INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company
WILFONG, Jr., Harry Dean, 1934-
CHRONOLOGICAL LENIENCY AND LINEAR TRANSFORMATION
OF THE SUPERVISORY APPRAISAL SCALE.

The Ohio State University, Ph.D., 1972
Psychology, industrial

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© Copyright by
Harry Dean Wilfong, Jr.
1972
CHRONOLOGICAL LENIENCY AND LINEAR TRANSFORMATION
OF THE SUPERVISORY APPRAISAL SCALE

DISSERTATION
Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Harry Dean Wilfong, Jr., A.B., M.A.

The Ohio State University
1972

Approved by

Adviser
Department of Psychology
PLEASE NOTE:

Some pages may have
indistinct print.
Filmed as received.

University Microfilms, A Xerox Education Company
PREFACE

This report is a partial attempt to summarize the results of analyses of supervisory rating trends and relationships from a fifteen year accumulation of data and heretofore diverse and unpublished government research reports. All conclusions, narrative interpretations, and opinions stated in this analysis are solely those of the author and in no way reflect or represent either official policy or statements of intent on the part of the Air Force Logistics Command, the Agency, or any other complement of the United States Federal Government.

It is impossible to give credit to all those who have in some manner contributed to this study. Considerable assistance was provided by the Air Force Logistics Command civilian personnel research staff whose persistent efforts provided much of the data compiled in this summary. Special acknowledgement is given to Brigadier General William Fullilove and William B. Rivers whose personal interest and stimulation greatly facilitated initial approval and implementation of the standard scoring system. Appreciation is extended to my graduate advisor, Dr. Robert Wherry, for his helpful assistance and guidance over the many years leading to the development of this analysis. Finally, special gratitude is extended to my wife and family whose patience and gentle encouragement eventually led to the completion of this effort.
TABLE OF CONTENTS

PREFACE ........................................................ ii
LIST OF TABLES .............................................. v
LIST OF FIGURES ............................................... vi
LIST OF PLATES ................................................ vii
LIST OF ABBREVIATIONS ............................................ viii

Chapter

I. INTRODUCTION .......................................... 1
   The Air Force Logistics Command
   Performance Rating Systems

II. REVIEW OF THE LITERATURE .............................. 15
   General Problem of Leniency
   Performance Dimensionality
   Differences between Raters
   Proposed Solutions

III. APPRAISAL RESEARCH WITHIN THE AFLC ............... 46
   The Personal Characteristics Appraisal (1961-1964)
   The Job Performance Appraisal (1965-1968)
   The Tailor-Made Appraisal Forms (Beginning 1968)
   Summary

IV. DEVELOPMENT, IMPLEMENTATION, AND INTRODUCTION
   OF THE LINEAR TRANSFORMATION ........................ 70
   Review of the Literature on Standard Scoring
   Statement of Problem
   Preliminary Service Test I
   Administrative Development
   Introduction to the Workforce
   Results of the Field Test
V. ANALYSIS AND CONCLUSIONS

Changes in Rater Behavior as Influenced by
Introduction of the Linear Transformation
Differences between Raters
Workforce Understanding and Acceptance
Stability of the Workforce
Summary

VI. DISCUSSION AND SUMMARY

APPENDIX

1. Personal Characteristics Appraisal
2. Job Performance Appraisal
3. Tailor-Made Appraisal
4. Standard Scoring Career Appraisal
5. Standard Supervisory Rating Instructions
6. Sample News Release
7. Revised Tailor-Made Appraisal
8. Revised Supervisory Rating Instructions
9. Appraisal Distribution Report
10. Endorsing Official's Distribution Listing
11. AFLC Chronological Raw Score f Distributions
12. Endorsing Officials: N < 7 and N < 10
13. Employees Corrected: Base Wide, by Activity

BIBLIOGRAPHY
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Elemental Patterns by Occupational Group-Personal Characteristics Appraisal</td>
<td>52</td>
</tr>
<tr>
<td>2. Correlations between Supervisory and Average Overall Ratings</td>
<td>54</td>
</tr>
<tr>
<td>3. Mean Ratings (PCA) by Year by Occupational Group</td>
<td>58</td>
</tr>
<tr>
<td>4. 1963-1965 Comparisons on Halo Influence</td>
<td>62</td>
</tr>
<tr>
<td>5. Air Force Tailor-made Appraiser Patterns</td>
<td>65</td>
</tr>
<tr>
<td>6. Basic Assumptions</td>
<td>91</td>
</tr>
<tr>
<td>7. Characteristics of Sample Promotion Registers</td>
<td>95</td>
</tr>
<tr>
<td>8. Sample $\bar{x}$ and $\sigma$ Ranges with Associated F-tests</td>
<td>96</td>
</tr>
<tr>
<td>9. Summary of Intercorrelations</td>
<td>98</td>
</tr>
<tr>
<td>10. Appraisal Computer Basic Computational Format</td>
<td>100</td>
</tr>
<tr>
<td>11. AFLC Comparative Raw and Standard Score Frequency Distributions following Introduction of Standard Scoring in 1969</td>
<td>111</td>
</tr>
<tr>
<td>12. Fisher's $g_1$ Statistic Showing Skewness of Ratings by Year</td>
<td>115</td>
</tr>
<tr>
<td>13. Analysis of Variance of Raw Appraisal Scores (Post-Transformation), Year by Activity</td>
<td>117</td>
</tr>
<tr>
<td>14. Analysis of Variance of Raw Appraisal Scores (Post-Transformation), Year by Form</td>
<td>118</td>
</tr>
<tr>
<td>15. Number of Employees rated on AFLC 679 Forms (1971)</td>
<td>121</td>
</tr>
<tr>
<td>16. Number of Endorsing Officials by Activity (1971)</td>
<td>122</td>
</tr>
<tr>
<td>17. Numbers of Endorsing Officials Rating More than Six Employees on Two Combinations of Forms</td>
<td>124</td>
</tr>
<tr>
<td>18. Analysis of Variance Summaries, Forms by Supervisors</td>
<td>126</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>Promotion Selection Ratios (i.e. Ratio of Total Eligibles to a Single Position Vacancy). Sample of 245 Promotion Actions Consummated during 1970</td>
</tr>
<tr>
<td>2.</td>
<td>Command Average Appraisal Score, 1961-1969</td>
</tr>
<tr>
<td>3.</td>
<td>Comparative Frequency Distributions, Operational (Personal Characteristics Appraisal) vs Experimental (Job Performance Appraisal), Ogden Air Materiel Area (N = 685)</td>
</tr>
<tr>
<td>4.</td>
<td>1969 Command-Wide Appraisal Distribution (Tailor-made)</td>
</tr>
</tbody>
</table>
LIST OF PLATES

<table>
<thead>
<tr>
<th>Plate</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Standard Scaling Example</td>
<td>99</td>
</tr>
<tr>
<td>Figure a. Raw scores</td>
<td></td>
</tr>
<tr>
<td>Figure b. Standard scores</td>
<td></td>
</tr>
<tr>
<td>II. AFLC Appraisal Distributions (1969-1971)</td>
<td>116</td>
</tr>
<tr>
<td>Figure a. AFLC comparative raw score appraisal distribution</td>
<td></td>
</tr>
<tr>
<td>Figure b. Chronological standard score distributions</td>
<td></td>
</tr>
<tr>
<td>III. Analysis of Variance Interactions</td>
<td>119</td>
</tr>
<tr>
<td>Figure a. Activity by years interactions</td>
<td></td>
</tr>
<tr>
<td>Figure b. Forms by years interactions</td>
<td></td>
</tr>
<tr>
<td>IV. Endorsing Official Rating Characteristics</td>
<td>125</td>
</tr>
<tr>
<td>Figure a. Average mean</td>
<td></td>
</tr>
<tr>
<td>Figure b. Average N</td>
<td></td>
</tr>
<tr>
<td>Figure c. Average o</td>
<td></td>
</tr>
<tr>
<td>Symbol</td>
<td>Definitions</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>AFLC</td>
<td>Air Force Logistics Command</td>
</tr>
<tr>
<td>AMC</td>
<td>Air Materiel Command—predecessor label for Air Force Logistics Command</td>
</tr>
<tr>
<td>WB/WG</td>
<td>Wage Board/Wage Grade; general pay title for blue collar employees paid under locality wage authority.</td>
</tr>
<tr>
<td>GS</td>
<td>General Schedule; white collar, professional, technical, clerical, and managerial employees.</td>
</tr>
<tr>
<td>JPA</td>
<td>Job Performance Appraisal</td>
</tr>
<tr>
<td>PCA</td>
<td>Personal Characteristics Appraisal</td>
</tr>
<tr>
<td>OCAMA</td>
<td>Oklahoma City Air Materiel Area, Tinker Air Force Base, Oklahoma</td>
</tr>
<tr>
<td>OOAMA</td>
<td>Ogden Air Materiel Area, Hill Air Force Base, Utah</td>
</tr>
<tr>
<td>SAAMA</td>
<td>San Antonio Air Materiel Area, Kelly Air Force Base, Texas</td>
</tr>
<tr>
<td>SMAMA</td>
<td>Sacramento Air Materiel Area, McClellan Air Force Base, California</td>
</tr>
<tr>
<td>WPAFB</td>
<td>Wright-Patterson Air Force Base, Ohio; tenant installation for Headquarters, Air Force Logistics Command</td>
</tr>
<tr>
<td>2750th ABW</td>
<td>2750th Air Base Wing; AFLC complement situated at Wright-Patterson Air Force Base</td>
</tr>
<tr>
<td>SBAMA</td>
<td>San Bernardino Air Materiel Area, Norton Air Force Base, California</td>
</tr>
<tr>
<td>WRAMA</td>
<td>Warner Robins Air Materiel Area, Robins Air Force Base, Georgia</td>
</tr>
<tr>
<td>MOAMA</td>
<td>Mobile Air Materiel Area, Brookley Air Force Base, Alabama (deactivated)</td>
</tr>
<tr>
<td>OER</td>
<td>Officer Effectiveness Report</td>
</tr>
<tr>
<td>JOSATE</td>
<td>Job Oriented Supervisor Appraisal and Training Technique</td>
</tr>
<tr>
<td>RIF</td>
<td>Reduction in Force</td>
</tr>
<tr>
<td>TRAEX</td>
<td>Training, Education, and Experience; Merit Promotion Ranking Factor</td>
</tr>
<tr>
<td>SSAN</td>
<td>Social Security Account Number</td>
</tr>
<tr>
<td>E-246</td>
<td>Air Force Logistics Command Computerized Merit Promotion and Skills Locator System</td>
</tr>
<tr>
<td>679 (Form)</td>
<td>Numeric designator for all series of AFLC supervisory appraisal forms</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Historically, and typically, a persistent problem associated with any supervisory appraisal is one of rater leniency. For the purposes of this investigation, the definition of leniency propounded by Sharon and Bartlett (1969) as "a significant shift in mean rating in the favorable direction from one rating condition to another" is accepted as the operating terminology. Acceptance of this definition carries with it measurement across time (i.e. chronology) and valence, in either favorable or unfavorable direction, with the direction of favorableness left to the perceptions of management.

The scope of investigation becomes one of examining the tendency of the "for-keeps" type of personnel appraisal (as opposed to the criterion reserved for research or predictive validation purposes) to decrease in sensitivity proportional to the length of time it has been in operational use. This latter phenomenon has been described under the various conceptual labels of "dynamic dimensionality" (Ghiselli, 1956a) "deliberate leniency" (Wherry, 1952), or simply "inflation" (Grappe, Alvord, and Poland, 1967). The most commonly reported affiliates of this chronological "creepage" or leniency effect are progressively decreasing variances and increasing means, increased negative skewness of the frequency distribution(s), and decreased utility of the rating as a personnel differentiating instrument for use in making management decisions.
Thus, movement of an appraisal distribution in a direction perceived as favorable to the employee, with larger numbers possessing higher numerical scores, results in a distribution of less and less value to the manager, as chronological leniency progresses. The resultant dilemma facing the industrial manager must be assessed within the context of his employee evaluation program. The manager whose appraisal is designed to diagnose individual training needs, foster supervisory/employee counseling and discussion, or jointly establish personal objectives has no particular problem with chronological leniency. He may have no need for any formalized evaluation system, simply because all employees are either known well enough by all managers, or informal information regarding current work efficiency is quickly and economically available.

The problem of chronological leniency thus falls within the particular domain of the manager of the large corporate, military, or government body. His appraisal program (or programs) must serve many objectives: identifying from among the total available workforce—oftimes widely dispersed geographically—those few individuals to be monetarily rewarded for high performance (i.e. merit increases, promotions, proficiency pay raises, other salary adjustments); identifying individual training and development needs; (presumably) to improve employee motivation and morale; integrating information about selected employee groups for the purposes of making the wide range of personnel management decisions.

The hierarchical organizational structure associated with the large corporate body further adds to the dilemma. Information required
by successively higher levels of management translates into the need for a compressed, quantifiable index of individual performance. The advent of computer banks of personnel management information systems further adds to the need for quantifiable, reliable summary information regarding employee performance.

As observed by Whisler and Harper (1962), performance measurement is always most difficult in large-scale activities where the relationship of individual effort to group output becomes quite remote and difficult to trace. As a means of realistically examining the dynamics of the overall problem, they suggest longitudinal studies within large industrial bodies extending for sufficient periods of time to learn something about organizational adaptation.

The large industrial unit also typifies the situation where new appraisal programs are periodically launched with vigor, enthusiasm, wide-spread publicity, and management enthusiasm. These same programs are also found to be subsequently abandoned after a few years of operational application with disgust (Taylor and Hastman, 1956), suspicion regarding the usefulness of such behavioral indices, and a search for a substitute performance measure—perhaps to once again adopt some similar type of overall "merit" evaluation. It is this phenomenon that is hopefully explored in this study, using a data base collected over twelve years of appraisal experience within a large government industrial-type organization, the Air Force Logistics Command (AFLC). The study covers four separate approaches actually applied over the twelve year period to the problem of merit rating, each new system
launched because the differentiating effectiveness of its predecessor had been neutralized through organizational adaptation and chronological leniency. It is hoped that a description of these trying and complex experiences will cast some light upon the effect of time, per se, as a moderating variable, and upon the more generalized nature of organizational adaptation (to a new stimuli in the form of a revised merit rating system or format). In this regard, the position established by Whisler and Harper (1962), op. cit., and Ronan and Prien (1966), recommending longitudinal experimental investigations of entire organizations with gradual working down toward sub-units and, finally, individuals is heartily endorsed.

The scope of the thesis is essentially descriptive in nature, with primary emphasis upon the effect of variously introduced rating systems upon overall rater behavior. Changes are related to organizational characteristics (where available) recorded over the same period of time.

The Air Force Logistics Command
(Adapted from Malouf and Gober, 1969)

The Air Force Logistics Command is responsible for the logistics support of the Air Force mission on a world-wide basis. During the period covered by this study, it employed from 144,000 employees in 1961 to the present level of 117,000. Unlike most military organizations, the civilian complement predominates, constituting approximately 90% of the total command population. These civil service employees are engaged in a wide variety of occupations with the majority involved in
the procurement, supply, transportation, and depot maintenance of operational weapons systems and other materiel resources.

Total funds managed annually are in excess of eight billion dollars, including operations and maintenance, the military assistance program, and central procurement funds. The command and its contractors modify or repair more than nine thousand aircraft per year, some 14,000 aircraft engines are overhauled, and the number of components and accessory items repaired annually is in excess of 2.8 million. Maintaining a world-wide spares inventory, AFLC has responsibility for more than 900,000 items in the Air Force inventory. The command provides services to 219 Air Force bases world-wide, to sixty-one countries under the Military Assistance Program, and to numerous attache offices and small organizations.

Organizationally structured on a regional servicing basis, the current organizational structure includes the Headquarters, five geographically dispersed Air Materiel Areas (AMAs), and two specialized repair facilities. Headquarters AFLC is located at Wright-Patterson Air Force Base, Ohio, with AMAs as follows: Oklahoma City Air Materiel Area (OCAMA) with 21,926 civilian employees, San Antonio Air Materiel Area (SAAMA)--22,299 employees, Warner Robins Air Materiel Area (WRAMA)--16,258 employees, Ogden Air Materiel Area (OOAMA)--16,638 employees, and Sacramento Air Materiel Area (SMAMA)--16,158 employees. Each AMA is centrally managed and is responsible for regional and world-wide support and management of certain assigned weapons systems or items of equipment. The workforce at each AMA represents a widespread sampling
of most of the civilian job series specialties described in the U.S. Civil Service Commission inventory, representing over 900 separate job categories.

Performance Rating Systems

As is true of most large governmental concerns (1965 survey conducted by the Public Personnel Association, reported in Lopez, 1968, Appendix B), performance rating within the AFLC serves many masters and is geared to the peculiar needs of various personnel programs. Although the primary area covered by this thesis is the performance rating used for making in-service merit promotion decisions, this application represents only a small portion of the total federal employee evaluation process. To a certain extent, some of these various performance review programs are overlapping and oftentimes present a complex and time-consuming administrative burden to the line supervisor. The average AFLC first level supervisor reports from one-fourth to one-fifth of available clock time performing duties related to personnel management, administration, and related employee-management activities (unpublished Air Force research report being prepared in final form at the time of this writing).

Some of the major federal employee evaluation programs are summarized below. Most of these are anchored to either public law or regulatory requirements emanating from the U.S. Civil Service Commission. As such, they represent the parameters within which agencies are permitted to devise and implement evaluation programs specific to their own requirements. Stated another way, research on appraisal systems
must be generally confined within these regulatory parameters unless the investigator is seeking data upon which he can optimistically propose a change in federal legislation.

Pay

Federal pay for the white collar and professional workers (around 50% of the typical AMA population) is set by congressional action. Basic compensation is determined by the graded classification to which an employee is appointed, generally disassociated from the performance of the incumbent. "In-grade" pay increases are essentially based on longevity (i.e., time-in-grade). Employee performance can be related to the awarding of an in-grade pay increase in that the incumbent must be certified to be performing at an "acceptable level of competence" by his official supervisor before movement to the next pay step can be permitted. "Quality salary increases" can also be awarded if the employee is adjudged to be consistently performing at an outstanding level of competence. Both of these in-grade performance review provisions were established through the Salary Reform Act of 1962, as amended. In actual practice, neither option has seen much practice: disallowing of the in-grade is exercised in approximately 1/10th of one percent of the cases eligible; the quality salary increase is awarded to less than one percent of the total workforce annually. For the large proportion of employees (the remaining 99 percent), the in-grade increase is automatic.

For the blue collar workforce, pay is regionally established based upon comparability wages payed by local industries employing sufficient
numbers of individuals performing duties in positions similar to those found at the federal activity. In-grade increases for the blue collar wage grade employees are based almost exclusively upon longevity, little influenced by incumbent performance.

Annual Performance Rating

The Annual Performance Rating Act of 1950, as amended, requires that all federal workers have their performance formally reviewed and recorded once annually by their immediate supervisor. Three rating categories are applied: Outstanding, Satisfactory, and Unsatisfactory. The overwhelming majority of employees are assigned the Satisfactory rating, with approximately two percent receiving the Outstanding rating and no recorded instances of the Unsatisfactory rating being awarded to any employee.

The annual performance rating serves the dual objectives of, at least once annually, forcing a supervisory/employee performance counseling session and compliance with a legislative requirement. The reward value of the formal annual performance is minimal. The Outstanding rating, where awarded, has some value only in reduction-in-force (RIF) situations, in that the possessor of an Outstanding is administratively awarded the equivalent of four additional years of tenure. Other benefits of the Act are largely intangible and arise primarily from positive actions that might emerge from the counseling session and the associated continuous performance evaluation process.
Assessment of Individual Training Needs

The AFLC civilian workforce is surveyed at least once annually to diagnose and document individual training and development needs. For the large segment of the civilian population, this survey takes the form of an informal employee/supervisory discussion where training is related to employee deficiencies and broken down into projected course requirements for the next two year period. This approach assumes that the optimum return from the training investment is best realized where an attempt is made to strengthen specific dimensions of individual weaknesses. Thus, the technique has a built-in performance review requirement in order for management to document specific training requirements.

For white collar employees at the middle and higher levels of management, a more formalized training needs survey is conducted where performance is evaluated against yearly self-established objectives in the tradition of Peter F. Drucker (1954). Past performance is jointly evaluated by the supervisor and the employee and performance objectives for the coming year are programmed along the same basis, with associated development needs identified and documented.

Promotion

Initial appointments to all federal agencies are processed under competitive examining procedures proscribed and administered by the U.S Civil Service Commission through the media of Interagency Boards of Civil Service Commission Examiners. The major flexibility remaining to
the discretion of the agencies falls within the realm of in-service placement.

The volume of employee turnover within the AFLC is such that inservice promotion accounts for the greater proportion of personnel actions, exclusive of initial appointment, totaling between 20,000 to 30,000 yearly promotion actions. Furthermore, since all employees within an AMA constitute the basic search pattern for merit promotion, typical selection ratios are quite large.

In order to evaluate, consider, and rank order this large competitive population, in 1961 the AFLC developed a computerized promotion system designed to implement provisions of the revised federal merit promotion plan. Operation of this mechanized skills locator and promotion ranking system has been described by Garlet (1965).

Briefly summarized, the promotion portion of the computer system first locates all qualified promotion eligibles AMA-wide, rank orders all eligibles in terms of their desirability for promotion to the next higher grade and generates a promotion register portraying personnel history information for referral to the selecting official(s). Ranking is accomplished by unit weighting three factors: appropriate test score, either aptitude or job knowledge; a qualitative/quantitative rating of each candidate's job relevant training, education, and experience; a rating of the individual's job performance by his immediate and second level supervisors. The rationale underlying this evaluation process is built around the following assumptions:

1. The problem is one of locating an employee in his proper position in multi-dimensional criterion space.
2. The agency is dealing with a truncated distribution of skills resulting from the employment pre-screening process.

3. Since employee records are available to the Agency, relevant "portions" can be extracted and quantified which represent selected dimensions of individual performance.

4. The large selection ratios (Figure 1) preclude manual manipulation of personal records, solicitation of candidates through the "announcement" process, and personal supervisory interviews of all eligible candidates. The data portrayed in Figure 1 represent a typical sample of 245 promotion actions consummated during 1969. It is to be noted that over 30 percent of these actions involved competitive populations, installation-wide, of over 100 eligibles for each vacancy.

5. Optimum placement validity is best achieved through unit weighting experience records, aptitude (tests) and performance and trait (appraisal) information. It should be mentioned that differential weighting of available personnel variables is permitted where supported by acceptable validation.

The promotion process is roughly analogous to the tri-dimensional occupational analysis concept suggested by Ghiselli (1956b) with the three dimensions of administrative characteristics, skills and knowledge and traits. Organization experience and related research on the appraisal dimension of this promotion process are more thoroughly described later in the thesis.

Although the selection procedure described above is followed in consummating the majority of AFLC civilian competitive actions, the machine ranking process is augmented by other screening and ranking
Fig. 1. PROMOTION SELECTION RATIOS (i.e. ratio of total eligibles to a single position vacancy). Sample of 245 promotion actions, consummated, during 1970.
procedures, particularly where higher-level or other critical management positions must be filled. The most common approach is to supplement the machine ranking of eligible candidates through intensive board/panel review of both employee records and individual panel interview.

Other Evaluation Requirements

In addition to the formalized employee evaluation programs outlined above, supervisors are also charged with various ad hoc performance review responsibilities in the following situations: evaluation and documentation against individual written performance standards to support a recommendation for an incentive award (over 50 various monetary and honorary award categories are available to select from); formal written evaluations required to substantiate individual adverse or disciplinary actions; "trial" and probationary period evaluations for employees either newly appointed to the agency (90-day follow-up evaluation) or the federal government (one-year probationary period evaluations). Employees assigned to formalized training programs also have their performance evaluated prior to progression to successively higher stages in the training process. These types of determinations are allowable exceptions to the competitive promotion program where the trainee progresses to any higher grade level.

The formal promotion appraisal detailed in this study is thus only a small portion of the total evaluation process carried out within the federal government. The total process must be regarded as continuous, composed of interlocking evaluation systems, subject to many changes in procedure, and time consuming to the manager.
The organizational environment within which the appraisal is administered is remarkably analogous to a rating situation described by Kingsbury (1925/26) almost 50 years ago in a highly departmentalized office organization. He found that:

1. Managers do not take the very considerable amount of time necessary to operate the system.

2. They do not know the details of work or workers outside their own department.

3. Divisions of the distribution curve are wholly foreign to the manager's ways of thinking.

4. Standards of performance were found to vary considerably between those departments having routine duties and those requiring a heterogeneous spread of tasks.

5. Carefully chosen descriptive adjectives and phrases (on the rating scale) were found to be ambiguous and to have different practical significance in different departments.
CHAPTER II

REVIEW OF THE LITERATURE

When viewed from the broadest possible point of view, most of the published literature covering the topics of ratings, supervisory appraisals and criteria at least touch upon the subject of leniency. A truly comprehensive review would accordingly cover the 500 plus articles abstracted by Mahler (1947), in addition to the estimated threefold increase in related literature subsequently issued.

In the interests of both relevancy and parsimony, the articles discussed in this chapter represent a selective review. Emphasis was placed on studies describing: ratings in the operating environment; reports examining rater leniency (particularly noted instances of chronological leniency); studies reviewing differences between rating supervisors or organizational levels of performance. Primary attention was devoted to those studies where an actual shift in criterion dimensions was reported to occur across various time intervals.

General Problem of Leniency

"Mankind in general is generous....There seems to be a lenient tendency in ratings." These statements represent conclusions arrived at by Kneeland (1929) almost 50 years ago. Using both professional shoppers and customers selected at random, ratings of sales clerks on a millimeter scale "shopper's rating guide" were found not only to
reflect generosity but also an increase in mean ratings between rating periods. No explanation other than the conclusion cited above was offered by Kneeland.

It is evident that the concept of leniency also embraces, and is confounded by, halo. When ratings on a merit (viz trait) rating scale begin to pile up at the favorable—to the ratee—end of the scale, the effects of leniency and halo become inseparable. Both have been labeled as bias or error factors in the assessment of human performance, beginning with the initial labeling of halo as a constant error by E. L. Thorndike (1920). As described by Thorndike: "The magnitude of the constant error of the halo, as we have called it, also seems surprisingly large, though we lack objective criteria by which to determine its exact size." His review of ratings might well have embraced leniency in addition to halo. Supporting evidence for this concept emerges from the several factor analyses of job performance criteria. Examples are Ewart, Seashore, and Tiffin (1941), Rush (1953), Moore (1953), Bryant (1956), and Turner (1960). The emergence of a single general factor arising from these studies variously labeled "ability to do the present job" (Ewart, et al), or "overall-all impression of the supervisor as a leader" (Moore), typically was found to account for the major portion of the performance variance and was usually attributable to the components of halo or bias.

To the extent that leniency was prevalent in the rating scales being studied, there is little doubt that its existence contributed greatly to the general factor(s) described. Although the concepts of
halo and leniency represent different attributes of rater behavior, the
two merge as distributions become more negatively skewed.

The problem of leniency was acutely highlighted during the onset
of World War II where military and civilian managers were faced with
mobilizing and differentiating between large groups of applicants. For
example, Sisson (1948) reports that "in 1940...suddenly discovered that
the years of regular efficiency reporting had provided no basis for the
important decisions that had to be made....Efficiency reports, instead
of showing the 150 best, showed only that of 4000 ground officers of
suitable general officer age, 2000 were superior and best....Selecting
authorities reluctantly fell back on personal knowledge." These state­
ments were graphically confirmed in Baier's (1951) classic réclame to
Travers (1951) "A Critical Review of the Validity and Rationale of the
Forced-Choice Technique." There is no doubt that the trend chart show­
ing the "percentage of Army captains receiving less than excellent
ratings on WD AGO Form 67, 1922 (approximately 72%)--1941 (less than
10%)" graphically documents the influence of chronological leniency.
As acknowledged by Baier, "Attention must be called to the fact that
despite what must be an extraordinary variety of directions to merit
rating forms, high negative skew and leptokurtosis are almost invariably
characteristic of the ultimate distribution. These stubborn character­
istics in fact, have served as motivation for the search for other than
the traditional rating techniques."

The same personnel selection dilemma was faced by other agencies
coincident with the onset of the war. Preston (1948) reported a similar
problem as encountered by the Army Air Forces Officer Selection
Committee charged with the integration of temporary air force officers into the regular army under Public Law 281. Again, "Past efficiency reports were found to be highly unreliable...and that efficiency reports rendered by certain commanding officers were consistently high or low."

Similarly, Weschler, Massarik, and Tannenbaum (1952) found the same situation to prevail for federal workers under the uniform efficiency rating system as described by the U.S. Civil Service Commission.

Theoretically, leniency has been handled within the confines of a variety of concepts. In a sense, criterion dimensionality was suggested by R. L. Thorndike (1949) in that performance was placed along an immediate, intermediate, and ultimate time continuum. His discussion placed leniency under both intrinsic—internal to the perceptions of the rater, and extrinsic—as influenced by events external to the rating situation, bias. This conceptualization was retained as discussed by Brogden and Taylor (1950a), with emphasis upon criterion contamination as influenced by extraneous elements. They defined a biasing factor as: "a variable, except errors of measurement and sampling error, producing a deviation of obtained criterion scores from a hypothetical 'true' criterion score."

Perhaps the most complete theoretical treatment of rater bias is that suggested by Wherry's "Theory of Rating" (1952), op. cit. None of the 46 theorems developed in this article address the topic of chronological leniency directly. Wherry has excluded from his formulae purpose, conscious, or deliberate biases that may be introduced into the rating situation by overt action on the part of the rater. However, it should be noted that one of the earlier formulations in the development
of Wherry's theoretical network [equation (6) page 4], integrating contamination due to environmental influences from Bellows (1941), does embrace the concept of criterion dimensionality. Broadly interpreted, this equation could include chronological leniency. Bellows alludes to the impact of change within the context of criterion instability: "The fact that so many test programs have been abandoned by management would suggest that the criteria used have not been adequate; or if adequate at the start of the program, work and worker situations have changed to such an extent that the criterion originally used for 'validation' lost its usefulness because of such changes."

As mentioned by Sharon and Bartlett (1969), op. cit., leniency has been referred to as a constant error in rating (Guilford, 1954; Ghiselli and Brown, 1955). The label of error is accepted as both theoretically and empirically supportable—the former as derived from Wherry (1952), op. cit., and the latter from uncountable reports, beginning with the findings of the Training Department, Hawthorne Works Western Electric Company (cited in Whisler and Harper, 1962) that "A rating should always be interpreted as having a 'zone of uncertainty.'" The label of leniency as a constant (error) in rating is not as well supported in the literature. Both Guilford (1954) and Ghiselli and Brown (1955) relate the constant error dimension of leniency to noted differences between rating officials. This attribute of rater behavior has been well documented in the literature (to be discussed in more detail in the next section) and is accepted as thoroughly substantiated. The other dimension of leniency error, namely fluctuations in rater tendencies over time, is neither as predictable nor as constant as the between-rater
component of the error factor. More complete discussions of this latter dimensionality factor are thoroughly covered by Ghiselli (1956a) op. cit., MacKinney (1967), and in the excellent development by Ronan and Prien (1966), op. cit.

Another point of contention is the somewhat rhetorical question of how best to measure or "track" leniency. One position is represented by Bass (1956) who pointed out that, in a selection situation, most of the poorer applicants are rejected and a negatively skewed distribution is the most adequate representation of on-the-job performance. On the other hand, Glickman (1955) suggests that approximating the normal curve not only provides more satisfactory ratings from a statistical standpoint but also more completely satisfies the motivation-stimulation functions of performance evaluation. He further proposes that a negatively skewed rating distribution is disadvantageous in that it over-emphasizes the importance of the "miscue," "mistake," or "bust." This position is essentially supported by Guion (1965), quoted as follows:

To be useful it (a criterion measure) must also be able to differentiate between good employees and those who are less good. If employees are so homogeneous in their quality, and if that quality is so high that it can be truly said that "only the best men work here," then there is no selection problem. If however, the quality is not, then even if the group is homogeneous, some means of discrimination must be found. Even among those who are clearly mediocre employees, some are more nearly satisfactory than others; predictors must be found which can make good predictions within this narrow range of talent in order to discover clues about the people who might be satisfactory.

There exists little doubt but what the normal distribution best approximates the "desired" configuration of ratings upon which to base administrative or personnel decisions from the selecting official's point of view. On the other hand, it is easily understood why certain
investigators examining overall organizational rating distributions, would accept the negatively skewed distribution as most characteristic of any pre-selected total workforce. Figure 4 is typical of this J-scale type overall distribution. However, where rating distributions are plotted for various sub-components of the workforce (a situation which more nearly equates to the operational personnel selection system) they once again assume a more normal shape. Thus, whether selecting from among colonels for promotion to general officers, office clerks for advancement to supervisor, or wage grade journeymen for promotion to the foreman ranks, a normal distribution should not only be expected but should also be reflective of the "true" continuum of performance resident within the sub-population(s). A similar conclusion was reported by Ronan and Prien (1966), op. cit., who indicated that individual performance variability is as much a characteristic of the individual as is an aptitude, personality trait, or other more commonly measured characteristic. Accordingly, at least for the purposes of this study, leniency is accepted as properly measured by any departure from the normal curve in a direction toward the favorable (to the employee) side of the scale.

