2. Confidential appraisals fail to deal with several of the root causes of inflation. A distrust of the system and the concern with leniency on the part of other raters are just two examples of probable causes of inflation left unaddressed.

3. Perhaps most important, confidentiality seems likely to evoke a negative response from officer personnel. For example, an AFHRL survey (Johnson, Meehan, & Wilkinson, 1976) found that the majority of respondents were opposed to a confidential evaluation of potential. Also, the majority believed that the proposed closed system would not really be closed. Olsen and Oakman (1979) report that confidential fitness reports appear to have little support among Coast Guard officers. In surveying a sample of naval officers for the effort described herein, Hearold et al. (1984) found similar objections. Such negative perceptions would undermine the effectiveness of a confidential system.

"Rating the Rater." This strategy refers to methods of statistically correcting or adjusting ratings so as to counterbalance the inflationary tendencies of individual raters. Such a strategy has, in one form or another, been advocated by several authors as a control for inflation (e.g., Brown, 1975; Codron, 1977; Bayless, 1981). Brown (1975) suggested computing a bias for each rater by comparing his ratings of his subordinates with all of the past marks received by these same officers. Codron (1977) has proposed a variation of rating the rater which he refers to as "rater standardization." His system has four major components:

1. Noncompulsory but closely monitored ranges of acceptable rating distributions for rater use. These ranges would be determined by personnel managers and could vary with grade.

2. A modification of rating forms to illustrate clearly a particular ratee's standing relative to his fellow officers and his rater's degree of leniency for the current rating cycle.

3. A report that traces the historical rating tendencies of each rater (to be collected, stored, and analyzed but released only to the individual).

4. A procedure for collecting, sorting, and summarizing data from individual evaluator histories for use in adjusting rater standards and reporting trends to selection boards.

Although "rating the rater" strategies appear promising, they are difficult to implement when they are most needed; that is, when performance appraisals are inflated. Since everyone rates highly, there is little information on which to differentiate raters. For example, Tupes and Kaplan (1965) compared ratings given by 1,790 Air Force officers during 1960-1961 to the mean OERs given to the same ratees by their superior officers during 1956-1959. They found that when situational differences, including year and form differences, were removed, only about six percent of the ratings deviated by as much as one OER point from one time to the other. Due to this small difference, they concluded that any systematic attempt to identify deviant raters and correct for their tendencies would not significantly improve the OER rating system.

It would seem, then, that, unless simultaneously implemented with an additional strategy to increase the spread of scores, correction for rater tendencies is unlikely to offset the effects of inflation. However, since a change of performance appraisal forms often results in a temporary decrease in inflation (Grappe, Alvord, & Poland, 1967), it might be useful to derive rater profiles when a new form is introduced. The Army took advantage of a change in forms in June 1979 by implementing DA Form 67-8-2 (Profile Report), which tracks the rating history of senior raters. This information is provided to both selection boards and the raters themselves. According to Bayless (1981), the profile report, supported by feedback to lenient raters, has been quite successful in curbing inflation.

In summary, when introduced in conjunction with a new form and enforced by headquarters, "rate the rater" strategies may help significantly in counteracting the effects of inflation. Of equal importance is the fact that they appear to be acceptable to the officer corps (Hearold et al., 1984).

A few years hence, it is anticipated that selection boards will be able to use interactive computer graphics, with terminals available to the board members during their deliberations, to analyze and display the FITREPs of the entire group of candidates for promotion or of selected subgroups of special interest. Thus, it would be easy to correct for rater leniency and for other confounding factors. It is not too soon to initiate research and development aimed at that goal.

Other Issues in FITREP Design

The strategies for controlling inflation discussed in the previous section represent only some of the issues that must be considered in designing an instrument for performance evaluation. Many authors (e.g., Haynes, 1978; Yager, 1981) have expressed the view that evaluations based on personal traits, as is partly the case with the Navy FITREP, are a major source of difficulty. Haynes emphasizes three main problems with the appraisal of personality factors:

The ambiguity of terms leads to appraisals which are biased by the appraisers subjectivity and are, therefore, usually unreliable and invalid. (For example, in one study which demonstrated the ambiguity of personality traits, definitions of "dependability" were obtained from 150 executives. There were 147 different concepts presented, with as many as six different definitions from one person.)

There is no general agreement as to which personality factors contribute, or to what degree, to an individual's performance.

Partly because they lack behavioral specifics, employees are generally unable to change personality traits, so that including them in an appraisal system leads to antagonism and defensiveness rather than improvement.

Burke (1972), in his discussion of the reasons for the poor performance of appraisal systems, also lists emphasis on personality traits as an important source of difficulty. Beer (1981) concurs, stating that feedback containing details of "what" and "how" is much more likely to be heard and considered than broad generalizations and is much more helpful to individuals who want to improve their performance. A report card type rating of traits is said to be "doomed to failure."

Behaviorally Anchored Rating Scales (BARS)

The trend in management has been away from appraisal based on personality and towards a focus on performance and results. As a consequence, several innovative approaches have been introduced. Of these, it appears that behaviorally-anchored rating scales (BARS), also referred to as behavior expectation scales (BES), have attracted the most attention. BARS were first proposed by Smith and Kendall (1963). According to Schwab, Heneman, and DeCotis (1975), developing BARS for a particular job typically consists of five steps:

1. Critical incidents. Subject matter experts (SMEs) who are familiar with the job describe incidents of effective and ineffective job behavior.
2. Performance dimensions. Collected incidents are clustered into smaller sets of performance dimensions.
3. Retranslation. A second group of SMEs is given the list of critical incidents and dimension definitions and asked to assign each incident to the dimension that it best describes. Incidents not reassigned to the original dimension by the second group of SMEs are eliminated. Typically, an incident is retained if 50 to 80 percent of the second group assigns it to the same dimension as did the first group.
4. Scaling incidents. Generally, the second group of SMEs is also asked to rate the behavior described in the incident. The average rating assigned the incident identifies the degree to which it represents effective performance on the dimension to which it is assigned. Incidents for which there is wide disagreement are excluded from the final instrument.
5. Final instrument. A subset of incidents (usually 6 or 7 per dimension) meeting the above criteria is used to develop behavioral anchors for the performance dimensions. The final BARS instrument usually consists of a series of vertical scales, one for each dimension, anchored by the retained incidents. The incident's location on the scale depends on the rating established in step 4.

Because of the detailed focus of BARS on behavior, some authors (e.g., Murphy, 1980) have recommended that they be included in military performance appraisal systems. The Coast Guard has recently adopted their use, and the Navy has made at least one effort to develop behaviorally-based scales (Borman, Dunnette, & Johnson, 1974). Several recent reviews of the literature, however, have raised questions about the overall value of BARS. The issues that appear most relevant to the current discussion are their cost and their psychometric value as measures of performance.

Kingstrom and Bass (1981) and Schwab et al. (1975), after conducting exhaustive reviews of the literature regarding the psychometric aspects of BARS, concluded that, despite their intuitive appeal, there is little reason to believe that BARS are superior to other evaluation instruments in terms of such important criteria as inflation, halo, spread of performance ratings, reliability, and validity. The same can be said for behavioral ratings in general (Bell, Hoff, & Hoyt, 1963; DeCotis, 1977; Kavanagh, 1971; Stagner, 1977; Massey, Mullins, & Earles, 1978; Borman & Dunnette, 1975). Landy and Farr (1980) report a continuing problem with identifying anchors for the central portions of the scales and a dispute concerning the usefulness of scales outside of the specific setting in which they were developed. Also, the evidence for the purported positive effects of behaviorally-based performance feedback and rater participation in scale construction is mixed at best (Friedman & Cornelius, 1976; Hom, DeNisi, Kinicki, & Bannister, 1982). Further, almost all of the researchers agree that BARS are expensive to produce. Landy and Farr (1980) conclude that "in general, the comparisons of the BARS method with alternative graphic methods make it difficult to justify the increased time investment in the BARS development procedure." At a more theoretical level, behavior ratings are likely to be influenced by the very trait inferences and judgments that they are designed to avoid, since memory for behaviors appears to be structured by general impressions (Murphy, Martin, & Garcia, 1982).

Although the evidence at this point seems to weigh against the use of BARS, several authors criticize such conclusions and still see promise in the method. For example, Bernardin and Smith (1981) maintain that much of the research on BARS published subsequent to the seminal article by Smith and Kendall (1963) has deviated from the original methodology. The Smith and Kendall procedure called for numerous observations of behavior to be made throughout the appraisal period, each individually scaled with the established anchors as a context. A summary rating based on these data was to be made at the end of the rating period. Unfortunately, several discussions of BARS have characterized them as a rating format in which the rater simply reads the dimension definitions and incidents at the end of a rating period and then marks the incident that represents the most "typically expected behavior." Bernardin and Smith feel that at least some of the critical research is thereby invalid. However, the implementation of the original method would thus be much more difficult.

