**Navy Personnel Research and Development Center**

San Diego, CA 92152-6800 TN 88-38 April 1988

(2)



# Developing Behaviorally Anchored Rating Scales for the Machinist's Mate Rating

DEPARTMENT OF THE NAVY
NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER
SAN DIEGO, CALIFORNIA 92152

3900
Ser 62/362
26 APR 1988

From: Commanding Officer, Navy Personnel Research and Development Center

Subj: **DEVELOPING BEHAVIORALLY ANCHORED RATING SCALES FOR THE MACHINIST'S MATE RATING**

Encl: (1) NPRDC TN 88-38

1. This research is part of the Navy's Job Performance Program and, in particular, the Navy Performance-Based Personnel Classification Project (Z17701.001). The research supports the Joint-Service Job Performance Measurement/Enlistment Standards Project. The Joint-Service Project has been mandated by Congress to link enlistment standards to job performance, which can be considered a landmark research thrust of the Armed Services. The present research has been funded under Program Element 63707N (Manpower Control System Development).

2. This report details the development of behaviorally anchored rating scales for use as indices of technical proficiency for first-term incumbents within the Machinist's Mate (MM) rating. The information it contains is intended to benefit the research and the operational MM communities. Ultimately, the outcome of the project will benefit the armed services, military and civilian research communities, and applied industrial organizational psychology in general.

JOHN J. PASS
By direction

Distribution:
Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics)
Chief of Naval Operations (OP-135L)
Defense Technical Information Center (DTIC) (2)

NPRDC TN 88-38 April 1988

# Developing Behaviorally Anchored Rating Scales for the Machinist's Mate Rating

Ronald Bearden
Navy Personnel Research and Development Center

Michael Wagner
Robert Simon
Dynamics Research Corporation

Reviewed by
Deborah Mohr Bruce

Approved by
John J. Pass, Ph.D.
Director, Personnel Systems Department



Navy Personnel Research and Development Center
San Diego, California 92152-6800

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED | 1b. RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY | 3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited. |
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S): NPRDC TN 88-38 | 5. MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a. NAME OF PERFORMING ORGANIZATION: Navy Personnel Research and Development Center; 6b. OFFICE SYMBOL (If applicable): Code 62 | 7a. NAME OF MONITORING ORGANIZATION |
| 6c. ADDRESS (City, State, and ZIP Code): San Diego, CA 92152-6800 | 7b. ADDRESS (City, State, and ZIP Code) |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (If applicable) | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER |
| 8c. ADDRESS (City, State, and ZIP Code) | 10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 63707; PROJECT NO Z1770; TASK NO Z1770.001; WORK UNIT ACCESSION NO |

11. TITLE (Include Security Classification)

Developing Behaviorally Anchored Rating Scales for the Machinist's Mate Rating

12. PERSONAL AUTHOR(S)
Bearden, R. M., Wagner, M., & Simon, R.

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Year, Month, Day) | 15. PAGE COUNT |
|---|---|---|---|
| Interim | FROM 1985 TO 1986 | 1988 April | 37 |

16. SUPPLEMENTARY NOTATION

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Behaviorally anchored rating scales, performance measurement, joint service project, Naval personnel, Theses. |
| 05 | 09 | | |

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


This study details the development of behaviorally anchored rating scales (BARS) for use by self, peer, and supervisory raters, which could be used for assessing the performance of Navy Machinist's Mates (MMs). Four steps were involved: (1) identifying and describing general and task level job performance dimensions, (2) generating performance examples and developing unique behavioral anchors for each of the dimensions, (3) constructing a set of rating forms and assembling them into rating administration packages, and (4) conducting a pilot test of the BARS on a sample of MMs. Results, while limited by sampling constraints suggest that both the general and task level performance rating scales are operating within acceptable limits. A recommendation that the BARS be included in a large scale field test to be administered to fleet MMs is presented.


| | | |
|---|---|---|
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED | |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL: Bearden, Ronald M. | 22b. TELEPHONE (Include Area Code): (619) 533-7732 | 22c. OFFICE SYMBOL: Code 621 |

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

## SUMMARY

### Problem

In 1982, Congress formally required the armed services to establish methods for measuring job performance and validating selection standards against them. The Navy's contribution to this coordinated effort is entitled Performance-Based Personnel Classification. Its objectives are to assess on-the-job performance and to improve the Navy's automated classification and assignment system (Kroeker & Rafacz, 1983) by including job performance information.

The Navy's approach focuses on direct measurement of technical proficiency, which follows the research strategy of the joint-service project. The purpose of this large scale effort is to develop job performance measures for first-term enlistees with four (or fewer) years of service and demonstrate their use as criteria for validating hands-on job sample tests (Laabs & Berry, 1987).

In the joint-service project, the job sample test has been adopted as the high fidelity measure of technical proficiency. Since these performance measures are extremely expensive to develop and administer, it is essential to construct economical job sample test substitutes such as performance ratings and symbolic simulations that correspond to the hands-on tests.

### Background

Behaviorally anchored rating scales (BARS) were selected for use as performance rating substitutes for job sample tests because the scales: (1) measure only observable job behaviors and do not attempt to measure attitudes, emotions, or personality, (2) have a greater level of user acceptance because the developmental process relies entirely upon incumbents, and (3) have been validated extensively and proven to be an accurate reflection of job performance.

### Objectives

The objectives of this research were to: (1) develop two sets of BARS for use by self, peer, and supervisory raters, which could be used as indices of technical proficiency within the Machinist's Mate (MM) rating, and (2) determine the effectiveness of the scales.

### Approach

These objectives were met by completing the following steps: (1) identifying and describing general and task level job performance dimensions, (2) generating performance examples and developing unique behavioral anchors for each of the dimensions, (3) constructing a set of rating forms and assembling them into rating administration packages, and (4) conducting a pilot test of the BARS on a sample of MMs.

### Results and Discussion

Both the general job performance dimensions and task level performance ratings scales appear to operate well. The entire judgment range of the scales are being used by all rater types; however, several task level scales exhibit a low difficulty index, leading to some consistently high ratings. Each rating scale appears internally consistent as illustrated by relatively high alpha coefficients and item-scale correlations. The size of

the standard deviations on individual general dimensions suggest that the differentiation among ratees was greater in the peer and supervisor ratings.

Conclusions

Both sets of BARS display reasonable statistical properties that are indicative of good judgmental criterion measures.

Recommendation

It is recommended that the BARS be included in a large scale field test package to be administered to fleet MMs to determine whether the BARS are a useful substitute for the hands-on test.

## CONTENTS

## LIST OF TABLES

## INTRODUCTION

### Problem

The general problem being addressed in this report is that the Navy, along with the other services, does not use a prediction of job performance to guide the classification and assignment of recruits. Instead, the services rely primarily on the prediction of success in entry-level job training where end-of-course grades are easy to obtain. However, in the majority of cases, these grades are based on tests that do not assess job skills, making them poor substitutes for a job performance criterion.

In 1982, Congress formally required the armed services to establish methods for measuring job performance and validating selection standards against them. The Navy's contribution to this coordinated effort is entitled Performance-Based Personnel Classification (PBPC). Its objectives are to assess on-the-job performance and to improve the Navy's automated classification and assignment system or CLASP (Kroeker & Rafacz, 1983) by including job performance information (Laabs, 1983).

The Navy's approach focuses on direct measurement of technical proficiency, which follows the research strategy of the joint-service project. The purpose of this large scale effort is to develop job performance measures for first-term enlistees with four (or fewer) years of service and to demonstrate their use as criteria for validation. Complete descriptions of the Navy Job Performance Measurement (JPM) program and the PBPC subproject are contained in Laabs and Berry, 1987.

While Congress and Department of Defense (DoD) have indicated that directly observed job performance is the preferred criterion for predictor validation, it is readily acknowledged that the institution of a service-wide performance testing program to support personnel selection would be prohibitively expensive. Accordingly, the Navy JPM program is investigating different measurement approaches with the goal of finding an economical way to routinely assess job performance.