**Performance Dimensionality**

Of interest to the above stated concept of leniency measurement are two separate types of studies tracking changes in performance across various intervals of time. The first sub-grouping encompasses reports of changes in overall performance or criterion reliability, usually in the form of correlation coefficients. The second, smaller series of
studies, actually record performance change over time in a specified direction, with the latter grouping typically reported as changes in rater frequency distributions, rating mean scores, or variances. The series of studies demonstrating variability in criterion reliability have been thoroughly reviewed and integrated into a conceptual framework by Ronan and Prien (1966), op. cit. They are of interest to this thesis to the extent that the variability of reliability is explained by a variety of individual, performance, measurement, and situational factors. The reporting of a decrease in criterion reliability, per se, does not however, provide many clues as to which of the above factors predominate. Consequently, only those studies were chosen for review which attempt to shed some light on the underlying causality of performance dynamics, particularly those where rater leniency was noted as a contributory variable.

Certain reports clearly demonstrate an improvement in performance across time, resulting from influences other than criterion instability. Hay (1947) reported a significant chronological increase in machine bookkeeping speed of posting rates from October, 1937, through June of 1946. The increase was attributed to increased efficiency resulting from the validity of applicant screening by aptitude testing.

The effect upon production of an entirely different influence—learning—was described by Ghiselli and Haire (1960). Following the earnings of 56 unexperienced taxicab drivers over an initial 18-month period of employment, they reported a linear increase in the measured rate of improvement. Significant differences occurred in average performance, variance, and rank order of the workers. It appears
reasonably safe to extend the generalized observations of this study to similar workforce situations where trainee, apprentice, or other novitiate sub-groupings are present.

Studies tracking ratings over time, as opposed to relatively purer production rates, show a much clearer effect of chronological leniency. Beginning with Kneeland's (1929), op. cit., two and three year follow-on studies of ratings collected from customers and professional shoppers, increased mean ratings and decreased variances were associated with rater generosity.

The degree of freedom allowed the rating official in executing his ratings has an apparent influence upon the extent to which chronological leniency is permitted to operate. For example, a forced distribution system will, by its very nature, inhibit leniency. This observation was born out in a study reported by Klores (1966). Ratings by immediate supervisors were distributed into five categories: Best, upper 10%; Next, 20%; Normal, 40% of the group; Next, 20%; Bottom, 10%. Immediate supervisory ratings were subjected to review and "adjustment" by higher levels of management. Covering ratings rendered by 21 different raters in a chemical corporation over a four year period, stable means were recorded for all four groups under study (i.e. professional, supervisory, non-professional, and clerical). An interesting side finding was that the second level higher management adjustment(s) reduced leniency error but not job level bias.

Another attribute of performance ratings which has been observed to influence leniency, particularly the rate of organizational acclimatization, is the stimulus complexity of the rating situation. The more
complex the rating instructions and translation into any resultant score to be used for making selection judgments, the less the susceptibility of the rating instrument to chronological leniency.

An example of this latter influence was documented by Probst (1947) covering over 18 years' experience with the Probst Service Report, described as the original check-list method for evaluating employee performance. The Service Report(s) contained over 100 descriptive items of character traits, work habits, and personal qualifications scored by a complex series of selectively job specific stencils. The general service reports developed were typically applied against positions found in the state and municipal civil service. Reliability of the Service Reports, as reported by Probst, was found to be quite high in over literally thousands of cases. However, in all seven of the examples cited in the text where direction of change across time was reported, change occurred in the favorable direction. Leniency was present, but moved at a much slower pace than traditionally recorded for the relatively less complex graphic rating scales.

A similar inhibited rate of chronological leniency has been demonstrated for forced-choice type rating scales. When comparing the forced-choice rating with other (graphic) scales, results have usually been in favor of the forced-choice scale, which showed less bias effects (Zavala, 1965). It was, of course, this attribute of the forced-choice scale to reduce the rater's ability to produce any desired outcome that led to its adoption in the first place (Sisson, 1948). "The rater who deliberately desires to manipulate his rating can undoubtedly do so. However, the forced-choice technique makes it a somewhat more difficult
the forced-choice scale has also shown greater resistance than the traditional graphic rating to rater bias and leniency when moving from the experimental to the "for keeps" environment (Taylor and Wherry, 1951; Berkshire and Highland, 1953).

In spite of innumerable accounts of disenchantment with graphic rating scales, typically centered around rater bias or leniency, there are surprisingly few actually recorded instances of chronological leniency over substantial periods of time. This same paucity of research was also noted by Ronan and Prien (1966), op. cit., who observed that: "This temporal aspect of job performance is one that has received very little attention, longitudinal studies over any time periods exceeding one year are the exception." The greater portion of longitudinal studies are reported out of military experiences with rating scales. One of the more comprehensive of these (Grappe, Alvord, and Poland, 1967), op. cit., summarized the results of analyses of rating trends and relationships from a twelve-year accumulation on officer effectiveness reports within the Air Force. In almost every category examined (e.g. regular versus reserve officer, rated versus non-rated, major command of assignment), inflationary trends were prevalent for each officer grade—second lieutenant through colonel. Despite several adjustments to both the rating format(s) and the overall evaluation system, the evaluation level in 1965 was found to be some 30% higher than in 1954. Only two exceptions to these chronological leniency trends were noted; both followed the introduction of form changes and were accompanied by initially depressed mean ratings for
short periods of time. These findings parallel those reported by other examiners of the military/governmental rating environment:

1. Baier (1951), op. cit., reporting on Army experience, provides graphic evidence that high negative skew and leptokurtosis are almost invariably characteristic of the "ultimate" (quotes provided) distribution. The graphic material provided by Baier emerged from a trend study of officer efficiency ratings for the period 1922-1945 (Staff, AGO, 1952a) which resulted in the finding that by 1945 almost all officers (i.e. 99%) were rated in the top two of five allowable rating categories.

2. Similar tendencies were reported for the enlisted efficiency report within the Army (Staff, AGO, 1950). Time trends noted in upgrading and halo were slight but detectable. Periodic checking was the method suggested to determine whether or not the value of the enlisted efficiency report is being reduced.

3. The annual efficiency rating system of the U.S. Civil Service Commission has also encountered similar difficulties. For example, within the Bureau of Reclamation in 1941, Yoder (1944) reported the mean rating for 1,307 construction workers as "very good." Only 0.1% of the employees were rated "unsatisfactory."

Experiences with graphic rating scales outside the federal government reflect essentially the same tendency to chronological leniency. Reporting on the results of a two-year graphic rating of professional health personnel, with rater identity held constant, Creswell (1963) attributed all significant differences which occurred to be solely a function of the rise of ratings on successive reports. She concluded that rating leniency simply increases with successive evaluations. This
finding was quite analogous to that recorded by DeWolff (1965) studying employee performance ratings in a large industrial concern over a three year interval. Based upon factor loadings, he found the three factors accounting for the major portion of the variance did not represent constant characteristics of the rated individuals but were specific to the year of the testing. In both studies it was unclear as to whether the rise in ratings was a consequence of rater fallibility (i.e. leniency) or in fact a valid parallel to ratee performance.

Chronological leniency has also been reported as an affiliate of selection instruments other than supervisory performance ratings. In studying the predictiveness of weighted application blank information over four subsequent years (1951 to 1954) it was found that managers were applying "body English" to the collection of application information such that applicants would fall in the desirable categories (Hughes, Dunn, and Baxter, 1956). The authors concluded that during the final year of review (1954), it was clear that application information was not differentiating to any significant degree between competitors. In this particular instance, field managers were given full authority for making appointments and could "manipulate" the application blank information free from administrative controls, reviews, or constraints.

From the studies reviewed in this section, it appears that chronological leniency is most commonly related; inversely to complexity of stimuli confronting the rater in the rating format; proportional to the degrees of freedom provided the rater in assigning numeric ratings along
a continuum from unstructured (typified by the graphic rating) to highly structured (viz. the forced distribution).

**Differences between Raters**

Much of the research, and a great deal of energy on the part of personnel officials, on the topic of supervisory ratings has been directed toward an attempt to get rating officials to evaluate employee performance in some consistent manner under a common interpretation (or usage) of performance standards. The literature is liberally sprinkled with statements such as: "Despite instructions and good intentions to the contrary, it is a fact that different raters apply different standards." (Adkins, 1947); or, "Differences in raters' standards remain the most serious problem in the use of single ratings." (Bayroff, Haggerty, and Rundquist, 1954).

As mean ratings shift favorably across annual rating periods (i.e. chronological leniency), the movement can be ascribed either to actual improvement in employee performance or a shift in rating standards. The latter may occur as evaluating officials, once aware of how other raters are evaluating their subordinates, increase their own ratings during the following rating cycle so as to obtain more favorable salary (or promotion potential) treatment for their own personnel. Thus, a manager's performance standards are constantly changing in a downward direction which leads to increased leniency in the rating system. The phenomenon has been labeled "adaptation" by Helson (1964) who defines adaptation as "changes in evaluation of old observations which occur as a result of new observations."
Another definition of the same difference was coined by Conrad (1932) who referred to spurious differences between the means or the standard deviation of two judges' ratings as the "personal equation." Although his initial review based upon army ratings collected in World War I concluded that the influence of the personal equation was nil, a later review (Conrad, 1933) suggested that the personal equation may be less of a "bogey" than usually considered. The findings of other earlier researchers tended to support Conrad's latter conclusion which referred to the personal equation effect as "the influence of spurious differences in means and standard deviations upon the correlation of ratings with an adequate criterion."

Kingsbury (1925/26), op. cit., based upon repeated conferences with managers, found standards of performance to vary considerably between departments having routine versus varied duties sufficient in their severity to require adjustments in the ratings where interdepartmental comparisons were necessary. In this instance, the adjustments were affected by company executives. The observation was confirmed by Kornhauser (1927) who found both mean and variance differences between various college instructors rating both graduate and undergraduate samples on a seven-trait graphic scale.

The findings of these earlier investigators reporting differences in rater standards have been subsequently confirmed through both commentary (Weschler, et al, 1952, op. cit.; Bittner, 1948) and research findings (Stockford and Bissell, cited in Fleishman, 1967; Preston, 1948, op. cit.; and Bayroff, et al, 1954, op. cit.). The latter study is of particular interest in that the authors were summarizing conclu-
sions based on thirteen Army Personnel Research Branch Reports studying officer rating methodology. Their conclusion to the effect that: "differences in raters' standards remain the most serious problem in the use of single ratings" is accordingly assigned significant weight in this review.

Similar findings have emerged from analyses of interdepartmental differences of recorded (i.e. rated) performance. The dichotomy—interrater versus interdepartmental—is essentially spurious in that both groupings rely on manager's reports of observed performance, within the perceptual constraints of the individual manager's (observer's) strategy of performance assessment. As reported by Lawshe and Balma (1966), studying interdepartmental rating differences in a steel mill, the average ability of employees to do their jobs undoubtedly does vary from department to department, but the supervisors' appraisal will be at least as great a variable. That bona fide differences in interdepartmental performance do exist was shown by Ferguson (1951). Correlating insurance agents' performance with the LLAMA Aptitude Index and the quality of district management, Ferguson concluded that there existed a substantial and significant change in test score validity according to the quality of district or agency management.

Such interdepartmental differences have been labeled a "constant" error by Tiffin and McCormick (1965) who caution that ratings should not be compared across rating groups unless they are adjusted for the differences through either mean score transformation (assuming equal variances) or a standard score transformation. By referring to rater differences shown for 10,000 men working in 14 departments, they further
state that "The difficulty of interpreting the significance of performance rating without reference to the department from which it was obtained may readily be seen." Support was provided by Hemphill and Sechrest (1952) who found it necessary to adjust effectiveness ratings to a standard score (mean = 50 with o = 10) because of marked differences found in the rating habits or biases of the rating officials (wing and squadron officers). An interesting finding emerging from this study was that a significant relationship was found between air crew bombing data (which showed no reliability) and the highly reliable adjusted supervisory ratings. Apparently the ratings were influenced by raters' knowledge of actual performance.

Early formation and stability of the "personal equation" or "generalized rating tendency" was reported by Wiley, Harber, and Giorgia (1959). Studying rater evaluations of rather objective material on well defined five-point scales, they nevertheless demonstrated individual tendencies that were consistent over a four hour interval. These personal tendencies were found to be characteristic of the individual with respect to both his rating mean and sigma, and generally consistent through nine different rating dimensions.

Although there exists little contradictory evidence on the subject of differences between raters, some studies have cast doubt over whether or not it is worthwhile to measure and correct for such differences. One example was cited by Tupes and Kaplan (1965) examining Average Difference Scores on Air Force Officer Effectiveness Reports (OER) for 1,790 different ratees. They concluded that: "even though a reliable measure of rater differences apparently can be obtained, the actual
magnitude of the differences is so small that the effect on the OER system is minimal. Thus, any effort to identify deviant raters and to correct their ratings would not be very effective in improving the OER system."

Other studies (Staff, AGO, 1952b; Staff, AGO, 1952c; and Mandell, 1956) approached the problem by examining the differential validity of ratings rendered by hard and easy raters. In all three studies, differences among the validity coefficients between the "lenient" and "hard" ratee groups were not found to be statistically significant. A possible explanation for these findings emerges from the study by Bayroff, et al (1954), op. cit., who also found no difference in validity between ratings made by hard, average, and easy raters. They also found that, even though different parts of the scale (graphic rating) were used (by different raters), the raters in the three groups placed the ratees in approximately the same rank order. This being the case, lack of differential validity would not be a surprising finding.

Perhaps the best means of summarizing this portion of the review is to echo the sentiments expressed by Barrett (1966) based on his analysis on the influence of supervisor's requirements on ratings: "Instead of attributing low inter-rater reliability to random errors of measurement, we must recognize that raters differ in their judgments largely because they differ in their basic management philosophy." Barrett suggests that someone (outside of the sphere of influence of the rating supervisor) must make the judgment that certain raters simply cannot be relied on and thus exclude their opinions (from the management selection or pay decision process.)
Proposed Solutions

Suggested remedies for the dual and interactive problems of chronological leniency and differences in rater performance standards are as varied and diverse as is much of the performance appraisal literature, per se. In fact, it is entirely possible that many of the innovative approaches to performance measurement were stimulated by the emergence of the two problems in pre-existing appraisal systems.

Typical of many earlier approaches to the problem of leniency/differences between rating officials is the "Annual Board of Review" method of resolving management differences of opinion. One of the earliest accounts of this approach is contained in Samuel Reyburn's Forward to Halsey's (1944) *Making and Using Industrial Service Ratings*. Reyburn, who claimed to be the first manager to apply a plan to teach, rate, and judge employees—Lord and Taylor New York dry goods store—attacked the problem of leniency during regularly held review meetings where corrections were agreed upon where tendencies to rate either too severely or leniently were noted. Halsey more tightly structured the board concept under his "merit system" by convening a formal "Board of Review" prior to each rating period. Objectives of the Board were to:

1. determine whether the standards of judgment used by the various raters were in reasonable harmony and relationship with each other;
2. the Board was not chartered with adjusting ratings of individual employees, rather raters and reviewers in that unit (being questioned) were asked to readjust their ratings in the light of the new standards;
3. in unusual cases of variance, the Board was authorized to administratively make such adjustments as deemed appropriate by subgroup.
Such administrative control procedures persist to the current time and are to be found in the military effectiveness report board review process or adjudication by higher review authority such as the panel review and approval of an "Outstanding Performance Rating" as described by the U.S. Civil Service Commission (Federal Personnel Manual, Chapter 430; revised 1969).

In the search for a common denominator or a common metric for combining sub-criterion measures, Brogden and Taylor (1950b) proposed the concept of "The Dollar Criterion--Applying the Cost Accounting Concept to Criterion Construction." Another example of a proposed substitute to the biases associated with merit rating was the Rank-Comparison method suggested by Bittner and Rundquist (1950). The approach of separation of the criterion group into sub-groups, ranking within sub-groups by a modified paired-comparison approach obviously neutralized inter-rater biases.

Inception of Flanagan's (1949) "critical incidents" approach to personnel evaluation was devised to obtain more adequate definitions of critical job requirements and to require systematic observation of actual job performance and behavior, thereby presumably obtaining greater standardization between various raters' standards of performance. His approach emerged from general impressions, reports, and incidental observations of ratings (and other criteria) studied under the Army Air Forces Aviation Psychology Program (Flanagan, 1948). Of interest is Flanagan's conclusion that: "Studies (of various criteria) have also shown that ratings of specific abilities and traits are seriously biased because the rater, in his observation of the individual's
performance, has been chiefly concerned with collecting evidence for some specific decision which he knows he is going to have to make."

Flanagan suggested focusing the rater's attention on the specific ability or trait on which it is desired that he render a report as a means of neutralizing certain of the rater's tendencies to bias.

The other major rating innovation emerging from the war years was the forced-choice approach developed essentially to combat extensive leniency associated with the predecessor system (Sisson, 1948, op. cit.; Baier, 1951, op. cit.). The relative resistance of the forced-choice approach to leniency has been well documented (Berkshire and Highland, 1953, op. cit.; Zavala, 1965, op. cit.). In spite of its statistical reliability and validity, the eventual demise of the forced-choice rating system introduced an extremely important variable for consideration in devising rating systems— that of rater and ratee acceptability and understanding of the rating process. Baier's (1951, op. cit.) statement to the effect "...that the principal disadvantage of Forced Choice is that the use of the technique had tended to be unacceptable to Army officers...." was based on two dimensions of unacceptability: first, selection and publication of the name "forced-choice," per se, was unfortunate in that raters perceived themselves as being "manipulated" by factors beyond their span of control and could thereby blame "the system" as the scapegoat for resultant questioned personnel actions; secondly, forced-choice ratings (i.e. the raw scores) were converted to standard scores, thereby further leading raters to the conclusion that their ratings were not "properly" represented by a particular standard score. This latter point is especially of interest to this study in
light of Baier's conclusion that "...it is believed that a great deal of the objection encountered by the forced-choice technique has been misdirected and the point to which objection is taken is basically the difficulty of reconciling relative and absolute standards."¹ Baier's observations have been echoed most articulately by Uhlaner (1968): "In the spring of 1945, the Personnel Research Section had begun work on development of a new officer efficiency report. The effort, which sought to achieve objectivity through forced-choice rating techniques, effectively concealing the resultant 'score' from the rating officer, has been termed abortive. From an operational standpoint, it was chiefly because officers preferred to know exactly what rating they were giving their subordinates. From the research standpoint, it was not, since the empirically derived scales produced a broader range of ratings and greater discrimination among individuals than had previous methods." This latter, extremely critical point emphasized by Uhlaner remains the historical "albatross" around the collective necks of applied industrial psychologists. Quite simply stated, any measurement system can be—and frequently has been—rejected, in spite of its statistical adequacy, through various forms of external intervention, political and union pressures, management and employee apathy or unacceptance, and a similar host of external critics.

¹Underlining added. It should be further mentioned that the forced-choice methodology discussed by Baier involved conversion of both raw score forced choice ratings and an overall evaluation of proficiency to a relative standard scale. It is unfortunate that Baier's figure 60.1 is unaccompanied by any direct evidence concerning variations in rater agreement.
Since World War II, a multitudinous and ingenious variety of rating techniques have been suggested (and sometimes applied) as solutions to the biasing influences of leniency and differences in rater standards. Most of these approaches have enjoyed limited usage and acceptance and have been infrequently supported by any type of longitudinal follow-up. A sampling of some of the approaches discussed in the literature is summarized in the following sub-sections:

1. Discriminate binary scoring was proposed and evaluated by Bass (1956) as a means of reducing leniency in merit ratings. Bass essentially suggested substituting a binary (1 or 0 depending on the strength of the rating item) rating instead of the arbitrary 5, 4, 3, 2, 1, etc. (outstanding to unsatisfactory continuum means of element scoring) as a means of differentiating between employees. Based on a limited experimental rating of 350 salesmen, Bass concluded that discriminate binary scoring is no better or worse than summated arbitrary weighted ratings and may offer a compromise solution.

2. A "role-playing" approach was tried by William H. Ward (1961) as a counteractive methodology to rater leniency. Rating officials were presented with the simulated situation of their moving (to another department or company) and being asked to nominate, in rank order 1 + N their successor(s). The suggested method reduced errors of leniency and stringency inherent in absolute ratings by forcing rank ordering, but did not address itself to inter-departmental or inter-rater differences.

3. A complex operations research analysis type approach to rating was suggested by Lamouria and Harrell (1963) and applied against four
heterogeniously constituted research departments. The proposed technique involved differential weighting of department activities (e.g. papers, publications, patents, process development, etc.) related to specifically stated objectives of the laboratory. Weighted scores achieved on the various activities were normalized for combination into a composite criterion. Applying the method to four research departments, the authors reported a "poor" correlation between the composite normalized department ratings and clinical ranking of the departments by the laboratory director.

4. A rather unique approach to merit rating which contains the attributes of both the "management by objectives" and the "management by exception" approaches to performance was proposed by Laney (cited in Dooher and Marquis, 1950). Titled "The Employee Progress Report Plan" the system rates employees on five general traits in four categories: has gone back (in performance); little or no change; has improved; and continued good performance. Personnel action would be initiated only for those employees in the two end categories and individual performance is discussed in terms of employee objectives and development needs. Interrater differences are minimized under the Plan by having each employee rated by three competent raters (which was also observed to be one of the weaknesses of the approach). As evaluated against the criterion of frequency of complaints in the Washington Gas Light Company, the Plan was adjudged to be an improvement over the predecessor rating system.

5. Although not specifically propounded as a counteractive to the problems of leniency and inter-rater differences, the various advocates
of the "management-by-objectives" are attempting to address the "Gestalt" of the manager/employee performance appraisal dilemma. Having its genesis in an entirely new look at the whole subject of management from the point of view of basic social values (Drucker, 1954, op. cit.), the approach has focused on the very interpersonal employee/supervisory goal setting environment as the vehicle for accomplishment. Championed by several reviewers in The Harvard Business Review (McGregor, 1957; Meyer, Kay, and French, 1965; Thompson and Dalton, 1970), in addition to other reviewers (Odiorne, 1963; Kellog, 1965), the approach has strong backing and does resolve the quantitative problems associated with leniency in that summary ratings are avoided all together. This absence of summary ratings could also be perceived as a disadvantage, particularly where large groups of employees must be administratively merged (quantitatively) for the purposes of affecting personnel actions; as opposed to appraisal designed to foster employee development—a goal well served by the management by objectives approach to performance assessment.

6. Various statistical techniques have been suggested as a means of neutralizing between-rater tendencies. Although those articles specifically addressing themselves to standard scaling will be discussed more thoroughly in the following chapter, some limited summarization is provided through discussion of the following critical advocates:

   a. The first application of the use of statistical techniques to eradicate between-rater differences was that of Patterson (1922/1923) with the Scott Company Graphic Rating. The technique was proposed based upon studies (of the Scott Company) which showed that tendencies
on the part of one foreman to rate all of his employees too high and of another to rate too low were not due to actual differences in the abilities of the various groups. The "leveling" technique applied was to divide each rating foreman's frequency distribution into five parts so that the highest 10% of the total scores were given a final letter rating of A, the next 20%—B, next 40%—C, next 20%—D, and lowest 10%—E. The conversion procedure thus converted actual ratings (on a graphic scale) into relative ratings and decidedly did counteract for between-supervisory rating tendencies. It is most unfortunate that few reports of actual usage of Patterson's suggested technique have appeared in the literature since 1922.

b. Balanovich (1946), based upon a factor analysis of a 14 item merit rating collected on 200 field service engineers widely dispersed geographically found wide differences between both the means and variances of rating supervisors. His proposed solution was the development of statistical quality control charts displaying rater tendencies (i.e. both $\bar{x}$'s and $\sigma$'s) and graphically portraying those raters "out of control" in either direction from the overall mean rating and mean variance. The quality control charts were shown to, and discussed with, assistant general managers who were, in turn, expected to control deviant rater tendencies (presumably during the next rating cycle). Thus, although no statistical "corrections" were actually applied to rater differences, Balanovich's approach was predicated on the premise that: "The statistical problems involved (i.e. between management and workers) are to establish and administer standards agreed upon so that in their application they are the same as they were originally intended to be."
c. A similar approach was suggested by Benge (1946) for use in making combination judgments regarding job evaluation, merit rating and incentive pay. His "JEM" system was built around a system of statistical tables to offset rater bias. "JEM's" were actually decile divisions of the rating distribution(s) with leniency measured by the average of each rater's ratings and dispersion measured by the rater's average deviation (with both measures converted into standard "JEM's" by reference to the appropriate table). Unfortunately, Benge's text contains no empirical descriptions or results of actual application of the "JEM" approach.

d. Rating 405 employees (by 21 supervisors) on four overall item measures of job success, C. E. Jurgensen (1950) reported a high rater reliability of $r = .88$ after converting each item rating into normalized standard scores with a mean $= 50$ and $\sigma = 10$. He concluded the approach to be a promising and reliable technique.

e. The most complete discussion of the techniques, advantages, and possible pitfalls of converting ratings to standard scores is that as described in Guilford (1954), op. cit. Guilford opens the possibility of applying linear transformations as a means of obtaining equivalent distributions where distributions of ratings of the same set of objects made by different raters differ with respect to their means and sigmas. The rather laborious linear transformation process developed by Guilford carries with it one requirement infrequently encountered in the actual operational rating situation: namely, roughly equivalent knowledge of all objectives to be rated (i.e. employees) by a sizable group of evaluators, with N preferable
Such a fortuitous situation is simply not encountered that frequently in the industrial shop except where the management environment would allow for peer evaluations.

Guilford further acknowledges that ratees may benefit or be discriminated against unduly because they happen to be in a certain group (if the linear transformation were to be applied non-differentially against all groups). To further elaborate, Guilford states:

"There is no simple, generally applicable solution to this problem. To the extent that any two or more raters have ratings in common sufficient to make the kind of study of ratings that was described above (i.e., multirater/multiratee), something can be done to make adjustments. Linear transformations taking care of differences in means as well as differences in standard deviations would be important in this kind of situation. If one is willing to make assumptions concerning comparability of subgroups of ratees, one extends the possibility of making inferences about the amounts of errors of different kinds." As discussed earlier in this chapter, with supporting opinions from Glickman (1955), op. cit., and Guion (1965), op. cit., it is felt that such assumptions can safely be made for many competitive industrial selection situations.

In an admirable attempt to simultaneously solve the three-fold problems of leniency, and intra- and extra-departmental differences, Ross (1966) proposed a rating process based upon the use of reference groups in man-to-man performance ratings. He tried a rating procedure requiring the rater to nominate out-of-department people whose job performance he knew well and ranking a subordinate on overall job
performance within the "out-of-department" reference group. An index was then computed from these man-to-man comparison rankings which was in turn, compared with ratings from an anchored rating scale for their validity in guiding salary determinations in a research and development organization. The conversion index ($\kappa$) proposed by Ross was essentially a unit weighted standard scale with a mean = 50 and $\sigma$ = 5. He concluded that the man-to-man comparison procedure was found to be as valid as the anchored ratings, although it was interesting to note that the mean of a one-year follow-up of the anchored rating scale ratings was higher than the previous year's mean rating for the same 68 subordinates. Ross's proposed technique is, of course, reminiscent of—and carries with it the same perils—the man-to-man rating scale introduced into the Officer's Training Camps, United States Army, in 1916, by Walter Dill Scott (cited in "Landmarks of Professional Psychology," Fryer and Henry, 1950). Both approaches are subject to the persistent problems of rater understanding and acceptance. There is little doubt that these factors eventually discredited the man-to-man rating scale of World War I, according to accounts by Rugg (1922) and D.G. Patterson (cited in Fryer and Henry, 1950, op. cit.). Patterson's account indicates that the method was too time-consuming, cumbersome, and difficult for the average officer to understand, and that the ratings were adjudged to be too unreliable.² It is believed reasonably safe to

²Unreliability of the man-to-man rating scale mentioned by Patterson was based on Rugg's earlier (1921/1922) studies at Camp Taylor of 3,000 "civil like" captains and 6,000 lieutenants promoted to captain where very low (i.e. average $r = .15$) correlations were found between test scores and the man-to-man ratings.
assume that Ross's proposed approach would encounter essentially the
same problems of rater resistance if extended for any length of time
within the actual industrial situation. Patterson's proposed solution
(which carries with it some of the same problems of rater salability)
was to avoid using extremely derogatory steps combined with a statis-
tical correction to remove the error of the "personal equation" by
transforming ratings of each rater into a common distribution.

Regardless of the underlying metric, methodology, technique, or
accompanying rationale, all systems of human performance measurement can
be regarded as two dimensional. All evaluation systems are dependent
upon the interactive attributes of the methodology, per se, and the
understanding and willingness of the administrators of the system to
make it work. For the various merit rating systems, this latter factor
takes the form of rater training, thereby attempting to obtain common
understanding regarding how the various dimensions of the absolute
merit scale (be they in the form of adjectival descriptive categories,
suggested percentage breakouts, "anchored" descriptive statements, or
other metric shredouts) should be universally applied by all evaluators.
That such rater "training" is typically unsuccessful is supported by
various reports of between-rater differences and the failure of merit
systems to withstand the onslaughts of chronological leniency as
previously documented in this chapter.

With evaluation systems other than the merit rating (or variants
thereof) the second factor alluded to above becomes one of rater/
managerial acceptance and cooperation. Problems of rater understanding
of the evaluation process become minimal the more tightly structured
the rating task becomes. For example, under a forced-distribution rating system, the responsibilities of the rating supervisor are self-explanatory and require little or no rater training. What is required (and very seldom obtained) is the cooperation and willingness of the rater to accept, judiciously apply, and continuously and effectively support the system to his subordinates. Raters seem to possess the persistent and peculiar quirk of assessing all rating systems from their vantage point as a rating supervisor rather than from their equally responsible position as a selecting, promoting, or pay awarding official. Regardless of the value of any rating system to effectively serve these latter functions of management, they persist in rejecting (or infrequently accepting) appraisal systems based solely on their perceptions of how much difficulty it (i.e. the system) presents them as a rating supervisor. McGregor's (1957), op. cit., observation that managers are uncomfortable when they are placed in the position of "playing God" is obviously relevant to such frequent rejection of various appraisal systems.
CHAPTER III

APPRaisal RESEARCH WITHIN THE AFLC

Since each appraisal system applied within the AFLC during the 1961-1969 time frame was accompanied by in-service research, this chapter is designed to summarize this collection of studies independent from the overall literature review. Three separate—but not necessarily independent—approaches were tried. Experience with each is described from developmental to operational stages with appropriate research summarized associated with each separate application. It is believed this chronology will serve as a useful background to, and a conceptual lead-in for, introduction of the linear transformation in 1969.

The first common application of a standard performance appraisal system within the AFLC was in 1961, coincident with agency-wide implementation of the Federal Merit Promotion Program. Prior to 1961 the various federal agencies were permitted to develop agency-specific promotion plans, with little regularized homogeneity of procedures. The overriding concept behind this "new" promotion program was in-service promotion of civilian workers in terms of the overall merits

---

The research studies reported in this chapter represent AFLC formalized reports typically unpublished, occasionally available through the Defense Documentation Center, Alexandria, Virginia. Copies of the studies or relevant extracts may be obtained from the Research Branch, Headquarters, Air Force Logistics Command.
of competitors leading, in turn, to selecting of those individuals "best qualified" for advancement. It should be reemphasized at this point that the awarding of a promotion serves both management and the individual as the primary mode of monetary recognition for sustained performance.

For several years prior to 1961, the primary supervisory appraisal used within the AFLC was based upon an overall prediction of promotion potential, with the rating official recording the degree of confidence he had in his rating. Both judgments were recorded on a graphic rating scale. By 1958 the sensitivity of this instrument had decreased to the point where it was practically worthless for differentiating between competitors. The overall command average was 18.0, out of a possible total score of 20, with a standard deviation of 2.10. Of the total population of 145,000 civilian employees, 38 percent had the maximum score of 20, with around 80 percent of employees in the higher professional and managerial grade levels also being awarded the perfect score.

Command experience with the three appraisal formats applied since 1961 is graphically portrayed in Figure 2. Average scores are plotted along the ordinate in terms of average proportional values since maximum scores attainable varied depending upon the total number of scorable items which varied for each form. Each of the three forms used during this time frame had the following administrative attributes in common:

a. The primary operational goals of each form were to differentiate between satisfactory and "highly qualified" groups of competitors and to generate a promotion certificate of the five to eight "best
Fig. 2. COMMAND AVERAGE appraisal score, 1961 - 1969.
qualified" individuals. Sublimated objectives were to globally discuss individual employee development needs, to counsel employees regarding present job behavior and to formalize the supervisory/subordinate performance discussion at least once annually.

b. Each employee was rated once yearly by his immediate supervisor of record with the rating formally confirmed by the second level supervisor.

c. It was presumed that the performance rating represented the most efficient means of reflecting supervisory judgment regarding current levels of job relevant performance.

d. The annual rating was assumed to be based upon twelve months' continuous evaluation of performance and was expected to be reasonably consistent with other required performance evaluation programs as described in Chapter I.

e. Primary reliance was placed upon management control, through continuous supervisory training, to counteract the traditional biases of halo, leniency, dependence upon memorable incidents, etc.

f. Employee complaints with their rating were resolvable only through informal discussion and appeal up through the third level of management.