All things considered, BARS are probably not significantly better as performance measures than are the type of graphic rating scales currently in use by the Navy. The present FITREP covers specific aspects of performance, as well as personality traits. Given the present state of appraisal research, there seems to be no compelling reason to change to a new format other than to recover ground lost to inflation. The present system could, however, be improved by providing expanded definitions of the personality traits in the appraisal worksheet. Suggestions are given below:

1. Analytic ability--Quality of thought. Ability to organize and integrate information, deal with problems critically and objectively, establish suitable priorities, and look at both short-range and long-range consequences.
2. Imagination--Resourcefulness, creativeness. Ability to devise alternative solutions to problems.
3. Judgment--Ability to make timely decisions of high quality based on information at hand.

4. Personal behavior--Demeanor, sociability, and public behavior. The extent to which an individual represents the Navy with dignity and sets an example of good conduct for subordinate personnel.

5. Forcefulness--Positive and enthusiastic performance of duty. The extent to which an individual exerts a positive influence on other individuals and on the Navy and shows commitment to action.

6. Military bearing--Smartness of appearance, correctness of uniform, physical fitness, and adherence to weight standards.

Further, the current form can be simplified by eliminating several blocks that have, in practice, little value for selection boards, such as blocks 77-79 ("weaknesses discussed") and 53-56 ("trend of performance").

Although there have been no real breakthroughs in the design of performance rating scales, new formats should be evaluated from time to time in an attempt to improve discrimination. In the survey of officer opinions (Hearold et. al., 1984), the sample of fleet officers was asked to evaluate five performance rating scales and select the one they would most prefer to have on the FITREP form (see Figure 1):

1. Current format (blocks 51 and 52).
2. Total range of officer value format.
3. Distance from average format.
4. Local distribution format.
5. Varying promotion rate format.

Thirty-two percent of the sample selected "the total range of officer value" format (#2 on Figure 1) compared to 29 percent for blocks 51 and 52 of the current FITREP, the second most preferred format. Although endorsement by fleet officers does not guarantee that the new format will help guard against inflation, it might prove helpful. Its inclusion on an experimental basis in the next FITREP is recommended, along with explicit instructions to prevent misuse or misunderstanding.

The FITREP Narrative

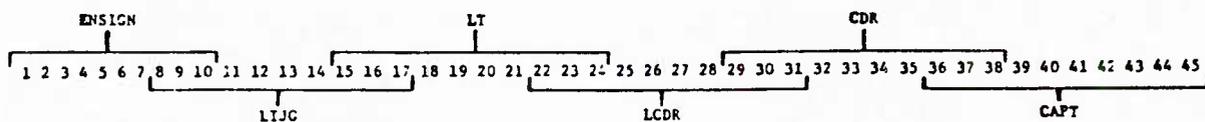
Unfortunately, when FITREP performance blocks fail to distinguish adequately between officers, as is the case when ratings are inflated, selection boards must often rely on "nuances, oddities, and subtleties" (Hearing, 1980). Thus, the FITREP narrative section gains added importance as board members seek information on which to base their decisions. The narratives themselves, however, too often are a reflection of the writer's skill rather than of the ratee's accomplishments. Further, research shows rather low agreement between judges reading such narratives, when the judges are asked for their assessment of relative performance (Coyle & Gorman, 1970).

There are at least two ways of improving the quality of FITREP narratives. The first is by restricting their length to half a page, thus requiring raters to report only the most noteworthy aspects of a ratee's performance. The second is to emphasize the reporting of specific accomplishments, events, or behaviors. These could, in large part, be derived from a list of accomplishments submitted by the ratee as part of the performance counseling process (submission of a list of accomplishments by ratees is discussed further on pages 17 and 21). These two changes would facilitate the task of selection boards by lessening the amount of reading and interpretation required when evaluating performance narratives.

1. Current Format. Shown below are blocks 51 and 52 on the current fitness report form. These two blocks are intended to represent an officer's overall contribution to the Navy. In the EVALUATION block (51), a rater marks a particular subordinate. In the SUMMARY block (52), the rater indicates all the ratings he/she has given to officers of the subordinate's grade. {Selected by 29%}

MISSION CONTRIBUTION	NOT OBS	HIGH				LOW			
		1	2	3	4	1	2	3	4
51 EVALUATION			X						
52 SUMMARY		2	2	1	0	1	0	0	0

2. Total Range of Officer Value Format. The scale running from 1 to 45 below is intended to represent the value of this officer in accomplishing the mission of the Navy, as compared with the other officers in the Navy. A rating outside the designated range for officers of his/her particular rank must be substantiated in writing and evidence cited. (For instance, a rating below 15 or above 24 for a Lieutenant requires substantiation.) Circle the number reflecting your rating of this officer. {Selected by 32%}



3. Distance from Average Format. First indicate with an "O" the box you believe to be appropriate for the average officer of the present officer's grade and length of service. Then place an "X" to indicate the present officer's performance of duty in comparison with the average officer you indicated. {Selected by 20%}

DUTY ASSIGNMENT	NOT OBS OR N/A	Outstanding Performance	Excellent Performance. Frequently Demonstrates Outstanding Performance	Very good Performance. Frequently Demonstrates Excellent Performance	Satisfactory Performance. Basically Qualified	Inadequate Performance. He is not Qualified (Adverse)
(a) Present Assignment						
(b) Shiphandling and Seamanship						
(c) Alismanship						
(d) Collateral Duties						
(e) As _____ Watch Officer						
(f) Technical Specialty (_____)						
(g) Command Potential or Ability						
(h) Administrative and Management Ability						

4. Local Distribution Format {Selected by 9.5%}

Overall Evaluation: (a) In comparison with other officers for his grade and approximate length of service, how would you designate this officer?
 (b) For this report period indicate in (b) how many officers of his grade you have designated in each category of (a).

	Not Observed	One of the Highly Outstanding Officers I Know	A Very Fine Officer of Great Value to the Service	A Dependable and Typically Effective Officer	An Acceptable Officer	Unsatisfactory (Adverse)
(a)						
(b)						

5. Varying Promotion Rate Format. I would promote this officer to the next higher grade if I were on a promotion board meeting next month to select for promotion the following percentage of officers in his/her grade. (Check only the smallest percentage that applies.) {Selected by 9.5%}

() () () () () () () () () ()
 Only Only Only Only Only Only Only Only Only
 1% 5% 10% 15% 20% 25% 40% 60% 80% 100%

Figure 1. Five possible performance rating scales rated by sample of fleet officers.

Larger Issues

Pervading this report is the finding that appraisal research, as well as the "innovative" methods and formats produced, has failed to live up to expectations. In industry, surveys continue to show both a widespread dissatisfaction with and a short life-span for appraisal systems (Teel, 1980; Cohen & Jaffe, 1982). It seems obvious that there are important factors that have been overlooked in most attempts to design new systems. User acceptability of the system is an example of a fundamental factor that has too often been ignored in the military. Several other issues should also be kept firmly in mind.

First, organizational factors can overwhelm any system (Zammuto, London, & Rowland, 1982). As Tenopyr and Oeltjen (1982) point out, too much research has been conducted on rating formats and not enough on the rating context. For example, an "up-or-out" promotion policy, such as that used in the American military, will guarantee inflationary pressure on any appraisal system regardless of the format used. Simply put, the likelihood of obtaining an honest evaluation is reduced if the result is harm to another individual (Kearney, 1978). Research into alternatives to the up-or-out system may, in the long run, be as productive as appraisal research per se. Recently, the Defense Officer Personnel Management Act (DOPMA) provided for limited suspension of "up-or-out." As the Navy finds itself increasingly dependent on the skills of highly specialized officers (e.g., computer technologists) who have traded intensive narrow training for broad experience, an increase in the number of up-or-out waivers may be necessary.

Second, tradition and habit can be severe stumbling blocks when implementing new systems. Behavior change in the desired direction should not be left to chance. Reward mechanisms, educational campaigns, and other strategies are often needed to bring about compliance. Individuals must be motivated if behavior goals are to be achieved (Bolt & Rummler, 1982).