In the joint-service project, the hands-on job sample test has been adopted as the high fidelity measure of technical proficiency. Since these performance measures are extremely expensive to develop and administer, it is essential to construct economical job sample test substitutes such as behaviorally anchored performance rating scales. If the resulting scales exhibit adequate statistical properties and relate well to job performance, they may be used in place of expensive hands-on tests.

### Background

Behaviorally anchored rating scales (BARS) were selected for use in the JPM program over other types of behavior based scales, such as Behavioral Observation Scales (Latham & Wexley, 1981) or Mixed Standard Scales (Blanz, & Ghiselli, 1972), for several reasons. First, BARS measure only observable job behaviors and do not attempt to measure attitudes, emotions, or personality. Second, the scales are understandable and have a greater level of acceptance to the users than other types of scales, because the process of scale development relies entirely on the incumbent population. Finally, such scales have been validated extensively through research in a variety of settings and proven to be an accurate reflection of job performance. (For further details of the advantages of using BARS, see Bernardin & Smith, 1981).

BARS contain two primary components that must be defined during scale development. The first component, performance dimension, focuses on major categories or dimensions of job behavior for a particular job. One scale is developed for each performance dimension. The more complex the job, the more dimensions that must be represented. For each dimension, job performance examples (i.e., examples of actual job behaviors) must be generated for several levels of performance ranging from above average to below average. The second component focuses on the specific behavior definitions on each scale. These definitions are known as behavioral anchors and are the performance statements representing a specific level of job performance effectiveness.

Objectives

The objectives of this research are to: (1) develop two sets of behaviorally anchored rating scales for use by self, peer, and supervisory raters to accurately assess technical proficiency within the Machinist's Mate (MM) rating and (2) determine the effectiveness of the scales by designing and testing rating administration packages.

## APPROACH

The steps shown in Table 1 represent the general approach to BARS development followed in this study, and are a variant of a procedure introduced by Smith and Kendall (1963). This approach was used to produce two sets of BARS and their associated administration packages for the MM rating. The first set of BARS consisted of four general job performance dimensions; the second consisted of a series of task level job performance dimensions. Six administration packages were designed for obtaining ratings from supervisors, peers, and job incumbents on both task level and general job dimension BARS.

**Table 1**

Steps in Developmental Approach

A. Develop General BARS

1. Review MM documentation.
2. Identify tentative general level dimensions.
3. Revise general level dimensions.
4. Generate job performance examples.
5. Develop behavioral anchors.
6. Construct general level BARS.

B. Develop Task Level BARS

7. Examine MM job sample test.
8. Generate job performance examples.
9. Develop behavioral anchors.
10. Construct task level BARS.
11. Pilot test both sets of BARS.

## Developing General Job Performance Rating Scales

The first step in developing the general job performance ratings scales involved examining the job functions of the MM rating to identify tentative job performance dimensions. Initial familiarization of the project staff with the MM rating was accomplished by examining (1) the Navy Occupational Task Analysis Program outputs on the MM rating, which provided detailed information (e.g., time to perform and difficulty level) on each task, (2) a preliminary version of the MM hands-on test (Kroeker, Laabs, Vineberg, Joyner, & Zimmerman, 1986), (3) MM Personnel Qualification Standards, and (4) the Manual of Navy Enlisted Manpower and Personnel Classifications and Occupational Standards. Additionally, the research team spent a full day participating in an actual MM job orientation aboard the USS JESSE L. BROWN (FF 1089). This orientation included a briefing by the ship's Chief Engineer about the primary job responsibilities of a typical first-term MM.

### Identifying, Reviewing, and Revising General Dimensions

Based on the review of the MM information sources, a number of tentative general job performance dimensions were identified and refined (see Table 2). The objective was to produce a set of dimensions that would serve as a starting point for discussions to be held at the Great Lakes Naval Training Center (NTC). This approach was similar to that used by Borman, Hough, and Dunnette (1976) to develop preliminary job performance dimensions. However, Borman et al. used job performance examples, rather than job descriptive material, as the input for the construction of preliminary job performance dimensions.

**Table 2**

Preliminary General Job Performance Dimensions for MM Rating

| Mechanical Aptitude/Ability |
|---|
| Effectively and efficiently maintains a variety of mechanical equipment. Always uses proper tools. When required, produces components that meet highest standards. |
| **Troubleshooting Skills** |
| Rapidly isolates causes of equipment malfunctions by systematically testing component systems. Understands how components of a system interrelate. Quickly prioritizes diagnostic and remedial actions. Can improvise effective solutions. |
| **Dedication/Dependability/Cooperation** |
| Commits self to success of mission. Encourages others to high performance. Completes task regardless of time or sacrifice. Can always be depended on to assist other crew members. Steady conscientious worker. Puts ship ahead of self. |
| **Safety** |
| Performs all maintenance functions in a safe manner, and checks to make sure that actions will not interfere with performance of other system components or other ship systems |

A group of eight subject-matter experts (SMEs) (i.e., individuals who have attained at least an E-7 pay grade in the MM rating) was then assembled at Great Lakes NTC to review and revise the set of preliminary general job performance dimensions. Each dimension was discussed and analyzed until the group reached a consensus. The resulting four general job dimensions are listed and defined in Table 3.

**Table 3**

Definitions of General Job Performance Dimensions for MM Rating

| Safety |
| --- |
| Has effective knowledge of all applicable safety programs and procedures. Performs all maintenance functions in a safe manner. Has safety awareness for self, others, and ship. |
| Mechanical Aptitude/Ability |
| Effectively and efficiently maintains a variety of mechanical equipment. Needs little or no supervision. Always uses proper tools when available. Knows location of special tools and measuring devices. In performing overhauls or fabrications, "gets it right" the first time. Understands how components of a system interrelate. |
| Technical Procedures |
| Follows appropriate engineering and casualty control procedures. Rapidly locates correct reference material to perform tasks and completes tasks in a timely manner. Promptly completes all paperwork associated with actions. |
| Adaptability/Dedication |
| Quickly adapts to shipboard living. Develops and maintains a positive attitude. Maintains flexibility to different job assignments and working conditions. Volunteers to help wherever needed. Can always be depended on to assist other crew members. Is a steady and conscientious worker. |

### Generating, Reviewing, and Editing General Performance Examples

After defining the job performance dimensions, a set of instructions appearing in Appendix A was created for use by two groups of eight SMEs in writing preliminary performance examples (PEs). The first group, located at Great Lakes NTC, was ask to generate six PEs for each dimension: two examples for each of three levels of performance (i.e., above average, average, and below average). The second group, located in San Diego, was asked to generate PEs pertinent to a list of MM duties provided by Navy Personnel Research and Development Center (NAVPERSRANDCEN) (see Appendix B). Each PE produced by both groups was then typed in random order onto retranslation forms. An example appears in Appendix C. Next, the forms were assembled into a package and used in a retranslation process in which performance examples produced by individuals are reviewed by a group of SMEs, and assigned to the most pertinent dimension and placed on a scale (see Smith & Kendall, 1963).

Following the guidelines set forth in the retranslation process, a different group of Great Lakes SMEs was asked to make two judgments about each PE. The first judgment was to select the general dimensions that was most pertinent to the PE. The second judgment involved placement of the PE along a continuum by rating the example on the dimension selected. A nine point Likert Scale was used to obtain these ratings (i.e., 9 = very high performance, 5 = fully adequate performance, and 1 = very low performance).

The performance examples were then sorted into groups corresponding to the general dimension to which they were most frequently assigned. PEs were eliminated if there was less than 80 percent agreement as to the dimension to which they were most pertinent. The mean rating of each PE assigned to a general dimension was computed and the PEs were then rank ordered. The PEs for each dimension were divided into four groups, which corresponded roughly to four levels of performance (i.e., extremely effective, effective, marginal, and ineffective). Next, three behavioral anchors corresponding to each of the four levels of performance for each dimension were developed. The performance examples were used to determine the content and level of the behavioral anchors being developed.

### Rating Behavioral Anchors on General Dimensions

Next, the 12 behavioral anchors constructed for each general dimension were evaluated to ensure that they accurately reflected the four levels of performance for each of the general dimensions. Each of the behavioral anchors for each general dimension was typed onto an index card. Then, SMEs were asked to group each set of 12 cards into four categories based on the performance level reflected in the statements and then to rank order the cards within each category.