Pre-implementation and operational research experiences with the three formats applied from 1961 through 1969 are summarized in the following sections.
The Personal Characteristics Appraisal (1961-1964)

Development

The first appraisal format (Appendix 1) developed for accommodation into the promotion field of the computerized Skills Locator System was built around a personal trait approach. This approach assumed that cognitive dimensions of promotable behavior were better reflected in the quantifiable test and experience components of the ranking system. It was further concluded that elicitation and priority-ordering of these unrecorded attributes would lead to greater acceptance if they came from the ranks of the AFLC supervisory corps.

Development, validation, and format construction of the resultant "Personal Characteristics Appraisal" (PCA) emerged from the two interlocking research studies, monitored by the Personnel Standards Branch, Civilian Personnel Division, Headquarters Air Materiel Command (AMC, the command title of the predecessor to the Air Force Logistics Command). These unpublished studies are summarized as follows:


The initial list of personal traits were elicited from open-ended questions administered to two independent samples of 110 supervisors, each assigned to Hill Air Force Base, Utah (O0AMA). Each supervisor rated a separate subordinate position with samples stratified by representative job series distributions at that activity. The open-ended form contained the two main headings of Negative and Positive Attitudes,
Abilities, Traits and Characteristics, with sub-headings under each heading as "Very Important" or "Desirable" positive or negative factors. A maximum of ten statements was permitted for each of the main headings. Supervisors were asked to describe attributes related to successful performance required of the position, not the incumbent. Responses were condensed, through consensual agreement of the Headquarters AMC research staff, into seven general grade groupings: four white collar and three Wage Grade. The number of trait statements finally selected was constrained by an administrative decision that the computer field could accommodate a total of only 35 three-digit positions with a maximum of 14 positions allowed for any derived trait cluster. The headquarters distillation process resulted in a final list of 27 attributes. No reliability or weighting data were reported. The final list of elements were subjected to further cross-check through replication of the original study at Brookley Air Force Base, Alabama. Eighty percent of the final 27 items were reported to be found common to the two separate activities. The items were again distilled through staff agreement into 27 separate trait statements with combinations of 14 traits similarly established for seven broad occupational groupings (Table 1). All traits on the form were to be completed for all employees, with only the 14 occupational specific traits to be scored for promotion ranking to any vacant position.
### TABLE 1

**ELEMENTAL PATTERNS BY OCCUPATIONAL GROUP—PERSONAL CHARACTERISTICS APPRAISAL**

<table>
<thead>
<tr>
<th>Element</th>
<th>White Collar</th>
<th>Blue Collar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clerks</td>
<td>Admin</td>
</tr>
<tr>
<td>1. Quality of Work Produced</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2. Volume of Work Produced</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>3. Cost Consciousness</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>4. Compliance with Policies and Procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Leadership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Oral Expression</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7. Written Expression</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8. Personal Appearance and Habits</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>9. Interest and Enthusiasm</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10. Authority Acceptance</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>11. Loyalty</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>12. Ability to Work Under Pressure</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>13. Self-Reliance</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>14. Organizing and Planning Ability</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>15. Problem Solving Ability</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>16. Dependability</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>17. Getting Along with People</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>18. Initiative</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>19. Attendance</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>20. Work Habits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Original and Creative Thinking</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>22. Cooperation</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>23. Safety Minded</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>24. Adaptability</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>25. Desire for Self Improvement</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>26. Responsibility Acceptance</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>27. Self Confidence</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
Study Number 2 - "A Study of Supervisory Appraisal Forms," Headquarters AMC, January 1960

The new form was tried experimentally in November 1959 along with the Air Force Form 77, Officer Effectiveness Report, and Air Force Form 75, Airman Performance Report. Ratings on all three experimental predictors were collected on representative samples of 1500 civilian employees of each of two subordinate AFLC activities, the Oklahoma City Air Materiel Area and Wright-Patterson Air Force Base, Ohio. The dependent variable was an evaluation of overall job performance rendered by five raters, each familiar with current job performance of each employee in the sample. The employees' immediate and second level supervisors were not accepted as evaluators of overall performance.

Any employee who had two or more of the five raters differing in their overall evaluation by more than three raw score points (on the overall performance scale from 0 to 20) was dropped from the sample. 2 The application of this procedure resulted in a final sample of 247 "agreed upon" cases at Wright-Patterson Air Force Base and 206 cases at Oklahoma City Air Materiel Area.

Correlation between the experimental forms and the overall average rating are shown in Table 2. All coefficients are significant at greater than the .01 level.

2 It should be pointed out that the application of such a procedure eliminates not only a great deal of essential variance from any resultant correlations but also excludes those employees whose performance must be incorporated into the analysis—particularly those at the higher levels where differentiation is essential for promotion ranking purposes.
TABLE 2

CORRELATIONS BETWEEN SUPERVISORY AND AVERAGE OVERALL RATINGS

<table>
<thead>
<tr>
<th>Format</th>
<th>WPAFB (GS) (N = 197)</th>
<th>WPAFB (WB) (N = 50)</th>
<th>OCAMA (Total) (N = 206)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Characteristics</td>
<td>.45</td>
<td>.56</td>
<td>.54</td>
</tr>
<tr>
<td>Officer Effectiveness Report</td>
<td>.46</td>
<td>.57</td>
<td>.63</td>
</tr>
<tr>
<td>Airman Performance Report</td>
<td>.40</td>
<td>.42</td>
<td>.50</td>
</tr>
</tbody>
</table>

The combined results of the experimental study and administrative considerations discussed previously led to adoption of the Personal Characteristics Format modified by adaptation of a five-point rating scale for each trait similar to that used in the Officer Effectiveness Report.

Research During Operational Phase

The Personal Characteristics Appraisal was included as an element of study in five U.S. Air Force-Air Force Logistics Command research reports during its application. The findings of these studies are summarized as follows:

1. The first study on the new form was initiated by Larkins (1963) to investigate the stability of the seven job profile clusters of the PCA over a two year period. Random (not matched) samples of 100 employees for each of the seven patterns were collected in 1962 at each of four AMA's and contrasted with similar distributions collected at the same activities during the first year of application—1961. Three conclusions emerged from this one-year follow-up study:
a. Mean ratings were more stable for non-supervisory than for supervisory appraisal patterns.

b. All distributions (both 1961 and 1962) were characterized by negative skew and high peaks.

c. Intensive rater training was emphasized as a remedial method for stabilizing PCA distributions.

2. Acting on recommendations emerging from the previous study, rater training was intensified on a trial basis in an attempt to counteract chronological leniency. Rater briefings and a printed discussion of factors involved in appraising employees (e.g. halo, leniency, utilizing the full range of the rating scale) were interpolated between successive appraisals of employees in 1962 under controlled conditions by Larkins (1964a). The study hypothesized that such rater training should affect rate-rerate appraisal distribution parameters of individual appraisal scores. Evaluation of changes in mean ratings, after training, showed that the means were not significantly reduced. In fact some samples, after training, reflected significant increases in mean ratings and significant decreases in variances. It was concluded that the strength of other influences on raters was too great to be overcome by the training.

3. In an attempt to obtain an independent, more objective evaluation of the Personal Characteristics Appraisal in the operating environment, a search was made for a job where bona fide production data existed and was consistently recorded on an individual basis. After an extended and exhaustive search by ten psychologists at 10 separate activities, covering the range of over 900 separate jobs, only
one job was discovered where quantitative and qualitative data had been systematically maintained. Accordingly, the operational appraisal (PCA) and an independent experimental appraisal form were correlated with data gleaned from production records on 39 subjects working as Electronic Equipment Repairer Helpers. Details of the search and resultant study were reported by Ward (1964). For both forms, the correlation was significant at greater than the .01 level. More importantly, the study produced the distressing conclusion that usable quantitative production data were, for all practical intents and purposes, non-existent for the very large proportion of civilian positions within the AFLC. A supervisors evaluation of performance was agreed upon to be the remaining acceptable substitute to formally apply within the context of the merit promotion process.

4. The Personal Characteristics Appraisal was included in an overall research review of the Air Force Merit Promotion by Seidman in 1963, along with other appraisal formats authorized for Agency use. The basic intentions of the study were to longitudinally trace employees advanced through merit promotion and to identify, through supervisory contacts, primary areas of concern for future research purposes. Both operational appraisal forms tested in the study were concluded to be satisfactory predictors of promotion to higher level positions at a minimal but satisfactory (i.e. < .05) level of significance. Evidence collected through supervisory interviews indicated dissatisfactions with the existent form(s) falling into two classes:

Class I--Problems in appraisal format: relevance of rating elements, rigidity of elements in the scale, the requirement
that all elements be rated against all employees regardless of job relevancy.

Class II—Problems inherent in the rater: halo, rater biases, leniency.

As Seidman pointed out, an interesting facet of this dichotomy is that by offering solutions to Class I problems, potentially greater errors in Class II problems may be permitted. That is, as rating officials are afforded more flexibility in selecting elements and "tailor-making" the system to meet individualized peculiarities of subordinate positions, wider dispersion would be expected to occur between raters' biases, standards of performance, and rating tendencies. Conversely, any actions taken to reduce between-raters' variance would tend to over-structure the rating situation and lead to increased supervisory resistance to the rating task.

5. The chronology of the four years' operational application of the PCA was described in a study by Wilfong (1965). Yearly mean ratings on each of the seven occupational patterns for each year are shown in Table 3.
TABLE 3

MEAN RATINGS (PCA) BY YEAR
BY OCCUPATIONAL GROUP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clerical</td>
<td>52.58</td>
<td>50.84</td>
<td>51.78</td>
<td>56.05</td>
<td>52.44</td>
</tr>
<tr>
<td>2. Administrative</td>
<td>53.36</td>
<td>53.84</td>
<td>54.63</td>
<td>58.82</td>
<td>53.56</td>
</tr>
<tr>
<td>3. Professional/Scientific</td>
<td>55.95</td>
<td>55.89</td>
<td>57.05</td>
<td>59.91</td>
<td>56.15</td>
</tr>
<tr>
<td>4. Managerial/Supervisory</td>
<td>57.10</td>
<td>58.19</td>
<td>57.41</td>
<td>62.28</td>
<td>57.37</td>
</tr>
<tr>
<td>5. Unskilled</td>
<td>45.40</td>
<td>49.67</td>
<td>46.70</td>
<td>52.30</td>
<td>45.67</td>
</tr>
<tr>
<td>6. Skilled</td>
<td>52.03</td>
<td>51.51</td>
<td>52.84</td>
<td>55.85</td>
<td>52.12</td>
</tr>
<tr>
<td>7. Wage Foreman</td>
<td>55.42</td>
<td>56.39</td>
<td>56.05</td>
<td>60.64</td>
<td>55.70</td>
</tr>
<tr>
<td>TOTAL</td>
<td>51.39</td>
<td>51.04</td>
<td>53.78</td>
<td>57.29</td>
<td>51.81</td>
</tr>
</tbody>
</table>

Differences between overall row and column means were tested with the Kruskal-Wallis One-Way Analysis of Variance. Differences between mean ratings, over years, were shown to be insignificant (p > .20).

Differences between overall means between occupational groups were found to be significant at p < .01. Differences between the individual group means, tested with the Mann-Whitney U Test, showed them to cluster in three general groups: (1) white-collar non-supervisory (groups 1, 2, and 3); (2) blue-collar non-supervisory (groups 5 and 6); (3) supervisory (groups 4 and 7).

The Job Performance Appraisal (1965-1968)

Responsive to increasing chronological leniency (Figure 2) and supervisory and employee resistance to the Personal Characteristics Appraisal (Seidman, op cit), an administrative revision was designed and tested in 1964 (Appendix 2). The new format was essentially a re-engineered version of the PCA with the following changes:

1. Rating elements were rewritten in a more job relevant style,
attempting to avoid the objectionable personal trait orientation of the predecessor version.

2. The 6-point rating scale was re-ordered on a percentage rather than an adjectival basis. The new rating scale, adopted from a Headquarters U.S. Air Force proposal, was also designed to increase the number of "above-satisfactory" categories in an attempt to counteract leniency.

3. The computer routine was reprogrammed to provide the rater with greater flexibility in the choice of job relevant elements. Under the new system he had his choice of rating any combination of 14 traits as opposed to being required to rate all 27 elements under the PCA system.

Development

Supervisory ratings on the experimental form (JPA) and the Personal Characteristics Appraisal were collected from immediate supervisors for random samples of 200 employees at each of four AFLC installations. (Total N = 685 finally accepted pairs of ratings). Comparisons between frequency distributions plotted for both forms (Figure 3) showed marked differences in favor of the revised form (Larkins, 1964b). Differences in both means and standard deviations were found to be significant far beyond the .001 level for all four installation samples. The differentiating results for separate grade level samples further were accepted as demonstrating the experimental superiority of the proposed appraisal format. Although admittedly obtained in the experimental environment,
Fig. 3. COMPARATIVE FREQUENCY DISTRIBUTIONS, Operational (Personal Characteristics Appraisal) vs Experimental (Job Performance Appraisal), Ogden Air Materiel Area (N = 685).
statistical findings of the study, coupled with an anticipated increase in face-validity of the revised format, led to its operational acceptance.

It should be noted at this point the questionableness of assuming that appraisal distributions derived from the experimental setting may retain their purity when moved to the actual "for keeps" situation. As shown by Taylor and Wherry (1951), the average of graphic ratings is much higher in the for keeps than in the experimental situation, with a correlated marked distortion of the shape of the distribution. This conclusion is also confirmed in Figure 2.

Research during Operational Phase

After implementation, the Job Performance Appraisal (JPA) was included as a subject of study in two separate research reports summarized as follows:

1. In an attempt to identify relative halo effect, Ward (1965) conducted a follow up study on a stratified sample of 96 individuals rated in 1963 on the PCA and again in 1965 on the JPA. Although identification control was built essentially around the employee, 80 percent of the final 1965 group ended up being rated by the same supervisor who had completed the 1963 PCA evaluation. Halo was estimated along four dimensions as shown in Table 4.
TABLE 4
1963-1965 COMPARISONS ON HALO INFLUENCE

<table>
<thead>
<tr>
<th>Dimension</th>
<th>PCA 1963</th>
<th>JPA 1965</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number of raters marking in a single column</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2. Number of raters marking 19 or more elements in a single column</td>
<td>43</td>
<td>20</td>
</tr>
<tr>
<td>3. Number of raters using no more than two columns</td>
<td>40</td>
<td>19</td>
</tr>
<tr>
<td>4. Number of raters marking 14 or more elements in highest scale category (halo and leniency measure)</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

With the exception of dimension number 4 which represents both halo and leniency, it was concluded that introduction of the new form successfully curtailed halo effect.

2. An evaluation of the Job Performance Appraisal was included in an Air Force validation study of aptitude tests used for promotion ranking (Trattner, 1968). The study covered test application for a sample of 787 employees assigned to eight separate job categories. The dependent variable was a first level supervisory evaluation on 20 tailor-made traits, aptitudes and abilities adjudged to be critical determiners of successful job performance. Raters were aware of the experimental, as opposed to operational, application of their evaluations. Predictor variables were 8 aptitude tests current in the Air Force inventory at that time and the operational appraisal (the JPA). The operational appraisal was introduced essentially as a check on criterion reliability. Results were:

a. The tests correlated higher with the special test research criterion than with the JPA, perhaps because the raters were
assured their research evaluations would have no bearing on the
subjects' career progression and under the assurance that the subjects
would not see the ratings.

b. The average correlation between the two ratings across
the eight job categories was .28.

c. The greatest variance in validity seemed to be related to
job group rather than test performance. If one job group correlated
highly with one predictor, it tended to correlate highly with the other
aptitude measures.

The Tailor-Made Appraisal Forms (Beginning 1968)

Emerging from conclusions derived from Seidman's early review
(1963) of the merit promotion process, the Air Force began exploring
the feasibility of devising "tailor-made" supervisory appraisals. This
approach assumed that decreasing the structure of the rating situation
would result in higher supervisory acceptance of the evaluation process,
with an increase in validity. The basic methodology for developing
"tailor-made" appraisals emerged from a study of quality performance
among Air Force scientists and engineers conducted by Seidman (1964).
The study, which examined a myriad of predictors of promotion potential,
produced the conclusion that: "The most valid reflection of management's
judgment concerning individual productivity and worth of any given time
can be achieved only through a carefully constructed supervisory
appraisal and trained, cooperative supervisors to carry out the program.
Design of the supervisory format should be formulated with the help of
using supervisors to insure communality of relevance of the job elements
on which employees are to be rated."
The element elicitation procedure emerging from this study was labeled the "Job Oriented Supervisory Appraisal and Training Technique" (JOSATE). Briefly summarized, JOSATE involved the development of rating elements using the following steps:

1. Convening a group of operating supervisors to identify critical job behaviors suitable for rating job performance. The group is instructed to think in terms of both relevant and observable characteristics of the job.

2. The resultant list of job duties and behaviors are regrouped into three levels of generalization:
   a. Diffuse—factors homogeneous across the entire spectrum of jobs.
   b. Restricted—factors homogeneous or restricted to a job grouping of various grade levels.
   c. Specific—factors specific to only a few grades or job series within a work family.

3. A rerating of the rephrased job elements using the J-Coefficient rating technique proposed by Primoff (1957). This step identifies elements at three levels: 0—unimportant and not present; 1—moderately important; 2—extremely important, present and observable.

4. Step 3 above is replicated using an independent sample of supervisors of the same job groupings, convened at another AFLC installation.

5. Common elements having the highest J-Coefficient values and cross-installation reliabilities are accepted as the final rating statements chosen for inclusion in the operational form.
The initial series of tailor-made forms were developed around the three general job groups of white collar non-supervisory, blue collar supervisory, and supervisory/managerial (Wilfong, 1965, op. cit.). Another grouping was implemented, from the elements derived by Seidman (1964), for scientific and engineering personnel. An excellent description of the developmental procedure for the blue collar supervisory profile is contained in a study by Baddley and Larkins (1967). The operational form emerging from this study is reprinted in Appendix 3. It should be noted that the basic rating scale format previously used for the JPA was retained in the tailor-made series of formats. The tailor-made concept was eventually expanded to include a variety of specialized occupations listed below in Table 5 (Department of the Air Force, 1966).

**TABLE 5**

**AIR FORCE TAILOR-MADE APPRAISAL PATTERNS**

<table>
<thead>
<tr>
<th>Occupational Group</th>
<th>No. of Rating Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Supervisory</td>
<td>27</td>
</tr>
<tr>
<td>Wage Board—Non-Supervisory</td>
<td>15</td>
</tr>
<tr>
<td>Class Act—Technical</td>
<td>16</td>
</tr>
<tr>
<td>Clerical</td>
<td>18</td>
</tr>
<tr>
<td>Scientists and Engineers</td>
<td>25</td>
</tr>
<tr>
<td>Auditors</td>
<td>30</td>
</tr>
<tr>
<td>Operations Analyst</td>
<td>26</td>
</tr>
<tr>
<td>Intelligence</td>
<td>35</td>
</tr>
<tr>
<td>Counter-Intelligence</td>
<td>33</td>
</tr>
<tr>
<td>Middle and High Level Managerial and Executive</td>
<td>30</td>
</tr>
</tbody>
</table>
Summary

The investigations summarized in this chapter are by no means free from the possibility of serious criticism. The studies were by-products of specific functional undertakings, analyses were not as uniform in plan and organization as could be desired and conditions were not always favorable to a straightforward, logical research approach. For these reasons, comparisons with other similar research studies may sometimes be occluded. In addition to the typically encountered problems of employee attrition, changes in mission and management, advancing technology with correlated encroaching mechanization, etc., the Command encountered other severe upheavals during the 1960's which had a definite influence on the appraisal program and many of the research studies cited in this chapter. Foremost among these changes were: a drastic change in command composition during the early 1960's; the close-out of four depots (San Bernadino Air Materiel Area, Mobile Air Materiel Area, Middletown Air Materiel Area and Rome Air Materiel Area) in the 1964/1965 time frame, which resulted, in turn, in mass transfers of both workload and personnel (some 38,000 employees were involved in the overall exercise); large hiring programs in 1966/1967 to accommodate mission requirements to support the Southeast Asia combat build-up; frequent changes in personnel policy which influenced directly attitudes toward and constraints surrounding the appraisal program. Consequently, most of the research was conducted under conditions which caused uneven cell frequencies, changes in personnel distributions, missing samples, and other similar problems mitigating against what would be regarded as straightforward and pure designs of experiment. Nevertheless, it is
believed that reasonable generalizations and confirmatory parallels can be made between these studies and those reported by other investigators.

1. Favorable findings with each form during experimental phases with subsequent successive deterioration during operational usage (i.e., increasing means and negative skew) confirm the findings of Taylor and Wherry (1951), op. cit.; Berkshire and Highland (1953), op. cit.; and Wherry's (1952), op. cit., Corollary (11a): "Ratings obtained under experimental conditions (i.e., to be used only to improve instruments, methods, or the like for the good of the organization) will be more accurate than those obtained under actual on-the-job conditions where resulting administrative action will or may affect the ratee."

2. Trait clusters emerging from development of the tailor-made series of forms under a modified J-coefficient approach confirm the factor analysis findings of Stogdill, Shartle, Wherry, and Jaynes (1955). Analyzing 46 variables for 45 different positions located in 47 different organizations, they found eight factors anchored to occupational clusters, substantiating the hypothesis that job behavior is essentially homogeneous regardless of specific job assignment.

3. Larkins' (1964a) experience with rater training replicates the finding of Stockford and Bissell (1949), op. cit., who also found no difference between untrained and trained supervisory groups in the leniency with which they rated their employees.

4. The introduction of changes in rating format, rating instructions and evaluation elements and scales were initially accompanied by decreased means and leptokurtosis of the overall distribution followed by increased chronological leniency with further usage. This finding
replicates almost exactly the same phenomena reported by Grappe, et. al. (1967), op. cit. To a certain extent, failure of changes in rating scale format to substantially modify rater behavior also support the findings of Blumberg, Soto, and Koethe (1966). They reported subjects to be the main source of variance in their study; with changes in scale format, numeric versus graphic, or singular versus matrix displays of rating objects to have little bearing on rater variance.

5. From reference to Figure 2, it appears that ratings were somewhat more lenient on the Personal Characteristics Appraisal than on the Job Performance Appraisal; the same finding as reported by Taylor, Barrett, Parker, and Martens (1958) in their series of studies on the relationship of scale formats to graphic ratings.

6. Employee and rater dissatisfactions with all three rating formats appear to have been centered around the two primary sources of: (a) differences between rating officials leading, in turn, to chronological leniency and a perceived penalization for those employees unfortunate enough to find themselves under either a "tough-minded" rater or a rater who (in his "naïveté) continued to rate in a somewhat objective manner, and; (b) raters tended to anticipate the actions that would be triggered by their ratings and consequently modified their ratings in an attempt to "beat" a system which has become both familiar and unpleasant. Both of these sources of dissatisfaction have been noted by other investigators (e.g. Whisler and Harper in Whisler and Harper, 1962, op. cit.; Weschler, et. al., 1952, op. cit.; Guion, 1965, op. cit.).
7. Samuel W. Ward's finding (1964), op. cit., that the only available performance measure was a supervisor's appraisal is reminiscent of the frequently encountered phrase in the literature to the effect that: "The best available criterion was a rating," Guion (1965, p. 96), op. cit. This finding is also supportive of the surveys of both company rating plans (Lopez, 1968, op. cit.; Benjamin, 1952) and surveys of the extent to which ratings are relied on as criteria (cited in Guion, 1965, p. 96, op. cit.) showing a range of from 80 percent to 90 percent reliance upon ratings.

8. The most significant (at least to this thesis) conclusions emerging from the nine years of rating experience within the AFLC as recorded in this chapter are: first, persistent susceptibility of the graphic scales used to chronological leniency, and; secondly, documented evidence that sole reliance upon management self-control of appraisal tendencies is both impractical and futile. Parallels with other researchers on these two factors have been previously discussed in Chapter II.
CHAPTER IV

DEVELOPMENT, IMPLEMENTATION, AND INTRODUCTION

OF THE LINEAR TRANSFORMATION

The year 1969 was marked by standard revised appraisal pronouncements by the U.S. Civil Service Commission (Federal Personnel Manual, Chapter 355, revised 1969). For the first time, all federal agencies were required to apply a supervisor's evaluation of performance for all employees; to use the appraisal in the promotion evaluation process (i.e. in rank ordering all employees eligible for advancement within an installation-wide area of consideration); to constitute the evaluation instrument around elements of performance relevant to the requirements of the job.

Although the AFLC system had previously incorporated these policy ingredients, the net effect, coupled with uncontrollable chronological creepage, was to leave management in an untenable situation. On one hand, an appraisal system which demands a graphic scaling approach is required and, on the other hand, sufficient documentary evidence had been gathered to indicate that any graphic scale applied within the AFLC rating environment rapidly renders itself impotent as a function of increased supervisory familiarity with how other supervisors rate (or have been rating) their subordinates and with the systemic related appraisal format, per se.
The command-wide frequency distribution for ratings collected on
102,000 employees in March of 1969 (on the tailor-made series of forms) is shown in Figure 4. Some 57,000 cases appear in the upper three score intervals (raw scores from 86 to 100), and over 4,000 employees were awarded the perfect score of 100. It is most interesting to note, however, that rating supervisors were eminently capable of differentiating between employees spread between what would be regarded as the "marginal" and "unsatisfactory" levels of performance.

The in-service promotion system was further aggravated (in 1969) by drastically decreased effectiveness and usage of the other two allowable promotion ranking factors—tests and experience. Test usage had been sharply curtailed during the preceding year by Headquarters, AFLC, restricting usage only to those occupations and grades where test validity had been empirically established. As a function of continuous manpower reductions in assigned strength, 1 the experience ranking factor became, for all intents and purposes, a constant in the ranking process, with most employees possessing the maximum allowable score. Thus, the highly truncated appraisal became the more critical factor in establishing promotion rank order. With such a large proportion of the workforce possessing approximately the same appraisal scores (e.g. 86 or higher), managers frequently were handed promotion registers containing hundreds of qualified candidates (reference is made to Figure 1),

1 As is typical with most corporations, any downward manning trend, and associated reductions-in-force (RIF's), newly hired employees and those with fewer years of experience are the first to be removed from the workforce. Hiring is also typically sharply curtailed. The remaining cadre is accordingly somewhat homogeneous with regard to tenure and years of qualifying experience.
Fig. 4 1969 COMMAND-WIDE Appraisal Distribution
(Tailor-made).
frequently with 50 or more employees possessing the perfect score. The resultant interview burden placed on the selecting supervisor obviously became both overly time consuming and costly.

From experience with the operating environment as summarized in Chapter III, it is evident that solutions to the dilemma are severely limited. Any variant of forced distribution, promotion by nomination or promotion by exception must be ruled out because first, these types of approaches are not sanctioned by the Civil Service Commission\(^2\); and secondly, because of anticipated management, employee, and union resistance. Man-to-man, ranking, rank-comparison, paired comparison, and forced-choice also are not regarded as plausible alternatives since they are difficult to translate into the applied employee counseling situation. It should be noted, parenthetically, that a usable profiling, evaluative, and diagnostic version of the forced-choice rating has been successfully developed and tested by Wherry (1959). Moreover, the very size of the promotion area of consideration—\(N = 15\) to 25 thousand—would render such types of rating approaches extremely cumbersome and expensive to administer within pre-established regulatory and computerized parameters.

The resultant management dilemma is peculiarly similar to that reported by Taylor and Manson (1951) who also found themselves in a

\(^2\)Chapter 335 of the Federal Personnel Manual, U.S. Civil Service Commission specifically proscribes against the use of tests, educational requirements, appraisals, and the like as a means of excluding candidates from merit promotion consideration, except where exclusions are allowed direct by the Central Office of the Commission. Consequently, the use of tests or appraisals as "screening" factors in the historical tradition of test theory is an extremely infrequent possibility for application against the U.S. federal civil service employee.
situation which practically precluded the use of anything but a graphic scale. Their solution (in a test rather than an operational environment) to "over-the-shoulder" audit supervisor's ratings, would be an economically unacceptable practice for AFLC usage. In the same vein, AFLC appraisal problems documented in the preceding chapter are by no means regarded as unique to either the Air Force or the federal government. Many of these common problems emerged from employee complaints, formal grievances, the in-house suggestion program, and appeals to higher authority. Using these sources as a data base, it was concluded by the Headquarters' research staff that the primary articulated objections to the predecessor graphic system revolved around differences between competitive rating sub-units (viz. supervisors). Further differences were noted associated with other organizational peculiarities such as job level, unit assignment, occupational status, and even geographical location. Organizational reviews of rating systems by other investigators have highlighted similar problems (in addition to parallel research findings previously summarized in Chapter III):

1. The overall picture is like the case history of merit rating as reported in the Atlantic Refining Company by Driver (contained in Dooher and Marquis, 1950, op. cit.). Driver also encountered the pre-existing pattern of 15 years of unsatisfactory experience with merit rating systems not sufficiently reliable. Driver's subsequent pragmatic approach, description of the multitude of problems and analyses of management/employee difficulties in merit rating are strikingly analogous to the AFLC situation.
2. Combining the observations rendered by various reviewers of company merit rating systems, the following communalities have also been reported:

   a. Merit-rating systems do not take into account, or correct for individual differences among raters (Truman Kelley, 1919; Lawshe and Balma, 1966; op. cit.).

   b. Standards of performance vary considerably between departments (Kingsbury, 1925/1926, op. cit.; Bills, cited in Dooher and Marquis, 1950, op. cit.).

   c. Carefully chosen descriptive adjectives and phrases are found to be ambiguous and to have different practical significance in different departments (Kingsbury, 1925/1926, op. cit.); and supervisors had no formal or standard performance records (Bills, in Dooher and Marquis, 1950, op. cit.).

   d. "It is believed that a great deal of the objection encountered...has been misdirected, and the point to which objection is taken is basically the difficulty of reconciling relative and absolute standards...," again quoting from Baier's (1951), op. cit., comments regarding the Army's forced-choice rating technique.

   The above cited common program difficulties appear to typify any rating process which presupposes the absolute equivalence of performance judgments rendered by different judges along a common scale. Thus, all individuals being awarded a score of 100, or 70, or zero—for that matter—are presumed to be situated at the same level on the performance continuum within the total organizational hierarchy, regardless of
differential standards of performance known to be applied by individual rating supervisors.

In exploring the AFLC interpersonal rating problem in more depth through interviews with immediate and reviewing supervisors, it was concluded that absolute scaling was completely alien to the actual work environment for reasons summarized as follows:

1. First, it was impossible for any member of the management corps at any AMA to be thoroughly familiar with the performance of all those employees--installation-wide--who met minimum qualification requirements for advancement to any specific vacancy under his span of control. Supervisors were typically unfamiliar with the performance of employees beyond their own work unit. Thus, each rating official selected different sections of the available appraisal scale to describe excellence within the context of his own rating situation.

   a. Lenient supervisors relied upon the perfect score to represent excellence (i.e. many perfect scores of 100) resulting in the typical J-type distribution. This type supervisor was found to be the typical rater.

   b. Other less lenient rating officials used scores less than 100 to describe excellent performance (such as 88 to 90), but never-the-less permitted scores to "pile-up" just below their perceived level of excellence resulting again in a negatively-skewed distribution.

   c. The truly "hard-nosed" supervisor attempted to use a far wider range of raw scores and occasionally achieved the positively-
skewed distribution optimum for promotion differentiation purposes.\(^3\) Such raters were usually found among the ranks of wage grade foremen or other large work units where work standards and quantifiable performance records were occasionally available.

d. When queried further, supervisors in all three of the above groupings would insist that the top rated employees in their subordinate unit(s) were superior to (or at least equal to) top rated employees in other organizations. Accordingly, propinquity, as discussed by David Kipnis (1960), is regarded as an important consideration in rater behavior, regardless of the standards of performance applied by different supervisors. The familiar is preferred and supported over the unfamiliar. Empirical support has been provided through a study conducted by Dorothy Kipnis (1957) who found that officers' evaluations of enlisted men on B-29 bomber crews were more lenient toward those men who work closest (actual physical distance) to the rating officer.

e. Conclusion: Different rating officials use different segments of the appraisal rating scale to describe "excellent" performance.