Third, there will always be inaccuracy and subjectivity in performance ratings, no matter what format is used. Borman (1978) has shown that raters disagree significantly, even in a nearly ideal environment for obtaining performance ratings. Further, raters and ratees will differ in their perceptions of the latter's performance. In general, employees tend to have an exaggerated view of their achievements (Meyer, Kay, & French, 1965; Thornton, 1980; Kerr, 1982). Ilgen, Peterson, Martin, and Boesch (1981) reported that, "even when the feedback was very straightforward and presented on a scale with which employees were very familiar, employees still overestimated their own performance." Hearold et al. (1984) found that more than half of their sample of over 300 officers judged themselves as being in the top 10 percent of officers of their rank. On the other hand, supervisors are influenced by such factors as likeability (Thorndike, 1949) and the extent to which the rater perceives his subordinates support him and his goals (Kipnis, 1960). Clarkin (1973) found that the "need to create a good impression" was more strongly related to Navy performance ratings than any other personality factor. There are many potential sources of error and conflict in performance ratings, some of which may be impossible to avoid or correct.

Assessment Centers

Another method of selecting individuals for promotion or special assignment is the use of assessment centers, where candidates are systematically observed and evaluated on their performance of a series of structured tasks or exercises. The participants, who must usually spend several days at the assessment center, are rated on a number of dimensions by trained assessors. Although caution is urged by some authors (e.g., Sackett, 1982;

Klimoski & Strickland, 1977), the evidence seems to indicate that assessment centers are at least as valid as are traditional means of evaluation (Cohen, Moses, & Byham, 1977; Cascio & Silbey, 1979). Possibly because of the number of positive research findings and the fact that the U.S. courts have endorsed the assessment center process as fair in a number of decisions, the method has been adopted at one time or another by over 1000 organizations (Cohen, 1980). There has also been periodic interest in military applications, particularly by the Army (Smith, 1978). There are, nevertheless, considerations that would appear to limit the usefulness of assessment centers.

Codron (1977) concluded that, given the requirement for a wide variety of expertise, extensive facilities, and considerable time and travel expenses, assessment centers would be a prohibitively expensive means of regularly evaluating officers being considered for promotion. Any use of the method would, from a practical standpoint, need to be restricted to relatively small numbers of individuals. The British, German, Australian, and Israeli armed forces, with their smaller numbers of personnel, all employ assessment center technology of one type or another (Farr, 1980; McKenna, 1979). Codron has proposed that an acceptable American variation might be to limit the assessment center approach to officers nominated for accelerated promotion. McKenna (1979) feels that assessment centers may be practical for commanders and captains, particularly for selection to initial or major command. The number of officers involved under such restrictions might be feasible from the cost standpoint. In summary, while assessment centers are not a realistic alternative to the fitness report, they might provide useful information under certain circumstances.

Performance Appraisal Interview

A major component of the performance evaluation process is the performance appraisal interview. Paralleling the perplexing inflation problem in evaluation is an equally intractable problem in supervision: providing meaningful guidance feedback to subordinates (Hood, 1980). The problem appears to have several sources. First, the form typically used in military appraisal tends to focus on the ratee's general characteristics rather than on job-specific behaviors. Second, at present, superiors have few incentives for investing their time and energy in what is usually considered an inherently unpleasant and often counterproductive task. Unfortunately, as ratees, military officers share the common human fault of regarding most criticism as being unwarranted.

The Need for Separate Systems

At present, Navy appraisal interviews are closely tied to the FITREP, both in function and in timing. Nevertheless, most authorities on performance appraisal recommend the use of dual systems--one for counseling and one for evaluation. Sashkin (1981) stated that research at General Electric "demonstrated beyond doubt that a clear separation of the incongruent judge and helper roles led to a more effective appraisal system in terms of employee satisfaction and performance improvement." Beer (1981), Burke (1972), Rilling (1980), Clarkin (1973), and Yager (1981) all stressed the need for separate performance review and promotion processes or systems. The following sections address two major issues that arise at this point if a separate counseling system is advisable: (1) what incentives will help to bring about a meaningful and useful performance counseling interview, and (2) what timing and format will produce optimum results.

The Need for Incentives

As noted earlier, military performance counseling efforts suffer from the same vulnerability that has undermined the various attempts to control inflation: No system can work unless it has the support of the officer corps. Regardless of how ideal a system may appear in theory, it will fail unless senior officers are motivated to invest time and energy in interactions with their subordinates. Unfortunately, reluctance to make this investment appears to be widespread. Beer (1981) has discussed the "vanishing performance appraisal":

In many organizations, supervisors report that they hold periodic appraisal interviews and give honest feedback, while their subordinates report they have not had a performance appraisal for many years or they have heard nothing negative. The appraisals conducted by the supervisor seem to "vanish." What probably happens is that supervisors, fearful of the appraisal process, have talked in very general terms to the subordinates, alluding only vaguely to problems.

This reluctance is understandable. As McGregor (1972) observes, it reflects an unwillingness to play God. Also, it permits supervisors to avoid immediate or unpleasant interpersonal friction, albeit often at the expense of the organization's goals. Separating the performance feedback aspects of appraisal from the evaluation process and making the forms less person-oriented and more job-oriented both seem to be important ways of reducing the aversiveness of the face-to-face interaction. There are, however, at least two other possibilities worth considering.

First, subordinate participation should be emphasized. The Army has overcome reluctance to hold formal appraisal interviews to some extent by making the initiation of the interviews a joint superior and subordinate responsibility.² Mandating this sharing of responsibility makes a meaningful exchange of views and expectations much more likely.

Second, the rater's job should be made easier. Sashkin (1981) identified ten characteristics of effective performance appraisal systems. His first criterion was whether or not managers are rewarded for developing their subordinates. The time and effort invested by a manager in coaching subordinates should directly benefit the manager. Similarly, Burke (1972), in an article titled, "Why performance appraisal systems fail," lists the absence of incentives for employee performance counseling as one reason. According to Burke, "if the organization says employee development is important but does not act accordingly, the manager will only pay lip service to this objective." Under the structure of the current Navy FITREP system, it appears, at least on the surface, that raters have little reason to engage in meaningful performance counseling. One of the widely overlooked benefits of such interactions, however, is that they help to familiarize superiors with the work of their subordinates. At a very minimum, such interviews should greatly facilitate the completion of the fitness report. To maximize this benefit, subordinates should be encouraged to submit, for use during the interview, written input concerning their accomplishments. Some commands already use this concept in the form of locally designed "brag sheets." Such input would be especially useful to the reporting senior when completing the narrative section of the FITREP. The procedure would also allow the ratee an opportunity to provide direct input into the rating process. A survey of naval officers by Clarkin (1973) found that over 80 percent desired more input to the FITREPs submitted on them by their superiors.

²Miller, J. Personal communication, 29 April 1983.

Choosing a Format: Management by Objectives

During the last decade, much interest has centered on the strategy of management by objectives (MBO), in which performance is assessed by comparing it with established goals. In the typical MBO procedure, an employee and his supervisor (1) agree on the employee's performance goals and (2) meet periodically to assess progress towards those goals. If necessary, appraisal criteria are revised from time to time. Frequent counseling, feedback, and supervisor/subordinate interaction, as well as an apparent high level of objectivity, are the key features of MBO systems. As objectives are accomplished, new ones are established. The Marine Corps, the Coast Guard, and the Army currently employ MBO-type methodologies as part of their appraisal systems.

The main advantages and disadvantages of MBO systems have been discussed by various authors (e.g., Bayless, 1981; McKenna, 1979; Beer & Ruh, 1976; Yager, 1981). The purported advantages are listed below:

1. Since they are performance-oriented, rather than trait-oriented, they minimize subjectivity.
2. They elicit commitment from the ratee in addition to providing him or her with feedback.
3. They help establish a strong superior/subordinate relationship.
4. They focus attention on future performance rather than on past failures.
5. They are flexible, nonzero sum systems.
6. They provide well chosen objectives, which can be good motivators.

MBO systems also have several disadvantages:

1. They increase the risk that performance may be viewed in too narrow a context.
2. They may be unrealistic, in that they try to establish accurate objectives a year in advance.
3. Some employees are said to be uncomfortable with setting their own goals.
4. They provide little basis on which to compare one individual with another.
5. They characteristically require large amounts of paperwork and excessive time to implement.