Behavioral anchors, which were retranslated into the wrong category by any individual, were reviewed, discussed with the SMEs, and revised as necessary.

Based on the data gathered at the Great Lakes NTC, a draft of the general job performance BARS was constructed by the contractors, and later submitted for editorial review by NAVPERSRANDCEN researchers. The resulting scales for the general dimensions, which included three behavioral anchors at each of four levels of performance, contained 10 points. The point to performance level configuration was organized as follows:

| | |
|---|---|
| 9 or 10: | Extremely effective |
| 6, 7, or 8: | Effective |
| 3, 4, or 5: | Marginally effective |
| 1 or 2: | Ineffective |

An example of a completed general level BARS is included as part of the rating instruments appearing in Appendix D.

### Developing Task Level Performance Rating Scales

Prior to the development of the scales, a group of tasks was identified by Kroeker et al. (1986) that were representative of the MM rating. These tasks consisted of two subsets of parallel tasks, one set that is performed in the generator room and the second set that is performed primarily in the engine room of FF 1052 class fast frigates. The purpose of this undertaking was to develop behaviorally anchored rating scales that correspond to these sets of tasks.

Examining the Hands-on Job Sample Test

The initial step in the development of the scales was to examine the latest version of the MM hands-on job sample test being developed by NAVPERSRANDCEN. This test included parallel tasks performed in both the engine and generator rooms (see Table 4).

**Table 4**

List of Generator and Engine Room Tasks

| | Generator Room | | Engine Room |
|---|---|---|---|
| 1. | Inspect, clean circulating water strainer (EOP) | 1. | Inspect, clean feedwater pump suction strainer (EOP) |
| 2. | Loss of vacuum in auxiliary condenser, casualty procedure (EOCC) | 2. | Loss of vacuum in main engine condenser, casualty procedure (EOCC) |
| 3. | Inspect, clean, pressurize, test lube oil strainers SSTG (EOP) | 3. | Shift, inspect, clean main engine lube oil strainers (EOP) |
| 4. | Hot bearing in turbogenerator casualty procedure (EOCC) | 4. | Hot bearing in main engine, casualty procedure (EOCC) |
| 5. | Sample and inspect lube oil (MRC) | 5. | Sample and inspect lube oil (MRC) |
| 6. | Align, start, stop main drain eductor (MRC) | 6. | Align, start, stop main drain eductor (MRC) |
| 7. | Major leak in SSTG lube oil system, casualty procedure (EOCC) | 7. | Major leak in main engine lube oil system, casualty procedure (EOCC) |
| 8. | Align, start, operate, stop fire pump (EOP) | 8. | Align, start, operate, stop fire pump (EOP) |
| 9. | Record temperature and pressure on various gauges | 9. | Record temperature and pressure on various gauges |
| 10. | Loss of turbogenerator lube oil pressure, casualty procedure (EOCC) | 10. | Loss of main engine lube oil pressure, casualty procedure (EOCC) |
| 11. | Clean and inspect duplex lube oil filter (PMS-MRC) | 11. | Start, operate, stop main lube oil pump (EOP) |
| 12. | Manufacture an 8-hole flange gasket and install in flange system | 12. | Manufacture an 8-hole flange gasket and install in flange system |
| 13. | Disassemble, inspect, and identify repair techniques and materials for globe valve (MRC) | 13. | Disassemble, inspect, and identify repair techniques and materials for globe valve (MRC) |
| 14. | Repack globe or gate valve using LP or HP packing (MRC) | 14. | Repack globe or gate valve using LP or HP packing (MRC) |

An examination of the individual tasks revealed that they generally fell into one of three categories: (1) engineering operational procedures (EOP), (2) preventive/repair maintenance, and (3) engineering operational casualty control (EOCC). EOP and preventive/repair maintenance tasks are similar in that they are routine tasks that are performed periodically in accordance with written procedures. An example of an EOP task is to inspect and clean circulating water strainers, while an example of a maintenance task is to manufacture an 8-hole flange gasket and install in flange system. EOCC tasks, however, require a swift response by technicians who have memorized the required controlling actions necessary to control a casualty. An example of an EOCC task is the hot bearing in main engine casualty procedure.

## Preparing Performance Examples and Constructing the Task Level BARS

During this step, task level BARS were constructed based on information obtained in interviews with Great Lakes SMEs. The development of these scales involved (1) preparing a form for generating performance examples, (2) developing behavioral anchors, and (3) constructing task level BARS.

Forms were developed for generating performance examples on task level dimensions for the engine and generator room tasks. These forms listed one task to a page, providing space for SMEs to write performance examples and assign scale values to each example. A nine point scale was used (e.g., 9 = very high performance, 5 = fully adequate performance, and 1 = very low performance).

A group of SMEs was assembled and each SME was given either the generator or engine room PE development form. Each SME was asked to develop a low, average, and high PE for each task.

The performance examples were to be used to construct a unique set of five behavioral anchors corresponding to five performance levels for each task. However, it became evident that the resulting anchors used for EOCC tasks would be quite different from those for EOP/maintenance tasks. Therefore, the performance examples were used to construct two sets of generic behavioral anchors: one for EOCC tasks and one for EOP/maintenance tasks. These anchors were then reviewed and revised by SMEs.

Next, the behavioral anchors were edited and compiled into two sets of scales for use with the EOCC and the EOP/maintenance tasks. Each scale within these sets contained nine points. Behavioral anchors were shown at five of the nine levels of performance. The point to performance level configuration was organized as follows:

9 Extremely effective
7 Very effective
5 Effective
3 Marginally effective
1 Ineffective

This group of task level dimensions and their descriptors were submitted by the contractors for editorial review and subsequently revised as necessary. The completed set of rating forms for task level BARS was then prepared for the pilot test. Examples of completed engine and generator room task level BARS are included as part of the rating instruments in Appendix D.

Pilot Testing of the Behaviorally Anchored Rating Scales

This step consisted of two activities. First, rating administration packages were designed. Second, the ratings packages were administered in a pilot test to MMs and their supervisors aboard two FF 1052 class ships.

Designing Rating Administration Packages

The rating administration packages were constructed so that general and task level ratings could be obtained from supervisors, peers, and job incumbents as summarized in Table 5. Within each package, ratings were administered in sequence following two rules.

1. Self ratings were obtained before peer ratings.

2. General performance ratings were obtained before task level ratings.

**Table 5**

Rating Administration Package Design

| | General Dimension Form | | | Task Level Form (Generator) | | | Task Level Form (Engine) | | |
|---|---|---|---|---|---|---|---|---|---|
| Rater Type | Self | Peer | Supervisor | Self | Peer | Supervisor | Self | Peer | Supervisor |
| Supervisor (Generator) | | | x | | | x | | | |
| Supervisor (Engine) | | | x | | | | | | x |
| Job Incumbent (Generator) | x | x | | x | x | | | | |
| Job Incumbent (Engine) | x | x | | | | | x | x | |

The rating scales were organized so that each left-hand page contained the dimension/task to be rated and the right-hand page contained a response scale where a rating on that dimension/task could be entered. Detailed instructions were provided to assist the rater in completing the scales. The raters were also given several suggestions on how to avoid potential sources of error when assigning their ratings (e.g., consider each dimension/task separately and read all of the descriptions of each performance level).

A pilot test of these rating administration packages[1] was conducted on a sample of 19 first-enlistment MMs, 10 engine room, and 9 generator room watch standers aboard

[1]Requests for copies of the rating administration packages should be made to NAVPERSRANDCEN.

two FF 1052 class fast frigates. Job incumbents were given the generator or engine room rating package depending on their recent space assignment. Additionally, each incumbent's immediate supervisor (typically an MM with an E-6 or above pay grade) and at least two co-workers were selected to provide ratings on the incumbent.

## RESULTS

The results of the pilot test, while limited by sampling constraints to nonstatistical analyses of data trends, provide some evidence on how well the general and task level BARS operate. A more complete assessment of the scales, including any required statistical tests, will be possible upon completion of the field testing of the BARS and the hands-on job sample performance measures.