2. Secondly, it was discovered that the supervisor's approach to the rating task was related to the size of the unit he was required to evaluate:

---

\(^3\)It should be noted at this juncture that the greater the proportion of employee and supervisory grievances, congressional inquiries and other forms of complaints came from employees assigned under raters in this category. Their primary concern, and rightfully so, was that they found themselves penalized (for promotion consideration) by being under a "tough" rating supervisor or a rater who attempted to honestly and objectively apply the prescribed standards of a rating.
a. For large employee groupings (i.e. \( N \geq 50 \)) the immediate rating official tended to rate along a nominal scale by dividing the total ratee population into manageable subgroupings, usually three—superior, satisfactory, and unsatisfactory (or marginal, depending upon the verbal inclinations of the individual). He would then insure proper rank ordering (i.e. ordinal) for employees assigned to the top, or superior, category. Less frequently this same ranking treatment would be afforded the bottom category primarily because of the potential of appeals emanating from those employees assigned the lowest ratings.

b. Supervisors of intermediate sized groups, \( N = 20 \) to 50 (which accounted for the largest proportion of ratee groups) characteristically rank-ordered the complete subordinate population.

c. Although some supervisors of smaller units (\( N < 20 \)) would continue to rely solely upon ordinal ranking, many raters in this category conscientiously attempted to differentially profile their subordinates on the various appraisal rating elements (within the constraints of any pre-established tendencies to leniency), thereby approaching the interval scale of measurement. It was the raters of employees in this category who were least susceptible to the influence of halo.

d. Almost invariably, second-level endorsing officials approached their review responsibilities along the same lines as referenced in paragraph 2.a. above by immediate supervisors of large employee groups. The other notable and extremely significant tendency of the endorsing or reviewing official was his unquestioning acceptance of the absolute ratings assigned by his subordinate supervisors. That is, he
simply was not discharging his responsibility to arbitrate or moderate differences between rating standards applied by his subordinate raters.

e. Conclusion: Appraisal ratings did not conform universally to any single scale of measurement; most ratings were established along an ordinal scale.

3. Almost to a man, appraising supervisors confided that they: first, established ratee rank order—placing in the highest positions those subordinates whose probability of appearing for promotion consideration during the coming 12 months was quite high (in effect nominating the most promotable); secondly, assigning raw appraisal scores to conform to the pre-determined rank ordering; finally, scoring various rating elements on the appraisal to substantiate both the pre-determined overall raw score and the rank ordering. In effect, they tended to rank on a "whole-man" concept rather than relying upon the uncertainty of differentially profiling sub-elements. In other words, raters felt most comfortable by reversing the rating procedure laid down by regulatory and other supervisory appraisal instructions. Conclusion: The most natural rater appraisal tendency was to rank order employees in terms of promotability, using a "whole-man" concept.

Based upon the foregoing conclusions, and reinforced by a decade of unsatisfactory experience with absolute merit scales unsuccessfully controlled by management, it was proposed that a possible solution would be to transform raw score supervisory rating distributions into a

---

4 The author is especially indebted to Mr. William Wright, Staff Psychologist, Sacramento Air Materiel Area, for his initial suggestions for problem resolution and for his assistance in early formulation of the research foundations for this thesis.
standard scale. Before further developing the formal statement of problem—with attendant assumptions and experiment hypotheses—it is believed appropriate, at this point, to more thoroughly investigate the available literature regarding application of standard scoring to employee rating situations.

**Review of the Literature on Standard Scoring**

Hopefully, without redundantly repeating articles previously reviewed in Chapter II, the literature on standard scoring breaks down into four general categories: those advocating or providing epistemological support for the application of standard scores; studies in the experimental setting; the very limited actual instances of reported industrial application; negative accounts or counter-advocates.

1. Supporting Opinions, Theory, Advocates

Theoretical support for the correcting of between-rater differences through statistical means as propounded by Patterson (cited in Fryer and Henry, 1950), op. cit., and Guilford (1954), op. cit., have been previously reviewed. Supporting positions have been established by the following:

a. Guilford (1965) once again emerges as a strong advocate for usage of a linear transformation of raw score values. To quote (Guilford, 1965, p. 510): "If modern psychology and education have taught us anything about measurement, they have amply demonstrated the fact that there are few, if any absolute measures of human behavior."

He further argues that "No informed person would think of using raw
scores as a basis for making direct comparisons among an individual's scale values with respect to trait values."

b. Additional support for the use of linearly transformed score values has been provided by Gulliksen (1950) who contrasts the convenience of a linear transformed scale with the various disadvantages of percentile and normalized scores. "Linear derived scores (including, of course, standard scores) have this very valuable property: the characteristics of the original distribution of gross scores are duplicated in the transformation scores. The indices of skewness and kurtosis for the distribution of linear derived scores, and both sets of scores, will have the same correlation with any other variable. Nonlinear transformation of gross scores will in general have indices of skewness, kurtosis, and correlation that are different from those of the original gross scores." Another advantage of the linear transformation, as pointed out by Gulliksen, is that it involves no assumption about the distribution of the population or the sample.

c. An interesting theory which substantiates some of the observations made regarding the AFLC rating environment has been propounded by Overall (1968). He conceives of raters as "linear transformation filters" such that, except for random errors of measurement, rating scores reflect within a linear transformation the levels or values of some underlying psychological characteristic of interest. Thus, different raters represent different filters through which the world is seen. Overall cautions that, in his concept of error, it is important to understand that there is no reality associated with the characteristic being measured except as projected through the linear
transformation filter which is the rater. Approaching the problem from the point of locating stimulus objects in multidimensional criterion space, Horan (1969) concluded that mean square transformations gave less deformation of the stimulus configuration than did a mean transformation (unless all observers in a group perceived the stimuli in perceptual spaces which were essentially the same).

d. Examining rating scales as used in making clinical judgments, J.B. Taylor (1968) argued that the faults of leniency, social desirability, idiosyncratic tendencies toward polarization (i.e. halo) are not a necessary concomitant of rating scales, per se. Rather, such faults spring from two sources: first, a failure to determine the unidimensionality of the universe being measured, and secondly, a reliance upon absolute rather than comparative judgments. His proposed solution would be to use an example-anchored approach analogous to that suggested by Ross (1966), op. cit.

e. Various other investigators also sanctify the transformation of inter-rater or inter-departmental differences under certain conditions and with qualifications. Joe Tiffin, typically a strong advocate of forced distributions of direct measurements (Whisler and Harper, 1962, p. 188, op. cit.), has mentioned (Tiffin, in Dooher and Marquis, 1950, op. cit.) that it is often necessary to make corrections for the department, plant experience, job experience, age, and certain other factors of this type that sometimes exercise a systematic influence on the ratings. Further, Tiffin and McCormick (1965), op. cit., state that ratings should not be compared across rating groups unless they are adjusted for the differences. The authors suggest a standard
score or mean score adjustment (assuming equal group variances if the latter adjustment were to be applied). Both Ghiselli and Brown (1955), op. cit., and R. L. Thorndike (1949), op. cit., allow that external sources of bias (either constant errors or differences between territories) may sometimes be corrected for statistically. Finally, Mâchin (1956) suggests that rating is a delicate operation which, to be successful, cannot be left with impunity to just anyone. He recommends that, to be efficacious, the rating method used must provide for a general statistical control of the results.

2. Applications in the Experimental Environment

Experimentation with variants of standard scoring conversion by Bolanovich (1946), op. cit.; Jurgensen (1950), op. cit.; and Ross (1966), op. cit., have been described earlier in Chapter II. In each instance, although the techniques were concluded to have promise, follow-on reports of actual application or even experimental replication have not been found in the literature.

Four studies were found where the researchers found it expedient and useful to convert ratings (which were used as dependent variables in the studies) to standard scales because of widely noted differences between criterion units/departments (Turner, 1960, op. cit.; McElwain, 1951; Hemphill and Sechrest, 1952, op. cit.; Barrett, 1966, op. cit.). In all four studies, conversion to standard scales was accomplished to increase reliability of the dependent variable and to exclude unnecessary biases or contaminations.
3. Applications in the Operating Environment

Outside of those few applications already cited (i.e. Benge, 1946, op. cit., and his proposed JEM system; Patterson, 1922-1923, op. cit., and the Scott Company Graphic Rating System; and Baier's, 1951, op. cit., report of standard scoring as merged with the Army's forced-choice rating process), a paucity of additional industrial uses of standard scoring was found (or rather not found). Only three additional cases emerged from reviewing the available literature:

a. Adelbert Ford (1931), motivated by a desire to relieve foremen of much of the worry in making ratings (in such a way that the scores-between-foremen may be compared on an equivalent basis) tried both percentile and mean score transformations. He found little difficulty in training foremen to make ratings but found it almost impossible to give them an adequate understanding of the significance of a probability curve or the desirability of raw score distribution.

b. Bills (cited in Dooher and Marquis, 1950), op. cit., described an attempted mean scale transformation applied in the Acme Steel Company. Individual supervisory ranking points were computed by the personnel office transforming the average rater's tendency to a standard mean scale having a $\bar{x} = 70$. The system further allowed adjustments (of no more than two ranking points) for seniority. The converted ranking scores were used for promotion advancement and reduction-in-force purposes. An interesting footnote by Bills quotes Joseph Tiffin's reaction to the system as follows: "The merit rating plan used by the Acme Steel Company follows procedures which avoid the hazards, dangers, and possible inequities of merit rating more
carefully than any plan I have had the opportunity to examine."

c. Saffirio (1960) describes a situation where the introduction of standardized statistical controls for personnel rating met with initial wide-spread resistance by corporate management. However, time, and the proved usefulness of such controls, brought about a more favorable acceptance of quantifiable evaluations.

4. Counter Advocates to Standard Scoring

In spite of what amounts to limited application of the technique, a few vocal and articulate critics have come out against transforming ratings to a standard scale.

a. Stevens and Wonderlic (1934) in devising the rationale for their original Stevens and Wonderlic trait rating system also state their non-sanctioning of statistical correction(s) of raw score distributions. Their nonconcurrence stemmed from a strong belief that all rating distributions by nature are negatively skewed in the industrial environment. This same belief has also been espoused by Bass (1956), op. cit.

b. Thomas Ryan (1945) also strongly criticizes statistical leveling on the basis that it is no safer to assume that the difference between raters is due to a real difference in rater's leniency than it is to assume that the differences (in both rating means and variances) are due to real differences in the quality of men. Ryan proposes that supervisors ought to be able to distinguish between employees on an absolute basis at at least three levels of overall value to the department (i.e. analogous to an "Inventory of Personnel" divided into the categories: outstanding, poor, average, trainees).
c. Perhaps the most vocalized critique of the standard scoring approach has been published by Thompson and Dalton (1970), op. cit., who refer to the process of statistical leveling variously as a "zero sum," "peer-comparison," or a "closed" rating system. Their criticisms are based upon a narrative review of the use of normal distribution ranking within a large electronic corporation^ implemented to prevent all ratings from "clustering" in a narrow above average range. Their follow-up (after the system had been in use for a few years) reported wide-spread dissatisfaction among both managers and employees. Adverse effects of the "zero sum" system were: a negative effect on management's thinking (toward employees?); forced invidious comparisons between employees; a resultant impoverishment of the manager's reward system; unintended (negative) consequences upon employee performance. The authors favor, as a substitute, an objective's approach to performance evaluation à la Peter F. Drucker (1954), op. cit.

Thus, in spite of what appears to be sound and strong theoretical support, with a few exceptions (the Scott Company's merit system and Bills' report of usage in the Acme Steel Company), standard scoring has not enjoyed wide-spread usage in the industrial world. Of course, as mentioned previously, it is to be noted that neither has any other single rating scheme been widely acclaimed or applied for other than short periods of time. Nevertheless, for reasons as stated earlier in

^It is unfortunate that the authors provide neither data nor appropriate references to permit further exploration of their claim (p. 151) that such approaches (i.e. performance ranking or standard scaling) have been gaining in popularity in recent years. Such a conclusion has not been supported through the review in this paper.
this chapter (to be expanded upon later), it was decided to try the linear transformation within the AFLC environment.

Statement of Problem

The formal problem is delineated by historical antecedent, augmented by formal and informal survey reports of supervisory and employee attitudinal expectations regarding the role of supervisory appraisal. Specifically, an approach is required which would meet organizational goals of differentiation for advancement, and which would simultaneously meet the criteria of workforce (including supervisory personnel) understanding and acceptance.

Within constraints imposed by the environment and regulatory parameters, and using the backdrop of the review of the literature cited in this chapter, a linear transformation built around an empirically acceptable rating unit is proposed for trial as a solution to the pre-described dilemma. Operational exploration of the trial solution is the basic researchable topic of this thesis. It is proposed that the rank-order established as a by-product of summary scoring of an elemental graphic rating scale can be translated into a usable and acceptable appraisal process through application of the linear transformation around the unit raw score distribution. The transformation formula selected for test is that described in Guilford (1965, p. 517), op. cit.:  

6 It should be mentioned that the Civil Service Commission both sanctions and recommends the use of standard scores (Attachment 335-1, Chapter 335, Federal Personnel Manual, issued 1969) where agencies find it necessary to combine promotion ranking variables on a unit weighting basis.
\[ A' = A \left( \frac{\sigma_e}{\sigma_o} \right) - \left| \left( \frac{\sigma_e}{\sigma_o} \right) M_o - M_e \right| \]

where:  
\( A \) = score on the obtained scale  
\( A' \) = score on the standard scale  
\( \sigma_e \) = expected standard deviation  
\( \sigma_o \) = observed standard deviation  
\( M_o \) = observed mean  
\( M_e \) = expected mean

The precise application of this formula will be devised after service test of the linear transformation.

**Perceived Advantages**

Within the constraints imposed by regulatory issuances, while nevertheless consistent with expectations of both the rater and the rated, such an approach is proposed as an acceptable methodology for trial for many reasons:

1. First, as previously mentioned by Gulliksen (1950), op. cit., the shape of the rating distribution is changed little when the scores are transformed, assuming the majority of rating officials attain an observed standard deviation reasonably close to the expected deviation. This particular attribute, in turn, permits the rater to "do his own thing," in terms of ordering the perceived distribution of talent within his rating units without being concerned that the standardization process will result in gross distortions in his rating distribution. This aspect of the linear transformation, where adequately explained to--and accepted by-- the workforce (including rating officials) should
be a logical and powerful counterargument to those objections raised by those disallowing any "zero-sum" appraisal system.

2. Secondly, the linear transformation does not demand an assumption that talent is "normally" distributed within all rating units. As proposed to be explained to the workforce during the trial period, the system assumes only that some good talent exists within all rating units. This particular assumption is most relevant to the operating environment in that it most closely approximates verbal statements of intent regarding merit promotion; that is, promotability is demanded to occur among only those best qualified for advancement.

3. Thirdly, the linear transformation of unit rank order forces no change in either the perceived level of individual performance (as reflected through summation of individual element scores) or in the interpersonal relationship between the rater and the ratee. Employee counseling can accordingly be conducted within the explanatory context of the supervisor's perceived evaluation of individualized performance as related to established standards of performance, theoretically disassociated from the controversial counseling topic of "promotability."

It must be acknowledged that this peculiarity of the suggested approach presents a very distinct rater/ratee escape in that the "Transformation System" could be verbally blamed as the scapegoat in determining individual promotability.

4. A fourth, indirect, value of the approach is that it carries with it an implied penalty to those supervisors continuing to rate their employees with excessive halo, including those rating with excessive leniency. If properly communicated to the supervisory cadre,
it could be demonstrated that the continuance of such behavior would result in adjustment of the truncated distribution to a "clustering" of standard scores around the expected mean. Faced with this knowledge, it is to be expected that most rating officials would develop a tendency to build more dispersion into their ratings, thereby avoiding "penalizing" their more outstanding subordinates.

5. Lastly, and more importantly, the linear transformation can be used to relieve the evaluating official of his anxieties over how other rating supervisors in competing organizations are rating their employees. If properly communicated to rating officials, the point could be strongly emphasized that he would no longer—under the standard scaling approach—need to be concerned about either "overrating" or "underrating" his subordinates. Further, it can be explained that the system would permit the better regarded employees in various organizations to have their performance evaluated on a reasonably equitable basis without regard to different standards of performance established by individual rating officials.

Assumptions

Translated into statements of basic assumptions, the above cited applied attributes of the linear transformation of unit rating distributions are summarized in Table 6. Each assumption is accompanied by the appropriate statement of a counter assumption.

Where feasible to test, the assumptions stated in Table 6 will be empirically examined during two proposed service tests of the proposed linear transformation rating.
TABLE 6

BASIC ASSUMPTIONS

<table>
<thead>
<tr>
<th>Basic Assumption</th>
<th>Counter Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Differences in rating tendencies exist</td>
<td>1. Differences in rating tendencies do not exist</td>
</tr>
<tr>
<td>2. These differences can be corrected</td>
<td>2. No need to correct</td>
</tr>
<tr>
<td>3. The best estimates of a rater's tendencies are his average rating and dispersion</td>
<td>3. No need for estimates</td>
</tr>
<tr>
<td>4. All organizations have some excellent performers</td>
<td>4. Some organizations do not have excellent performers</td>
</tr>
<tr>
<td>5. Performance can be measured via a graphic rating scale</td>
<td>5. Same, but maintain that scores retain absolute value because of assumption #1</td>
</tr>
<tr>
<td>6. The System can be communicated to and understood (to some degree by the workforce)</td>
<td>6. System cannot be communicated or understood to some degree</td>
</tr>
<tr>
<td>7. The workforce will experience a normal level of turnover; overall and between departments</td>
<td>7. The workforce will remain static</td>
</tr>
</tbody>
</table>

Test 1 — Limited manual exploration of assumptions 1 and 3, necessitated by unavailability of data regarding individual rater tendencies below major organizational level in the computerized skills locator system. The initial service test also will hopefully aid in developing administrative procedures for application during the operational test phase.

Test 2 — Because of the magnitude of the proposed change, and the concomitant impact upon a workforce of over 120,000 employees, a management decision was made to operationally service-test the proposed
approach—within the environment of the entire command—for one year before permitting application of the transformed scores in constituting promotion registers. The basic objectives of the command-wide service test were to test the acceptability of assumptions 1 and 6 (with assumption 6 regarded by management as the more critical of the two; and to collect data upon which to "sell" the system to supervisors, employees, higher authority, and external critics.

Explanations of the relevance and importance of each of the assumptions stated in Table 6 within the AFLC personnel context are as follows:

Assumption No. 1: It should be explained that since differences between rating tendencies of the immediate and second-level reviewing officials have been (or are to be) administratively obliterated, differences to be examined are those existing between second-level (or endorsing) supervisors. Quantitative, rather than subjective evidence of the type presented earlier in this chapter, is required.

Assumption No. 2: Essentially self-explanatory; since we are dealing with a mathematical system, it is evident that this assumption can be safely met. It is not considered necessary to empirically test this assumption other than to establish that a linear transformation significantly changes promotion rank ordering over the pre-existing rating system; otherwise there would exist no need to modify the system.

Assumption No. 3: The critical issue here is to arrive at a decision as to which is preferable: a linear transformation using both the rater's mean and standard deviation, or a simple mean transformation. If relatively equal variances can be demonstrated to exist
between various endorsing officials, then a mean score transformation would suffice (reference Tiffin and McCormick, 1965, op. cit.).

Assumption No. 4: Since application of the transformation would be applied solely to establish promotion ranking, it is believed desirable to assume only that some excellent performers exist in all organizations. Under the linear transformation, it is neither necessary nor desirable to assume a normal distribution of talent in all rating units. In fact, as pointed out by Gulliksen (1950), op. cit.: "The linear transformation involves no assumptions about the distribution of the population or the sample."

Assumption No. 5: Essentially a rhetorical and untestable assumption. What is required here is acceptance of the ranking by-product emerging from graphic raw scores as an acceptable distribution around which a linear transformation can be computed. The issue boils down to what type rating "best" reflects employee performance: an absolute graphic rating score or some relational rank ordering.

Assumption No. 6: Regarded as the ultimate criterion of the worthiness of any appraisal system. If eventually unacceptable to management, the resultant demand to devise yet another (more acceptable) rating process constitutes sufficient evidence that it (the existing system) has proven unsatisfactory. The qualified statement of this assumption has been made because it is believed that it is neither practical nor possible to expect that the entire workforce fully comprehend all of the intricate workings and manipulations involved in a linear transformation of raw appraisal scores within the confines of a computerized skills locator and promotion ranking system; particularly
if an understanding of the exact computations involved in computing a
standard deviation are demanded to be understood by the workforce.
What should reasonably be expected is a general understanding of the
intent of the system, coupled with a reasonable level of trust in the
integrity of the administrators of the linear transformation process.

Assumption No. 7: A reasonable level of employee turnover to
insure the viability of assumption number 4 and to give rating officials
reoccurring "head-room" so as to avoid being pushed into "chronological
leniency." On the other hand, if the workforce remains static and
employees continue to improve in performance with experience and
training, the rating supervisor has little alternative to raising his
raw score ratings yearly simply as a means of sustaining employee
incentive and morale.

Preliminary Service Test 1

Method

Constrained by time limitations and the necessity to manually
compile and compute raw and standard score distributions for second-
level raters, the initial test was limited to a sampling of six promo-
tion registers at three AFLC installations. For each register, using
an installation-wide area of promotion consideration, means and
standard deviations were computed for those endorsing officials whose
employees appeared on the promotion profile; the registers were recon-
stituted using standard appraisal scores with an $\bar{x}_e = 70$ and $\sigma_e = 10$;
and comparisons were made between the original (i.e. absolute scores)
and reconstituted (i.e. standard scores correcting for differences
between endorsing officials). The promotion register has a particular advantage as a vehicle for service test in that it represents a sample of employees essentially free from the contaminating effects of grade, job assignment, and geographical location. The differences remaining between competitors are those attributable primarily to inter-rater and inter-individual performance effects. The test was a pure "paper exercise" not involving rating supervisors or employees.

Results and Conclusions

Descriptive characteristics of the six sample registers selected for test are shown in Table 7.

<table>
<thead>
<tr>
<th>Hypothetical Position to be Filled (series &amp; grade)</th>
<th>Installation</th>
<th>Employees Different on Endorsing Register</th>
<th>No. Endorsing Officials</th>
<th>Factors Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft Sheetmetal Worker (W-2892-10)</td>
<td>SMAMA</td>
<td>112</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Wage Foreman (F-2854-06)</td>
<td>SMAMA</td>
<td>71</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Contract Negotiator (GS-1102-12)</td>
<td>OOAMA</td>
<td>61</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Computer Specialist (GS-334-11)</td>
<td>OOAMA</td>
<td>59</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Civil Engineer (GS-810-13)</td>
<td>WPAFB</td>
<td>28</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Clerk-Stenographer (GS-318-05)</td>
<td>SMAMA</td>
<td>29,c</td>
<td>29</td>
<td>3</td>
</tr>
</tbody>
</table>

*"a---Synonymous with promotion register. Series and grades are standardized U.S. Civil Service Commission jobs categories.

*b---Two ranking factors = appraisal score plus Experience Ranking (TRAEX). Three ranking factors = Appraisal + TRAEX + appropriate test score.

*c---Actual register contained over 200 promotion eligibles situated under 110 different endorsing supervisors. Number studied was reduced (randomly) to the more manageable number of 29.
TABLE 8
SAMPLE $\bar{x}$ AND $\sigma$ RANGES WITH ASSOCIATED F-TESTS

<table>
<thead>
<tr>
<th>Promotion Register</th>
<th>$\bar{x}$ range</th>
<th>$\sigma$ range</th>
<th>$a$ larger</th>
<th>$a$ smaller</th>
<th>$F$</th>
<th>$df$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft Sheetmetal Worker</td>
<td>36.7 to 72.9</td>
<td>11.4 to 23.7</td>
<td>561.69</td>
<td>129.96</td>
<td>4.32</td>
<td>112, 59</td>
</tr>
<tr>
<td>Wage Foreman</td>
<td>36.7 to 75.1</td>
<td>7.1 to 19.8</td>
<td>392.04</td>
<td>50.41</td>
<td>7.78</td>
<td>32,262</td>
</tr>
<tr>
<td>Contract Negotiator</td>
<td>69.4 to 83.3</td>
<td>6.7 to 15.6</td>
<td>244.30</td>
<td>45.02</td>
<td>5.43</td>
<td>14, 36</td>
</tr>
<tr>
<td>Computer Specialist</td>
<td>64.3 to 82.8</td>
<td>8.6 to 14.9</td>
<td>320.77</td>
<td>73.96</td>
<td>4.33</td>
<td>57, 20</td>
</tr>
<tr>
<td>Civil Engineer</td>
<td>68.0 to 90.6</td>
<td>1.6 to 16.4</td>
<td>269.91</td>
<td>2.72</td>
<td>99.24</td>
<td>1, 21</td>
</tr>
<tr>
<td>Clerk-Stenographer*</td>
<td>35.5 to 79.2</td>
<td>6.6 to 20.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*a-- F-ratio not computed.

Ranges in means and standard deviations between the various endorsing officials are shown in Table 8. Ranges are shown against a scale with a maximum obtainable raw score of 100.7 Degrees of freedom (i.e. N-1) represent frequencies of the unit(s) under the endorsing.

7It will be noted from reference to Appendix 3 that no space is provided for supervisory computation of total appraisal scores. The actual computational formula

$$\left(\frac{\Sigma \text{element scores}}{5 \times \text{no. elements rated}} \times 100\right)$$

was found to be much more accurately calculated by the computer, by key punching individual elemental scores, than by the rating supervisor. Several studies conducted at various AFLC installations showed an average supervisory computational error rate between 20 and 30 percent.
Officials. The F ratio applied was grossly estimated by

\[ F = \frac{\text{larger variance}}{\text{smaller variance}} \]

for each promotion register (from Guilford, 1965, p. 192), op. cit. All F ratios are significant beyond the .01 level.

Conclusions:

(1) Significant inter-rater differences do occur for employees from different organizations competing for the same position vacancies (in terms of both means and sigmas of the endorsing officials) large enough to justify transformation to a common scale.

(2) Different rating units actually use different segments of the appraisal scale to describe excellence of performance.

(3) The assumption of homogeneity of unit variance cannot safely be met, thereby leading to the conclusion that a standard scale transformation is to be preferred over a simple mean score transformation.

Rearranged (standard score with \( \bar{x}_e = 70 \) and \( \sigma_e = 10 \)) rank orderings of candidates on each of the six promotion registers were contrasted with the original rank ordering. Intercorrelations (Pearsonian) between both sets of rankings between appraisal scores only and total rankings, which would include both TRAEX and test scores as appropriate, are shown in Table 9. Surprisingly enough, few usable inferences can be drawn from the intercorrelations shown in Table 9. About the only conclusion that can be made is that the two appraisal scores, as represented by absolute versus linearly transformed values, are not representing parallel dimensions of performance (i.e. average \( r = .81 \), accounting for some 65% of common variance between the two
### TABLE 9
SUMMARY OF INTERCORRELATIONS

<table>
<thead>
<tr>
<th>Promotion Register</th>
<th>N</th>
<th>Ranking Factors</th>
<th>Total Ranking ( r_{\text{absolute}} \times LT )</th>
<th>Appraisal Ranking Only ( r_{\text{absolute}} \times LT )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage Foreman</td>
<td>71</td>
<td>2</td>
<td>.91</td>
<td>.92</td>
</tr>
<tr>
<td>Computer Specialist</td>
<td>59</td>
<td>2</td>
<td>.90</td>
<td>.90</td>
</tr>
<tr>
<td>Civil Engineer</td>
<td>28</td>
<td>2</td>
<td>.84</td>
<td>.75</td>
</tr>
<tr>
<td>Aircraft Sheetmetal Worker</td>
<td>112</td>
<td>3</td>
<td>.93</td>
<td>.64</td>
</tr>
<tr>
<td>Contract Negotiator</td>
<td>61</td>
<td>3</td>
<td>.96</td>
<td>.85</td>
</tr>
<tr>
<td>Clerk-Stenographer</td>
<td>29</td>
<td>3</td>
<td>.90</td>
<td>.79</td>
</tr>
</tbody>
</table>

*Where total ranking score = TRAEX + Test Score (percentile, with three ranking factors) + Appraisal (either absolute or linearly transformed). Since neither test nor TRAEX scores changed, they represent a constant value in computing the intercorrelations.

measures). Thus, the linear transformation did not replicate the pre-existing absolute appraisal at a level of correlation high enough to indicate that they were duplicative measures.

One additional side benefit emerged from service-test one, namely a classic portrayal of entirely divergent standards of performance exercised by endorsing officials supervising units containing employees who competed for the same vacant position (that of Aircraft Sheetmetal Worker). The distributions of ratings in both these organizations, under absolute and linear transformation conditions, are shown in Plates 1a (absolute) and 1b (transformed). Needless to say, endorsing supervisor "A" typifies the lenient rater and supervisor "B" represents the "hard-nosed" rater. When interviewed, both managers insisted that their "better" workers were equally competent and deserving of promo-
Fig. a.

Fig. b.
tion. These figures later proved extremely valuable in demonstrating (and "selling") the linear transformation to managers, employees, and higher-level approving authorities.

Administrative Development

Based partially on the results of the first service test, the computerized skills locator system was reprogrammed to accommodate the one year, command-wide test of the linear transformation as described in the following steps:

1. Means and standard deviations were computed for employees rated on each of the five tailor-made forms under the span of control of endorsing (second-level) supervisors. Tabulations were ordered around the matrix header shown in Table 10.

TABLE 10
APPRAISAL COMPUTER BASIC COMPUTATIONAL FORMAT

<table>
<thead>
<tr>
<th>Endorsing Official</th>
<th>Form</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>Social Security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account No. (SSAN)</td>
<td>N</td>
<td>S</td>
<td>M</td>
<td>N</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>111 XXX XXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>222 XXX XXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>333 XXX XXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Installation Total</td>
<td>. .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = Number employees rated
S = Standard deviation
M = Mean rating
2. Where \( N > 6 \) for any cell (i.e. common appraisal form) in the matrix in Table 10, the linear transformation would be made around that distribution, re-entered into the master record for each employee, and the transformed score used for promotion ranking purposes (for that employee until the next rating cycle). Where \( N \leq 6 \) for employees rated on the same form under one endorsing official, the transformation would be made around the total distribution (all forms) for that endorsing official. Where \( N \leq 6 \) total for any endorsing official, transformations would be made using the installation-wide distributions on the specific form (e.g. all clerical). Example:

\[
\begin{align*}
\text{if } N < 6 & \quad \rightarrow \quad N_{\text{Form ABCD or E}} \\
\text{if } N_A \leq 6 & \quad \rightarrow \quad N_{T(\text{SSAN})} \\
\text{if } N_T > 6 & \quad \rightarrow \quad N_{T(\text{Form ABCD or E})}
\end{align*}
\]

Several administrative determinations are represented by the computer program as described above.

First, the endorsing official was chosen as the primary unit of statistical control because regulatory necessity demanded a negotiated raw score on the graphic-tailor-made form between first and second levels of supervision; it was discovered through inspection of sample distributions that larger, more representative ratee groups fell under the second level of supervision; studies with the operating environment
(Wilfong, 1964)\(^3\), concluded that a performance review one level removed from the employee renders more objective and effective ratings; finally, the employee group localized under the second level of supervision most closely approximated the cooperative work unit within an Air Materiel Area.

Secondly, correction around all employees rated on elements common to the same form was adjudged appropriate as the first level of transformation since this group represented the most homogeneous application of supervisory performance standards. Transformation around the total group evaluated by the endorsing official was concluded to be the next appropriate level of adjustment presupposing the acceptability of assumption 3, Table 5. Final correction around the total installation distribution (where \(N < 6\) total for any endorsing official) was administratively established as the only remaining acceptable level of correction. Thus, the best acceptable distribution for transformation was established for all employees rated on common elements (specifically, the same rating form) under an endorsing rater, second best as all employees rated by an endorsing official, and least acceptable, the installation-wide distribution of all employees rated on the same format (where \(N < 6\) under a common endorser).

\(^3\)This study showed that experimental ratings of second level supervisors correlated higher with scores on a job knowledge test for electronic employees than the ratings of first level supervisors. Similar findings were reported by Whitla and Tirrell (1954). Counter-conclusions have been cited by those favoring the propinquity theory (Ferguson, 1949; David Kipnis, 1960, op. cit.; Dorothy Kipnis, 1957, op. cit.; Wherry, 1952, Theory (4), p. 11, op. cit.). The final decision was more heavily weighted by research findings specific to the operational work environment.
Introduction to the Workforce

As mentioned earlier, because of the mass change in rater behavior anticipated to be caused by changeover from an absolute to a linear transformation appraisal system, a management decision had been made to field-test the new system for one year prior to actual implementation. It was further decided that, while individual promotions during the service test period would continue to be based on the pre-existing absolute graphic scale values, the workforce would nevertheless be instructed to constitute new ratings (in July of 1969) for all Command employees around assumptions attendant to the linear transformation process. The underlying rational behind this determination was predicated on the belief that rater behavior would only be advantageously modified where rating officials were properly acquainted with the attributes of the impending rating system and its associated probable impact upon individual employee promotability. It was expected that both systems would be run in parallel between July 1969 and March 1970 (when a new annual rating cycle was scheduled) with the intervening period used to further indoctrinate the workforce, collect operational service test data, and obtain approval from higher authority for operational use of the standard scoring system.

New appraisal forms were redesigned (see Appendix 4), using five of the tailor-made elemental appraisal patterns shown in Table 5; standard instructions were prepared for dissemination to all rating
supervisors (Appendix 5) and the system was introduced (on a service-test basis) in July, 1969. To avoid the contaminating influences otherwise associated with experimental try-out of any new rating system, all rating officials were instructed that their absolute ratings would be converted to standard scores and actually used in constituting promotion registers once the field-test was concluded and sometime before the next annual rating cycle.