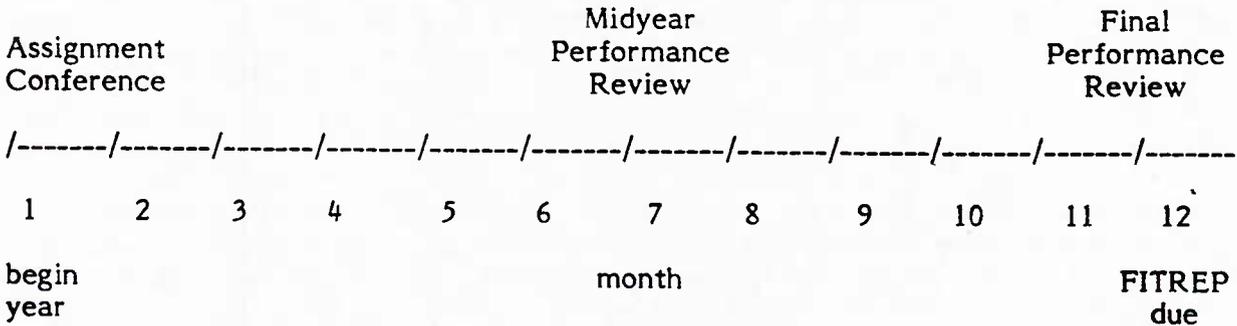
While research has provided strong support for goal setting (Locke, Saari, Shaw, & Latham, 1981), MBO systems per se have not been clearly shown to be especially effective (Levinson, 1970; Sokolik, 1978; Aplin, Schoderbek, & Schoderbek, 1979; Ford & McLaughlin, 1982). After analyzing 185 studies, Kondrasuk (1981) concluded that research support for MBO was inversely related to the rigor of the research. While all the case studies reviewed were favorable, actual experiments provided mixed support at best. He concluded that, although MBO can be effective, questions remain about the circumstances in which it is likely to succeed.

Perhaps most relevant to our interests are the experiences of the other services with MBO. Murphy (1980) concluded that MBO-based performance counseling goals frequently are not met in the Marine Corps, primarily because officers are generally reluctant to counsel their subordinates. Unless such reluctance can be overcome, it is doubtful that any system will succeed.

The popularity of MBO, despite its disadvantages, suggests that it is perceived as filling a need for structure and explicitness in supervisory relations, albeit MBO fills the need poorly. The present authors feel that a significant amount of supervisory dissatisfaction with subordinate performance may result from subordinates not adequately understanding their duties and priorities as perceived by the supervisor. MBO provides a formalized--perhaps too formalized--attempt to avert such misunderstandings. A highly simplified version could provide many of the same benefits.

To determine the need for enhanced supervisor-subordinate agreement on duties and priorities, the sample of 300 naval officers surveyed (Hearold et al. 1984) was asked whether they felt the counseling process should include a formal procedure for clarifying the exact nature and priorities of a subordinate's duties. Over 80 percent responded affirmatively (definitely yes, 50%; probably yes, 32%). A follow-up question was asked concerning the scheduling of both the proposed assignment conference and a proposed performance review:

"Assume that formal discussions of assignments (so that both rater and ratee understand explicitly what is expected of the ratee) and periodic reviews of an officer's performance are to be conducted on one or more occasions during each fitness report cycle. On the timeline below, put an "A" where you think the formal assignment conference(s) should be scheduled, and a "P" where you think the formal performance review(s) should take place."



A clear majority (62%) of the respondents indicated that the assignment conference should be scheduled 12 months prior to the FITREP due date. A substantial plurality (42%) preferred that the formal performance review be scheduled 6 months prior to the FITREP submission. While such circumstances as change of command or reassignment of an officer may require some rescheduling, this could be easily provided for in implementing instructions.

DISCUSSION AND CONCLUSIONS

Inflation

One purpose of this review has been to determine why so many past attempts to halt inflation have failed. A prime source of difficulty has been negative reactions by the officer corps to the introduction of rating formats that they perceived as being ambiguous, unjust, and, in some cases, threatening. Although efforts to design and implement appraisal programs have been uniformly well-intentioned, there has been a consistent failure to give adequate consideration to the "human side" of appraisal.

Of the strategies for controlling inflation reviewed, "rating the rater" (rater profiles) appears to have the most merit. This method has recently been implemented by the Army and Coast Guard with at least temporary success. Its introduction is best carried out in conjunction with other system revisions, such as new forms and procedures, to take advantage of the reduced inflation.

Several additional points should be considered with respect to rater profiles. Due to the Privacy and Freedom of Information Acts, the raters themselves could request information on rating tendencies, once it becomes a matter of official record. If the profiles facilitated comparisons between raters, one likely outcome would be a sudden rating increase for "low" raters. High raters, being less conspicuous, would be less inclined to change their standards. The problem of inflation might therefore be compounded, at least initially.

The type of profile used by the Army provides for no interrater comparisons (Figure 2). Since the profile displays both the total distribution of marks given by an individual rater and the rating he has given to a particular subordinate, a rating can be viewed in the context of the rater's general tendencies. To be successful, such a system might require that feedback be given to serious inflaters. Marking a reasonable distribution of scores could itself be used as a performance factor by selection boards when a rater's own turn to be evaluated arose. Other enforcement mechanisms, such as letters of reprimand, could also be employed to correct reporting seniors who overrate seriously and consistently.

A further issue is the question of which block or blocks should be involved in the profile. For the Army, the "evaluation of potential" block is involved. The Navy has no such rating factor. Getting a useful spread of ratings should be easier on a new factor than on one contaminated by previous rating tendencies. An "evaluation of potential" block, introduced on a new form, would facilitate the successful development of rater profiles in the Navy.

Other Issues in FITREP Design

As Landy and Farr (1980) state, "After more than 30 years of serious research, it seems that little progress has been made in developing an efficient and psychometrically sound alternative to the traditional graphic rating scale." The type of appraisal form currently used in the Navy, with its combination of traits, behaviors, and narrative, has not been significantly improved upon, despite the theoretical appeal of proposed alternatives such as BARS. Nevertheless, the current form can be improved by (1) including better definitions in the appraisal worksheet and (2) shortening the narrative section, with emphasis on specific accomplishments. In addition, the new "total range of officer value" format should be adopted, on a trial basis, to evaluate its effects on inflation.

Performance Appraisal Interviews

The performance appraisal literature almost uniformly agrees that performance counseling and evaluation should be separate processes. Ideally, these processes should be supported by separate documents. The Army and the Coast Guard have recently adopted this strategy. However, there is less agreement as to the performance counseling document itself and the way it is used.

As noted earlier, there is abundant evidence that MBO has not lived up to initial expectations. Rather than employing MBO, it seems desirable that the Navy implement a relatively simple Assignment Conference Form (see Figure A-3). In completing this form, the superior and subordinate should reach a clear understanding as to the subordinate's duties and priorities. The first page of the proposed form requires a listing of duties in order of priority. (Understanding between superiors and subordinates on job duties can positively influence the perceived fairness and accuracy of performance evaluation (Landy, Barnes, & Murphy, 1978)). This section, which should be completed at the beginning of the rating period, allows the rater and ratee to exchange views and expectations. The conference and the completion of the form would be both a mandated and a shared responsibility. A midyear performance review should be held to assess progress, discuss ways to improve performance, and, if necessary, revise duties. Near the end of the rating period, the subordinate would complete and submit the second part of the form, which lists the specific accomplishments achieved in the context of his duties. The subordinate and supervisor then have a final meeting. Soon afterward, when the reporting senior completes a FITREP on the officer, he will have at his disposal a document describing specific accomplishments on which to base his ratings. When he actually submits the FITREP, he will also send the prioritized list of the officer's duties to personnel headquarters, where it will provide valuable input to selection boards.

Conclusions

Performance appraisal is an area fraught with many problems--for the raters, the ratees, and the administrative users. Given that there will never be a perfect system, the two most important conclusions to emerge from this study are:

1. The problems in military performance appraisal result primarily from attitudinal factors, rather than from psychometric issues. It has been found repeatedly that acceptance by the officer corps is essential for the success of a performance appraisal system.
2. Reliance on a single instrument and occasion for both performance evaluation and performance counseling purposes is ill-advised.

RECOMMENDATIONS

Based on results of this effort and those of Hearold et al. (1984), it is recommended that the Navy's FITREP system be modified as follows:

1. Implement a beginning-of-year assignment conference and midyear assignment review conference between the ratee and the reporting senior, to be held 12 and 6 months prior to the FITREP completion date. These interviews are intended to ensure mutual and clear understanding of the subordinate's duties and priorities. A proposed assignment conference worksheet has been designed to facilitate and document these meetings (Figure A-3).
2. Revise the appraisal worksheet by providing expanded definitions of the traits.
3. Revise the current FITREP form by (a) reducing the amount of space for the narrative, (b) requiring that the narrative describe specific accomplishments, (c) implementing an "evaluation of potential" section, (d) deleting blocks 53-56 and 77-79 ("trend of performance" and "weaknesses discussed"), and (e) including the "total range of officer value" scale on an experimental basis.
4. Develop rater profiles for the "evaluation of potential" section, with a feedback and enforcement mechanism for dealing with flagrant inflators.
5. Introduce all changes with a significant educational campaign, beginning several months prior to actual system changes.
6. Initiate preliminary research directed toward developing an interactive computer graphics system that would enable selection boards to make on-line inquiries of a data base consisting of all FITREP data for ratees.
7. Make more use of provisions in the recently enacted DOPMA enabling selective waiver of the "up-or-out" system. These provisions should be broadened to permit a larger range of exceptions to up-or-out. Such policy modifications will become increasingly important as larger numbers of officers become involved in narrow but vitally important areas of specialization (e.g., computer technology).