Borman et al. (1976) have documented that each rating method has its own strengths and weaknesses according to the configuration shown in Table 6. These methods may produce different results largely due to differences in the opportunities afforded each type of rater to observe the ratees and the impartiality of each type of rater.

**Table 6**

Primary Strengths and Weaknesses of Rating Methods

| Method | Major Strength | Major Weaknesses |
|---|---|---|
| Supervisor | Familiar with rating process. | Often subject to halo bias (tendency to base all ratings on an overall impression rather than identify an individual's strengths and weaknesses). |
| Peer | Good opportunity to view ratee's performance on all aspects of job. | Reluctance to rate their peers as less than outstanding (don't want to be seen as backstabbing). |
| Self | Best opportunity to observe job performance.<br>Can identify one's own strengths and weaknesses. | Subject to leniency error. |

### Rater Bias

A detailed examination of the ratings was made to identify the extent of three potential sources of bias: leniency, restriction in range, and halo.

1. Leniency bias. Leniency is a response bias that occurs when a rater assigns consistently high ratings to individuals, regardless of their "true" performance levels. Leniency bias can be identified when raters select only the upper values of the rating scale.

Table 7 displays the means and standard deviations for self, peer, and supervisor ratings for both engine and generator room groups. For the general dimensions (10 point scale), the mean scale ratings of all three methods ranged from 6.7 to 7.0, about three scale points from the top of the scale. This apparent lack of an upper ceiling assignment would suggest that the raters did make effective use of the judgment range of the general dimension scales. However, the supervisors ratings averaged one-third scale point below those of the peer and self ratings, suggesting that peer and self ratings were slightly more lenient than supervisor ratings.

This same pattern was reflected in the task level rating (9 point scale) in which mean supervisor ratings (7.25) were somewhat less lenient than self (7.65) or peer (7.85) ratings. Task level ratings were, on the average, less than 1.5 scale points from the top of the scale (7.7). While this was quite close to the top of the scale, individual task means did vary quite a bit. For supervisors, the task means ranged from 4.7 to 9.0. For peers, the task means ranged from 6.2 to 8.8. For self ratings, the task means ranged from 6.1 to 8.7. Thus, it appears that raters also made effective use of the judgment range of the task level scales, at least for the difficult tasks.

2. Restriction of range bias. A restriction of range error occurs when a rater is unable to differentiate among ratees of differing abilities on a given performance dimension. An indirect assessment of how well raters are able to differentiate among ratees can be made by examining the standard deviations of ratings on each general dimension/task.

Table 7 contains these standard deviations. Peer and supervisor ratings show more dispersion (median SD = 1.5 and 1.5 (untabled), respectively) than do self ratings (median SD = 1.3 (untabled)) for the general dimensions. However, there is little or no difference in dispersion across rater types for the task level dimensions.

3. Halo bias. This form of response bias occurs when a rater formulates an overall judgment about a ratee's performance and then assigns the same rating level on all performance dimensions. The magnitude of inter-item correlations on the general and task level dimensions provides an indication of the degree of halo error present in the ratings. Very high inter-item correlations suggest raters are grouping dimensions together and failing to differentiate performance on them, while lower inter-item correlations indicate that raters are treating dimensions individually and attending to within-ratee differences in performance on the dimensions.

Table 8 contains the inter-item correlations on the general dimensions for self, peer, and supervisor ratings, respectively. The median correlations were .73 for peer ratings, .57 for supervisor ratings, and .38 for self ratings.

Table 9 contains the inter-item correlations on the task level dimensions. The median correlations were .60 for peer ratings, .43 for supervisor ratings, and .44 for self ratings. These results indicate that peer ratings had the lowest level of differentiation among dimensions/tasks while self ratings had the greatest differentiation among dimensions/tasks. Thus, self ratings seem to provide the most information on the strengths and weaknesses of individual ratees while peer ratings provided the least information of this sort. However, these results cannot be statistically analyzed to determine whether they are in agreement with earlier studies (e.g., Klimoski & London, 1974) that demonstrated self-ratings show less halo effects.

**Table 7**

Item and Scale Means and Standard Deviations for Whole Group and Three Types of Raters

| Dimensions | Group Mean | Group s.d. | Self Mean | Self s.d. | Peer Mean | Peer s.d. | Supervisor Mean | Supervisor s.d. |
|---|---|---|---|---|---|---|---|---|
| General | | | | | | | | |
| 1 | 7.0 | (1.6) | 7.3 | (.9) | 7.0 | (1.8) | 6.8 | (1.6) |
| 2 | 7.4 | (1.7) | 7.7 | (1.3) | 7.5 | (1.7) | 7.1 | (1.9) |
| 3 | 6.0 | (1.9) | 5.6 | (1.0) | 5.9 | (2.0) | 6.3 | (2.3) |
| 4 | 7.4 | (2.0) | 7.5 | (1.8) | 7.8 | (1.9) | 6.5 | (2.0) |
| Scale Mean | 6.9 | (1.5) | 7.0 | (.9) | 7.0 | (1.6) | 6.7 | (1.6) |
| Task level Engine: | | | | | | | | |
| 1 | 6.2 | (2.0) | 6.2 | (2.3) | 6.7 | (1.7) | 4.7 | (2.0) |
| 2 | 6.9 | (1.8) | 6.9 | (2.0) | 7.3 | (1.8) | 6.0 | (1.5) |
| 3 | 6.8 | (2.0) | 6.5 | (2.3) | 7.0 | (2.0) | 6.5 | (1.8) |
| 4 | 6.9 | (1.4) | 7.2 | (1.1) | 7.2 | (1.4) | 6.2 | (1.5) |
| 5 | 8.2 | (1.0) | 7.8 | (1.2) | 8.3 | (.9) | 8.2 | (1.0) |
| 6 | 8.5 | (.9) | 8.5 | (1.7) | 8.6 | (.9) | 8.2 | (.9) |
| 7 | 6.6 | (2.1) | 7.0 | (1.2) | 6.6 | (2.6) | 6.2 | (1.6) |
| 8 | 8.6 | (.9) | 8.4 | (1.3) | 8.8 | (0.5) | 8.2 | (1.0) |
| 9 | 8.6 | (.8) | 8.7 | (1.7) | 8.6 | (.8) | 8.5 | (1.1) |
| 10 | 7.0 | (1.5) | 7.0 | (1.8) | 7.1 | (1.5) | 6.6 | (1.2) |
| 11 | 7.3 | (1.4) | 7.2 | (1.4) | 7.4 | (1.4) | 7.1 | (1.5) |
| 12 | 7.9 | (1.6) | 8.4 | (1.1) | 8.2 | (1.2) | 6.7 | (2.0) |
| 13 | 6.1 | (2.3) | 6.7 | (1.8) | 6.2 | (2.2) | 5.3 | (2.7) |
| 14 | 6.9 | (2.1) | 6.5 | (2.2) | 6.8 | (2.2) | 7.3 | (1.6) |
| Scale Mean | 7.4 | (1.1) | 7.2 | (1.1) | 7.5 | (1.1) | 7.1 | (.9) |
| Task level Generator | | | | | | | | |
| 1 | 7.7 | (1.2) | 7.8 | (.7) | 8.0 | (1.2) | 6.9 | (1.1) |
| 2 | 8.0 | (1.2) | 7.9 | (1.1) | 8.3 | (1.1) | 7.2 | (1.0) |
| 3 | 7.3 | (2.1) | 7.4 | (1.9) | 7.8 | (1.9) | 6.1 | (2.2) |
| 4 | 7.4 | (1.9) | 7.6 | (1.2) | 7.8 | (1.9) | 6.2 | (2.0) |
| 5 | 8.7 | (.8) | 8.7 | (.7) | 8.6 | (.8) | 8.7 | (.9) |
| 6 | 8.6 | (.8) | 8.9 | (.3) | 8.6 | (.9) | 8.3 | (1.0) |
| 7 | 7.3 | (1.7) | 7.6 | (1.3) | 7.7 | (1.8) | 6.4 | (1.3) |
| 8 | 8.8 | (.6) | 8.9 | (.3) | 8.7 | (.7) | 8.9 | (0.3) |
| 9 | 8.8 | (.6) | 8.9 | (0.3) | 8.7 | (.7) | 9.0 | (0.0) |
| 10 | 7.5 | (1.2) | 7.4 | (1.3) | 8.0 | (.8) | 6.4 | (1.3) |
| 11 | 7.8 | (1.3) | 7.9 | (1.6) | 8.1 | (1.2) | 7.0 | (1.1) |
| 12 | 8.3 | (.9) | 8.4 | (1.1) | 8.3 | (1.0) | 8.1 | (.7) |
| 13 | 7.5 | (1.7) | 7.6 | (1.2) | 7.8 | (2.0) | 6.8 | (1.2) |
| 14 | 8.2 | (1.6) | 8.6 | (.7) | 8.1 | (1.9) | 8.3 | (1.5) |
| Scale Mean | 8.0 | (1.0) | 8.1 | (.7) | 8.2 | (1.1) | 7.4 | (.8) |