Pre-rating training emphasis was directed to the following points.

1. First, top level management was thoroughly briefed on the proposed change. The introduction of any mass modifier of human behavior first demands understanding and acceptance of the top executive levels. Consequently, acceptance of the proposed standard scoring concept was solicited, and eventually secured, at the Congressional, Air Staff, Civil Service Commission, and Command levels.

2. The secondary level of training was directed toward rating and endorsing officials. Media used included personal briefings, base newspaper releases (a typical example appears in Appendix 6), closed circuit television, individual letters to all rating officials, Commander's Calls, and regulatory issuances. Primary dimensions of the linear transformation emphasized during the supervisory briefings were:

It should be noted that the title change to "Career Appraisal" and the addition of a narratively-completed Part 2 to the Supervisory, Scientific and Engineering and Technician Forms represented a departure from previous editions of the appraisal forms. This modification was introduced as an aid to an AFLC "Career Management Board" review and selection process which was to be implemented simultaneously with publication of the revised appraisal forms.
a. Corrections will be made around rating distributions and policies established by the endorsing official. Separate transformations will be made for each of the five basic rating forms, controlled around the endorsing official.

b. Endorsing officials are responsible for insuring that the distribution of scores for all employees under their span of review are aligned with the overall performance levels of subordinates; that the best worker has the highest score and that scores are distributed along the entire scale, accurately reflecting performance from the most effective to the least effective.

c. Endorsing officials are cautioned against clustering final ratings around any one point on the rating scale in order to distinguish the better from the average employee with further caution against artificially forcing the ratings into a predetermined statistical distribution(s).

d. Ratings are to be anchored to standards of performance established for individual positions.

3. The third level of indoctrination embraced the entire workforce; describing basic attributes of the revised appraisal system, influence upon the promotion process and expectations regarding anticipated shifts in individual raw score ratings.

**Results of the Field Test**

Few changes to the system actually emerged from the six-months service test of the standard scoring system. The field-test was designed essentially as an administrative review, as a check on accuracy.
of the computer programming and as an initial "sounding-board" for eliciting supervisory and employee reaction. Accordingly, no methodological design of experiment was prepared nor were statistical analyses or conclusions attempted.

Primary results emerging from the field-test, changes made to the system, and accompanying rationale are summarized in the following sub-paragraphs.

1. Part 2 (the career evaluation and "secret" portion) was dropped from the appraisal forms for Scientists and Engineers, Supervisors, and Class Act Technical personnel. This portion of the form(s) was expanded upon and incorporated into a formalized AFLC career management program for all employees at grades GS-11 and above, and Wage Foremen grades (AFLC Manual 40-2).

2. The minimum cell frequency around which linearly transformed scores were computed was raised in all cases from \( N < 6 \) to \( N < 10 \). This change was introduced for two basic reasons:

   a. It was established, through machine simulation, that the maximum transformed standard score that a single individual can achieve is a direct function of group size (permitting unit \( \bar{x} \) and \( \sigma \) to vary infinitely). This function is given by:

\[
S_{max} = 70 + 10 \sqrt{N-1}
\]

---

5Appreciation is expressed to Dr. D. L. Henry, Chief, Operations Analysis Office, Headquarters AFLC, for his assistance in devising this formulation and conducting the computer simulations.
Evaluating the function for all integers from 1 to 101, given an assignment of scores optimally favorable to the elevation of a single score, shows that group size \( > 10 \) permits the possible occurrence of standard scores \( \geq 100 \) (where \( \bar{x}_e = 70 \) and \( \sigma_e = 10 \)).

b. The practical effect of the change, in turn, permitted satisfactory explanation (to all employees) that they all stood a chance of achieving the maximum obtainable standard (presuming, of course, that the endorsing official would be willing to suffer the administrative consequences of "sacrificing" the majority of his subordinates in order to optimize the promotability of a single employee, and recognizing that under a standard scoring system, optimum standard scores would be infrequently attained nor required in the actual competitive promotion system.

3. The label of the standardization process was changed from "transformation" to "standard scoring." It was discovered that the former label connoted a negative affect to employees and supervisors alike in that they inferred that "someone" (either the personnel office or the computer) was tampering with, or otherwise manipulating, their initially assigned raw scores. Although this misconception obviously stemmed from the failure on the part of members of the workforce to make the mental shift from an absolute to a relational scoring system, the label change was nevertheless made. It was felt that the term "standard score" had the further advantage of better describing the nature of the actual change between the two systems.

4. The standard scoring formula was changed to \( \bar{x}_e = 100 \) with \( \sigma_e \) remaining = 10. This change was initiated responsive to management
feelings that employees would be more receptive to converted standard scores which always resulted in a numerical increase over their pre-established absolute or raw appraisal score values.

5. The basic rating scale against which each tailor-made element was evaluated was changed from percentile categories (Appendices 3 and 4) to a scaling header which anchored only low and high points of the scale as related to standards of performance established for each rating element by individual first level supervisors. This modification had the advantage of permitting rating supervisors to change their standards of performance concomitant with any realignments in their rating unit. The revised format is shown in Appendix 7 and accompanying revised regulatory instructions to the rating officials are shown in Appendix 8.

6. Two machine listings were prepared for installation civilian personnel office use in describing and explaining the standard scoring system to employees and supervisors:

a. A computer master listing, showing the \( \bar{x}, \sigma, \) and \( n \) for each form for each endorsing official (i.e. the computer analog to the matrix portrayed in Table 10). A sample page of this listing is shown in Appendix 9.

b. An "endorsing official's distribution listing" showing raw and standard score frequency distributions by form and total for each endorsing official (the latter identified by Social Security Account Number and Organizational Assignment Code). A sample page from this listing is replicated in Appendix 10. In addition to its use in employee and supervisory counseling, the endorsing officials'
printout was used to numerically show the endorsing official the actual consequences of his rating behavior—in standard score terms.

7. The influence of introduction of the standard scoring system had a drastic influence upon rater behavior, at least for the first year's application during the service test period. This change is shown in Figure 5, contrasting command-wide raw score distributions under the absolute scoring system (pre-transformation, ratings collected in March 1969) and the standard scoring system (post-transformation, ratings collected in July 1969).
Raw data from which the plots in Figure 5 were computed are shown in Table II, along with frequency and proportion of standard scores. The latter was introduced as a check to programming accuracy and to test goodness of fit against the normal \( z \) distribution. As tested by the Kolmogorov-Smirnov one-sample test (Siegel, 1956, pp. 47 to 52), the obtained standard scoring distribution was significantly different from the normal \( z \) distribution at the \( p < .01 \) level, hence the Null hypothesis must be rejected (largest obtained \( D = .0251 \); \( D \) required at \( p = .01 - .0051 \), with \( df = 102419 \))^6. An administrative adjustment was made to the computer program arbitrarily adjusting all observed standard scores > 100 to = 100, and similarly standard scores < 40 were not permitted to occur.

---

6 It should be observed that with \( df = 102419 \), almost a perfect curve fit would be required in order to accept the Null hypothesis.
TABLE 11
AFLC COMPARATIVE RAW AND STANDARD SCORE FREQUENCY DISTRIBUTIONS
FOLLOWING INTRODUCTION OF STANDARD SCORING IN 1969

<table>
<thead>
<tr>
<th>Score Interval</th>
<th>Pre-Standardization</th>
<th>Post-Standardization</th>
<th>Standard Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Score</td>
<td>%</td>
<td>Raw Score</td>
</tr>
<tr>
<td>100-96</td>
<td>19,567</td>
<td>18.23</td>
<td>4,897</td>
</tr>
<tr>
<td>95-91</td>
<td>20,141</td>
<td>18.77</td>
<td>10,160</td>
</tr>
<tr>
<td>90-86</td>
<td>17,714</td>
<td>16.61</td>
<td>12,784</td>
</tr>
<tr>
<td>85-81</td>
<td>12,732</td>
<td>11.86</td>
<td>14,556</td>
</tr>
<tr>
<td>80-76</td>
<td>8,820</td>
<td>8.22</td>
<td>14,723</td>
</tr>
<tr>
<td>75-71</td>
<td>5,215</td>
<td>4.86</td>
<td>12,963</td>
</tr>
<tr>
<td>70-66</td>
<td>4,897</td>
<td>4.66</td>
<td>8,960</td>
</tr>
<tr>
<td>65-61</td>
<td>3,477</td>
<td>3.24</td>
<td>7,028</td>
</tr>
<tr>
<td>60-56</td>
<td>2,882</td>
<td>2.69</td>
<td>5,182</td>
</tr>
<tr>
<td>55-51</td>
<td>1,981</td>
<td>1.85</td>
<td>3,663</td>
</tr>
<tr>
<td>50-46</td>
<td>2,142</td>
<td>2.00</td>
<td>2,289</td>
</tr>
<tr>
<td>45-41</td>
<td>1,969</td>
<td>1.83</td>
<td>1,904</td>
</tr>
<tr>
<td>40-36</td>
<td>1,793</td>
<td>1.67</td>
<td>1,383</td>
</tr>
<tr>
<td>35-31</td>
<td>1,279</td>
<td>1.19</td>
<td>796</td>
</tr>
<tr>
<td>30-26</td>
<td>1,244</td>
<td>1.16</td>
<td>448</td>
</tr>
<tr>
<td>25-21</td>
<td>895</td>
<td>.83</td>
<td>291</td>
</tr>
<tr>
<td>20-16</td>
<td>498</td>
<td>.46</td>
<td>184</td>
</tr>
<tr>
<td>15-11</td>
<td>38</td>
<td>.04</td>
<td>15</td>
</tr>
<tr>
<td>10-6</td>
<td>18</td>
<td>.02</td>
<td>6</td>
</tr>
<tr>
<td>5-1</td>
<td>10</td>
<td>.01</td>
<td>1</td>
</tr>
</tbody>
</table>

N 107,312 102,233 102,420
EX 8,660,878 7,702,602 7,191,200
EX^2 733,929,208 603,372,540 515,284,220
X 80.71 75.34 70.21
\sigma 18.03 15.01 10.09
CHAPTER V

ANALYSIS AND CONCLUSIONS

Using the three year (1969-1971) field test experience with the standard scoring system as a data base, the following analyses are appropriate:

1. Examination of the extent to which chronological leniency was neutralized by inhibitory attributes of the standard scoring system as reflected by changes in overall raw score distributions.

2. Re-examination of assumptions stated as corollaries to (successful) operational application of the linear transformation, referring back to Table 6, Chapter IV:
   a. Assumption No. 1 - Differences in rating tendencies exist. This assumption can be subjected to a more rigorous test since information regarding endorsing officials' individualized rating tendencies is stored and can be retrieved from the mechanized system.
   b. Assumption No. 6 - The system can be communicated to and understood (to some degree) by the workforce. This assumption to be tested through a mixture of clinical/administrative exploration of employee and supervisory perceptions regarding the system.
   c. Assumption No. 7 - The workforce will experience a normal level of turnover, overall and between departments. Stability to be inferred from trends in workforce attrition.
Presuming an "acceptable" understanding of the intentions and operations of the standard scoring system on the part of rating and endorsing officials, it would be predicted that introduction of the linear transformation should have an inhibitory effect upon rater inclinations to chronological leniency. The primary attribute of the system expected to inhibit leniency is an explanation that "any rater failing to achieve any dispersion of raw scores will result in 'clustering' his scores around the mean of the standard scale (see Appendix 4)."

Likewise, a tendency to "clustering" of scores around any raw score point will translate into a similar distribution of scores around the mean of the standard scale. These statements imply, in turn, that rating officials indulging in extreme tendencies to leniency would be, in effect, "penalizing" their more outstanding subordinates.

The extent to which introduction of the standard scoring system inhibited chronological leniency, as related to previous appraisal systems, is best represented graphically in Figure 6 (which is essentially a three year extension of command raw appraisal averages as previously shown in Figure 2, Chapter III). It is evident that the linear transformation had no appreciable inhibitory effect on supervisory tendencies toward chronological leniency, and it must be concluded that the strength of other forces (upon the rating/endorsing officials) had a far greater influence upon rater behavior than did any built in counter-influences inherent in the standard scoring approach.
Fig. 6. COMMAND AVERAGE APPRAISAL SCORE, 1961 - 1971.
Command-wide raw score distributions under the standard scoring system, by year, are reflected in Figure a, Plate II. Frequency data, by raw score interval, from which this chart was plotted are contained in Appendix 12. The relative asymmetry of these three rating distributions is shown in Table 12. Skewness as measured by Fisher's $g_1$ statistic (Fisher, 1958, p. 52), showed all distributions to be significantly, progressively skewed in the negative direction.¹

<table>
<thead>
<tr>
<th>Year</th>
<th>$g_1$</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969 (Post Transformation)</td>
<td>.087</td>
<td>102,233</td>
</tr>
<tr>
<td>1970</td>
<td>.092</td>
<td>101,931</td>
</tr>
<tr>
<td>1971</td>
<td>.114</td>
<td>98,765</td>
</tr>
</tbody>
</table>

Standard score distributions for the same three year period are shown in Figure b, Plate II. As expected, the distributions became more pronouncedly leptokurtic over time, reflecting the progressively negative skew already demonstrated for the raw scores from which they were derived.² With additional time, where ratings continue to show inflationary tendencies, further leptokurtic tendencies are expected to prevail. These tendencies lead to two conclusions:

¹ Again, since the frequencies involved are so large, almost a complete fit would be required to statistically demonstrate symmetry. For example, using the N of 1969 of 102,233, a $g_1 \leq .00059$ would be required to fall within the standard error of the sampling variance as formulated by Fisher (1958, p. 75), op. cit.

² It was believed inappropriate to test these distributions for kurtosis under the fourth power, derived from the fourth cumulant of the population (Fisher's measure of $g_2$, 1958, p. 73), op. cit., since standard score values below 70 or above 130 had been administratively adjusted to the $\pm 3.00 \sigma$ level(s).
Fig. a. AFLC COMPARATIVE RAW SCORE appraisal distributions

Fig. b. CHRONOLOGICAL standard score distributions
1. Supervisors were decidedly not statistically "manipulating" subordinate raw scores in a direction advantageous to their more promotable employees despite numerous claims and complaints to the contrary.

2. It can be reasonably concluded from the preceding trend charts that the behavior of rating officials was not directed toward the primary use of the supervisor's appraisal as a device for merit promotion differentiation. Rather, they were apparently behaving purposely in a contraindicated direction, so as to maximize the frequency of occurrence of high absolute appraisal scores for their employee group. Such behavior must be regarded as purposive in nature inasmuch as all rating officials were thoroughly preindoctrinated in, and freely acknowledged their understanding of the resultant "negative" impact of such tendencies upon subordinate promotional opportunities.

Similar tendencies prevailed at each AFLC activity as reflected through the analysis of variance of raw score distributions for the three year period 1969 through 1971 shown in Table 13.

**TABLE 13**

**ANALYSIS OF VARIANCE OF RAW APPRAISAL SCORES (POST TRANSFORMATION), YEAR BY ACTIVITY**

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity (A)</td>
<td>5</td>
<td>879118.00</td>
<td>5111.00</td>
</tr>
<tr>
<td>Year (B)</td>
<td>2</td>
<td>1275581.00</td>
<td>7416.00</td>
</tr>
<tr>
<td>Activity by Year (A×B)</td>
<td>10</td>
<td>54932.00</td>
<td>319.00</td>
</tr>
<tr>
<td>Within Sets</td>
<td>302911</td>
<td>172.00</td>
<td></td>
</tr>
</tbody>
</table>
All F-ratios were found to be significant at $p < .01$, with the main effect of years contributing the substantially greater portion of variance. Interaction effects are portrayed in Figure A, Plate III.

From this graph, it is interesting to note that the rate of acceleration in mean appraisal scores was much less for those activities which began in 1969 with a relatively high installation-wide mean rating. Conversely, those activities starting out with depressed means (notably SAAMA, COAMA, and SMAMA) reflected the greatest proportional increase over the three year period. It should be also mentioned, parenthetically, that these same three activities had historically shown greater management control over rater leniency in the years preceding introduction of the linear transformation (as measured by lower mean activity appraisal scores).

Essentially the same results were found when analyzing variance components of different appraisal forms (i.e. homogeneous job groupings) and post transformation raw score distributions by years (Table 14).

### Table 14

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form a (A)</td>
<td>4</td>
<td>243383.00</td>
<td>1345.00</td>
</tr>
<tr>
<td>Years (B)</td>
<td>2</td>
<td>1371254.00</td>
<td>7576.00</td>
</tr>
<tr>
<td>Form by Year (A×B)</td>
<td>8</td>
<td>30101.00</td>
<td>166.00</td>
</tr>
<tr>
<td>Within Sets</td>
<td>303085</td>
<td>181.00</td>
<td></td>
</tr>
</tbody>
</table>

*aThere were five separate forms in use, each representing an occupational grouping: Wage Board Nonsupervisory, Class Act Technician, Supervisory, Scientific and Engineering, and Clerical.*
PLATE III

ACTIVITY

Fig. a. ACTIVITY BY YEARS INTERACTION

APPRAISAL FORM

Fig. b. FORMS BY YEARS INTERACTION
Interaction effects, forms by years, are plotted in Figure b, Plate III.

The primary notable effect was initial escalation for all forms during the first year's application. Other differences can be traced to traditional patterns of rater behavior such as the tendency to consistently rate blue collar (Form A) and technical (Form B) personnel lower than clerical (Form E) and supervisory (Form C).

In all instances—forms, activities and overall—the tendency to chronological leniency was essentially linear in nature, with movement across time accounting for the greater proportion of variance. Some indication of the populations involved, form by activity, is shown in Table 15, for all employees rated during the March 1971 rating cycle.

**Differences between Raters**

One of the basic intentions of the program was that of capturing and retaining information about standardized dimensions of individual raters on tape or disk storage and retrieval systems. Such capability was expected to provide a more systematic examination of Assumption No. 1, with specific attributes of rater behavior related to a variety of external factors such as organizational size, type of employee supervised, attrition, and other similar measures related to organizational performance.

The total number of endorsing officials by activity in 1971 is shown in Table 16. The average endorsing official was found to have 24 employees falling under his span of control.
### TABLE 15

NUMBER OF EMPLOYEES RATED ON AFLC 679 FORMS (1971)

<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>Forms</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>OCAMA</td>
<td>12,042</td>
<td>6,817</td>
</tr>
<tr>
<td>OOAMA</td>
<td>7,185</td>
<td>5,521</td>
</tr>
<tr>
<td>SAAMA</td>
<td>11,256</td>
<td>7,284</td>
</tr>
<tr>
<td>SMAMA</td>
<td>7,239</td>
<td>5,175</td>
</tr>
<tr>
<td>WRAMA</td>
<td>6,556</td>
<td>5,152</td>
</tr>
<tr>
<td>2750th</td>
<td>2,143</td>
<td>3,840</td>
</tr>
<tr>
<td>TOTAL</td>
<td>46,421</td>
<td>33,789</td>
</tr>
<tr>
<td>% OF TOTAL</td>
<td>47%</td>
<td>34.21%</td>
</tr>
</tbody>
</table>


Of the total number of 4,048, only 67% had populations $\geq 10$, permitting standardization around either the total distribution for that endorsing official or standardization around the form-specific cell distribution. The remaining 1,336 second-level raters had units with $N < 10$, forcing standardization around the installation-wide total form specific distributions for employees in these organizations. This frequent occurrence of this latter situation was regarded as undesirable in that it forced a violation of Assumption No. 2, namely that differences between rating officials can be corrected for.

The actual number of endorsing officials, by activity, where unit size $N < 10$ or $N < 7$ is shown in Appendix 12. Although some 33% of the endorsing officials were faced with unit standardization around installation-wide norms, only 5.34% of the total employee population were, in turn, thusly affected (see Appendix No. 13). This peculiarity is basically a function of organizational structure where overlaying strata of supervision resulted in units (particularly at the Headquarters and special service organizational levels) where second-level administrative
supervision extended to only two, three, or four subordinate individuals, usually clerical or lower level supervisory.\(^3\)

Distributions of means, group sizes, and sigmas for the population of 983 endorsing officials at OQAMA and OCAMA attained during 1971 are shown in Plate IV. Endorsing officials where total group size was less than 10 were excluded from these distributions. The J-shaped N curve (Figure b, Plate IV) naturally follows average group size. The somewhat normal \(\sigma\) curve was somewhat surprising in that a greater concentration of occurrences at the lower sigma ranges (i.e. \(\sigma < 5.00\)) would be expected where unit means increase and unit sizes are smaller. The rectangular shape of the \(\bar{x}\) distribution was not believed to be unusual considering the fact that supervisors had sometimes been earlier instructed that an attained raw mean of 70, with a standard deviation of 10, would produce "best" results under the standardization process.

One possible means of evaluating the extent to which differences between performance standards established by individual rating officials remain stable, regardless of which rating form is used, is to examine variance in those cases where sufficient numbers of different types of employees are rated under the same endorsing official. Such a test also would permit verification of the desirability of standardizing around

\(^3\)Another significant error factor emerged during this portion of the analysis. It became routine to reexamine key-punch accuracy of all social security account number entries for endorsing officials showing only one subordinate within his unit. The vast majority of these cases turned out to be transposition errors in key punching one digit of the social security account number.
the total distribution of an endorsing official where $N_{separate form} < 10$. Using the same sampling of endorsing officials' distribution listings (see Appendix 10) from OQAMA and OCAMA, a total of 126 endorsing officials were found to have rated heterogeneous groupings of employees in sufficient numbers on two separate forms to permit variance testing as shown in Table 17. Only twenty endorsing officials were found to have rated groups where $N > 6$ on three separate forms (Forms B, C, and E).

TABLE 17

<table>
<thead>
<tr>
<th>Form</th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
<th>Form E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Wage Board</td>
<td>-</td>
<td>31</td>
<td>25</td>
<td>None</td>
<td>10</td>
</tr>
<tr>
<td>B Technical</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>C Supervisor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>None</td>
<td>13</td>
</tr>
<tr>
<td>D Scientific &amp; Engineering</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>E Clerical</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The absence of forms' combinations A*D, C*D and E*D were not regarded as unusual considering the small proportional population of scientists and engineers (Form D) within the command and the operational improbability of these professionals being organizationally co-located with other of the four groups for which separate appraisal forms existed.

Analysis of variance summaries for the available combinations are shown in Table 18. Computational procedures followed were those as described in Winer (1962, p. 241) for an unweighted means analysis.
Fig. a. AVERAGE MEAN.

Fig. b. AVERAGE N.

Fig. c. AVERAGE O
TABLE 10

ANALYSIS OF VARIANCE SUMMARIES, FORMS BY SUPERVISORS

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forms A×B</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>30</td>
<td>9224.07</td>
<td>101.64**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>2265.75</td>
<td>16.20**</td>
</tr>
<tr>
<td>S×F</td>
<td>30</td>
<td>139.84</td>
<td>1.54</td>
</tr>
<tr>
<td>Within Cell</td>
<td>1314</td>
<td>90.75</td>
<td></td>
</tr>
<tr>
<td><strong>Forms A×C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>24</td>
<td>908.27</td>
<td>9.46**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>7451.45</td>
<td>19.51**</td>
</tr>
<tr>
<td>S×F</td>
<td>24</td>
<td>381.94</td>
<td>3.97**</td>
</tr>
<tr>
<td>Within Cell</td>
<td>1435</td>
<td>96.04</td>
<td></td>
</tr>
<tr>
<td><strong>Forms A×E</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>9</td>
<td>809.00</td>
<td>7.68**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>163.67</td>
<td>1.92</td>
</tr>
<tr>
<td>S×F</td>
<td>9</td>
<td>177.92</td>
<td>1.69</td>
</tr>
<tr>
<td>Within Cell</td>
<td>371</td>
<td>105.29</td>
<td></td>
</tr>
<tr>
<td><strong>Forms B×C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>27</td>
<td>832.69</td>
<td>7.96**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>1985.55</td>
<td>17.68**</td>
</tr>
<tr>
<td>S×F</td>
<td>27</td>
<td>112.29</td>
<td>1.07</td>
</tr>
<tr>
<td>Within Cell</td>
<td>646</td>
<td>104.54</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forms B×D</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>11</td>
<td>609.62</td>
<td>5.61**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>179.52</td>
<td>2.62</td>
</tr>
<tr>
<td>S×F</td>
<td>11</td>
<td>68.36</td>
<td>.63</td>
</tr>
<tr>
<td>Within Cell</td>
<td>373</td>
<td>108.53</td>
<td></td>
</tr>
<tr>
<td><strong>Forms B×E</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>19</td>
<td>1255.35</td>
<td>13.05**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>89.79</td>
<td>.93</td>
</tr>
<tr>
<td>S×F</td>
<td>19</td>
<td>169.26</td>
<td>1.76*</td>
</tr>
<tr>
<td>Within Cell</td>
<td>601</td>
<td>96.16</td>
<td></td>
</tr>
<tr>
<td><strong>Forms C×E</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>12</td>
<td>140.36</td>
<td>2.08</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>1</td>
<td>1019.26</td>
<td>21.88**</td>
</tr>
<tr>
<td>S×F</td>
<td>24</td>
<td>46.11</td>
<td>.68</td>
</tr>
<tr>
<td>Within Cell</td>
<td>345</td>
<td>67.55</td>
<td></td>
</tr>
<tr>
<td><strong>Forms B×C×E</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors (S)</td>
<td>20</td>
<td>790.60</td>
<td>14.11**</td>
</tr>
<tr>
<td>Forms (F)</td>
<td>2</td>
<td>78.60</td>
<td>.93</td>
</tr>
<tr>
<td>S×F</td>
<td>40</td>
<td>84.18</td>
<td>1.50</td>
</tr>
<tr>
<td>Within Cell</td>
<td>773</td>
<td>56.04</td>
<td></td>
</tr>
</tbody>
</table>
With the singular exception of ratings on employee groups C (supervisory) and E (clerical), all combinations showed significant differences to exist between different endorsing officials. Further, excepting those situations where Form C was coupled with another rating group (i.e. AxC, BxC, and CxE), differences between supervisors accounted for the greater proportion of variance. Different standards of performance apparently were actually being applied in all cases where subordinate groups of supervisors were found to exist. Since interaction affects all approached unity, they were not plotted. The differentiating effectiveness of the various forms held up for two out of the three combinations including supervisory sub-groupings. In the single instance where three sub-groupings existed (BxCxE), between-forms differences were less than unity, with the primary contribution to variance being between-raters differences.

Workforce Understanding and Acceptance

From the initial inception of the concept of standard scoring, difficulty was anticipated in effectively securing understanding and acceptance of the concept from the workforce, particularly the supervisory corps whose cooperation would be essential. As expected, the first year's implementation of the system (including the service test period, was accompanied by waves of trauma, verbal accusations of "foul-play," confusion, and a host of misconceptions regarding both the
intent and the administration of the standard scoring process. Initial protests emerged from among the ranks of rating officials, many of whom full well understood the intent of the new system, after having been thoroughly indoctrinated on the workings of the system and exposed to several "briefings" and group meetings on the subject.

It was also recognized that a systematic quantitative assessment of worker understanding, acceptance, and evaluation of the new system (through the media of a standardized attitude questionnaire, for example) could not be effectively conducted for a variety of reasons:

1. As a function of workforce reductions, a generalized "negative affective set" tended to permeate all attitude questionnaires administered by the command research office. This generalized set tended to neutralize the diagnostic effectiveness of any affectively-loaded questionnaire.

2. A similar negative valence had been built up over a period of years toward the Federal Merit Promotion Program itself, as reflected through various formal Air Force and Civil Service Commission inspections of AFLC field level civilian personnel office operations. To the extent that the appraisal was perceived as an integral part of the merit process, it too would be influenced by these same historical perceptions.

3. A lack of supervisory cooperation in administering predecessor appraisal systems had already been thoroughly documented.

---

Earlier action had been initiated in an attempt to "pre-condition" employees to standard scoring by reporting test scores results to individual employees in a standard scoring format. Furthermore, it was presumed that most of the Air Force officer corps would be familiar with the concept of "stanine."
The conclusions described in this section were accordingly abstracted from a variety of informational sources, foremost among which were: staff visits to field installations with accompanying individual and group interviews with literally hundreds of supervisors and employees; reviews of individual and group complaints, requests for staff review, and unsolicited suggestions for system's improvement; input from field psychologists and other civilian personnel office staff members; formal survey team reports.

Survey Team Findings

Formal Air Force Civilian Survey Team reviews of acceptance of the appraisal system at the operational level were particularly valuable. The formal team approach offered an external, systematic examination of any specified program area in accordance with prescribed group and individual interviewing and data collection sampling procedures (Air Force Manual 40-5, 1965).

Relying primarily upon these formal Air Force survey reports of AFLC field level civilian personnel programs, attitudes towards the standard scoring system were found to vary with the organizational level interviewed:

1. Employee Attitudes
   a. Although all members of the workforce acknowledged receiving copies of locally released newsletters explaining the standard scoring system, the majority voiced the opinion that the system was too complicated for most employees to understand.
b. Employees felt, in general, that supervisors were directing their attention more toward identifying promotables and rating employees to insure their eligibility, with a corresponding tendency not to use the appraisal for what employees perceived to be its intended purpose—that of truly measuring employee performance.

c. Employees felt that with the advent of the "leveling process" (a label frequently applied by employees) attempts were frequently and continually made by rating officials to "beat the system." They also expressed the opinion that, in some instances, ratings were changed or "manipulated" between the supervisor and the civilian personnel office. They felt that supervisors deliberately made adjustments to insure promotion consideration of favored individuals.

d. Employees resented the considerable power placed in the hands of the second level supervisor as an "endorser." They felt the second level supervisor was too often unaware of the detailed operations and performance of the rated employee.

e. Employees did believe that the system offered a protection against strict or rigid rating supervisors, but that this favorable attribute was offset to some extent by the base-wide area of consideration and the vulnerability of the system to "manipulation."

2. First and Second Level Supervisory Reaction

a. Supervisors expressed the general feeling that they did not believe they understood operation of the system too well, particularly the more complicated mathematical aspects such as the application of standard deviations.
b. Supervisors claimed that the major objectives of the system, those of better adherence to rating criteria and adjustment of differences between raters, were not being achieved. Many of them felt compelled to rate individuals with an "appropriate curve" in mind (despite instructions to the contrary printed in bold-face type on the reverse of each appraisal form—see Appendix 7). They persisted in a belief that each endorsing official had to show a certain percentage of "goods" and "poors" in order to "make it come out right."

c. They strongly questioned the assumption (Number 4, Table 6, Chapter IV) that all activities function with the same command-wide effectiveness percentage.

d. Supervisors believed that the leveling process has eliminated some problems, particularly under a base-wide area of consideration. However, the majority favored more limited areas of promotion consideration and a return to the old, absolute appraisal rating system.

3. Top Management Reaction

a. Although top management understanding of the system and how it operates was found to be much better than with employees and supervisors, it was believed to be still not clearly understood at the top levels.

b. Top management reported fewer complaints under the system (in contrast to predecessor appraisals) but attributed this phenomenon to a lack of familiarity on the part of employees as to how the system operates.
c. They expressed the general opinion that, under the present expanded base-wide area of consideration, the leveling process or some other similar medium would be necessary, particularly in view of the current emphasis placed on supervisory appraisals in the merit promotion process and in light of significant differences between individual rating supervisors.

d. Top management admitted to studying and working the system (always for the "protection" against the other guy). They cited instances where ratings were "programmed," before being formally submitted, by those more knowledgeable with the system, with concomitant adjustments made by the "knowledgeables" to obtain optimum desired results.

4. Reactions of Employee Groups (Unions)

Contacts with officials of local unions possessing exclusive recognition rights elicited the following classical reaction:

They felt the operating details of the system needed better explanation. The union is opposed to supervisory appraisals and the system generally. They say it may be slightly better than former systems but is still in need of significant improvement. They feel it provides too much reliance upon supervisor's judgment for promotion. They believe there should be less weight in the hands of one or two individuals (i.e. first and second-level supervisors) and more emphasis on experience, tests, and seniority.

One survey team summarized field-level reaction to the system after a two-week inspection, as follows:

During this review, management officials and supervisors alike expressed dissatisfaction and distrust with the standard scoring method. Other than isolated inequities and personal dislike of the system, however, the real basis for this criticism is difficult to isolate. In fact, a comprehensive examination of
the present system provides substantial evidence the standard scoring technique is achieving the intended objective by eliminating most of the inequities resulting from lack of uniformity among rating officials which is characteristic of most other supervisory appraisal methods. In addition, with rare exception, employees with most promotion potential have the higher ratings under the standard scoring system. On the basis of these findings and in the absence of supportable evidence to the contrary, the wide-spread criticism appears to be without factual foundation. Furthermore, the standard scoring process has not had a "fair trial" to date [written in June 1971]. Accordingly, management officials and supervisors need to be better informed on the objectives of the standard scoring process, provide their full support to the system in order to establish a realistic basis for meaningful evaluation of its effectiveness, and demonstrate this support by offering objective explanations to their employees when requested.