REFERENCES

- Aplin, J. C., Schoderbek, C. G., & Schoderbek, P. P. Tough minded management by objectives. Human Resource Management, Summer 1979, 9-13.
- Bayless, R. F. Developing an officer appraisal system to meet the needs of the 80s (Rep. No. 0175-81). Maxwell Air Force Base, AL: Air Command and Staff College, Air University, May 1981.
- Beer, M. Performance appraisal: Dilemmas and possibilities. Organizational Dynamics, Winter 1981, 24-36.
- Beer, M., & Ruh, R. A. Employee growth through performance management. Harvard Business Review, July-August 1976, 59-66.
- Bell, F. O., Hoff, A. L., & Hoyt, K. B. A comparison of three approaches to criterion measurement. Journal of Applied Psychology, 1963, 47(6), 416-418.
- Bernardin, H. J., & Buckley, M. R. Strategies in rater training. Academy of Management Review, 1981, 6, 205-212.
- Bernardin, H. J., & Pence, E. C. Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 1980, 65, 60-66.
- Bernardin, H. J., & Smith, P. C. A clarification of some issues regarding the development and use of behaviorally-anchored rating scales (BARS). Journal of Applied Psychology, 1981, 66(4), 458-463.
- Blakelock, R. A. An analysis of the impact of the new OER system on the officer corps using a Lewin-based model (Rep. No. GSM/SM/76D-27). Wright-Patterson Air Force Base, OH: Air Force Institute of Technology, December 1976.
- Bolt, J. F., & Rummler, G. A. How to close the human performance gap. Management Review, January 1982, 38-44.
- Borman, W. C. Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 1978, 63(2), 135-144.
- Borman, W. C., & Dunnette, M. D. Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60(5), 561-565.
- Borman, W. C., Dunnette, M. D., & Johnson, P. D. The development and evaluation of a behaviorally-based naval officer performance assessment package (final report). Minneapolis, MN: Personnel Decisions Inc., July 1974.
- Bottenberg, R. A. Relationships among factors in new officer effectiveness report system (AFHRL-TR-78-40). Brooks Air Force Base, TX: Computational Sciences Division, Air Force Human Resources Laboratory, August 1978.
- Brown, W. G. A prediction for OER rating spread (Rep. No. 0335-77). Maxwell Air Force Base, AL: Air Command and Staff College, Air University, April 1977.

- Brown, W. T. OER inflation, quotas, and rating the rater (Rep. No. 5522). Maxwell Air Force Base, AL: Air War College, Air University, April 1975.
- Burke, R. J. Why performance appraisal systems fail. Personnel Administration, May-June 1972, 32-40.
- Cascio, W. F. Applied psychology in personnel management. Reston, VA: Reston Publishing Co., Inc., 1978.
- Cascio, W. F., & Silbey, V. Utility of the assessment center as a selection device. Journal of Applied Psychology, 1979, 64(2), 107-118.
- Clarkin, J. J. An investigation of a performance evaluation process (Doctoral dissertation). Harvard University, 1973.
- Codron, B. J. The USAF OER system revisited: Modifying the controls (Rep. No. 0485-77). Maxwell Air Force Base, AL: Air Command and Staff College, Air University, May 1977.
- Cohen, B. M., Moses, J. L., & Byham, W. C. The validity of assessment centers (DDI Monograph 2). Pittsburgh, PA: Developmental Dimensions, Inc., 1977.
- Cohen, S. L. Validity and assessment center technology: One and the same? Human Resource Management, Winter 1980, 2-11.
- Cohen, S. L., & Jaffee, C. L. Managing human performance for productivity. Training and Development Journal, December 1982, 94-100.
- Coyle, H. S., & Gorman, C. D. The utility of OER word pictures as discriminators (AFHRL TR 70-23). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, July 1970.
- DeCotis, T. A. An analysis of the external validity and applied relevance of three rating formats. Organizational Behavior and Human Performance, 1977, 19, 247-266.
- Dunne, E. J. A reasoned approach to officer evaluation (TR 77-7). Wright-Patterson Air Force Base, OH: Air University, Air Force Institute of Technology, August 1977.
- Farr, M. J. 16th International Symposium on Applied Military Psychology. London, England: Office of Naval Research, London Branch, December 1980.
- Ford, R. C., & McLaughlin, F. S. Avoiding disappointment in MBO programs. Human Resource Management, Summer 1982, 44-49.
- Friedman, B. A., & Cornelius, E. T. Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. Journal of Applied Psychology, 1976, 61(2), 210-216.
- Grappe, L. B., Alvord, R. W., & Poland, J. V. Air Force officer performance evaluation: Rating trends and relationships from 1954 through 1965 (PRL-TR-67-12). Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, October 1967.

- Haynes, M. G. Developing an appraisal program, Parts I and II. Personnel Journal, January 1978, 14-19, and February 1978, 66-107.
- Hearing, G. Fitness report finesse. Naval Institute Proceedings, 1980, 106(1), 34-38.
- Hearold, S., Larson, G. E., Rimland, B., & Lahey, R. A. Officer performance evaluation systems: Officer survey (NPRDC Tech. Rep. 85-7). San Diego: Navy Personnel Research and Development Center, November 1984.
- Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. Effectiveness of performance feedback from behaviorally-anchored rating scales. Journal of Applied Psychology, 1982, 67(5), 568-576.
- Hood, W. T. Performance appraisal as a military development aid. U.S. Naval Institute Proceedings, December 1980, 104-105.
- Ilgen, D. R., Peterson, R. B., Martin, B. A., & Boeschen, D. A. Supervisor and subordinate reactions to performance appraisal sessions. Organizational Behavior and Human Performance, 1981, 28, 311-330.
- Jacobcik, J. D. The new OER system and behavior (Rep. No. 1410-76). Maxwell Air Force Base, AL: Air Command and Staff College, Air University, May 1976.
- Johnson, C. A., Meehan, J., & Wilkinson, R. E. Officer effectiveness report development--1971 through 1972 (AFHRL-TR-76-61). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, September 1976.
- Kavanagh, M. J. The content issue in performance appraisal: A review. Personnel Psychology, 1971, 24, 653-668.
- Kearney, W. J. Improving work performance through appraisal. Human Resource Management, Summer 1978, 15-23.
- Kerr, S. Substitutes for leadership. Talk given at the Navy Personnel Research and Development Center, San Diego, CA, 14 July 1982.
- Kingstrom, P. O., & Bass, A. R. A critical analysis of studies comparing behaviorally-anchored rating scales (BARS) and other rating formats. Personnel Psychology, 1981, 34, 263-289.
- Kipnis, D. Some determinants of supervisory esteem. Personnel Psychology, 1960, 13, 377-391.
- Kleiman, L. S., & Durham, R. L. Performance appraisal, promotions, and the courts: A critical review. Personnel Psychology, 1981, 34, 103-121.
- Klimoski, R. J., & Strickland, W. J. Assessment centers: Valid or merely prescient? Personnel Psychology, 1977, 30, 353-361.
- Kondrasuk, J. N. Studies in MBO effectiveness. Academy of Management Review, 1981, 6(3), 419-30.