Table 8

Inter-item Correlations

| Dimension[a] | | 1 | 2 | 3 | 4 | Scale |
|---|---|---|---|---|---|---|
| Self Ratings on General Job Dimensions[a] (N = 29) | | | | | | |
| | 1 | 1.00 | | | | |
| | 2 | .20 | 1.00 | | | |
| | 3 | .40 | .61 | 1.00 | | |
| | 4 | .09 | .36 | .40 | 1.00 | |
| Scale | | .48 | .76 | .80 | .76 | 1.00 |
| Peer Ratings on General Job Dimensions (n = 56) | | | | | | |
| | 1 | 1.00 | | | | |
| | 2 | .81 | 1.00 | | | |
| | 3 | .64 | .65 | 1.00 | | |
| | 4 | .83 | .85 | .54 | 1.00 | |
| Scale | | .92 | .93 | .80 | .90 | 1.00 |
| Supervisor Ratings on General Job Dimensions (n = 26) | | | | | | |
| | 1 | 1.00 | | | | |
| | 2 | .54 | 1.00 | | | |
| | 3 | .33 | .59 | 1.00 | | |
| | 4 | .36 | .65 | .63 | 1.00 | |
| Scale | | .66 | .86 | .83 | .84 | 1.00 |

[a]Dimensions and dimension numbers are defined in Table 3 on page 4.

## Table 9

Inter-item Correlations

| Task[a] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Self Ratings on Engine Room Task[a] Dimensions (n = 10)** | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .67 | -- | | | | | | | | | | | | |
| 3 | .66 | .38 | -- | | | | | | | | | | | |
| 4 | .63 | .64 | .81 | -- | | | | | | | | | | |
| 5 | .68 | .36 | .60 | .44 | -- | | | | | | | | | |
| 6 | .21 | -.20 | .24 | .02 | .77 | -- | | | | | | | | |
| 7 | .70 | .41 | .86 | .78 | .72 | .38 | -- | | | | | | | |
| 8 | .54 | .23 | .33 | -.05 | .72 | .58 | .40 | -- | | | | | | |
| 9 | .33 | -.03 | .47 | .11 | .86 | .81 | .53 | .76 | -- | | | | | |
| 10 | .79 | .71 | .80 | .80 | .64 | .09 | .88 | .36 | .36 | -- | | | | |
| 11 | .68 | .65 | .70 | .69 | .87 | .45 | .83 | .54 | .66 | .87 | -- | | | |
| 12 | .54 | -.08 | .45 | .13 | .32 | .15 | .33 | .34 | .34 | .23 | .09 | -- | | |
| 13 | .45 | .18 | .18 | .35 | .23 | .13 | .10 | -.08 | .01 | .07 | .07 | .60 | -- | |
| 14 | .67 | .48 | .53 | .79 | .29 | -.04 | .45 | -.19 | -.11 | .50 | .33 | .43 | .77 | -- |
| Scale | .89 | .64 | .86 | .82 | .78 | .32 | .89 | .48 | .50 | .88 | .85 | .49 | .40 | .66 |
| **Peer Ratings on Engine Room Task[a] Dimensions (n = 29)** | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .77 | -- | | | | | | | | | | | | |
| 3 | .77 | .77 | -- | | | | | | | | | | | |
| 4 | .68 | .63 | .81 | -- | | | | | | | | | | |
| 5 | .39 | .38 | .44 | .35 | -- | | | | | | | | | |
| 6 | .57 | .24 | .43 | .25 | .27 | -- | | | | | | | | |
| 7 | .60 | .47 | .50 | .54 | .14 | .21 | -- | | | | | | | |
| 8 | .21 | .15 | .26 | -.06 | .53 | .33 | .03 | -- | | | | | | |
| 9 | .25 | -.02 | .20 | .20 | .23 | .79 | .01 | .17 | -- | | | | | |
| 10 | .78 | .80 | .84 | .76 | .55 | .22 | .64 | .23 | .05 | -- | | | | |
| 11 | .75 | .77 | .83 | .75 | .60 | .43 | .60 | .32 | .29 | .89 | -- | | | |
| 12 | .55 | .34 | .40 | .38 | .41 | .41 | .13 | .14 | .23 | .40 | .33 | -- | | |
| 13 | .90 | .68 | .68 | .56 | .34 | .58 | .48 | .15 | .30 | .72 | .70 | .68 | -- | |
| 14 | .66 | .60 | .52 | .59 | .50 | .04 | .40 | .10 | -.12 | .73 | .61 | .70 | .72 | -- |
| Scale | .92 | .81 | .87 | .79 | .55 | .50 | .66 | .25 | .25 | .91 | .90 | .60 | .88 | .77 |

[a]Task and task numbers are defined in Table 4.

Table 9 (Continued)

| Task[a] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Supervisor Ratings on Engine Room Task[a] Dimensions (n = 13) | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .87 | -- | | | | | | | | | | | | |
| 3 | .84 | .75 | -- | | | | | | | | | | | |
| 4 | .77 | .64 | .61 | -- | | | | | | | | | | |
| 5 | .17 | .40 | .61 | .09 | -- | | | | | | | | | |
| 6 | .46 | .61 | .35 | .45 | .23 | -- | | | | | | | | |
| 7 | .56 | .68 | .66 | .30 | .31 | .26 | -- | | | | | | | |
| 8 | .19 | .73 | .31 | .35 | .46 | .65 | .39 | -- | | | | | | |
| 9 | .29 | .70 | .48 | .42 | .65 | .65 | .21 | .91 | -- | | | | | |
| 10 | .70 | .85 | .69 | .73 | .34 | .54 | .75 | .70 | .62 | -- | | | | |
| 11 | .59 | .60 | .75 | .36 | .78 | .35 | .49 | .26 | .45 | .48 | -- | | | |
| 12 | .38 | .73 | .49 | .40 | .57 | .67 | .33 | .94 | .94 | .67 | .37 | -- | | |
| 13 | .27 | .42 | .63 | -.01 | .94 | .13 | .46 | .34 | .47 | .30 | .83 | .44 | -- | |
| 14 | .28 | .57 | .25 | .86 | .40 | .21 | .52 | .75 | .37 | .79 | .60 | .78 | -- | |
| Scale | .82 | .88 | .97 | .58 | .74 | .63 | .60 | .76 | .84 | .82 | .79 | .82 | .70 | .79 |
| Self Ratings on Generator Room Task[a] Dimensions (n = 9) | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .32 | -- | | | | | | | | | | | | |
| 3 | .47 | .64 | -- | | | | | | | | | | | |
| 4 | .78 | .34 | .72 | -- | | | | | | | | | | |
| 5 | .09 | -.05 | .30 | .52 | -- | | | | | | | | | |
| 6 | -.13 | -.40 | -.11 | .17 | .88 | -- | | | | | | | | |
| 7 | .44 | .41 | .52 | .62 | .62 | .44 | -- | | | | | | | |
| 8 | .44 | .67 | .86 | .78 | .35 | -.13 | .44 | -- | | | | | | |
| 9 | .44 | -.04 | .09 | .17 | -.18 | -.13 | .44 | -.13 | -- | | | | | |
| 10 | .41 | .48 | .73 | .44 | .31 | .13 | .76 | .41 | .41 | -- | | | | |
| 11 | .09 | .43 | .77 | .60 | .40 | -.03 | .26 | .90 | -.26 | .26 | -- | | | |
| 12 | .31 | .36 | .70 | .79 | .83 | .48 | .65 | .81 | -.18 | .43 | .78 | -- | | |
| 13 | .32 | .15 | .56 | .67 | .67 | .47 | .85 | .47 | .47 | .59 | .47 | .70 | -- | |
| 14 | .29 | -.07 | .33 | .59 | .89 | .80 | .80 | .29 | .29 | .49 | .27 | .73 | .87 | -- |
| Scale | .53 | .52 | .87 | .85 | .64 | .27 | .83 | .82 | .20 | .74 | .70 | .88 | .83 | .71 |

[a]Task and task numbers are defined in Table 4.