Common Misconceptions/Misunderstandings

The most frequently noted misunderstandings, related explanations and accompanying assumptions are summarized below. Most of these refer back to an unacceptance of one of the basic assumptions as described previously in Table 6, Chapter IV. In spite of wide publication of this listing throughout the command on many occasions, many of these same misconceptions were still broadly persistent at the time of this writing.

1. **Misunderstanding**: The system places an endorser's scores into a "bell shaped" (normal distribution). (Assumption No. 3)

   **Explanation**: The system does not normalize or otherwise drastically change the basic shape of the rating distribution established by the endorsing official. Rather, it retains the basic shape and moves the distribution essentially intact to the standard scale.
2. **Misunderstanding:** An employee possessing a raw score of 90 under one endorsing official should receive the same standard score as any other employee also receiving a raw score of 90. (Assumption No. 1)

**Explanation:** Basically this is a plea for the use of the absolute score given by a supervisor. The above statement would have been accurate under the old system where an individual's relative standing within the rating unit had no influence upon the appraisal score. The old system pre-supposed the absolute nature of a raw appraisal score (i.e. all scores of 90 reflect exactly the same level of performance regardless of the rating tendencies of different supervisors).

3. **Misunderstanding:** The standard scoring system is based on the assumption that talent is equally (normally) distributed within each rating unit; that for each high score there must also exist a compensating low score. (Assumption No. 4)

**Explanation:** The standard scoring system assumes only that some talent exists within all rating units and that such talent can best be identified by neutralizing differences between performance standards (or rating "sets") applied by individual endorsing officials. As demonstrated in the example in paragraph 1, the basic distribution of ratings established by the endorsing official is not significantly distorted when his scores are computed to the standard scale.

4. **Misunderstanding:** Under the present system (scale from 40 to 100, later converted to 70 to 130), a score in the low 80's (81, 82, 83) places an employee out of reach for merit promotion consideration.

**Explanation:** Any statements or inferences made about the "competitiveness" of an appraisal score without simultaneous
consideration of the effect of test score, TRAEX, and other promotion consideration variables are unrealistic and attach more importance to the appraisal than it actually possesses. It must be emphasized that introduction of the standard score has changed the mental reference point as to what constitutes a "competitive" score. Based upon their experiences with the previous system, most employees seem to think that only standard scores of 96 or higher are "good" scores. Previously, 19,567 (18.2%) employees had scores of 96 or higher whereas now only 113 (or 0.1%) have standard scores in the same range. There is a need to change the mental set of supervisors and employees regarding the range of available scores under the new system. In fact, review of samples of command-wide promotion registers shows that standard scores from 60 on up are quite competitive.

5. **Misunderstanding:** Where, through selective management practices, an organization has managed to corner the talent within a specific skills area or where "elite" groups exist, they are penalized under the standard scoring system. (Assumption No. 4)

**Explanation:** The actual occurrence and/or existence of a truly elite group of employees is both extremely difficult to document and unlikely to occur in actual operating situations. In order to substantiate the existence of such a unit would demand that all employees within the unit possess equivalent high appraisal and TRAEX scores and that all supervisors of employees with similar, competing skills universally agree that employees in the supposedly elite group are truly superior in all aspects of their performance. Furthermore, it must be recognized that a claim to so-called eliteness is also a request for
exemption from competition (i.e. employees in elite units deserve an "edge" in promotion consideration). It is obvious that acceptance of such claims would subvert the intent of merit promotion by competition.

6. Misunderstanding: Employees localized in small units (i.e. smaller than 10) are penalized under the standard scoring system.

   Explanation: It must be first stressed that, regardless of the size of the rating unit, 50% of the employees under an endorsing official actualize standard scores of 100 (new mean) or higher. Thus, as explained above in item 4, employees within all units have equal assurance of receiving competitive scores. In actual practice, it can be shown that top rated employees in small units (size 10 to 20) can achieve standard scores in the 80's, and in some cases in the 90's. As presently applied, no penalty is being attached to employees in small groups.

7. Misunderstanding: The standard scoring system neutralizes the authority of the immediate supervisor in: (a) reviewing and evaluating the performance of subordinates; (b) determining the exact promotion standing of his employees since the standard score cannot be computed exactly in advance.

   Explanation: (a) The present system in no way alters the basic authority of the supervisor in evaluating employee performance. A standard AFLC requirement has always existed for joint immediate/endorsing official concurrence on the official supervisor's appraisal. This joint evaluation of performance, in addition to the joint decision as to relative standing of employees within the rating unit, remains as the prime determiner of the standard score. Movement of a raw score to a
standard scale in no way alters the supervisor's perception of individual levels of performance. (b) The promotion standing of an employee on any specific promotion register is dependent on a multitude of variables in addition to the appraisal which change frequently and do not remain constant over time or for all registers for which the employee qualifies.

**Stability of the Workforce**

The integration of Assumption No. 7, Table 6, Chapter IV, into the supporting corollaries was considered critical to both the conceptual and operational acceptability of the standard scoring model. Further exploration of this assumption—the workforce will experience a normal level of turnover, overall and between departments (during the application of standard scoring)—is required at this point.

Conceptually, sustainment of a "normal" attrition level is essential to successful application of the system in order to insure that changes observed in mean rating values represent the effects of chronological leniency, per se, rather than group changes in performance over time. The latter type of change has been integrated into MacKinney's (1967), op. cit., inductive treatment of the assessment of performance change. By way of example, any substantive influx of trainees during actual application of the standard scoring model would mitigate against meaningful interpretation of group change effects without relating observed average changes to the differences in change observed among individuals within groups. A heavy concentration of trainees in the workforce could mean that changes in recorded performance could reflect
increases in the rate of skill acquisition or the insufficiency of ratings collected during training to reliably carry over into the job situation. The former situation has been documented by Fleishman and Fruchter (1960) and Fleishman and Hempel (1954), with the latter problem described by Brown and Ghiselli (1952).

Within the AFLC context, normal turnover is defined at 12% of the workforce per annum, using data collected during the base years 1961 through 1967. Turnover, as measured by these data represents all losses to the command which result in the creation of a position vacancy, which in turn, translate into bona fide recruitment or other type placement actions.

The other very pragmatic concern attached to this assumption is that movement between departments occur with sufficient frequency to permit stabilization of performance standards (within a rating unit). If movement were not to occur, static situations would likely emerge where employee performance could increase steadily over time, providing little option to the rating official other than to increase his observed performance raw score values as a means of sustaining employee "motivation." Such behavior could result, in turn, in a decreased probability of competitive promotion standing for employees ranked highest in such units.

Available management indices of total command workforce movement occurring during the three-year application of the standard scoring system are shown in Table 19. All variables listed represented average fiscal year (July to June) values.
TABLE 19
COMMAND WORKFORCE MOVEMENT INDICES (1969-1971)

<table>
<thead>
<tr>
<th>Index</th>
<th>1969</th>
<th>1970</th>
<th>1971</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned Strength</td>
<td>119,234</td>
<td>112,142</td>
<td>107,376</td>
</tr>
<tr>
<td>Separation Rate (Total)</td>
<td>8.92</td>
<td>9.51</td>
<td>8.58</td>
</tr>
<tr>
<td>Separation Rate (Resignations Only)</td>
<td>4.66</td>
<td>4.30</td>
<td>2.79</td>
</tr>
<tr>
<td>College Graduates Hired</td>
<td>748</td>
<td>369</td>
<td>136</td>
</tr>
<tr>
<td>Wage Board Apprentices Assigned</td>
<td>2,502</td>
<td>705</td>
<td>255</td>
</tr>
<tr>
<td>Total Promotions</td>
<td>18,766</td>
<td>11,473</td>
<td>5,476</td>
</tr>
<tr>
<td>&quot;Competitive&quot; Promotions</td>
<td>8,235</td>
<td>4,524</td>
<td>2,147*</td>
</tr>
</tbody>
</table>

*Estimated command totals based upon samples collected from two AMA's.

All indices typify a decreasing workforce situation: resignations became depressed as workers attempted to hold onto the equity represented in their jobs; promotional opportunities became sharply curtailed; numbers of trainees hired and assigned became fewer as available vacancies were filled by surplus personnel or cancelled. From these data it must be concluded that Assumption No. 7 did not hold up during the three year period standard scoring was applied within the command. Both organizational and skills levels remained static under a generalized downward manning trend.

**Summary**

Conclusions emerging from the analysis of data accumulated during the three year operational field test of the standard scoring appraisal system can be summarized as follows:
1. Chronological leniency was demonstrated to predominate gross appraisal distributions, over time, uninhibited by counter-influences introduced by linear transformation of endorsing officials' raw score rating distributions.

2. Differences between rating officials were found to continue to account for greater rating variance than differences between the tailor-made series of appraisal forms.

3. Workforce understanding of the linear transformation was concluded to be progressively less than desirable moving from the top management to the worker levels. Lack of acceptance of the standard scoring system was adjudged to be sufficiently strong enough to place systems continuance in jeopardy.

4. The period of time embraced by the service test was accompanied by a continuous downward manning trend mitigating against stabilization of within units performance standards.
CHAPTER VI
DISCUSSION AND SUMMARY

The results of this study--statistically, graphically, chronologically--indicate that chronological leniency is a most persistent and powerful force to be reckoned with in the operational administration of a graphic rating scale. The phenomenon was found to be related to rater habituation, organizational dynamics, linear reoccurring after each insertion of a forms revision, and uninfluenced by intensive rater training. Stabilization of between-raters differences through introduction of the linear transformation had no appreciable impact upon rater tendencies to chronological leniency. Perhaps a definition by example, of chronological leniency, is in order at this point:

Any given rating supervisor, faced with an unfamiliar stimulus situation regarding the rating (either in the form of a new rating format or revised administrative instructions) of performance of subordinate behavior, is initially presented with a "mathematical guessing game." In order to optimize either promotional or incentive pay opportunities for his ratee group, he is faced with either relating to rating standards established by the predecessor rating system or conscientiously complying with the administrative demands of the newly introduced rating system. Given this pre-conditioned set, findings such as reflected in Figure 6, Chapter V, are not in the least surprising. After the initial exposure to the new rating situation (again under any absolute rating system), the rating official begins to receive feedback regarding the consequences of his--up to this point in time--reaction to the new stimulus situation as reflected by numeric rating scores:

1. If he was fortunate enough to "guess" a raw score distribution above the overall attained mean, his subordinates
accordingly receive more than their expected proportional frequency of rewards (e.g. promotional opportunities, pay increases, increased retention or tenure).

2. If he attempted to objectively and conscientiously follow administrative instructions accompanying the introduction of the new rating system(s), his subordinate group accordingly suffers the consequences as a by-product of his objective behavior by receiving less than their proportional share of "competitive" absolute rating scores. Supervisors thusly attempting to abide by the rules of the system, are soon made aware of this fact (usually by the negative input provided by disadvantageously affected subordinates) and accordingly are "forced" to modify their rating behavior by increasing mean ratings during the next available rating cycle. Such modification is typically accompanied by decreased variances (for that rater), decreased standards of performance, and again involves guesswork regarding the extent to which other supervisors of employees in competitive units are also engaging in tendencies toward chronological leniency.

3. Those rating officials who find themselves in the position of achieving mean ratings far below the "competitive" absolute values either engage in overreactive behavior by drastically decreasing their rating standards during the next available rerating opportunity or "sticking to their guns," thereby opening the invitation to group employee appeal (a situation which led to introduction of the linear transformation responsive to congressional, employee group, and individual complaints in the first place).

Given the foregoing combination of probable outcomes, findings reported in the preceding chapter regarding tendencies toward chronological leniency are considered not only predictable but also congruent with the patterns of rating behavior habitually established within the AFLC organizational environment. These conclusions are peripherally documented in other reports notably among which are the studies cited by Baier (1951, op. cit.), and Grappe, et al (1967, op. cit.).

The phenomenon of leniency has been examined in detail by other investigators, listing a variety of contributory considerations leading to the "motivational" complex underlying the lenient rater. An excel-
lent summarization of these considerations has been developed by Bass (1956, op. cit.), drawn from an earlier article by Glickman (1955, op. cit.):

The lenient rater is motivated by several considerations:

1. He may feel that anyone under his jurisdiction who is rated unfavorably will reflect poorly on his own worthiness.

2. He may feel that anyone who could have been rated unfavorably had already been discharged from the organization.

3. He may feel that a derogatory rating will be revealed to the ratee to the detriment of relations between rater and ratee.

4. He may rate leniently in order to win promotions for his men and therefore indirectly increase his future control of his subordinates by earning a reputation as a superior with "influence upstairs."

5. He may be projecting.

6. He may feel it necessary to always approve others in order to gain approval for himself.

7. He may be operating on the basis: "Whoever associates with me is meritorious; therefore I am meritorious."

8. He may rate leniently because there exists in our culture a response set to approve rather than disapprove.

Of the considerations listed above, only items number 2, 3, and 8 (which is strikingly analogous to Kneeland's 1929 conclusion that "mankind in general is generous," op. cit.) approach the motivational causality contributory to chronological leniency as described in this study. In a very pragmatic frame of reference, rater tendencies toward leniency as described in this chronology were associated with four observable operational variables:
1. The rating official felt obligated to exercise his loyalty to a subordinate group—as reflected by absolute ratings—rather than to the very impersonal object represented by the regularized, mandatory, leveling "system" directed by higher authority. Interview results revealed that the rating official felt far more comfortable in blaming the "standard scoring system" for subordinate insufficiencies rather than accepting the ethical responsibilities (or consequences) of rating as a direct result of his own behavior. Such findings substantiate the propinquity hypothesis of supervisory esteem as promulgated by David Kipnis (1960, op. cit.). R. L. Thorndike (1949, op. cit.) has expressed a similar supporting opinion:

The loyalty of superior to subordinate may equal or surpass his loyalty to the distant and impersonal agency which has directed that ratings be made. Furthermore, he may feel himself in a sharply competitive situation with other subgroups of the organization. The ratings he gives to his men may, therefore reflect his eagerness to keep them contented and to win promotion for them on the one hand, and his concern that his own ratings come up to those of his competitors on the other. When this is true, instructions and exhortations from the agency using the ratings fall upon deaf ears and the ratings crowd the upper end of the scale.

2. Within the constraints established by mandatory bureaucratic necessity, rating officials found themselves literally caught in an administrative "squeeze play." On one hand the rating official had to acknowledge existence of the linear transformation by achieving a "reasonable" dispersion in his ratings, and on the other hand, he was
well aware of employee appraisal appeal rights\(^1\) to which he might be subjected where assigned elemental ratings were perceived (by the employee) to be less than expected. Faced with this dilemma, the rating official invariably choose in favor of the employee. Toward the latter years of application of the standard scoring system, supervisors often admitted to "rating their subordinates as high as possible and letting the system do their 'leveling' for them." Perhaps, as Ghiselli and Brown (1955, op. cit.) state: "Ratings which are intended to be measures of worker proficiency should not also be expected to serve the purpose of motivating the worker. If the rater has in mind this dual function, then the validity of the rating as a proficiency index will suffer.... In effect the supervisor is using reproof or praise in an attempt to motivate the worker to higher levels of performance. If ratings are used for such a purpose, they are useless for evaluation."

3. The decreasing frequency of competitive promotion actions (Table 19, Chapter V) made it easier for rating officials to resolve their dual role as a selecting official versus that of a rating official. The fact that few promotion situations actually occurred made it easier for the supervisor to reject his responsibilities (for merit promotion differentiation) as a selecting official and identify more closely with his role as an evaluator (by elevating his ratings). Psychological

\(^1\)During the service test, on 15 September 1971, the Air Force issued a formal employee grievance procedure for appealing supervisory appraisals of performance (Air Force Regulation 40-771, 1971). The establishment of a formal appeal procedure, where none had been previously permitted, opens up a vehicle which through its mere existence will be used.
distance between perceptions required of the two roles widened. Had promotion activity increased, psychological distance could very well have narrowed as the supervisor became visually aware of the consequences of his rating behavior through frequent examination of promotion registers. This problem was also accentuated by other organizational affiliates of a downward manning environment (Table 19), which further strengthened the supervisor's loyalty to the work unit. Hence "best" supervisory rating behavior became that which was found to be most conformable to the habits of judgment which the rater used in his own situation (Kingsbury, 1925/26, op. cit.).

4. A final contributory factor leading to chronological leniency was the administrative environment within which the system was required to operate. As pointed out during the interview series summarized in Chapter V, it is obvious that the linear transformation is necessary only where large areas of consideration are required (i.e. installation-wide involving over 20 thousand employees). In situations where promotion consideration might be limited only to the organization possessing the vacancy (i.e. the familiar employee group), statistical leveling between standards of performance applied by different supervisors would neither be necessary nor desirable.

One of the basic points of dissatisfaction with the old Federal Uniform Efficiency Rating System as described by Weschler, et al (1952, op. cit.) seems pertinent to this issue:

Second, the rating procedures have been surrounded with many rules and regulations which do not permit the appearance of valid performance ratings. These rules and regulations—which have been devised primarily to protect the individual employees against
discrimination, favoritism, and unjust treatment [such as installation-wide areas of consideration]—have been incorporated into the system in the interest of good morale and fair play and succeed in preventing injustices which the individual employee may at one time have had to face. However, when attention is focused upon obtaining ratings which accurately reflect the supervisor's estimates of his ratees' performance, the desired results do not seem to be realized.

A distinction must be made in the purposes for which an efficiency rating system is utilized. If the system is instituted as a device for stimulating effective supervisor-employee relations—i.e., to help the employee learn from the supervisor about his weaknesses in a permissive and helpful atmosphere—the present rules and regulations are likely to be unnecessary. If, on the other hand, the rating system is tied in with factors such as promotions, demotions, or reductions in force, then the present rules and regulations seem to prevent the appearance of a realistic distribution of valid ratings. Since the latter is at present the case [also analogous to the AFLC regulatory situation], it is not surprising that the collective efficiency ratings of various government agencies show a preponderance of high ratings and a negligible number of "fair" or "unsatisfactory" ratings.

It is also quite evident that the organizational environment, per se, contributed greatly to the findings reported in this study. To begin with, a study of this scope could only have been conducted in a large industrial setting, with data storage capacity available over many years and where large homogeneous groups are available for analysis. Secondly, organizational adaptation to various appraisal systems such as described in Chapter III had a definite influence upon results achieved during the linear transformation test period. Research along these lines has been strongly recommended by several authors (Whisler and Harper, 1962, op. cit.; Ronan and Prien, 1966, op. cit; MacKinney, 1967, op. cit.; Wallace, 1965; and Viteles, 1932). Above all, a large organization is required. Both Ronan and Prien and MacKinney promise further exploration in studying the complex interaction of situational
variables as related to the larger context of multidimensional job performance measurement. Similar in-depth analyses are also planned as an extension of the present study. Now that unit performance measures can be captured on historical disk storage and individual rater tendencies can also be tracked, across time, meaningful relationships between unit performance variables, supervisory tendencies and individual aptitude and performance measures can be sampled widely.

Such a capability for unit-individual performance interaction analysis should provide valuable information to augment data supplied by other researchers and to lead to a much needed practical theory of work. Several studies have documented the need for such substantiation:

1. Lawler (1967), a proponent of the multitrait-multirater approach to job performance measurement, has integrated organizational characteristics into his conceptual schematic. Beginning with the objective characteristics of the rating system; with the interactive variables of organizational characteristics and individuals; filtering through attitudes toward fairness and acceptability of the system; eventual validity of the resultant ratings can be inferred. As related to this thesis, the moderator effects of organizational characteristics and perceived fairness of the system have been demonstrated as definite influencing factors in this chronology.

2. Culmination of the series of studies initially conducted within the Navy and later replicated in an external industrial organization (Taylor, Parker and Ford, 1959), concluded by expressing a suspicion that situational factors are more important than rating format in
determining the predictability of ratings. If one includes organizational adaptation over time as one of these situational factors (i.e. chronological leniency), then this hypothesis has been substantiated through the findings reported in this study.

3. Ghiselli (1969) has suggested that structural properties of an organization bear heavily upon the efficacy of merit as a basis for filling vacant positions. In general, his thesis proposed that advancement on the basis of merit alone is found to be most efficient in organizations which have large spans of control, which are homogeneous, and are flat (in terms of supervisory span of control). Such performance correlates to gross organizational indices could easily be examined as an extension of the present study.

4. Although the results of this study did not examine supervisory effectiveness as related to rating behavior (Kirchner and Reisberg, 1962), such analyses could easily be retrieved from the master social security account number matrix retained as part of the linear transformation computer program.

To the extent that the findings of this study can be generalized from the operational supervisory rating setting across to those situations utilizing ratings as criteria for personnel research validation purposes, several concerns rise to the surface.

1. First, those researchers investigating the dynamic nature of criteria (Ghiselli, 1956a, op. cit.; Ronan and Prien, 1966, op. cit.; MacKinney, 1967, op. cit.) should be cautious in reviewing the nature of performance change by partialing out the effects of chronological
leniency, wherever it can be documented as a relatively pure variable. Prien (1966) has made the point that criteria of job performance are not only dynamic during the learning and development period but also are related to the dynamic nature of the functions of an organization. In the present AFLC organizational setting, organizational variables and rater habituation were found to render the graphic rating almost valueless as an effective criterion measure. Only neutralization of individual rater tendencies through the linear transformation led any meaningful differentiation to the merit process.

2. The very emergence of chronological leniency as a gross error factor in supervisory merit rating could definitely mask certain of the inferences drawn from the various studies attempting to assess the convergent and discriminant validity of ratings (Campbell and Fiske, 1959; Gunderson and Nelson, 1966; Fiske and Cox, 1960, and Lawler, 1967, op. cit.). The level of precision required to assess multitrait-multirater intercorrelation matrices would not be available where merit ratings had been severely "biased" across time by chronological leniency.

3. The much broader concerns—"Criteria for What?" (Wallace, 1965, op. cit.), "Validity for What?" (Jenkins, 1946), and "Criteria for Criteria" (Weitz, 1961)—are also impacted by the emergence of chronological leniency. Where ratings continue to be used as the best available criterion for validation of independent variables, a distinct likelihood as indicated by both Vroom (1964) and Guion (1965, op. cit.), the contaminating influences of chronological leniency should be examined along with other possible sources of error variance in the criterion. This caution would definitely embrace applied industrial
research settings where merit ratings are collected along an absolute scale, without correction for between-raters differences. Situations involving nominal or ordinal scaling of the dependent variable should be relatively freer from the contaminating effects of chronological leniency. No solutions to the above questions have been offered through this study. In fact, the emergence of chronological leniency should prove to be a very real confounding variable in those longitudinal studies attempting to establish meaningful and interpretable predictive validities.

In light of the inability of the linear transformation correction to appreciably counteract chronological leniency, few administratively acceptable alternatives appear available to management. Emphasis is placed upon administrative—and hence economically acceptable—for reasons as stated: by Kingsbury (1925/26, op. cit.); "The rating-scale ought to be thought of as a practical administrative device, and not as a scientific measuring instrument"; and by Guion (1965, op. cit.); "If criterion development reaches a point at which it continually interferes with the regular production activities in order to maintain the necessary day-by-day records, and if these records do not serve any other useful function, then the testing tail has indeed begun to wag the production dog."

One could speculate that continuance of the linear transformation system in the AFLC environment would eventually result in a plateauing of tendencies toward chronological leniency as reflected by group absolute score distributions. Such a leveling off of the Army rating system was shown by Baier (1951, op. cit.). However, essentially linear
escalation of raw score values on the Air Force OER were graphically demonstrated by Grappé, et al (1967, op. cit.). In any event, speculation regarding future occurrences in the absence of data is seldom productive. In the final analysis, management will be forced to decide the issue whether or not to continue the system based upon perceived returns (in terms of discriminating effectiveness of the measure) versus negative affective reactions (on the part of both the rater and the ratee).

In addition to the cautionary remarks already offered in this chapter, it appears that once again, data has been provided as an argument against operational use of, and the unceasing search for, a unitary common measure of job performance (Toops, 1944; Seashore, Indik, and Georgopoulos, 1960; Wallace and Weitz, 1955; among a host of others). The suggested alternative thusly open to management is to break the merit promotion universe down into homogeneous subgroups and establish, over sufficient periods of time, differential molecules of predictive validities. Advent of third generation computers also opens up a potential for rapid manipulation of individual performance variables as related to organizational objectives criteria interactions. Given the appropriate research environment, even the final utopian industrial research complex as suggested by Toops (1959) may be approached.

Within the existing cultural and attitudinal environment, it appears, however, that future rating approaches will follow much tamer lines than those suggested above. For the rank and file worker,
seniority may very well become the primary determiner of promotability. For the executive, technical, and professional worker, formal assessment approaches (Bray and Grant, 1966) or extensive panel interview processes specific to homogeneous career groupings are expected to predominate. Above all, what is needed is an awareness of the fallibility of human judgments, the resultant limitations imposed upon ratings used in industrial placement and the always surprising fickleness of rater behavior. The findings of the decade of research chronicled in this report reveal the gross imprecision of our supervisory rating process and reconfirm the persistence of man's humanity to man and his loyalty to the cooperative work unit.
APPENDIX 1

Personal Characteristics Appraisal (1961-1964),
All elements were rated for every employee. Seven occupational patterns were selectively applied against the position for which employees were being considered as shown in Table 1, Chapter III.
<table>
<thead>
<tr>
<th>PERSONAL CHARACTERISTICS APPRAISAL</th>
<th>DESCRIPTION OF PERSONAL CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Employee is neat and clean, especially in his personal appearance. This enables him to have a favorable physical impression and be pleasant to be around.</td>
<td></td>
</tr>
<tr>
<td>2. Employee has a definite interest in his work. He desires to put in full days' work exerting his full effort as much as possible.</td>
<td></td>
</tr>
<tr>
<td>3. Employee respects the authority of his superiors. He does not resent taking orders or argue with his supervisor to excess.</td>
<td></td>
</tr>
<tr>
<td>4. Employee is faithful and loyal to his employer. He supports others when they need help and does not run them down behind their backs.</td>
<td></td>
</tr>
<tr>
<td>5. Employee has the ability to hold up under pressure of deadlines; i.e., he does not become easily upset so that it affects his work.</td>
<td></td>
</tr>
<tr>
<td>6. Employee is sufficiently independent to be able to perform his duties without requiring an excessive amount of supervision. He does not lean too heavily on others—whether they be co-workers or supervisors.</td>
<td></td>
</tr>
<tr>
<td>7. Employee has the ability to put things into logical order so that a task can be most effectively accomplished. He thinks ahead. He does not do things haphazardly i.e. without organization.</td>
<td></td>
</tr>
<tr>
<td>8. Employee has an analytical mind and uses good judgment in making decisions. When considering a problem he takes all factors into consideration and comes up with a workable solution.</td>
<td></td>
</tr>
<tr>
<td>9. Employee is dependable and reliable. He does a good job and there is no need to continually check his work. He will not &quot;slow down&quot; when his supervisor is absent.</td>
<td></td>
</tr>
<tr>
<td>10. Employee works with people harmoniously. He has a personality which enables him to get along with people.</td>
<td></td>
</tr>
<tr>
<td>11. Employee goes ahead on his own whenever this is desirable. He is a self-starter.</td>
<td></td>
</tr>
<tr>
<td>12. Employee is dependable in his attendance and prompt in reporting necessary absences.</td>
<td></td>
</tr>
<tr>
<td>13. Employee keeps his work area clean, neat and orderly.</td>
<td></td>
</tr>
<tr>
<td>14. Employee has the ability to come up with new ideas and suggestions. This requires imagination and creativity.</td>
<td></td>
</tr>
<tr>
<td>15. Employee expresses himself orally in a clear, concise, pleasant and effective manner. If his job actually calls for the making of speeches he does this well.</td>
<td></td>
</tr>
<tr>
<td>16. Employee has the ability to express himself in writing in a clear, concise, and effective manner.</td>
<td></td>
</tr>
<tr>
<td>17. Employee cooperates with whomever he comes into contact. He doesn't always insist on getting his way. He is a good team worker. He will do what is asked of him by his supervisor; he will be helpful with his fellow workers, etc.</td>
<td></td>
</tr>
<tr>
<td>18. Employee has the ability to obtain cooperation of others. He causes other people to want to follow his suggestions or advice because they respect him.</td>
<td></td>
</tr>
<tr>
<td>19. Employee is safety-minded. He follows safe practices.</td>
<td></td>
</tr>
<tr>
<td>20. Employee is willing to try new ideas. Once a change has been made he adjusts to the new situation or idea without becoming upset.</td>
<td></td>
</tr>
<tr>
<td>21. Quality of work produced.</td>
<td></td>
</tr>
<tr>
<td>22. Volume of work produced.</td>
<td></td>
</tr>
<tr>
<td>23. Employee desires to get ahead; he is ambitious. He always wants to improve his performance.</td>
<td></td>
</tr>
<tr>
<td>24. Employee is willing to learn and actually has a positive desire to acquire new knowledge that pertains to his job.</td>
<td></td>
</tr>
<tr>
<td>25. Employee has the ability and willingness to assume full responsibility for any task assigned to him.</td>
<td></td>
</tr>
<tr>
<td>26. Employee thinks for himself. He is able to make his own decisions.</td>
<td></td>
</tr>
<tr>
<td>27. Employee is confident in his own abilities. He does not feel inferior.</td>
<td></td>
</tr>
<tr>
<td>28. Employee has the ability to maintain his composure and to remain calm under pressure. He does not become easily upset, even under trying conditions.</td>
<td></td>
</tr>
<tr>
<td>29. Employee has the ability to handle a variety of duties and problems without becoming frustrated or upset.</td>
<td></td>
</tr>
<tr>
<td>30. Employee has an excellent and accurate record in reporting necessary absences.</td>
<td></td>
</tr>
</tbody>
</table>
INSTRUCTIONS FOR THE SUPERVISOR OF THE EMPLOYEE BEING RATED

1. PURPOSE: This form is designed for use in conjunction with both manual and automatic data processing skills locator and merit promotion plan operations. For use with manual systems the card order numbers may be ignored as well as the control information such as clock number and date of birth. The form is also designed for use with both the Air Force Merit Promotion and Annual Performance Rating Programs.

2. SUPERVISOR'S APPRAISAL OF PERSONAL CHARACTERISTICS: The 27 Personal Characteristics listed on the reverse side of this form have been found by Air Force supervisors to be very important to success in a great many Air Force occupations. Your appraisal of the extent to which this employee possesses each of these characteristics will be used to assist in evaluating his potential for promotion or reassignment into various occupations. Although all of the listed characteristics are not appropriate to any one position, it is necessary that the employee be appraised on all traits so that he can be considered for all occupations for which he may be qualified.

The following procedure will aid you in making valid and reliable appraisals:

a. Carefully read each characteristic listed on this form and decide how the employee compares with the standard of performance for his job. Then rate the employee by placing an "X" in the column which best describes him. Each rating should be based upon observation of the trait in the employee's day-to-day work. If you have not directly observed the characteristics in the employee's day-to-day work, and you are not certain how he compares with the performance standard of his job, make an estimate based on all available information including contacts with previous supervisors and higher level supervisors.

b. Consider each characteristic independently of all others; the employee may be superior on one trait and inadequate on another.

c. Be sure to rate all characteristics.

3. The immediate supervisor will complete the Date of Appraisal block as of the date of his signature signifying completion of either or both of the ratings required. The second level supervisor's signature certifies that he has reviewed and approved the Personal Characteristics Appraisal, and that any changes in the ratings have been thoroughly discussed with the immediate supervisor.

4. You will discuss the content of completed appraisal with the employee. When the evaluation concerns deficiencies weaknesses or developmental needs you will discuss with him the need and methods for improvement or development.

5. When this appraisal form has been completed and signed by the appropriate supervisors it will be returned to the Civilian Personnel Office.
APPENDIX 2

Job Performance Appraisal (1965-1968)
### Description of Personal Characteristics

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating Options</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Quality of Work Produced—Is completed work products consistently</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>2. Volume of Work Produced—Is time schedule met at high rate of</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>3. Cost Consciousness—Does the employee conserve time and money?</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>5. Leadership—Is employee actively cooperative with supervisor</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>6. Oral Expression—Does the employee express himself as clear,</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>7. Written Expression—Does the employee write letters,</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>8. Personal Appearance and Habits—Is the employee neat and clean</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>10. Authority Acceptance—Does the employee accept the authority of</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>11. Loyalty—Is employee faithful and loyal to the employee?</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>12. Ability to Work Under Pressure—Is the employee able to operate</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>13. Self-Reliance—How independent is the employee?</td>
<td>E D C B A M</td>
<td></td>
</tr>
<tr>
<td>14. Organizing and Planning—Does he put things in a logical order so</td>
<td>E D C B A M</td>
<td></td>
</tr>
</tbody>
</table>

---

**Signature of Immediate Supervisor:**

**Signature of Second Level Supervisor:**
INSTRUCTIONS FOR THE SUPERVISOR OF THE EMPLOYEE BEING RATED

1. PURPOSE: This form is designed for use in conjunction with both manual and automatic data processing skills locator and merit promotion plan operations. For use with manual systems, the card column numbers may be ignored as well as the control information, such as, check number and date of birth.