- Landy, F. J., Barnes, J. L., & Murphy, K. R. Correlates of perceived fairness and accuracy of performance evaluation. Journal of Applied Psychology, 1978, 63(6), 751-754.
- Landy, F. J., & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87(1), 72-107.
- Levinson, H. Management by whose objectives? Harvard Business Review, July-August 1970, 125-134.
- Locke, E. A., Saari, L. M., Shaw, K. N., & Latham, G. P. Goal setting and task performance. Psychological Bulletin, 1981, 90(1), 125-152.
- Massey, R. H., Mullins, C. J., & Earles, J. A. Performance appraisal ratings: The content issue (AFHRL TR 78-69). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1978.
- McGregor, D. An uneasy look at performance appraisal. Harvard Business Review, September-October 1972, 133-138.
- McIntyre, R. M., Smith, D. E., & Hassett, C. E. Accuracy of performance ratings as affected by rater training and perceived purposes of rating. Journal of Applied Psychology, 1984, 69, 147-156.
- McKenna, R. B. Alternative approaches to the U.S. Navy officer evaluation system. Monterey, CA: Naval Postgraduate School, March 1979.
- Meyer, H. H., Kay, E., & French, J. R. P. Split roles in performance appraisal. Harvard Business Review, January-February 1965, 123-129.
- Murphy, J. W. Use of behaviorally-anchored rating scales (BARS) to complement the management by objectives (MBO) and fitness report components of the Marine Corps performance evaluation system. Fort Leavenworth, KA: U.S. Army Command and General Staff College, June 1980.
- Murphy, K. R., Martin, C., & Garcia, M. Do behavioral observation scales measure observation? Journal of Applied Psychology, 1982, 67(5), 562-567.
- Neary, T. H. USAF OERs in the 1980s: A question of direction. Maxwell Air Force Base, AL: Air Command and Staff College, Air University, May 1978.
- New breed of workers. U.S. News and World Report, 3 September 1979, 35-38.
- Olsen, R. C., & Oakman, J. C. Officer performance appraisal in the Coast Guard: An analysis of the fitness reporting system. Monterey, CA: Naval Postgraduate School, December 1979.
- Pappageorge, J. G. Evaluation, inflation, and order of merit (MIRM 74-2). Carlisle Barracks, PA: Army War College Strategic Studies Institute, March 1974.
- Phillips, T. D. Evolution of the Air Force officer evaluation system: 1968-1978 (Rep. No. 1855-79). Maxwell Air Force Base, AL: Air Command and Staff College, Air University, April 1979.

- Rilling, D. S. Personnel performance appraisal--a need for change. Marine Corps Gazette, April 1980, 64, 48-52.
- Robins, A. R., & Seeley, L. C. Effects of mandatory showing of ratings to rated officers: Phase II--First Lieutenants (Research memo No. 56-21). The Adjutant General's Office, Department of the Army: Personnel Research Branch, Personnel Research and Procedures Division, July 1956.
- Rossi, L. A., Pappajohn, C., Penny, H. F., Bassham, A. F., Bussey, C. D., Delandro, D. J., Doctor, H., Druit, C. A., Fountain, C. D., Horst, T. G., McGraw, R. M., Mitchell, F. H., Sanders, M. D., Sands, T. A., Cauglin, T. M., & Malone, D. M. Executive appraisal (confidence in the officer evaluation system). Carlisle Barracks, PA: U.S. Army War College, May 1974.
- Russell, J. F. The new USAF officer evaluation system: Problems and recommended solution (Rep. No. 204). Maxwell Air Force Base, AL: Air War College, Air University, April 1977.
- Sackett, P. R. A critical look at some common beliefs about assessment centers. Public Personnel Management Journal, 1982, 11, 140-147.
- Sashkin, M. Appraising appraisal: Ten lessons from research for practice. Organizational Dynamics, Winter 1981, 9, 37-50.
- Schwab, D. P., Heneman, H. G., & DeCotis, T. A. Behaviorally-anchored rating scales: A review of the literature. Personnel Psychology, 1975, 28, 549-562.
- Seeley, L. C. Effects of mandatory showing of efficiency ratings (Research memo No. 54-43). The Adjutant General's Office, Department of the Army: Personnel Research Branch, Personnel Research and Procedures Division, December 1954.
- Shears, L. Performance rating: A data retrieval bibliography (PRASD Rep. No. 261). San Diego: U.S. Naval Personnel Research Activity, December 1964.
- Smith, K. H. Evaluation of leadership at the Fort Benning assessment center (TR-78-B11). Alexandria, VA: Army Research Institute, November 1978.
- Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.
- Sokolik, S. L. Feedback and control--the hollow in MBO practice. Human Resource Management, Winter 1978, 23-28.
- Spool, M. D. Training programs for observers of behavior: A review. Personnel Psychology, 1978, 31, 853-888.
- Stagner, R. On the reality and relevance of traits. The Journal of General Psychology, 1977, 96, 185-207.
- Tate, G. W. Behavioral effects of officer evaluation in the U.S. Army. Carlisle Barracks, PA: U.S. Army War College, May 1978.

- Taylor, E. K. Accuracy and manipulatability of graphic and forced-choice rating forms (Research Note Number 52-59). The Adjutant General's Office, Department of the Army: Personnel Research Branch, Personnel Research and Procedures Division, August 1952.
- Teel, K. S. Performance appraisal: Current trends, persistent progress. Personnel Journal, April 1980, 296-316.
- Tenopyr, M. L., & Oeltjen, P. D. Personnel selection and classification. Annual Review of Psychology, 1982, 33, 581-618.
- Thorndike, R. L. Personnel Selection. New York: John Wiley and Sons, 1949, 151.
- Thornton, G. C. Psychometric properties of self-appraisals of job performance. Personnel Psychology, 1980, 33, 263-271.
- Tupes, E. C., & Kaplan, M. N. A preliminary investigation of rater differences in officer effectiveness reports (PRL-TR-65-13). Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, July 1965.
- Vintson, S. P. Merit ratings of officers in the armed forces. (Thesis) University of Texas, August 1959.
- Yager, E. A critique of performance appraisal systems. Personnel Journal, February 1981, 129-133.
- Yankelovich, D. We need new motivational tools. Industry Week, 6 August 1979, 61-68.
- Zammuto, R. F., London, M., & Rowland, K. M. Organizational and rater differences in performance appraisal. Personnel Psychology, 1982, 35, 643-658.
- Zedeck, S., & Cascio, W. F. Performance appraisal decisions as a function of rater training and purpose of appraisal. Journal of Applied Psychology, 1982, 67, 752-758.

APPENDIX

EXAMPLES OF PERFORMANCE RATING FORMS

	Page
Current Report on the Fitness of Officers	A-1
Current Appraisal Work Sheet	A-3
Proposed Assignment Conference Form	A-7

REPORT ON THE FITNESS OF OFFICERS

1. NAME (LAST, FIRST, MIDDLE)		2. GRADE		3. DESIG.		4. SSN		
5. ACUTRA/TEMAC <input type="checkbox"/>		6. UIC		7. SHIP/STATION		8. DATE REPORTED		
OCCASION FOR REPORT				PERIOD OF REPORT				
9. PERIODIC <input type="checkbox"/>		10. DETACHMENT OF REPORTING SENIOR <input type="checkbox"/>		11. DETACHMENT OF OFFICER <input type="checkbox"/>		12. FROM: _____		
13. TO: _____		BASIS FOR OBSERVATION						
14. REGULAR <input type="checkbox"/>		15. CONCURRENT <input type="checkbox"/>		16. SPECIAL <input type="checkbox"/>		17. OPS CDR <input type="checkbox"/>		
18. CLOSE <input type="checkbox"/>		19. FREQUENT <input type="checkbox"/>		20. INFREQUENT <input type="checkbox"/>				
21. EMPLOYMENT OF COMMAND (CONTINUED ON REVERSE SIDE OF RECORD COPY)							22. DAYS OF COMBAT	
23. REPORTING SENIOR (LAST NAME, FI, MI)			24. TITLE		25. GRADE		26. DESIG.	
						27. SSN		
28. DUTIES ASSIGNED (CONTINUED ON REVERSE SIDE OF RECORD COPY)								
SPECIFIC ASPECTS OF PERFORMANCE (TYPE IN OCR CODE LETTER FROM WORK SHEET)								
29. GOAL SETTING & ACHIEVEMENT		30. SUBORDINATE MANAGEMENT & DEVELOPMENT		31. WORKING RELATIONS		32. EQUIP & MATERIAL MANAGE.		
						33. NAVY ORGAN. SUPPORT		
34. RESPONSE IN STRESSFUL SITUATIONS		35. EQUAL OPPORTUNITY		36. SPEAKING ABILITY		37. WRITING ABILITY		
WARFARE SPECIALTY SKILLS (FROM OCR WORK SHEET)								
38. SEAMANSHIP		39. AIRMANSHIP		40. WATCH STANDING		41.		
						42.		
44. SUBSPECIALTY CODE		REQUIRED BY BILLET		UTILIZATION		(WORK SHEET CODE)		
		45. YES <input type="checkbox"/>		46. NO <input type="checkbox"/>		47. FREQUENT <input type="checkbox"/>		
						48. INFREQUENT <input type="checkbox"/>		
						49. NONE <input type="checkbox"/>		
						50. PERFORMANCE		
MISSION CONTRIBUTION		NOT OBS		HIGH		MID		
						LOW		
						MARG. UNSAT*		
51. EVALUATION								
52. SUMMARY								
TEND OF PERFORMANCE								
53. FIRST REPORT <input type="checkbox"/>		54. CONSISTENT <input type="checkbox"/>		55. IMPROVING <input type="checkbox"/>		56. DECLINING* <input type="checkbox"/>		
DESIRABILITY (TYPE IN OCR CODE FROM WORK SHEET)								
57. COMMAND		58. OPERATIONAL		59. STAFF		60. JOINT/OSD		
						61. FOREIGN SHORE		
RECOMMENDATION FOR PROMOTION				RANKING FOR EARLY PROMOTION				
62. EARLY <input type="checkbox"/>		63. REGULAR <input type="checkbox"/>		64. NO* <input type="checkbox"/>		65. NUMBER RECOMMENDED		
						66. RANKING		
PERSONAL TRAITS (TYPE IN OCR CODE FROM WORK SHEET)								
67. JUDGMENT		68. IMAGINATION		69. ANALYTIC ABILITY		70. PERSONAL BEHAVIOR		
						71. FORCEFULNESS		
						72. MILITARY BEARING		
73		74		75		76		
WEAKNESSES DISCUSSED?								
77. NONE NOTED <input type="checkbox"/>		78. YES <input type="checkbox"/>		79. NO* <input type="checkbox"/>		80. NOT DESIRED		
						81. ATTACHED		
82. SIGNATURE OF OFFICER EVALUATED: (IAW BUPERS INST. 1611-12-SERIES). "I ACKNOWLEDGE THAT I HAVE SEEN THIS REPORT, HAVE BEEN APPRISED OF MY PERFORMANCE AND RIGHT TO MAKE A STATEMENT."								
85. SIGNATURE OF REPORTING SENIOR								
83. DATE FORWARDED		84.						
87. SIGNATURE OF REGULAR REPORTING SENIOR ON CONCURRENT AND CONCURRENT/SPECIAL REPORT								
86. DATE FORWARDED								