Table 9 (Continued)

| Task[a] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peer Ratings on Generator Room Task[a] Dimensions (n = 26) | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .81 | -- | | | | | | | | | | | | |
| 3 | .89 | .81 | -- | | | | | | | | | | | |
| 4 | .77 | .66 | .91 | -- | | | | | | | | | | |
| 5 | .59 | .51 | .57 | .59 | -- | | | | | | | | | |
| 6 | .83 | .60 | .86 | .86 | .83 | -- | | | | | | | | |
| 7 | .89 | .85 | .94 | .87 | .52 | .80 | -- | | | | | | | |
| 8 | .91 | .86 | .92 | .86 | .47 | .76 | .92 | -- | | | | | | |
| 9 | .82 | .90 | .83 | .72 | .33 | .57 | .86 | .93 | -- | | | | | |
| 10 | .58 | .66 | .63 | .43 | .25 | .35 | .70 | .54 | .61 | -- | | | | |
| 11 | .82 | .84 | .91 | .80 | .53 | .74 | .89 | .89 | .85 | .67 | -- | | | |
| 12 | .53 | .68 | .61 | .66 | .48 | .55 | .64 | .65 | .65 | .41 | .49 | -- | | |
| 13 | .89 | .71 | .93 | .86 | .59 | .88 | .89 | .87 | .73 | .55 | .79 | .63 | -- | |
| 14 | .87 | .75 | .93 | .93 | .65 | .91 | .89 | .91 | .76 | .45 | .79 | .74 | .95 | -- |
| Scale | .92 | .86 | .98 | .92 | .64 | .88 | .96 | .95 | .86 | .63 | .90 | .71 | .94 | .96 |
| Supervisor Ratings on Generator Room TAsk[a] Dimensions (n = 11) | | | | | | | | | | | | | | |
| 1 | -- | | | | | | | | | | | | | |
| 2 | .55 | -- | | | | | | | | | | | | |
| 3. | .80 | .36 | -- | | | | | | | | | | | |
| 4 | .83 | .33 | .86 | -- | | | | | | | | | | |
| 5 | -.03 | .06 | .31 | .03 | -- | | | | | | | | | |
| 6 | .55 | .15 | .84 | .65 | .42 | -- | | | | | | | | |
| 7 | .78 | .50 | .73 | .89 | .09 | .61 | -- | | | | | | | |
| 8 | .27 | .74 | .16 | .35 | -.10 | .09 | .61 | -- | | | | | | |
| 9 | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | -- | | | | | |
| 10 | .85 | .50 | .62 | .77 | -.42 | .38 | .70 | .35 | ++ | -- | | | | |
| 11 | .16 | .09 | .29 | .45 | .00 | .54 | .64 | .61 | ++ | .21 | -- | | | |
| 12 | .26 | -.32 | .58 | .69 | .04 | .67 | .51 | .04 | ++ | .29 | .65 | -- | | |
| 13 | -.24 | -.23 | .05 | .10 | .52 | .39 | .25 | .23 | ++ | -.42 | .70 | .51 | -- | |
| 14 | .08 | -.04 | .48 | .21 | .95 | .54 | .20 | -.16 | ++ | -.27 | .06 | .26 | .49 | -- |
| Scale | .76 | .41 | .91 | .90 | .35 | .85 | .90 | .41 | ++ | .59 | .58 | .65 | .34 | .49 |

[a]Task and task numbers are defined in Table 4.
++ Item 9 had zero variance in this scale.

## Scale Reliability

The extent to which items within a scale function as one is a measure of the internal consistency and hence, reliability of the scale. Table 10 contains scale reliability coefficients (Smith & Kendall, 1963) for all scales and rater types. The first column represents the internal consistency among the four general dimension ratings across each rater type. Columns 2 and 3 represent the internal consistency among the 14 task ratings across each rating type for both engine and generator room tasks. Although the scale coefficients in column 1 are somewhat lower than those appearing in columns 2 and 3, the median coefficient at .90 is still quite high indicating that the scales exhibit a high degree of internal consistency. Additionally, it is important to note that the differences in magnitude of the coefficients can partly be explained by the differences in the number of items used in computing the coefficients in column 1 versus columns 2 and 3. If the general ratings coefficients were stepped up to 14 "items," these differences would diminish.

**Table 10**

Reliability Coefficients for all Scales and Rater Types

| | Scale | | |
|---|---|---|---|
| Rater Type | General Job Performance | Task Level Performance | Task Level Performance Generator Room |
| Self | .64 | .90 | .90 |
| Peer | .91 | .92 | .97 |
| Supervisor | .81 | .88 | .87 |
| Whole group | .85 | .91 | .94 |

Table 11 displays for all rater types correlation coefficients between ratings on items of each scale and that scale's mean rating. This table clearly shows that, for the most part, low correlations only occur on very easy items (those that have low standard deviations and high means (see Table 7)). Thus, almost all items that exhibit large variations across ratees are highly correlated with their scales.

**Table 11**

Item and Scale Correlation Coefficients for
Whole Group and Three Types of Raters

| Dimensions | Group r | Self r | Peer r | Supervisor r |
|---|---|---|---|---|
| General | | | | |
| 1 | .69 | .25 | .85 | .46 |
| 2 | .80 | .52 | .87 | .74 |
| 3 | .61 | .65 | .65 | .64 |
| 4 | .70 | .37 | .82 | .69 |
| Task level Engine | | | | |
| 1 | .82 | .88 | .91 | .76 |
| 2 | .72 | .53 | .77 | .77 |
| 3 | .84 | .81 | .83 | .96 |
| 4 | .68 | .79 | .76 | .43 |
| 5 | .54 | .74 | .51. | .59 |
| 6 | .40 | .24 | .45 | .26 |
| 7 | .57 | .88 | .55 | .62 |
| 8 | .29 | .39 | .23 | .38 |
| 9 | .27 | .45 | .21 | .54 |
| 10 | .84 | .94 | .89 | .66 |
| 11 | .82 | .81 | .88 | .68 |
| 12 | .51 | .41 | .55 | .53 |
| 13 | .67 | .26 | .84 | .54 |
| 14 | .59 | .55 | .70 | .53 |
| Task level Generator | | | | |
| 1 | .84 | .48 | .91 | .71 |
| 2 | .72 | .44 | .84 | .33 |
| 3 | .90 | .81 | .97 | .86 |
| 4 | .87 | .81 | .89 | .86 |
| 5 | .48 | .60 | .61 | .27 |
| 6 | .78 | .24 | .87 | .82 |
| 7 | .93 | .78 | .95 | .87 |
| 8 | .77 | .81 | .94 | .39 |
| 9 | .60 | .17 | .85 | xx |
| 10 | .57 | .67 | .59 | .50 |
| 11 | .77 | .61 | .89 | .50 |
| 12 | .68 | .85 | .68 | .61 |
| 13 | .83 | .79 | .93 | .23 |
| 14 | .76 | .68 | .95 | .37 |

xx = a coefficient could not be computed.

## CONCLUSIONS

Preliminary results suggest that both the general job performance dimensions and the task level performance rating scales are operating within acceptable limits as indicated by mean rating levels that consistently fell below the top of the scale. Several tasks do appear, however, to be of low difficulty, leading to some consistently high ratings and low variability. This outcome should not be totally unexpected given the variety of the task difficulty levels existing on the job sample test.