2. SUPERVISOR'S APPRAISAL OF JOB PERFORMANCE: The 27 elements listed on the reverse side of this form have been found by Air Force supervisors to be very important to success in a great many Air Force occupations. Your appraisal of the extent to which this employee performs in each of these areas, based upon current performance, will be used to evaluate his potential for promotion or reassignment into various occupations.

3. RATING THE EMPLOYEE: Rate the employee on EACH element by circling the letter in the column which is most descriptive of his performance. Each employee will be rated by comparing his performance with the requirements of the job and with all incumbents of his job series and grade with whom you have worked or supervised. Uniformity between supervisors in standards for rating the extent to which the individual employee possesses each element is essential. Therefore, appraisal scale categories appearing on the reverse side of this form must be carefully applied as follows:

- A rating of "E" on an element indicates CONSISTENTLY OUTSTANDING AND EXCEPTIONAL PERFORMANCE and places the employee among the TOP 5% of all employees in that job series and grade.

- A rating of "D" on an element indicates a performance level which is consistently above satisfactory and often outstanding, and places the employee among about 75% of all employees in the job series and grade who are in the lower portion of the top 15%.

- A rating on "C" on an element indicates that employee performance is consistently above satisfactory and places the employee among about 15% of all employees in the job series and grade who are in the lower portion of the top quarter.

- A rating of "B" on an element indicates that employee performance is usually satisfactory and sometimes above the satisfactory level and places the employee among those in the same job series and grade in the upper third quarter.

- A rating of "A" on an element indicates satisfactory performance and places the employee among the majority or about 50% of all employees of that job series and grade.

- A rating of "M" on an element indicates that the employee's performance on that element is MARGINAL, or below the normally expected level for employees in the same job series and grade level. You should consider withholding a within-grade increase from any employee rated MARGINAL on a substantial number of job-related elements.

4. Be sure to circle only one letter for each of the 27 elements. If revisions are made to the ratings for any of the elements, be sure to leave only one column marked for each element.

5. The immediate supervisor is responsible for appraising the employee on the 27 elements. He will complete the Date of Appraisal block as of the date of his signature. The second level supervisor's signature certifies that he has reviewed and approved all ratings, and that any changes made have been thoroughly discussed with the immediate supervisor. Changes made by the second level supervisor will be accepted as the official rating. After the completed appraisal has been signed by the second level supervisor, the immediate supervisor is required to discuss the contents of the completed appraisal with the employee.

6. When this appraisal form has been completed, signed by the appropriate supervisor, and discussed with the employee, it will be returned to the Civilian Personnel Office. Appraisal on ALL 27 elements must be completed, or the form will be returned to the rating supervisor.

<table>
<thead>
<tr>
<th>ELEMENT</th>
<th>RATING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PURPOSE</td>
<td></td>
</tr>
<tr>
<td>2. SUPERVISOR'S APPRAISAL OF JOB PERFORMANCE</td>
<td></td>
</tr>
<tr>
<td>3. RATING THE EMPLOYEE</td>
<td></td>
</tr>
<tr>
<td>4. Be sure to circle only one letter for each of the 27 elements. If revisions are made to the ratings for any of the elements, be sure to leave only one column marked for each element.</td>
<td></td>
</tr>
<tr>
<td>5. The immediate supervisor is responsible for appraising the employee on the 27 elements. He will complete the Date of Appraisal block as of the date of his signature. The second level supervisor's signature certifies that he has reviewed and approved all ratings, and that any changes made have been thoroughly discussed with the immediate supervisor. Changes made by the second level supervisor will be accepted as the official rating. After the completed appraisal has been signed by the second level supervisor, the immediate supervisor is required to discuss the contents of the completed appraisal with the employee.</td>
<td></td>
</tr>
<tr>
<td>6. When this appraisal form has been completed, signed by the appropriate supervisor, and discussed with the employee, it will be returned to the Civilian Personnel Office. Appraisal on ALL 27 elements must be completed, or the form will be returned to the rating supervisor.</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 3

Tailor-made Appraisal—General Supervisory (beginning in 1968). Five tailor-made forms were eventually applied using this same format:

- AFLC Form 679-A  Wage Board Nonsupervisory
- AFLC Form 679-B  Technician
- AFLC Form 679-C  Supervisory
- AFLC Form 679-D  Scientist and Engineer
- AFLC Form 679-E  Clerical (added July 1969)
### Supervisor's Appraisal - Supervisory

#### Rating Elements

1. Has the fundamental knowledge and skills required to succeed at the job.
2. Stays abreast of new developments in his subject knowledge areas.
3. Plans and organizes his approach to problems.
4. Is flexible in adjusting to changing conditions.
5. Can absorb a great deal of information from various sources on different topics and combine the information to solve problems.
6. Is creative in approach and solution to problems.
7. Is willing to try new principles and techniques in his work.
8. Meets deadlines with acceptable products.
9. Persistes in attempting to solve a problem rather than going on to something else.
10. Organizes and expresses his thoughts clearly and effectively when speaking.
11. Organizes and expresses his thoughts clearly and concisely in writing.
12. Anticipates needs and plans ahead.
13. Can work cooperatively with people outside his immediate office.
15. Has knowledge of organizational environment and the procedures required to accomplish the work load.
16. Has a desire to learn new skills and seeks self-improvement.
17. Establishes work priorities which meet organizational needs.
18. Carries out assignments under adverse conditions.
19. Meets the requirements of his job in terms of personal appearance and habits.
20. Handles situations in a poised, confident manner.
21. Complies with regulations, directives and orders.
22. Effectively plans work of subordinates.
23. Constructively reviews the work of subordinates.
24. Encourages subordinates to improve skills and acquire new knowledge.
25. Keep superior and subordinates informed of new development to project accomplishment and mission.
26. Utilizes subordinates effectively.
27. Trains subordinates to become effective employees.

#### Table

<table>
<thead>
<tr>
<th>Rating Elements</th>
<th>E</th>
<th>D</th>
<th>C</th>
<th>B</th>
<th>A</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Immediate Supervisor

**Signature:**

**Date:**

#### Second Level Supervisor

**Signature:**

**Date:**

---

**AFLC FORM JAN 50, 679C**

---

**SUPERVISORY**
INSTRUCTIONS FOR THE RATING SUPERVISOR

1. PURPOSE: This form is designed for use in conjunction with manual and automatic data processing skills invent and merit promotion plan operations. Your appraisal of the extent to which this employee performs in each of the rating areas will be used to evaluate his potential for promotion or reassignment into various occupations. For annual ratings, your appraisal should be based upon observed performance during the past 12 month period.

2. USE OF THE FORM: The factors found on the reverse side of this form have been found by Air Force supervisors to be important for success in supervisory occupations. Your appraisal of these factors will be used to evaluate the employee's potential for promotion. Please review all of the elements listed on the appraisal and select those relevant to the employee's job duties. CIRCLE THE LETTER "N" BESIDE EACH ELEMENT THAT IS NOT RELEVANT TO HIS JOB. On all relevant items, compare the employee's performance with the requirements of the job and with all incumbents of his job series and grade with whom you have worked or supervised. You must rate the employee on a minimum of 7 elements.

3. RATING THE EMPLOYEE: Rate the employee on all relevant elements by circling the letter which is most descriptive of his performance in accordance with the following scale:

   - A rating of "E" indicates consistently outstanding and exceptional performance on that factor and places the employee among the top 1% of all employees in that series and grade.
   - A rating of "D" indicates performance consistently above satisfactory and often outstanding and places the employee among the top 5% of employees with respect to that factor.
   - A rating of "C" indicates performance on that factor is consistently above satisfactory; the employee is considered among the top 25% of employees on that factor.
   - A rating of "B" on a factor indicates performance which is usually satisfactory, sometimes above satisfactory, and places the employee among the top 45% of employees.
   - A rating of "A" indicates satisfactory performance on that factor and places the employee among the majority of all employees of that job series and grade.
   - A rating of "M" indicates employee performance which is marginal, or below the normally expected level for employees in the same grade and series, and places the employee among the lowest 25% on that factor.

4. Be sure to circle only one letter for each of the elements rated. If revisions are made to the ratings for any of the factors, be sure to leave only one column marked.

5. The immediate supervisor is responsible for rating the employee on the appropriate elements. The second level supervisor's signature certifies that he has reviewed and approved all ratings, and that the reasons for any changes made have been thoroughly discussed with the immediate supervisor. Changes made by the second level supervisor will be accepted as the official rating. After the completed appraisal has been signed by the second level supervisor, the immediate supervisor is required to discuss the contents of the completed appraisal with the employee. He will complete the Date of Appraisal block as of the date of this discussion with the employee. The immediate supervisor's signature certifies that this discussion has been accomplished.

6. When this appraisal form has been completed, signed by the appropriate supervisors, and discussed with the employee, it will be returned to the Civilian Personnel Office. All elements must be rated (marked with an "N" for non-relevant elements or with an appropriate score for relevant elements) or the form will be returned to the rating supervisor.
APPENDIX 4

# CAREER APPRAISAL—CLASS ACT TECHNICAL

(TO BE COMPLETED FOR ALL BS NON-Clerical, Non-Supervisory Employees)

<table>
<thead>
<tr>
<th>NAME (LAST, FIRST, MIDDLE INITIAL):</th>
<th>ORGIN CODE</th>
<th>DATE DUE</th>
<th>APPRAISAL SCALE</th>
</tr>
</thead>
</table>

**GRADE AND PAY SCHEDULE**

<table>
<thead>
<tr>
<th>CLOCK NUMBER</th>
<th>DATE OF BIRTH</th>
<th>DATE OF APPEARAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8</td>
<td>9 10 11 12 13 14 15 16</td>
<td>17 18</td>
</tr>
</tbody>
</table>

**PERIOD OF SUPERVISION**

FROM: | TO: |

**RATING ELEMENTS**

1. Organizes and expresses his thoughts clearly when speaking.  
   **Score:** 8 4 3 2 1 0 N 19

2. Adjusts effectively to new and changing situations; shows a willingness to try out new ideas or operations.  
   **Score:** 8 4 3 2 1 0 N 20

3. Amount of work produced  
   **Score:** 8 4 3 2 1 0 N 21

4. Exhibits good practical judgment  
   **Score:** 8 4 3 2 1 0 N 22

5. Plans and organizes approach to job duties, special projects and problems.  
   **Score:** 8 4 3 2 1 0 N 23

6. Requires minimal supervision, works independently to the extent required by his job.  
   **Score:** 8 4 3 2 1 0 N 24

7. Work products and work performance meet quality standards.  
   **Score:** 8 4 3 2 1 0 N 25

8. Stays abreast of new developments in his subject knowledge areas.  
   **Score:** 8 4 3 2 1 0 N 26

9. Organizes and expresses his thoughts clearly and concisely in writing.  
   **Score:** 8 4 3 2 1 0 N 27

10. Energetic in assuming additional duties; goes ahead on his own whenever desirable.  
    **Score:** 8 4 3 2 1 0 N 28

11. Confident in his ability and knowledge of job.  
    **Score:** 8 4 3 2 1 0 N 29

12. Considers all important aspects of a problem and comes up with the best solution.  
    **Score:** 8 4 3 2 1 0 N 30

13. Self-improvement: desire to increase general scope of knowledge.  
    **Score:** 8 4 3 2 1 0 N 31

14. Accepts responsibility  
    **Score:** 8 4 3 2 1 0 N 32

15. Understands and effectively applies the provisions of regulations, directives and instructions.  
    **Score:** 8 4 3 2 1 0 N 33

16. Gets along well with supervisors and co-workers.  
    **Score:** 8 4 3 2 1 0 N 34

**COMPUTING FINAL CONVERTED SCORE**

(Use the formula below)

\[
\text{CONVERTED SCORE} = \frac{\text{Total Score}}{\text{No. Elements Rated}} \times 100
\]

**COLUMN SUB-TOTALS**

<table>
<thead>
<tr>
<th>Total Score</th>
<th>No. Elements Rated</th>
<th>Converted Score</th>
</tr>
</thead>
</table>

**AFLC FORM 679B**

| PREVIOUS EDITIONS OF THIS FORM ARE OBSOLETE. |

TECHNICAL (CAREER)
1. Rating Elements. The immediate supervisor is responsible for rating the employee on individual rating elements appearing on the appropriate Career Appraisal. Each element will be reviewed as to its relevancy to the employee's job duties. The latter "N" will be circled beside each element that is not judged relevant to the requirements of the employee's current job. Each employee must be rated on a minimum of 7 elements. The employee will be rated on all relevant elements by circling the number which is most descriptive of his performance in accordance with the following scale:

- A rating of "0" indicates employee performance which is marginal, or below the normally expected level for employees in the same grade and series, and places the employee among the lowest 2% on that factor.
- A rating of "1" indicates satisfactory performance on that factor and places the employee among the majority of all employees of that job series and grade.
- A rating of "2" on a factor indicates performance which is usually satisfactory, sometimes above satisfactory, and places the employee among the top 48% of employees.
- A rating of "3" indicates performance on that factor which is considerably above satisfactory; the employee is considered among the top 23% of employees.
- A rating of "4" indicates performance often outstanding and places the employee among the top 9% of employees with respect to that factor.
- A rating of "6" indicates consistently outstanding and exceptional performance on that rating element and places the employee among the top 3% of all employees in that series and grade. Any employee who has more than half of the rated elements checked in the outstanding column must have his rating justified in the narrative portion of the career appraisal. IN ADDITION, ALL RATING OFFICIALS MUST BE MADE AWARE OF THE FACT THAT CLUSTERING ALL SUBORDINATE RATINGS IN THE UPPER END OF THE SCALE (COLUMNS 4 AND 5) ACTUALLY HAS THE RESULT OF PENALIZING THESE EMPLOYEES WHEN THE SCORES ARE TRANSFORMED IN THE SKILLS LOCATOR SYSTEM.

2. Computing Total Score. The Career Appraisal has been designed so that evaluating and endorsing officials can compute a final converted appraisal score for each individual rated. This block is also provided so that endorsing officials can take a look at the rating distributions for employees under their span of control. It is most important that the endorsing official insure that the converted scores are aligned with the overall performance levels of subordinates, that the best worker has the highest score, and that the scores are evenly distributed along the entire scale (as opposed to all of them being clustered at the top). Final converted scores are computed as follows:

- Compute column sub-totals by summing all numbers circled in that column.
- Add all sub-totals to produce a Total Raw Score.
- Compute the Final Converted Score by dividing the Total Raw Score by 5 times the number of elements rated (all non-ratable elements do not count). Since this produces a percentage value, the figure must be multiplied by 100 to yield a whole score.

3. Disposition of Part I. Once the individual elements on Part 1 have been rated, and a final converted score computed, individual element scores will be transposed to Part 2. Part 1 may then be detached, used for the supervisor's discussion with the employee and filed permanently with the employee's AF Form 971, "Supervisor's Record of Employee."
<table>
<thead>
<tr>
<th>NAME (Last, First, Middle Initial)</th>
<th>ODM CODE</th>
<th>PERIOD OF SUPERVISION</th>
<th>DATE OF BIRTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FROM:</td>
<td>TO:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DAY</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**SCORES ON RATING ELEMENTS**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | CONVERTED | SCORE |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

**PROMOTION POTENTIAL** (Check one of the following):

A ☐ Outstanding growth potential. Clearly ahead of contemporaries.

B ☐ Performing well in present job. Should be considered along with contemporaries.

C ☐ Does not demonstrate potential for promotion at this time.

**IMMEDIATE SUPERVISOR'S EVALUATION, SUPPORTING COMMENTS:**

---

**SOCIAL SECURITY ACCOUNT NUMBER**

**TYPED NAME, GRADE**

**DATE**

**SIGNATURE**

**ENDORSING OFFICIAL'S COMMENTS:** (To be completed by the immediate supervisor of the rating official):

---

**SOCIAL SECURITY ACCOUNT NUMBER**

**TYPED NAME, GRADE**

**DATE**

**SIGNATURE**
CAREER APPRAISAL (PART 2)

1. General. The information contained in this portion of the appraisal will be used by selecting officials and promotion boards to evaluate individual potential for promotion to specific vacancies. THE INFORMATION CONTAINED IN THIS PART OF THE EVALUATION IS NOT TO BE SHOWN TO THE EMPLOYEE NOR CO-WORKERS. Rating scores from Part 1 will be transposed to appropriate spaces in Part 2. The score to be entered in Part 2 for each element rated will be the number circled beside that element in Part 1.

2. Evaluating Promotion Potential. Based upon his evaluation of employee performance as reflected through the elemental ratings, the immediate supervisor checks that block on the front of the Career Appraisal which is most indicative of the employee's potential for progression to the next higher grade level. General guidelines for completing this section are:

   It is expected that fewer than 20% of the employees in any given organisational component would be rated as "Outstanding growth potential—Clearly ahead of contemporaries." Employees placed in this category should have their evaluation clearly supported both through their elemental ratings and in the narrative portion of the appraisal.

   Employees placed in the "Does not demonstrate capability for promotion at this time" category should also have their evaluation clearly supported in the narrative. Where appropriate, the narrative could also include recommendations regarding career placement possibilities for such individuals. Is realignment to another type of duty or career area indicated? Should the employee be identified for specific skills training or updating?

3. Immediate Supervisor's Comments. The appraising official (immediate supervisor) will use this space to record his comments concerning the employee's readiness for career development and progression, supported by specifics and reasons. The long range potential of the individual should be estimated along with suggested paths of training and development related to these goals. Other outstanding or marginal aspects of performance, as shown in the elemental ratings, should also be discussed in this portion of the appraisal. The supervisor will date and sign in the appropriate spaces and insure that his name, grade and social security account number are typed in the proper blocks.

4. Endorsing Official's Comments. The appraising official's immediate supervisor will review the Career Appraisal and furnish any additional comments which he believes are pertinent and add significance to the overall career appraisal process. It is further responsibility of the endorsing official to review the total distribution of ratings made by all his subordinate supervisors to insure that the rank ordering is consistent between organisations and that sufficient spread exists between individual ratings. As a general "rule-of-thumb", no two employees should be rated with the same total converted score. Once the endorsing official reviews all appraisals, he will date and sign in the appropriate spaces on the form. Endorsing supervisor's name, grade and social security account number will be typed in appropriate blocks.

5. Processing Completed Forms. When the Career Appraisal has been completed, signed by the appropriate supervisors, and discussed with the employee, it will be detached from Part 1 and returned to the servicing Civilian Personnel Office in a sealed envelope. Immediate supervisors are responsible for insuring that all information in Part 2 is complete, accurate and current.

6. Discussion of Appraisal with Employee. Once both parts of the evaluation have been completed and approved by the endorsing official, the results will be discussed with the employee. Since the evaluation is an integral part of the total career appraisal and promotion process, the discussion should concentrate upon employee goals, performance and past accomplishments related to these goals, training, educational or rotational needs, overall career development objectives and performance requirements related to these objectives.
APPENDIX 5

Standard instructions distributed to all supervisors

--AFLC Supplement 10 to Air Force Manual 40-1,
Section 335, Attachment 1, dated 1 July 1969.
INSTRUCTIONS FOR COMPLETION OF
AFLC CIVILIAN SUPERVISORY AND CAREER APPRAISALS

1. General Policy. It is a requirement of this Command that supervisory appraisals be collected for civilian employees as one of the primary factors used for identifying and ranking those candidates highly qualified for promotion. Corollary to this basic objective are the following additional statements of policy and appraisal philosophy:

a. The appraisal form serves as a formal document for recording supervisory judgments regarding individual attributes of performance, as opposed to informal or unrecorded opinions.

b. The appraisal is relied upon as the primary estimate of skill, performance and behavior levels for use in determining individual promotability.

c. The immediate supervisor, with his judgment of employee performance, is relied upon as the most accurate and valuable source of individual performance information. It is the responsibility of higher levels of management (that is, endorsing officials) to insure a reasonable degree of rating consistency between subordinate raters.

d. Appraisals will be based upon evaluations of specific factors or elements of significant relevance to the requirements of the job occupied by the incumbent. Career development needs, goals, and suggested developmental assignments will be proposed on the basis of these job relevant evaluations.

e. As a vital ingredient in determining future promotability, evaluating and endorsing officials will devote sufficient time to the appraisal process to insure its accuracy, reliability, and validity.

f. The supervisory appraisal for merit promotion is a recapitulation of the supervisor's judgment of employee performance exhibited since the preceding rating period. As such, appraisal results are expected to interact with and be relevant to other employee assessment and evaluation programs. Employees highly rated on the supervisor's appraisal are expected to be among the group of individuals nominated for performance and other job related types of awards. Conversely, employees receiving such awards are expected to possess high supporting supervisors' appraisals. No single action can better destroy the credibility of the supervisory appraisal system, and the evaluating officials themselves, than inconsistencies between the various evaluation programs.

g. The official appraisal scale appearing on the supervisor's appraisal is designed as a general guide for the assessment of each rating factor. The scale is not to be applied as a mandatory overall rating distribution nor are the overall total scores required to be reconfigured in any specified statistical distribution. It is expected, however, that adequate dispersion exist between individuals to provide for meaningful discrimination between competitors. Rarely would an employee so excel in performance that he attains a perfect score of 100.

h. Overall levels of employee performance are expected to remain relatively stable over a 12 month period. Where substantial changes do occur, they will be supported by reappraisals submitted during the open season on a management by exception basis.

2. Applicability. Appraisal forms described in the following subparagraphs are approved for use in effecting in-service promotion actions within the AFCL.

a. AFCL Form 679-A, "Supervisor's Appraisal—Wage Board Nonsupervisory," will be completed for all nonsupervisory Wage Board personnel.

b. AFCL Form 679-B, "Career Appraisal—Technician," will be completed annually for all nonclerical General Schedule employees through grade GS-13. Exempted are those employees covered by a separate career appraisal, as required either in this attachment or by higher authority. Where this form is used for employees in grades GS-4 and below, the
narrative portions of part 2 will not be completed.

c. AFLC Form 679-D, "Career Appraisal, Supervisory," will be completed for all supervisory personnel, both General Schedule and Wage Board.

d. AFLC Form 679-D, "Career Appraisal, Scientist and Engineer," will be completed annually for all nonsupervisory employees assigned to scientist and professional engineering job series.

e. AFLC Form 679-E, "Supervisor's Appraisal - Clerical." To be completed for all General Schedule nonsupervisory secretarial, stenographic and clerical employees (identified by use of the appropriate term in the employee's job title).

3. Transformation of Appraisal, Scores For Use In Merit Promotion. In order to reduce the influence of differences between rating standards applied by individual evaluating supervisors, all appraisal scores will be transformed within the skills locator system to a standard scale having an average score of 70, with ratings distributed normally around 70 from a low of 40 to a high of 100. This procedure will enable top rated employees in one organization to receive promotion consideration on a basis equal to that given top ranked employees in other organizations, regardless of the rating standards applied by different evaluating supervisors. Basic attributes of the transformation procedures are summarized below:

a. Corrections will be made around the rating distributions and policies established by the endorsing official. Separate transformations will be made for each of the five basic rating forms, controlled around the endorsing official.

b. Endorsing officials are responsible for insuring that the distribution of scores for all employees under their span of review are aligned with the overall performance levels of subordinates, that the best worker has the highest score and that scores are distributed along the entire scale, accurately reflecting performance from the most effective to the least effective.

c. All forms, including part 1 of the career appraisals, have been structured to provide for supervisory computation of final converted (not transformed) total scores. This step in the rating process has been inserted so that endorsing officials may compare all subordinates rated on the same form (for example, all clerical, etc.) to insure that each individual is rated with a different total score. The total converted scores for all subordinates rated on the same forms should reflect their rank order, from the highest to the lowest.

As shown in the example in figure 1, final converted scores are to be computed as follows:

(1) Compute column subtotal by summing all numbers circled in that column (two fives were circled for elements 6 and 20 = a subtotal of 10).

(2) Add all subtotals to produce a Total Raw Score (10 + 20 + 18 + 8 + 4 (+ 0) = 60).

(3) Compute the Final Converted Score by dividing the Total Raw Score by 5 times the number of elements rated (nonratable elements do not count). Since this produces a percentage value, the figure must be multiplied by 100 to yield a whole score. Thus, 60 + (5 x 22) x 100 = 54 (round to nearest whole number).
<table>
<thead>
<tr>
<th>RATING ELEMENTS</th>
<th>APPRAISAL SCALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Collecting background information before and during performance of work</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>2 Designing and conducting research studies</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>3 Reviewing and evaluating research findings</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>4 Writing research reports</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>5 Knowledge of fundamental technical information</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>6 Knowledge of current technical literature</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>7 Understanding of scientific or engineering techniques and principles and their application</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>8 Applying scientific or engineering techniques and principles</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>9 Creativity in applying research findings in the development of new methods, materials, or equipment</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>10 Planning and directing research studies and programs of other professionals</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>11 Effectively relating ideas and research findings to management and co-workers</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>12 Getting along with supervisor</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>13 Getting along with colleagues</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>14 Getting along with non-technical personnel</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>15 Formulating new research proposals</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>16 Meeting deadlines</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>17 Constructively reviewing research plans and research reports prepared by other professionals</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>18 Planning and carrying out problems of facilities, budget and personnel</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>19 Desire to learn and improve</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>20 Effective and general presentation of scientific results</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>21 Personal appearance</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>22 Knowledge of organization and their mutual</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>23 Camaraderie and cooperation with other personnel</td>
<td>5 3 2 1 0</td>
</tr>
<tr>
<td>24 Flexibility in adjusting to changing conditions</td>
<td>5 3 2 1 0</td>
</tr>
</tbody>
</table>

**Computed Final Converting Score**

*See instructions below*

**Column Sub Totals**

<table>
<thead>
<tr>
<th>Column Sub Totals</th>
<th>Total Score</th>
<th>Number of Ratings Noted</th>
<th>Converted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60</td>
<td>3</td>
<td>4.2</td>
</tr>
</tbody>
</table>

**Figure 1—Career Appraisal—Scientists and Engineers**
CAREER APPRAISAL (PART I)

1. Rating Elements. The immediate supervisor is responsible for rating the employee on individual rating elements appearing on the appropriate Career Appraisal. Each element will be reviewed as to its relevancy to the employee's job duties. The letter "N" will be circled beside each element that is not judged relevant to the requirements of the employee's current job. Each employee must be rated on a minimum of 7 elements. The employee will be rated on all relevant elements by circling the number which is most descriptive of his performance in accordance with the following scale:

A rating of "0" indicates employee performance which is marginal, or below the normally expected level for employees in the same grade and series, and places the employee among the lowest 1% on that factor.

A rating of "1" indicates satisfactory performance on that factor and places the employee among the majority of all employees of that job series and grade.

A rating of "2" on a factor indicates performance which is usually satisfactory, sometimes above satisfactory, and places the employee among the top 25% of employees.

A rating of "3" indicates performance on that factor which is considerably above satisfactory; the employee is considered among the top 45% of employees.

A rating of "4" indicates performance often outstanding and places the employee among the top 10% of employees with respect to that factor.

A rating of "5" indicates consistently outstanding and exceptional performance on that rating element and places the employee among the top 3% of all employees in that series and grade. Any employee who has more than half of the rated elements checked in the outstanding column must have his rating justified in the narrative portion of the career appraisal.

In addition, all rating officials must be made aware of the fact that clustering all subordinate ratings in the upper end of the scale (columns 4 and 5) actually has the result of penalizing these employees when the scores are transformed in the skills locator system.

2. Computing Total Score. The Career Appraisal has been designed so that evaluating and endorsing officials can compute a final converted appraisal score for each individual rated. This block is also provided so that endorsing officials can take a look at the rating distributions for employees under their span of control. It is most important that the endorsing official insures that the converted scores are aligned with the overall performance levels of subordinates, that the best worker has the highest score, and that the scores are evenly distributed along the entire scale (as opposed to all of them being clustered at the top). Final converted scores are computed as follows:

Compute column sub-totals by summing all numbers circled in that column.

Add all sub-totals to produce a Total Raw Score.

Compute the Final Converted Score by dividing the Total Raw Score by 5 times the number of elements rated (all non-ratable elements do not count). Since this produces a percentage value, the figure must be multiplied by 100 to yield a whole score.

3. Disposition of Part I. Once the individual elements on Part I have been rated, and a final converted score computed, individual element scores will be transposed to Part 2. Part 1 may then be detached, used for the supervisor's discussion with the employee and filed permanently with the employee's AP Form 971, "Supervisor's Record of Employee."
4. Completion of the Career Appraisal (AFLC Forms 679B, C and D). Part 2 of the AFLC career appraisal has been designed primarily as an adjunct to Promotion Board and career counseling operations.

   a. Where completed for employees normally considered for promotion through board operations (GS-11 and above and equivalent WF grades), narrative portions of the career appraisal will be directed toward augmentation of the elemental ratings (part 1) specifically as they relate to consideration for promotion to the next higher grade level. Where appropriate, the evaluating supervisor should also include comments regarding suggested rotational assignments, additional familiarization training and other specific developmental needs of the employee. Part 2 of the career appraisal will be centrally filed in the servicing civilian personnel office for use in promotion board and referral operations.

   b. Part 2 of the career appraisal completed for employees below the grade level of referral to promotion boards (GS-9 and below) should emphasize long range career goals, near term development needs related to these goals, and an explanation of specific rating elements (part 1) shown as outstanding or particularly unsatisfactory. At this grade level, special attention should be devoted to developing employee strengths rather than concentrating on areas of weakness. Portions of the narrative addressing themselves to employee developmental needs should be jointly agreed to by the supervisor and the employee. Part 2 of the career appraisal for employees at grades GS-9 and below will be maintained in the Civilian Personnel Office for referral to selecting officials along with the list of top ranked eligibles on a promotion certificate.

5. Supervisory Discussion with the Employee. It is the policy of this Command that results of the supervisory appraisal process be openly discussed with the employee. This does not imply that the rating sheet must be covered item by item with the employee, but rather, that he be given a general review of his relative standing. The discussion should be structured to insure employee understanding of the appraisal program, its application in the promotion process and the interaction between the supervisory appraisal for merit promotion and other evaluation and career development programs (that is, the annual performance rating, performance awards, and specific career development programs). Enough time should be set aside for the discussion to insure that all of the employee’s questions can be heard and answered. Unanswered or unresolved questions regarding the merits of the rating will be referred to the endorsing official and questions pertaining to appraisal regulations or procedures referred to the servicing personnel management specialist. Above all, the discussion should attempt to resolve areas of supervisory/employee misunderstanding, to relieve any employee anxieties or misconceptions regarding the appraisal situation and to relate career goals to the current performance level of the employee.

6. Supervisory Orientation and Training:

   a. All newly assigned supervisors must receive, as a minimum, formal instruction on the following aspects of the appraisal program:

      (1) General orientation training on the overall Command program, basic objectives, and relation to career management programs.

      (2) Specific instruction regarding the techniques and theories of employee counseling as they relate to the appraisal discussion environment.

      (3) A full explanation of the transformation process, with examples developed from local data.

   b. All supervisors, with particular attention to endorsing officials, must be annually reacquainted with the appraisal process, including program changes introduced since the last rating cycle.
APPENDIX 6

Sample of installation Civilian Personnel Office
News Release announcing advent of the standard
scoring procedure.
SUPERVISORY APPRAISALS

The Supervisory Appraisal used in the Merit Promotion Program is the record on which supervisors judge employee performance. It is one of the primary factors relied upon for determining an employee's standing on a Merit Promotion Roster.

Critical reviews have been made of the Air Force Logistics Command (AFLC) appraisal system over the past several years. These reviews reveal two major problem areas.

- First, the continuing rise in the average scores given to employees indicate that supervisors prefer to rate their employees as being nearly equal to each other in performance. This means that most employees receive very similar scores. The effectiveness of the appraisal as a ranking factor in the promotion program has been reduced because the supervisor has not made sufficient effort to distinguish his outstanding employees from his satisfactory ones. As an example, a supervisor may have ten employees and be readily able to rank them one through ten based on their performance. However, when reflecting his ratings on an appraisal form, he rates all ten employees within a 5-point spread, perhaps from 90 to 95. The difference in appraisal scores between the top employee with 95 points and the remaining nine employees is so slight that the top rated employee is not distinguished for merit promotion purposes.

- The second problem which has plagued past appraisal systems concerns the "high", "average" and "low" raters. Some supervisors tend to rate all their employees high, others rate low, and still others rate in the middle range of the appraisal scale. If we would question three different supervisors who might fall into these categories, we would find that each of them feels that he is honestly applying his best judgment to the appraisal process and, in his opinion, is ranking his employees fairly. At the same time, however, all three supervisors would agree that the top rated employee(s) of each supervisor should be in competition with the top rated employee(s) of the other two supervisors. Under the old appraisal system, however, the best employees of the supervisors who tended to rate "low" were unable to compete with the top or even average employees of a supervisor who rated "high" because the raw numerical score of the appraisal form itself was used for ranking purposes.