Figure A-1. Report on the Fitness of Officers

21. EMPLOYMENT OF COMMAND (Continued)

26. DUTIES ASSIGNED (Continued)

28. COMMENTS. Particularly comment upon the officer's overall leadership ability, personal traits not listed on the reverse side, and estimated or actual performance in combat. Include comments pertaining to unique skills and distinctions that may be important to career development and future assignment. A mark in boxes with an asterisk (*) indicates adversity and supporting comments are required.

1. NAME (Last, First, Middle)		2. GRADE	3. DESIG.	4. SSN
5. ADDRESSES	6. UNIT ID CODE	7. SHIP OR STATION		8. DATE REPORTED
OCCASION OF REPORT		PERIOD OF REPORT		
9. <input type="checkbox"/> PERIODIC	10. <input type="checkbox"/> DETACHMENT OF REPORTING SENIOR	11. <input type="checkbox"/> DETACHMENT OF OFFICER	12. FROM:	13. TO:
TYPE OF REPORT		BASES FOR OBSERVATION		
14. <input type="checkbox"/> REGULAR	15. <input type="checkbox"/> CONCURRENT	16. <input type="checkbox"/> SPECIAL	17. <input type="checkbox"/> OPS ODR	18. <input type="checkbox"/> GLESE
				19. <input type="checkbox"/> PROMOTED
21. EMPLOYMENT OF COMMAND (Summarize major activities of command during this reporting period; do not use code names, unusual or obscure abbreviations.)				22. DAYS OF COMBAT
23. REPORTING SENIOR		24. TITLE	25. GRADE	26. DESIG.
		27. SSN		

GENERAL INSTRUCTIONS

This Appraisal Work Sheet is designed to serve two purposes: First to assist in preparation of the OCR Sheet; secondly, and second to provide guidelines for the performance appraisal discussion.

All conditions made in this report shall be in comparison with officers of the same grade, competitive category (i.e., unrestricted line, supply corps with supply corps, etc.), and approximately the same grade when you have been.

Due to space limitations of the OCR Sheet, it is necessary to describe evaluation marks assigned in the various ranges, but corresponding OCR code letters for certain items. To prevent transcription error, an OCR code letter has been provided for each of these items to facilitate transfer of the information to the OCR Sheet.

General comments are required in item 21. Any mark in boxes with an asterisk (*) indicates authority and supporting comments are required in item 26. The officer receiving adverse marks must be informed of such and be given the opportunity to make a statement.

28. DUTIES ASSIGNED. (Identify principal duties assigned, primary subaltern duties and watch qualifications, including number of months assigned each during the period of report. Indicate dates of period of unavailability due to hospitalization, temporary additional duty, and leave and travel between duty stations. For reports based on other than "Class Observation," indicate after primary duty the number of months duties assigned followed by the number of months duties were performed physically approved from the reporting senior, if available (e.g., 6M). These dates along with command station should be discussed with the officer early in the reporting period. (Refer to BUPERS form 1411.12 Series for OCR Copy Entry.)

SPECIFIC ASPECTS OF PERFORMANCE. (The following items are specific aspects of performance. Each aspect has the sub-items listed below it to assist in defining it and to provide guidelines for the performance appraisal discussion. As you evaluate the officer on each performance aspect, review the sub-items and place marks in the appropriate boxes to the right of each sub-item. The marks in the sub-items boxes are guidelines only. This work sheet is retained by the reporting senior for the performance appraisal discussion. It is not forwarded with the OCR Sheet. After reporting the sub-items and indicating strengths and areas requiring greater emphasis, select appropriate blocks from the table below and place the appropriate transcription code letter in the OCR code letter box to the right of each item for transfer to the OCR Sheet, (Items 29 through 37.)

TRANSCRIPTION CODE

	N	A	B	C	D	E	F	G	H	I	
RANGE	N/A/ N.C.	HIGH				MID			LOW		
									EMPH	UNSAT*	

	N/A/ N.C.	A MODERATE STRENGTH	AH ASSET	EMPH GREATER EMPHASIS	OCR CODE LETTER
29. GOAL SETTING AND ACHIEVEMENT					<input type="checkbox"/>
A. PURSUES PROFESSIONAL DEVELOPMENT IN POSITIVE AND FORCEFUL MANNER.					
B. IS MEETING PROFESSIONAL QUALIFICATION STANDARDS IN TIMELY FASHION.					
C. DEMONSTRATES HIGH STANDARDS OF EXCELLENCE IN PERSONAL ACHIEVEMENTS.					
D. DEFINES REALISTIC GOALS.					
E. DEVELOPS PLANS AND PRIORITIES.					
F. INVOLVES SUBORDINATES IN PLANNING.					
G. RESPONDS POSITIVELY TO CHANGING CIRCUMSTANCES.					
H. EFFECTIVELY ACHIEVES GOALS.					
NOTE: FOR COMMANDERS, COMMANDING OFFICERS AND OFFICERS-IN-CHARGE, THE FOLLOWING CONSIDERATIONS IN ADDITION TO THE ABOVE WILL BE MADE AS APPROPRIATE.					
I. DEVELOPS UNIT GOALS CONSISTENT WITH THE OBJECTIVES OF HIGHER AUTHORITY.					
J. ESTABLISHES REALISTIC PRIORITIES AMONG UNIT GOALS AND ACHIEVES A PROPER BALANCE OF EFFORT EXPENDED BETWEEN SHORT TERM AND LONG TERM OBJECTIVES.					
K. EFFECTIVELY UTILIZES AVAILABLE RESOURCES IN UNIT GOAL ATTAINMENT.					
30. SUBORDINATE MANAGEMENT AND DEVELOPMENT					<input type="checkbox"/>
A. ESTABLISHES EQUITABLE AND CONSISTENT POLICIES.					
B. CONSIDERS THE IDEAS AND SUGGESTIONS OF SUBORDINATES.					
C. IS EFFECTIVE IN PERSONAL SUPERVISION.					
D. DELEGATES AUTHORITY APPROPRIATE WITH SUBORDINATES' CAPABILITIES.					
E. UNDERSTANDS, USES AND ENCOURAGES EFFECTIVE PRINCIPLES OF LEADERSHIP.					
F. DEMONSTRATES THE ABILITY TO INSPIRE AND LEAD.					
G. DISPLAYS INTEGRITY AND MORAL COURAGE.					
H. ACHIEVES POSITIVE RETENTION RESULTS.					
I. ENSURES PROPER POSITION MANAGEMENT, I.E., HIERARCHICAL STRUCTURING OF THE ORGANIZATION.					
J. ENSURES CIVILIAN POSITIONS ARE PROPERLY CLASSIFIED; I.E., COMPLETE AND ACCURATE, BUT UNLIMITED POSITION DESCRIPTIONS THAT HAVE BEEN SUBMITTED TO PROPER AUTHORITIES FOR CLASSIFICATION AND HAVE BEEN PERIODICALLY REVIEWED AS REQUIRED BY REGULATIONS. PROMPT ACTION TO ELIMINATE ANY OVERGRADED SITUATIONS.					
NOTE: FOR COMMANDERS, COMMANDING OFFICERS AND OFFICERS-IN-CHARGE, THE FOLLOWING CONSIDERATIONS IN ADDITION TO THE ABOVE WILL BE MADE AS APPROPRIATE.					
K. PLACES SUBORDINATES IN CHALLENGING SITUATIONS TO DEVELOP THEIR ABILITIES.					
L. EFFECTIVELY ACHIEVES PERSONNEL READINESS OBJECTIVES INCLUDING RETENTION.					
M. DEVELOPS OR MAINTAINS EFFECTIVE AND PROFESSIONAL QUALIFICATION PROGRAMS WHICH RESULT IN A HIGH LEVEL OF READINESS AMONG ALL PERSONNEL INCLUDING NAVAL RESERVE AUGMENTATION PERSONNEL.					