Also, the rating scales appear internally consistent as illustrated by relatively high alpha coefficients and moderate item-scale correlations. The strength of inter-item correlations indicates that the differentiation of raters among general dimensions/tasks was highest for self raters and lowest for peer raters. Finally, the size of the standard deviations on individual general dimensions indicated that the differentiation among ratees was greater in the peer and supervisor ratings.

These results suggest that BARS may provide a useful means of assessing job performance. They are easier to develop and less time-consuming to administer than are hands-on job tests. Because BARS development relies heavily on the expertise of SMEs in the occupational specialty, the resulting scales are understandable and acceptable to raters. They are written in a language used by job incumbents and their supervisors and are based on real world performance examples. Finally, BARS emphasize job performance and technical proficiency to a much greater extent than do traditional rating scales that measure personality traits or characteristics.

However, before any final conclusion can be drawn regarding the ultimate usefulness of BARS for the JPM project, a full scale test must be conducted. Data must be collected from an adequate sample of first-term MMs. The relative accuracy of self, peer, and supervisor ratings in assessing job proficiency can be determined by comparing these BARS scores to hands-on performance test scores. In addition, such comparisons can also allow us to determine whether BARS can serve as a valid substitute for costly hands-on performance testing in support of efforts to validate selection criteria.

## RECOMMENDATION

It is recommended that the BARS be included in a large scale field test package to be administered to fleet MMs to determine whether the BARS are a useful substitute for the hands-on job sample test.

## REFERENCES


Bernardin, H., J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales. Journal of Applied Psychology, 65, 458-463.

Blanz, W., & Ghiselli, E. E. (1972). The mixed standard scale: A new rating system. Personnel Psychology, 25, 185-199.

Borman, W. C., Hough, L. M., & Dunnette, M. D. (February 1976). Development of behaviorally based rating scales for evaluating the performance of U.S. Navy recruiters (NPRDC TR 76-31). San Diego: Navy Personnel Research and Development Center.

Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59, 445-451.

Kroeker, L. P., Laabs, G. J., Vineberg, R., Joyner, J. R., & Zimmerman, R. (In review). Developing a hands-on performance measure for the Machinist's Mate (MM) rating. San Diego: Navy Personnel Research and Development Center.

Kroeker, L. P., & Rafacz, B. A. (November 1983). Classification and assignment within PRIDE (CLASP): A recruit assignment model (NPRDC TR 84-9). San Diego: Navy Personnel Research and Development Center.

Laabs, G. J. (1983). Performance testing in support of Navy personnel classification and assignment. In C. R. Curran (Chair), Job based and performance testing. Paper presented at the symposium meeting of the American Psychological Association, Washington, DC.

Laabs, G. L., & Berry, V. (August 1987). The Navy job performance measurement program: Background and inception (NPRDC TR 87-34). San Diego: Navy Personnel Research and Development Center.

Latham, G. P., & Wexley, K. N. (1981). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley Publishing Co.

Smith, P. C., & Kendall, L. H. (1963). Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

# APPENDIX A

## INSTRUCTIONS TO SUBJECT-MATTER EXPERTS FOR GENERATING PERFORMANCE EXAMPLES

# INSTRUCTIONS TO SUBJECT-MATTER EXPERTS FOR GENERATING PERFORMANCE EXAMPLES

## Introduction

Dynamics Research Corporation has a contract with the Navy to develop behaviorally based job performance rating scales for the MM rating. We have found that the most efficient way of developing such scales requires us first to have persons knowledgeable about the job generate "behavior incidents" describing the performance of persons on the job being studied. That is where you come in.

We want you to generate, both in discussions and in writing, a number of performance examples describing the performance of Navy MMs based on your experience with the job. We have found that this is the best way to build performance rating scales that make sense to those using the scales and that are comprehensive in terms of covering the whole job. During the workshop sessions, we will be helping you to generate these performance examples that will form the building blocks of the performance rating scales for the MM rating. On the next few pages, we describe in more detail what we mean by "performance examples" and provide some examples.

## HOW TO WRITE JOB PERFORMANCE EXAMPLES

To write a job performance example or incident, try first to remember what someone in an MM job actually did or failed to do that made him effective or ineffective in performing his duties. These incidents can be examples of extremely effective, ineffective, or even average job performance. The important thing is that the example is described specifically as it happened.

When writing a performance example, describe only what you saw or what the person did, not what you inferred from the action. For example, if you were writing an incident about a person, instead of saying that a person carried out tasks in an unsafe manner, describe what he did that made you feel he was acting in an unsafe manner such as "did not use a safety shield while cleaning tube and strainer baskets," "tried to stop a freon leak with his hand, causing frostbite," or "when cleaning throttle board, sprayed condensate on main pump indicator electric panel, causing a shock." All of these are behaviors, but they are very different actions. We are asking you, then, to specifically describe not the traits, but the behaviors that you have observed in your experience with the MM job.

The characteristics of a good performance example are:

1. It concerns the actions of a person who is in the job under job consideration.

2. It tells what the person did (or did not do) that made you feel he was effective or ineffective in his job.

3. It is brief, to the point, and does not go to great lengths to specify the consequences of what the person did.

The next page contains some hypothetical incidents that illustrate how to write performance examples.

- An MM, during casualty control drills, tried to engage the jacking gear while it was still turning.

- MM failed to check shaft alley for a complete 4 hour watch

- MM overhauled scullart machine while underway.

- MMs completely painted out engine room while on watch in order to receive more free time in port.

- MM worked around the clock for 36 hours replacing spring bearing while underway.

- MM made lesson plans for use by his division to educate personnel in the aspects of the distilling plants and water treatment.

- MM removed steam traps from heating steam lines.

- MM replaced ship's whistle diaphragm while ship is underway.

- MM rebuilt an SSTG circ. water pump with very little supervision from work center supervisor.

Select correct tool for job. Dedicates self to safety and security of others and of the ship.

APPENDIX B

A PARTIAL LIST OF GENERAL MACHINIST'S MATE DUTIES

## A PARTIAL LIST OF GENERAL MACHINIST'S MATE DUTIES

A series of duty titles are listed for the MM rating. These duty titles were used by NAVPERSRANDCEN to generate some preliminary performance examples.

1. Propulsion Plant

   This category includes routine maintenance, operation and casualty procedures in the propulsion plant area, which includes securing the plant, operating the jacking gear, reduction gears, main/auxiliary shaft, steam turbine and steam generators.

2. Auxiliary Propulsion Equipment

   This area includes maintenance operation and casualty procedures for the auxiliary boiler, fresh water systems, seawater circulating systems, the main feed system, the make up feed system, air ejector systems, condensate systems, and diesel engines.

3. Auxiliary Systems

   This area includes operation, maintenance, and casualty procedures for the fireman system, auxiliary cooling water systems, the ship's heating system, hot water heaters, drain systems, bilge and tank stripping system, and the ships whistle.

4. Auxiliary Steam Systems

   This area includes operation, maintenance, an casualty procedures for the auxiliary steam systems, steam reducing stations, distilling plants, water treatment, the distillate transfer system, and high pressure/low pressure drain systems.

5. Air Systems

   This area includes operation, maintenance, and casualty procedures for the air systems (high pressure, low pressure, etc.), air compressors, oxygen/nitrogen systems, ventilation/revitalization systems, and MTB blow systems.

6. General Mechanical Maintenance

   This area includes cleaning, inspecting, replacing, rebuilding, testing etc. in various areas. It includes pumps, valves, flexible coupling hoses, expansion points, systems filters/strainers, sea chest, lagging, pipe supports, foundation bolts, taking soundings, tanks/voids, checking for leaks and corrosion, removing corrosion, painting, manufacture parts or gaskets, calibrating gauges, monitor alarm panels and maintaining pressure and temperature readings.

7. Air Conditioning/Refrigeration/Ship Service

   This area includes maintenance, operations, and casualty procedures for air conditioning units, refrigerator, laundry equipment, and galley and scullery equipment.