To resolve these problems it has become necessary to: (1) find a way to cause the spread of appraisal scores to be increased among employees so that it could be used to help identify the most promotable employees on promotion rosters and (2) realign the scores of the employees rated by "high", "average" and "low" raters so that the most promotable employees in each grouping would be in competition with each other. (continued)
Headquarters AFLC plans to use a mathematical technique applied to the automated skills locator system to alleviate these problem areas. To reduce the influence of differences between the rating patterns of different supervisors, all employee appraisal scores will be adjusted to a standard scale. The scale has an average score of 70 and a range of points from 40 to 100. The score the employee receives on his supervisory appraisal will be adjusted to this standard scale and this adjusted score will be used for merit promotion. This means that the top ranked employees of the "low" rater will have an adjusted score which will be nearly equal to the adjusted score of the top employee rated by the "high" rater. The new system will also require supervisors to show differences between their outstanding and satisfactory employees rather than rank them all close together. If the supervisor does not make such a distinction, all of the scores of his employees may be transformed to the "average" appraisal score. This requirement will make the adjusted appraisal score more meaningful for merit promotion purposes because there will be a significant spread of appraisal scores. Because of these adjustments, most employees will notice a decrease in their appraisal score this year as compared to previous years. Hq AFLC recognizes that the new approach may not automatically eliminate all problems in this complex area, but they are very hopeful that it will make the appraisal scores more valid than they have been and more equitable in their affect upon promotions.

THE ENDORSING OFFICIAL

The basic responsibility for insuring that all aspects of Hq AFLC's appraisal policy are carried out rests primarily with the endorsing official who is the second level reviewing supervisor. The endorsing official and his subordinate raters must insure that there are greater differences in the scores between their most effective employees and their least effective employees. Most employees have good points as well as shortcomings in their performance. Therefore, a difference in ratings can reasonably be expected. Since performance evaluations are also made to assist employees in developing their capabilities, periodic discussions throughout the course of a rating period as well as the formal discussion period at the time of rating are absolutely essential. Above all, the discussion should attempt to resolve areas of supervisory/employee misunderstanding, to relieve any employee anxiety or misconception regarding the appraisals, and to relate the employee's career goals to his current performance level. As a vital ingredient in determining the future promotability of employees, Hq AFLC requires supervisors to devote sufficient time to the appraisal process to insure its accuracy and reliability.
APPENDIX 7

Revised tailor-made appraisal format—wage-board nonsupervisory; issued February 1970.
### PERFORMANCE APPRAISAL – WAGE BOARD NONSUPERVISORY

#### RATING ELEMENTS

1. **Requires Minimum Supervision. Works independently to the extent required by the job.**
   - Level: 4
   - Comments: N/A

2. **Meets time schedules and maintains production.**
   - Level: 4
   - Comments: N/A

3. **Operates effectively under pressure of deadlines.**
   - Level: 4
   - Comments: N/A

4. **Cooperates well with others and is a good team worker.**
   - Level: 4
   - Comments: N/A

5. **Adjusts effectively to new and changing situations; shows a willingness to try out new ideas or operations.**
   - Level: 4
   - Comments: N/A

6. **Actively tries to improve performance by gaining new knowledge pertaining to the job.**
   - Level: 4
   - Comments: N/A

7. **Work products and work performance meet quality standards.**
   - Level: 4
   - Comments: N/A

8. **Attempts to minimize error and reduce waste.**
   - Level: 4
   - Comments: N/A

9. **Stays abreast of new developments in job related subject knowledge areas.**
   - Level: 4
   - Comments: N/A

10. **Willingly accepts full responsibility for any task assigned.**
    - Level: 4
    - Comments: N/A

11. **Minimizes risks with good safety practices.**
    - Level: 4
    - Comments: N/A

12. **Diligent in attendance, reports absences promptly.**
    - Level: 4
    - Comments: N/A

13. **Soloves problems encountered in assignments and, when needed, can independently troubleshoot.**
    - Level: 4
    - Comments: N/A

14. **Takes good care of all tools, material and equipment.**
    - Level: 4
    - Comments: N/A

15. **Follows policies, procedures and supervisory instructions.**
    - Level: 4
    - Comments: N/A

---

**Endorsing Official**

<table>
<thead>
<tr>
<th>ORGN</th>
<th>SSAN</th>
<th>Typed Name, Grade</th>
<th>Date</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**First Level Supervisor**

<table>
<thead>
<tr>
<th>ORGN</th>
<th>SSAN</th>
<th>Typed Name, Grade</th>
<th>Date</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**AFLC FORM 679A**

© 1970 AFL-CIO

Previous editions are obsolete

Part I – To be sent to Civilian Personnel
INSTRUCTIONS FOR THE RATING SUPERVISOR

1. General. The Af LC Performance Appraisal is to be completed in two parts: Part I is the official copy of record and is to be forwarded to the Civilian Personnel Office; Part II is the supervisor's copy and is to be detached, used for the supervisor’s discussion with the employee and filed with the employee’s AF Form 971, “Supervisor’s Record of Employee.” Part II should be completed first in pencil, reviewed with the endorsing official and the final ratings transposed in ink to Part I. The period of supervision block shows the entire period of time the employee has been assigned under the immediate supervisor.

2. Rating Elements. The immediate supervisor is responsible for rating the employee on individual rating elements appearing on the reverse side of this form. Each element will be reviewed as to its relevancy to the employee’s job duties. The letter “N” will be circled beside each element that is not judged relevant to the requirements of the employee’s current job. Each employee must be rated on a minimum of 7 elements.

A rating of “0” indicates a level of employee performance which is consistently below the normally expected level for employees in the same grade and series and which coincides with the low point on the supervisor’s performance standard for that rating element.

A rating of “5” indicates performance on that factor which is consistently at the high point on the supervisor’s performance standard and represents exceptional and consistently outstanding behavior.

The intermediate rating points represent levels of performance between the high point of “5” and low point of “0” and are to be evaluated consistent with standards of performance established by the rating supervisor. It is further expected that each rating supervisor take full advantage of the entire rating scale.

3. Computing Total Score. The appraisal has been designed so that evaluating and endorsing officials can compute a final converted appraisal score for each individual rated. This block is also provided so that endorsing officials can take a look at the rating distributions for employees under their span of control. It is most important that the endorsing official insure that the converted scores are aligned with the overall performance levels of subordinates, that the best worker has the highest score, and that the scores are distributed along the entire scale to the extent warranted by differences in performance (as opposed to all of them being clustered at any one point on the scale). Converted scores are computed as follows:

Compute column sub-totals by summing all numbers circled in that column.

Add all sub-totals to produce a Total Raw Score.

Compute the Converted Score by dividing the Total Raw Score by 5 times the number of elements rated (all non-ratable elements do not count). Since this produces a percentage value, the figure must be multiplied by 100 to yield a whole score. DURING THE EMPLOYEE’S DISCUSSION IT MUST BE EXPLAINED THAT THE DATA ON THE FORM WILL BE PROCESSED BY THE COMPUTER TO DEVELOP A STANDARD SCORE FOR RANKING FOR MERIT PROMOTION PURPOSES. IT SHOULD FURTHER BE EXPLAINED THAT THE STANDARD SCORE MAY DIFFER FROM THE CONVERTED SCORE ON THE BOTTOM OF THE FORM, BUT THAT THE STANDARD SCORE IS DIRECTLY BASED ON THE CONVERTED SCORE, PLUS THE EMPLOYEE’S STANDING AMONG OTHER EMPLOYEES WITHIN THE RATING UNIT.

4. The Endorsing Official. The endorsing official is responsible for reviewing the total distribution of ratings made by all subordinate supervisors to insure that rank ordering is consistent between organizations and that sufficient spread exists between individual ratings to realistically reflect observed differences in performance. Changes made by the endorsing official are accepted as the final official rating. Endorsing officials must insure that final ratings are not clustered around any one point on the rating scale in order to distinguish the better from the average employee. Additionally, endorsing officials must insure that ratings are not artificially forced to conform to any predetermined statistical distribution. Once the endorsing official reviews all appraisals, date and signature spaces on the form will be completed. THE ENDORSING OFFICIAL’S SOCIAL SECURITY ACCOUNT NUMBER MUST BE ACCURATELY RECORDED IN THE APPROPRIATE BLOCK ON PART I OF THE APPRAISAL.

5. Disposition of Completed Appraisal. After the completed appraisal has been signed by the second level supervisor, the immediate supervisor will discuss the results with the employee. The date of appraisal block will be completed as of the date of the discussion with the employee. The supervisor will review the completed form for accuracy and completeness, detach Part I and forward to the servicing Civilian Personnel Office. Part II will be retained for filing with the employee’s AF Form 971.
APPENDIX 8

Revised standard instructions distributed to all supervisors—APLC Supplement 1 (Change 1) to Air Force Manual 40-1, Section 335, Attachment 1, dated 17 September 1970.
INSTRUCTIONS FOR COMPLETING AFLC
CIVILIAN PERFORMANCE APPRAISALS

1. General Policy. It is a requirement of this Command that performance appraisals be collected for civilian employees and used as one of the primary factors for identifying and ranking candidates for promotion. Corollary to this basic objective are the following additional statements of policy and appraisal philosophy:

a. The appraisal form serves as a formal document for recording supervisory judgments regarding individual attributes of performance, as opposed to informal or unrecorded opinions.

b. The appraisal is relied upon as the primary estimate of current performance for use in determining individual promotability.

c. The immediate supervisor's judgment of employee performance is relied upon as the basic source of individual performance information. It is the responsibility of endorsing officials to insure a reasonable degree of rating consistency between subordinate raters.

d. Appraisals will be based upon evaluations of specific factors or elements of significant relevance to the requirements of the job occupied by the incumbent.

e. As a vital ingredient in determining future promotability both evaluating and endorsing officials will devote sufficient time to the appraisal process to insure its accuracy and relevance.

f. The performance appraisal used for merit promotion is a recapitulation of the supervisor's judgment of employee performance exhibited since the preceding rating period. As such, appraisal results are expected to interact with and be relevant to other employee assessment and evaluation programs. Employees highly rated on the supervisor's appraisal are expected to be among the group of individuals nominated for performance and other job related types of awards. Conversely, employees receiving such awards are expected to possess high supporting supervisors' appraisals.

g. The appraisal scale guideline appearing on the performance appraisal is prescribed as a general guide for the assessment of each rating factor and is to be related to individual standards of performance established for the position. The degree of attainment of such standards are presumed to be ordered on a scale from a low to a high degree of accomplishment. All scale points, including intermediate levels, should be evaluated based on consistent habitual performance rather than occasional outstanding or marginal examples of behavior.

2. Applicability. Appraisal forms described in the following subparagraphs are approved for use in effecting in-service promotion actions within the AFLC:

a. AFLC Form 679A, "Performance Appraisal—Wage Grade Nonsupervisory," will be completed for all nonsupervisory Wage Grade personnel.

b. AFLC Form 679B, "Performance Appraisal—Class Act Technical," will be completed annually for all nonclerical General Schedule employees through grade GS-13. Excepted are those employees covered by a separate career appraisal, as required either in this attachment or by higher authority.

c. AFLC Form 679C, "Performance Appraisal—Supervisory," will be completed for all supervisory personnel, both General Schedule and Wage Grade.

d. AFLC Form 679D, "Performance Appraisal—Scientist and Engineer," will be completed annually for all nonsupervisory employees assigned to scientist and professional engineering job series. (See Occupational Coverage of Scientists and Engineers).

e. AFLC Form 679E, "Performance Appraisal—Clerical," will be completed for all General Schedule nonsupervisory typist and stenographic employees and employees assigned in the GS-312, 318, and 322 job series.

3. Computation of Standard Appraisal Scores for use in Merit Promotion. In order to reduce the influence of differences between rating standards applied by indi-
individual evaluating supervisors, all appraisal scores will be computed within the skills locator system to a standard scale having an average score of 100 with a range from 70 to 130. This procedure will enable top rated employees in one organization to receive appraisal scores consistent with top rated employees in other organizations, regardless of the rating standards applied by different evaluating supervisors. Basic attributes of the standard scoring procedure are summarized below:

a. The AFLC standard scoring system is primarily a scaling and ranking technique used along with test and experience information in constituting promotion profiles. The standard score is based directly upon the supervisor's evaluation of performance and the employee's standing among other employees within the rating unit. Although the standard score will differ from the original converted score, it in no way alters the immediate and endorsing supervisor's descriptions of performance as reflected through individual elemental ratings. It is the elemental rating which is to be applied by the supervisor in discharging his responsibilities in employee counseling, training, and performance review.

b. Standard scores will be computed around the rating distributions and policies established by the endorsing official. Separate computations will be made for each of the five basic rating forms, controlled around the endorsing official. The rating approved and signed by the endorsing official is accepted as the official rating. The endorsing official is identified as the immediate supervisor of the rating supervisor. The rating supervisor is defined as one having full administrative and technical supervisory responsibility over the work of an employee.

c. Where structurally unique organizations exist, local management officials in consultation with and upon the approval of the Civilian Personnel Officer can locally establish levels of endorsement to increase the reliability and relevance of the performance appraisal. Variances granted for unusual organizational structures involving large numbers of employees should be confirmed by AFLC' (DPCB). The level for endorsement should be publicized by each activity.

d. All forms have been structured to provide for supervisory computation of converted (not standardized) total scores. This step in the rating process has been provided so that endorsing officials may review the extent to which employees under their span of control are meeting performance requirements. As shown in the example in figure 1, final converted scores are to be computed as follows:

1. Compute column subtotal by summing all numbers circled in that column (two fives were circled for elements 6 and 20 = a subtotal of 10).

2. Add all subtotals to produce a Total Raw Score (10 + 20 + 18 + 8 + 4 (+ 0) = 60).

3. Compute the Final Converted Score by dividing the Total Raw Score by 5 times the number of elements rated (non-ratable elements do not count); since this produces a proportion, the figure must be multiplied by 100 to yield a percentage value. Thus, $60 \div (5 \times 22) \times 100 = 55$ (round to nearest whole number.)

4. Supervisory Discussion with the Employee. It is the policy of this Command that results of the performance appraisal be openly discussed with the employee. The immediate supervisor will discuss the results of the appraisal with the employee. This discussion with the employee will not be conducted until after the performance appraisal has been reviewed and signed by the endorsing official. The discussion should be structured to insure that the employee understands the standards of performance required by his supervisor. Enough time should be set aside for the discussion to insure that all of the employee’s questions can be heard and answered. Unanswered or unresolved questions regarding the merits of the rating will be referred to the endorsing official and questions pertaining to appraisal regulations or procedures referred to the appropriate servicing personnel specialist. Above all, the discussion should attempt to resolve areas of supervisory/employee misunderstanding, to relieve any employee anxieties or misconceptions re-
Regarding the appraisal situation, and to relate career goals to the current performance level of the employee. It is expected that supervisory-employee discussions of work performance will occur frequently throughout the rating period. Feedback to the employee about the extent to which his work is meeting performance standards established by management should be continuous.

5. Supervisory Orientation and Training:

a. All newly assigned supervisors must receive, as a minimum, formal instruction on the following aspects of the appraisal program:

(1) General orientation training on the overall Command program, basic objectives, and relation to career management programs.

(2) Specific instruction regarding the techniques and theories of employee counseling as they relate to the appraisal discussion environment.

(3) A full explanation of the standardization process, with examples developed from local data.

b. All supervisors, with particular attention to endorsing officials, must be annually reacquainted with the appraisal process including program changes introduced since the last rating cycle.
APPENDIX 9

Appraisal Distribution Report. Sample of Master Matrix Printout collected from the Civilian Personnel Office, WPAFB.
<table>
<thead>
<tr>
<th>SSAM</th>
<th>FREQ</th>
<th>679/A DEV</th>
<th>MEAN</th>
<th>FREQ</th>
<th>679/B CEV</th>
<th>MEAN</th>
<th>FREQ</th>
<th>679/C DEV</th>
<th>MEAN</th>
<th>FREQ</th>
<th>679/D CEV</th>
<th>MEAN</th>
<th>FREQ</th>
<th>679/E DEV</th>
<th>MEAN</th>
<th>FREQ</th>
<th>679/F TOTAL</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>040-30-9817</td>
<td>0.60</td>
<td>0.00</td>
<td>2</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>0.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>041-28-2251</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>042-12-3139</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>042-40-4271</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>98.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>043-18-3115</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>83.00</td>
<td>3</td>
<td>6.92</td>
<td>83.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>043-22-3943</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>0.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>043-14-3322</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>047-12-3903</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>048-26-0017</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>82.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>050-27-2951</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>82.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>051-22-7678</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>87.25</td>
<td>4</td>
<td>14.31</td>
<td>88.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>053-23-4488</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>054-19-3672</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>3</td>
<td>3.54</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>056-36-4664</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>17.00</td>
<td>77.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>057-31-4669</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>057-07-2949</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>10</td>
<td>1.83</td>
<td>97.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>058-14-9158</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>058-17-2613</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>1.80</td>
<td>98.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>058-16-4224</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>059-31-4614</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>88.63</td>
<td>3</td>
<td>7.26</td>
<td>95.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>063-09-6646</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>4.00</td>
<td>92.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>063-24-2407</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>064-32-7781</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>069-14-7690</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>070-26-0793</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>99.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 10

Endorsing Official's Distribution Listing Sample printout collected from Ogden Air Material Area. Alpha headers refer to specific appraisal form.

C-SCR Raw Score
SD-SCR Standard Score for (individual[s]) possessing raw standard score
<table>
<thead>
<tr>
<th>C-SCR</th>
<th>PRLW</th>
<th>A-SCR</th>
<th>FREQ</th>
<th>B-SCR</th>
<th>FREQ</th>
<th>C-SCR</th>
<th>FREQ</th>
<th>D-SCR</th>
<th>FREQ</th>
<th>E-SCR</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>117</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>108</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>109</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>106</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>102</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>98</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>97</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>96</td>
<td>1</td>
<td>97</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>93</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>89</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>86</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>0</td>
<td>00</td>
<td>1</td>
<td>83</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>0</td>
<td>00</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>6</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| SUMS  | 00   | 398   | 970  | 00   | 930   | 1898 |
| SUMS-5C| 00   | 31834 | 79032| 00   | 47146 | 158062|
| MEAN  | 0.00 | 79.60 | 80.83| 0.00 | 88.33 | 82.52|
| STD-DV| 0.00 | 6.37  | 7.21 | 0.00 | 7.41  | 7.90 |</p>
<table>
<thead>
<tr>
<th>SCORE INTERVAL</th>
<th>Pre-Standardization 1969</th>
<th>Post-Standardization 1969</th>
<th>1970</th>
<th>1971</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>100-96</td>
<td>19,567</td>
<td>18.23</td>
<td>4,897</td>
<td>4.73</td>
</tr>
<tr>
<td>95-91</td>
<td>20,141</td>
<td>18.77</td>
<td>10,160</td>
<td>9.24</td>
</tr>
<tr>
<td>90-86</td>
<td>17,714</td>
<td>16.61</td>
<td>12,784</td>
<td>12.50</td>
</tr>
<tr>
<td>85-81</td>
<td>12,732</td>
<td>11.66</td>
<td>14,556</td>
<td>14.24</td>
</tr>
<tr>
<td>80-76</td>
<td>8,820</td>
<td>8.02</td>
<td>14,723</td>
<td>14.40</td>
</tr>
<tr>
<td>75-71</td>
<td>5,215</td>
<td>4.66</td>
<td>12,963</td>
<td>12.68</td>
</tr>
<tr>
<td>70-66</td>
<td>4,897</td>
<td>4.66</td>
<td>8,960</td>
<td>8.76</td>
</tr>
<tr>
<td>65-61</td>
<td>3,477</td>
<td>3.24</td>
<td>7,028</td>
<td>6.87</td>
</tr>
<tr>
<td>60-56</td>
<td>2,882</td>
<td>2.69</td>
<td>5,182</td>
<td>5.07</td>
</tr>
<tr>
<td>55-51</td>
<td>1,981</td>
<td>1.85</td>
<td>3,663</td>
<td>3.58</td>
</tr>
<tr>
<td>50-46</td>
<td>2,142</td>
<td>2.00</td>
<td>2,289</td>
<td>2.24</td>
</tr>
<tr>
<td>45-41</td>
<td>1,969</td>
<td>1.83</td>
<td>1,904</td>
<td>1.86</td>
</tr>
<tr>
<td>40-36</td>
<td>1,793</td>
<td>1.67</td>
<td>1,383</td>
<td>1.35</td>
</tr>
<tr>
<td>35-31</td>
<td>1,279</td>
<td>1.19</td>
<td>796</td>
<td>0.78</td>
</tr>
<tr>
<td>30-26</td>
<td>1,244</td>
<td>1.16</td>
<td>448</td>
<td>0.44</td>
</tr>
<tr>
<td>25-21</td>
<td>895</td>
<td>0.83</td>
<td>291</td>
<td>0.28</td>
</tr>
<tr>
<td>20-16</td>
<td>498</td>
<td>0.46</td>
<td>184</td>
<td>0.18</td>
</tr>
<tr>
<td>15-11</td>
<td>38</td>
<td>0.04</td>
<td>15</td>
<td>0.01</td>
</tr>
<tr>
<td>10-6</td>
<td>18</td>
<td>0.02</td>
<td>6</td>
<td>0.01</td>
</tr>
<tr>
<td>5-1</td>
<td>10</td>
<td>0.01</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>N</td>
<td>107,312</td>
<td></td>
<td>102,233</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>8,660,878</td>
<td></td>
<td>7,702,602</td>
<td></td>
</tr>
<tr>
<td>X^2</td>
<td>733,929,208</td>
<td></td>
<td>603,372,540</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>80.71</td>
<td></td>
<td>75.34</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>18.03</td>
<td></td>
<td>15.01</td>
<td></td>
</tr>
<tr>
<td>ACTIVITIES</td>
<td>TOTAL NO. OF ENDORSERS</td>
<td>N &lt; 7</td>
<td>%</td>
<td>N &lt; 10</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>OCAMA</td>
<td>745</td>
<td>163</td>
<td>21.88</td>
<td>207</td>
</tr>
<tr>
<td>OOAMA</td>
<td>578</td>
<td>105</td>
<td>18.17</td>
<td>148</td>
</tr>
<tr>
<td>SAAMA</td>
<td>902</td>
<td>243</td>
<td>26.94</td>
<td>299</td>
</tr>
<tr>
<td>SMAMA</td>
<td>586</td>
<td>127</td>
<td>21.67</td>
<td>166</td>
</tr>
<tr>
<td>WRAMA</td>
<td>620</td>
<td>126</td>
<td>20.32</td>
<td>184</td>
</tr>
<tr>
<td>2750th ABW</td>
<td>617</td>
<td>268</td>
<td>43.44</td>
<td>332</td>
</tr>
<tr>
<td>TOTAL</td>
<td>4,048</td>
<td>1,032</td>
<td>25.49</td>
<td>1,336</td>
</tr>
</tbody>
</table>
## APPENDIX 13

### NUMBER OF EMPLOYEES CORRECTED BASE WIDE BY ACTIVITY

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11,815</td>
<td>45</td>
<td>.38</td>
<td>A</td>
<td>11,815</td>
<td>102</td>
<td>.86</td>
</tr>
<tr>
<td>B</td>
<td>6,730</td>
<td>141</td>
<td>2.10</td>
<td>B</td>
<td>6,730</td>
<td>256</td>
<td>3.80</td>
</tr>
<tr>
<td>C</td>
<td>1,997</td>
<td>107</td>
<td>5.36</td>
<td>C</td>
<td>1,997</td>
<td>209</td>
<td>10.47</td>
</tr>
<tr>
<td>D</td>
<td>460</td>
<td>5</td>
<td>1.09</td>
<td>D</td>
<td>460</td>
<td>13</td>
<td>2.83</td>
</tr>
<tr>
<td>E</td>
<td>1,344</td>
<td>113</td>
<td>8.41</td>
<td>E</td>
<td>1,344</td>
<td>180</td>
<td>13.39</td>
</tr>
<tr>
<td>TOTAL</td>
<td>22,346</td>
<td>411</td>
<td>1.84</td>
<td>TOTAL</td>
<td>22,346</td>
<td>760</td>
<td>3.40</td>
</tr>
</tbody>
</table>

### OCAMA

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7,244</td>
<td>21</td>
<td>.29</td>
<td>A</td>
<td>7,244</td>
<td>102</td>
<td>1.41</td>
</tr>
<tr>
<td>B</td>
<td>5,527</td>
<td>106</td>
<td>1.92</td>
<td>B</td>
<td>5,527</td>
<td>206</td>
<td>3.73</td>
</tr>
<tr>
<td>C</td>
<td>1,584</td>
<td>88</td>
<td>5.56</td>
<td>C</td>
<td>1,584</td>
<td>190</td>
<td>12.00</td>
</tr>
<tr>
<td>D</td>
<td>483</td>
<td>13</td>
<td>2.69</td>
<td>D</td>
<td>483</td>
<td>14</td>
<td>2.90</td>
</tr>
<tr>
<td>E</td>
<td>1,041</td>
<td>81</td>
<td>7.78</td>
<td>E</td>
<td>1,041</td>
<td>134</td>
<td>12.87</td>
</tr>
<tr>
<td>TOTAL</td>
<td>15,679</td>
<td>309</td>
<td>1.95</td>
<td>TOTAL</td>
<td>15,679</td>
<td>646</td>
<td>4.07</td>
</tr>
</tbody>
</table>

### OAAMA

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11,267</td>
<td>132</td>
<td>1.17</td>
<td>A</td>
<td>11,267</td>
<td>189</td>
<td>1.68</td>
</tr>
<tr>
<td>B</td>
<td>7,295</td>
<td>171</td>
<td>2.34</td>
<td>B</td>
<td>7,295</td>
<td>371</td>
<td>5.09</td>
</tr>
<tr>
<td>C</td>
<td>2,024</td>
<td>214</td>
<td>10.57</td>
<td>C</td>
<td>2,024</td>
<td>323</td>
<td>15.96</td>
</tr>
<tr>
<td>D</td>
<td>406</td>
<td>21</td>
<td>5.17</td>
<td>D</td>
<td>406</td>
<td>28</td>
<td>6.90</td>
</tr>
<tr>
<td>E</td>
<td>1,343</td>
<td>99</td>
<td>7.37</td>
<td>E</td>
<td>1,343</td>
<td>176</td>
<td>13.10</td>
</tr>
<tr>
<td>TOTAL</td>
<td>22,335</td>
<td>637</td>
<td>2.85</td>
<td>TOTAL</td>
<td>22,335</td>
<td>1,087</td>
<td>4.87</td>
</tr>
</tbody>
</table>

### SAAMA

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7,288</td>
<td>48</td>
<td>.66</td>
<td>A</td>
<td>7,288</td>
<td>99</td>
<td>1.36</td>
</tr>
<tr>
<td>B</td>
<td>5,199</td>
<td>73</td>
<td>1.40</td>
<td>B</td>
<td>5,199</td>
<td>196</td>
<td>3.77</td>
</tr>
<tr>
<td>C</td>
<td>1,372</td>
<td>123</td>
<td>8.97</td>
<td>C</td>
<td>1,372</td>
<td>198</td>
<td>14.43</td>
</tr>
<tr>
<td>D</td>
<td>312</td>
<td>6</td>
<td>1.87</td>
<td>D</td>
<td>321</td>
<td>13</td>
<td>4.05</td>
</tr>
<tr>
<td>E</td>
<td>914</td>
<td>94</td>
<td>10.28</td>
<td>E</td>
<td>914</td>
<td>152</td>
<td>16.63</td>
</tr>
<tr>
<td>TOTAL</td>
<td>15,094</td>
<td>344</td>
<td>2.28</td>
<td>TOTAL</td>
<td>15,094</td>
<td>658</td>
<td>4.36</td>
</tr>
</tbody>
</table>

### SMAMA

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6,608</td>
<td>17</td>
<td>.26</td>
<td>A</td>
<td>6,608</td>
<td>103</td>
<td>1.56</td>
</tr>
<tr>
<td>B</td>
<td>5,157</td>
<td>148</td>
<td>2.87</td>
<td>B</td>
<td>5,157</td>
<td>326</td>
<td>6.32</td>
</tr>
<tr>
<td>C</td>
<td>1,450</td>
<td>89</td>
<td>6.14</td>
<td>C</td>
<td>1,450</td>
<td>173</td>
<td>11.93</td>
</tr>
<tr>
<td>D</td>
<td>377</td>
<td>5</td>
<td>1.33</td>
<td>D</td>
<td>377</td>
<td>29</td>
<td>7.69</td>
</tr>
<tr>
<td>E</td>
<td>1,193</td>
<td>123</td>
<td>10.31</td>
<td>E</td>
<td>1,193</td>
<td>211</td>
<td>17.69</td>
</tr>
<tr>
<td>TOTAL</td>
<td>14,785</td>
<td>382</td>
<td>2.58</td>
<td>TOTAL</td>
<td>14,785</td>
<td>842</td>
<td>5.70</td>
</tr>
</tbody>
</table>

### WRAMA

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2,146</td>
<td>79</td>
<td>3.68</td>
<td>A</td>
<td>2,146</td>
<td>205</td>
<td>9.55</td>
</tr>
<tr>
<td>B</td>
<td>3,865</td>
<td>262</td>
<td>6.78</td>
<td>B</td>
<td>3,865</td>
<td>499</td>
<td>12.91</td>
</tr>
<tr>
<td>C</td>
<td>764</td>
<td>96</td>
<td>12.57</td>
<td>C</td>
<td>764</td>
<td>136</td>
<td>17.80</td>
</tr>
<tr>
<td>D</td>
<td>248</td>
<td>35</td>
<td>14.11</td>
<td>D</td>
<td>248</td>
<td>50</td>
<td>20.16</td>
</tr>
<tr>
<td>E</td>
<td>1,194</td>
<td>295</td>
<td>24.71</td>
<td>E</td>
<td>1,194</td>
<td>389</td>
<td>32.56</td>
</tr>
<tr>
<td>TOTAL</td>
<td>8,217</td>
<td>767</td>
<td>9.33</td>
<td>TOTAL</td>
<td>8,217</td>
<td>1,279</td>
<td>15.57</td>
</tr>
</tbody>
</table>

### 2750th ABW

<table>
<thead>
<tr>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
<th>Form</th>
<th>No. Appraised</th>
<th>No. Corrected Base-Wide</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46,368</td>
<td>342</td>
<td>.74</td>
<td>A</td>
<td>46,368</td>
<td>800</td>
<td>1.73</td>
</tr>
<tr>
<td>B</td>
<td>33,773</td>
<td>901</td>
<td>2.67</td>
<td>B</td>
<td>33,773</td>
<td>1,854</td>
<td>5.49</td>
</tr>
<tr>
<td>C</td>
<td>9,191</td>
<td>717</td>
<td>7.80</td>
<td>C</td>
<td>9,191</td>
<td>1,229</td>
<td>13.37</td>
</tr>
<tr>
<td>D</td>
<td>2,295</td>
<td>85</td>
<td>3.70</td>
<td>D</td>
<td>2,295</td>
<td>147</td>
<td>6.41</td>
</tr>
<tr>
<td>E</td>
<td>7,029</td>
<td>805</td>
<td>11.45</td>
<td>E</td>
<td>7,029</td>
<td>1,242</td>
<td>17.67</td>
</tr>
<tr>
<td>TOTAL</td>
<td>98,656</td>
<td>2,850</td>
<td>2.89</td>
<td>TOTAL</td>
<td>98,656</td>
<td>5,272</td>
<td>5.34</td>
</tr>
</tbody>
</table>

Source: Appraisal Distribution Report E246.74P1 (1971)


Bittner, R. H. Developing an employee merit rating procedure. Personnel Psychology, 1948, 1, 403-432.


Driver, R. S. A case history of merit rating. Personnel, 1940, 16, 137-162.


Ferguson, L. W. The effect upon appraisal scores of individual differences in the ability of supervisors to appraise subordinates. Personnel Psychology, 1949, 2, 377-382.


Ghiselli, E. E. Dimensional problems of criteria. *Journal of Applied Psychology*, 1956, 40, 1-4. (a)


Horan, C. B. Multidimensional scaling: combining observations when individuals have different perceptual structures. Psychometrika, 1969, 34, 139-165.


Kipnis, Dorothy M. Interactions between members of bomber crews as a determinant of sociometric choice. *Human Relations*, 1957, 10, 263-270.

Kirchner, W. K., & Reisberg, D. J. Differences between better and less effective supervisors in appraisal of subordinates. *Personnel Psychology*, 1962, 15, 295-302.


Méchin, D. Comment notrez-vous votre personnel? (How do you rate your personnel?) Travail et M é thodes, 1956, 103, 30-33. (Abstract)


Rugg, H. Is the rating of human character practicable? *Journal of Educational Psychology*, 1922, 13, 81-93.


Sisson, D. G. Forced-choice—the new Army rating. Personnel Psychology, 1948, 1, 365-381.


Taylor, E. K., & Hastman, R. Relation of format and administration to the characteristics of graphic rating scales. Personnel Psychology, 1956, 9, 181-206.

Taylor, E. K., & Manson, Grace E. Supervised ratings—making graphic scales work. Personnel, 1951, 27, 504-516.


Toops, H. A. A research utopia in industrial psychology. Personnel Psychology, 1959, 12, 182-227.


Wherry, R. J. An evaluative and diagnostic forced-choice rating scale for servicemen. *Personnel Psychology*, 1959, 12, 227-236.


Whitla, D. K., & Tirrell, J. E. The validity of ratings of several levels of supervisors. *Personnel Psychology*, 1964, 6, 461-466.