Figure A-2. Current appraisal work sheet.

26. ABILITY TO SPEAK IN AN EFFECTIVE MANNER OCR CODE LETTER

27. ABILITY TO WRITE IN AN EFFECTIVE MANNER OCR CODE LETTER

WARFARE SPECIALTY SKILLS (Demonstrated proficiency and knowledge in executing his warfare specialty. Indicate evaluation in each applicable skill area by selecting appropriate block from the scale and place the transcription code letter in the OCR code letter box.)

TRANSCRIPTION CODE											
N	A	B	C	D	E	F	G	H	I		
N.A./N.O.		HIGH			MID			LOW			
										Strong/Usual*	
RANGE											

38. OCR CODE LETTER

39. OCR CODE LETTER

40. OCR CODE LETTER

41, 42, 43. (LEAVE BLANK ON OCR FORM)

SUBSPECIALTY (Complete if applicable)

SUBSPECIALTY CODE (Enter subspecialty code from lowest OOCR)

SUBSPECIALTY REQUIRED BY BILLET (Indicate whether billet is subspecialty coded)

44. YES 45. NO

SUBSPECIALTY UTILIZATION (Indicate degree of utilization of subspecialty)

46. FREQUENT 47. INFREQUENT 48. SOMETIMES

SUBSPECIALTY PERFORMANCE (Indicate evaluation of subspecialty performance by selecting appropriate block from the scale in WARFARE SPECIALTY SKILLS and place the transcription code letter in the OCR code letter box.)

49. OCR CODE LETTER

MISSION CONTRIBUTION (Evaluate the officer's performance with regard to contributions to the unit's mission. Include effective integration of personnel and the mission and completion of assigned tasks.)

EVALUATION (Indicate evaluation by placing an "X" in appropriate box and provide supporting comments in section 82 emphasizing how well the officer contributed to mission accomplishment while effectively integrating personnel and the mission.)

SUMMARY DISTRIBUTION (Enter the total of all officers of this rank and competitive category marked in each corresponding block of item 51 on reports submitted by you on this occasion. Enter 0 for none. Include this officer in the summary. Include Regular, Concurrent, and Special Reports submitted by you on officers of this rank on this occasion.)

RANGE	NOT OBS	HIGH			MID			LOW		
									Strong/Usual*	
51. EVALUATION										
52. SUMMARY										

TREND OF PERFORMANCE (Since last report)

53. Past Report 54. Constant 55. Improving 56. Deteriorating*

DESIRABILITY (Indicate your attitude toward having this officer under your command in the following categories of assignment. Select the transcription code letter, corresponding to the relevant description, and place it in the OCR code letter box.)

57. Command 58. Joint/DEP 59. Operational 60. Foreign Shore 61. Staff

DESIRABILITY TRANSCRIPTION CODE

N.A./N.O.	PARTIC. DESIRE	PREFER	PLEASED	SAT.	PREFER NOT
N	A	B	C	D	E

RECOMMENDATION FOR PROMOTION (As a consequence of this officer's achieved performance and potential during this evaluation period I would recommend (indicate recommendation by "X" in appropriate box.))

IF THIS OFFICER IS RECOMMENDED FOR EARLY PROMOTION (rank with all officers of the same grade and competitive category recommended for early promotion during the reporting period. Indicate total number of such officers in the "number recommended" box. Indicate this officer's ranking in the "ranking" box. If the officer is 1 of 3 place a "2" in the "ranking" box. If 1 of 1, place a "1" in the box.)

62. Early Promotion 63. Regular Promotion 64. No Promotion* 65. Number Recommended 66. Ranking

PERSONAL TRAITS (How do you rate this officer in exhibiting the following qualities? Indicate evaluation by selecting appropriate block from the scale in WARFARE SPECIALTY SKILLS and place the transcription code letter in the box provided.)

67. Judgment (Sound reasoning; develops logical conclusions)

68. Imagination (Resourcefulness; creativity; constructive planning)

69. Analytical ability (Logical discrimination between assumption and fact)

70. Personal behavior (Demeanor; tactability and public behavior)

71. Forcefulness (Positive and enthusiastic performance of duty)

72. Military bearing (Smartness of appearance; correctness of uniform; physical fitness; adherence to weight standards must be taken into consideration)

(LEAVE BLANK ON OCR FORM) 73. 74. 75. 76.

WEAKNESSES (Significant weakness should be discussed with the officer. Are there any?) 77. None Noted (Significant) 78. Yes 79. No*

80. STATEMENT (The officer shall either indicate that a statement is attached or the drawer not to make a statement.)

81. SIGNATURE OF OFFICER EVALUATED (NAV SUPERS INST 1411.12 apply)

82. DATE FORWARDED (Date reporting senior signed and forwarded report.) 83. INACTIVE NAVAL RESERVE ONLY 84. SIGNATURE OF REPORTING SENIOR

85. DATE FORWARDED (Date regular reporting senior signed and forwarded concurrent/special report.) 86. SIGNATURE OF REGULAR REPORTING SENIOR ON CONCURRENT AND CONCURRENT/SPECIAL REPORT.

87. COMMENTS (Particularly comment upon the officer's overall leadership ability, personal traits not listed above, and estimated or actual performance in combat. Include comments pertaining to unique skills and characteristics that may be important to career development and future assignments. A mark in boxes with an asterisk (*) indicates advisory and supporting comments are required.)

Figure A-2 (Continued)



PART II. TO BE RETAINED BY
THE REPORTING SENIOR. NOT TO
ACCOMPANY THE FITNESS REPORT

1. To be completed by the rated officer. List on this form your accomplishments during the rating period, and arrange to meet with your senior officer several weeks before the end of the rating period to review your performance. Submit this document at that time.

Signature of rated officer

Date

Figure A-3. (Continued).

DISTRIBUTION LIST

Military Assistant for Training and Personnel Technology (OSUDS) (R&AT)
Deputy Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Chief of Naval Operations (OP-01B7) (2), (OP-987H)
Chief of Naval Material (NMAT 01M), (NMAT 05), (NMAT 0722)
Commander, Naval Military Personnel Command (NMPC-00), (NMPC-013C), (NMPC-32)
(5)
Commanding Officer, Naval Aerospace Medical Research Laboratory
Chief of Naval Research (Code 200), (Code 400), (Code 440), (Code 442), (Code 442PT)
Officer of Naval Research, Detachment Pasadena (Code 00A), (Code N-21), (Code N-5)
Chief of Naval Technical Training (Code N-6)
Chief of Naval Air Training
Commanding Officer, Naval Training Equipment Center (Technical Library) (5), (Code 1)
Commandant, Coast Guard Headquarters
Commanding Officer, U.S. Coast Guard Research and Development Center, Avery Point
Superintendent, Naval Postgraduate School
President, National Defense University (3)
Director of Research, U.S. Naval Academy
Secretary-Treasurer, U.S. Naval Institute
Defense Technical Information Center (DDA) (12)

DUDLEY KNOX LIBRARY - RESEARCH REPORTS



5 6853 01073638 2

U214983

STEEL SHARPS

DEPARTMENT OF THE NAVY

NAVY PERSONNEL RESEARCH AND
DEVELOPMENT CENTER
SAN DIEGO, CALIFORNIA 92152

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE \$300



POSTAGE AND FEES PAID
DEPARTMENT OF THE NAVY
DOD-316