8. Hydraulics/Fuel Oil/Lube Oil

This area includes maintenance, operation, and casualty procedures for the steering hydraulic systems, main and vital hydraulic system, the external hydraulic system, the main lube oil system, oil sampling, the oil purification and transfer system, and the waste and oil separator system.

# APPENDIX C

# RETRANSLATION FORMS

## RETRANSLATION FORMS

Level of Performance Shown by Incident

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| Very Low | | | | Fully Adequate | | | | Very High |

| | Category | Level of Performance |
|---|---|---|
| 1. MM did not know location of start/stop switches for fire pumps on DC deck. | _______ | _______ |
| 2. MM failed to check alignment of operating fire pump to burn up due to suction valve being closed. | _______ | _______ |
| 3. MM operated STM to laundry bypass because no one knew how to fix reducer. | _______ | _______ |
| 4. MM investigating reason for poor output from evaporators discovered demisteds clogged and replaced same. | _______ | _______ |
| 5. MM failed to check shaft alley for a complete 4 hour watch causing major flooding. | _______ | _______ |
| 6. MM relagged STM lines in laundry to bring ambient temperature down. | _______ | _______ |
| 7. MM cleaned Mn condensor headers of kelp, fish, and plastic bags. | _______ | _______ |
| 8. MM removed, calibrated, and reinstalled all gauges on Mn throttle board. | _______ | _______ |
| 9. MM replaced flange shields on lube oil system. | _______ | _______ |
| 10. MM used wire wheel to shine deck plates. | _______ | _______ |
| 11. MM color coded VLV wheels on various systems. | _______ | _______ |
| 12. MM relagged approximately 20 feet of Mn STM piping. | _______ | _______ |
| 13. MM, while rigging bilge and stripping PMP down escape trunk, did not utilize proper equipment, which caused PMP to become out of alignment when PMP dropped down escape trunk. | _______ | _______ |

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| Very Low | | | | Fully Adequate | | | | Very High |

| | Category | Level of Performance |
|---|---|---|
| 14. MM replaced impeller and casing wearing rings on aux circ. pump but during reassembly failed to support casing causing shaft to become bent. | ______ | ________ |
| 15. MM overhauled Leslie constant press PMP gov 12 times before finally getting it to work correctly underway--man had no previous experience in this type of repair. | ______ | ________ |
| 16. MM properly replaced first stage discharge VLV on HP air compressor as cause of casualty for low output by checking gauges only. | ______ | ________ |
| 17. MM had to rewrite work request three times due to improper phrasing in remarks section. | ______ | ________ |
| 18. MMs completely painted out engine room while on watch to receive more free time in port. | ______ | ________ |
| 19. MM put wrong size casing basket in Mn cond. pmp. | ______ | ________ |
| 20. MM incorrectly sounded Fd bottom giving excessive usage of water for 2 hours. | ______ | ________ |
| 21. MM repacked STM VLV leaving last ring practically out of packing gland causing it to bend when tightened down. | ______ | ________ |

APPENDIX D

EXAMPLES OF GENERAL AND TASK LEVEL BARS

## General Level BARS

### Safety

Has effective knowledge of all applicable safety programs and procedures. Performs all maintenance functions in a safe manner. Has safety awareness for self, others, and ship.

#### 9 or 10
**Extremely Effective Performance**

Constantly checks work area to make sure there is no danger to self or others. Notices dangerous situations such as exposed wires or leaks. Takes immediate action and stays with it until problem has been rectified and danger has passed. Frequently checks fire and safety equipment.

Uses chemicals and equipment with respect and care to ensure the well-being of self and others. Always uses proper safety devices and makes sure that others do too. When proper equipment or tools are not available, always chooses substitutes with safety awareness and checks with supervisor.

Makes all necessary checks on equipment before bringing it on line. Always tags out system when required.

#### 6, 7, or 8
**Effective Performance**

Routinely makes sure that work area does not present danger to self and others. Notices danger and immediately reports to supervisor. Occasionally checks fire and safety equipment.

Uses chemicals and equipment properly. Uses proper safety devices and protective equipment. When proper equipment or tools are not available, usually chooses substitute with safety awareness and checks with supervisor.

Checks equipment before bringing it on line. Tags out system when performing maintenance operations which require it.

#### 3, 4, or 5
**Marginal Performance**

Notices serious safety problems in work area and reports to supervisor. Inspects fire and safety equipment when told to do so.

With some supervision, will use chemicals and equipment properly. Usually uses proper safety devices and protective equipment. When proper equipment or tools are not available, will usually check with supervisor before attempting to make a substitution.

Usually remembers to check out equipment before bringing it on line. Occasionally has to be reminded to tag out system.

#### 1 or 2
**Ineffective Performance**

May notice a safety problem or danger but may not report it to supervisor. Does not carry out required inspections of fire and safety equipment.

Uses chemicals and equipment carelessly or improperly, causing possible danger to self and to others. May fail to use devices such as safety shield, safety harness and ear protection. When proper equipment or tools are not available, may make bad substitutions and/or fails to check with supervisor.

Fails to adequately check out equipment before bringing it on line. Does not wait until system is ready or drained before starting certain procedures. Often fails to tag out a system.

## Level Task BARS (Generator)

### Inspect, Clean Circulating Water Strainer (EOP)

MM is directed to start the ship's service generator. He must begin by inspecting and cleaning the circulating water strainers.

9 - Needs little or no supervision in performing task. Thoroughly knows and understands technical procedures. Always uses proper tools--will never make bad substitutions. Extremely safety conscious. Closely monitors all required parameters while performing tasks. Always obtains and uses manual appropriately.

8

7 - Needs very little supervision in performing task. Knows and understands technical procedures. Uses proper tools--will rarely make a bad substitution. Safety conscious. Monitors all required parameters while performing tasks. Obtains and uses manual appropriately.

6

5 - Needs little supervision in performing task. Pretty much knows and understands technical procedures. Uses proper tools--will rarely make a bad substitution. Safety conscious. Usually monitors all required parameters while performing tasks. Obtains and uses manual appropriately.

4

3 - Needs some supervision in performing task. To some degree, knows and understands technical procedures. Usually uses proper tools--occasionally may make a bad substitution. Fairly safety conscious. May need prompting to correctly monitor required parameters on routine tasks. Usually obtains manual but may need prompting to use it correctly.

2

1 - Needs constant supervision in performing task. Does not know or understand technical procedures. May use improper tools if left unsupervised. Not very safety conscious. May fail to monitor one or more required parameters. May fail to either obtain proper manual or use it correctly.

## Task Level BARS (Engine)

### Loss of Main Engine Lube Oil Pressure Casualty Procedure (EOCC)

MM is standing lower level watch while the ship is underway at standard speed. 1 Alpha main lube oil pump is the standby pump and 1 Bravo is the emergency pump. The low lube oil pressure alarm goes off. The MM must perform the supplementary actions to determine why the alarm went off.

9 - Memorizes all controlling, immediate and supplementary actions and can carry them out in a quick and efficient manner. Always maintains self-control. Extremely safety conscious. Closely monitors all required parameters while performing tasks. Needs little or no supervision. Can be relied on to carry out EOCC efficiently and effectively.

8

7 - Memorizes and carries out all controlling and immediate actions but may need to consult manuals for supplementary actions. Maintains self-control. Safety conscious. Monitors all required parameters while performing tasks. Needs very little supervision. Will perform all EOCC steps correctly.

6

5 - Memorizes and carries out most controlling and immediate actions. Needs to use manual for supplementary actions. Usually maintains self-control. Safety conscious. Usually monitors all required parameters while performing tasks. Needs little supervision to ensure that all EOCC steps have been correctly performed.

4

3 - Occasionally forgets an immediate or controlling action; must then check manual. Needs to use manual for supplementary actions. May lose self-control for brief periods during casualty. Fairly safety conscious. May need prompting to correctly monitor a parameter while preforming task. Needs supervision at certain points of the EOCC procedure to ensure that all EOCC steps have been correctly performed.

2

1 - Does not know immediate or controlling actions. Needs to check manual before beginning task. Tends to lose self-control during casualty. Not very safety conscious. May fail to monitor one or more required parameters while performing task.