

Lindenwood University

Digital Commons@Lindenwood University

Theses

Theses & Dissertations

10-1979

Problems Associated with the Use of Performance Appraisals in Promotion Decisions in the Federal Personnel System

Donald J. Detzel

Follow this and additional works at: <https://digitalcommons.lindenwood.edu/theses>



Part of the Business Commons

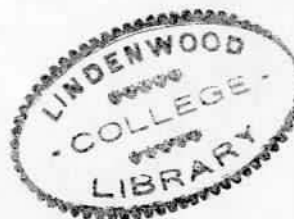
PROBLEMS ASSOCIATED WITH THE USE OF PERFORMANCE
APPRAISALS IN PROMOTION DECISIONS IN THE FEDERAL
PERSONNEL SYSTEM

BY
DONALD J. DETZEL

FACULTY SPONSOR: ALAN BREITLER, PH. D.
FACULTY ADMINISTRATOR: CRAIG EISENDRATH, PH. D.

OCTOBER 1979

SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS, LINDENWOOD COLLEGES



Thesis
D486p
1979

TABLE OF CONTENTS

I. INTRODUCTION 1
Background
Purpose and Significance of the Study

II. REVIEW OF LAWS AND REGULATIONS 6
Federal Civil Service Systems
Equal Employment Opportunity Applications

III. REVIEW OF RESEARCH LITERATURE. 16
Area 1: Reliability and Validity of
Performance Ratings
Area 2: Behaviorally Based Ratings
Area 3: Rater Errors and Rater Training
Area 4: Predictive Validity
Area 5: Racial and Sexual Bias

IV. STUDY OF RACIAL, SEXUAL AND AGE BIAS IN
PERFORMANCE RATINGS. 26
Methodology
Discussion

V. CONCLUSION 36

APPENDIX A Documentation for Quality Increase 42

APPENDIX B Excerpt form Removal Action (ASCS) 44

APPENDIX C FORM AD 434, Performance Rating. 48

APPENDIX D Population Distribution -- Performance
Appraisal Study -- ASCA (Washington
Office) 1978 50

APPENDIX E Model of ASCS Use of Performance
Appraisals in Promotion Decisions. 55

BIBLIOGRAPHY 57

INTRODUCTION

Background: The judgment of the performance of an individual has probably existed from the beginning of man. Such judgments have naturally occurred out of a perceived necessity of showing gratitude for services rendered, motivating toward a common objective, promoting to greater authority and responsibility, punishing for failure to provide the value expected, or if nothing else, just human tendency to differentiate. The necessity of such judgments seems to be greater the more organized a group becomes and the more complex the organization. A typical example is the military hierarchy which was one of the first organizations to perceive the necessity for judging the performance of the individual and establishing formal recognition of performance. Thus, performance appraisals are an inherent part of mankind's organizational behavior in that they provide a basis for many decisions affecting the employees of an organization.

Similarly, within the Executive Branch of the Federal Government the performance appraisal has provided the basis or partial basis for many types of personnel actions. While the documentation of such appraisals has taken many forms over the years, today the performance appraisal in most Federal Agencies is usually documented by a performance rating form using multiple traits and rating scales. The traits are usually vague adjectives of behavior, and/or sometimes of attitudes, which can apply

to hundreds and literally thousands of jobs, grades, and individuals. The rating scales are normally unanchored. For the most part, the ratings are periodically required (usually on a quarterly, semi-annual, or annual basis) and are invariably tied by policy into a formal system of rewards and punishments.

The system of rewards and punishments may take many forms. Rewards are manifested through salary increases, promotions, monetary and non-monetary awards, and influence and status within the organization. Punishments may take the form of demotions, salary decreases, loss of influence and status, disciplinary action or removal. Of these two types of personnel actions, (i.e., rewards and punishments) the trend is mostly toward positive or reward oriented actions rather than negative or punishment oriented actions. Performance evaluations provide the basis for this system.

An interesting phenomenon within this framework of rewards and punishments is that while the rating form, with its traits and scales, is tied into the system as a matter of policy, it is not procedurally used as the basis for justifying the vast majority of the rewards and punishments. An exception to this is in the area of competitive promotions where the Office of Personnel Management requires, as one of the criteria, that a comparison be made of the evaluations of candidates and used as one factor in ranking the candidate for promotion. The comparison is usually achieved through the use of a rating scale. Otherwise, the justifications for the rewards or punishments are

separate and discrete narrative documentations or oral declarations which are unrelated to the rating scales and which are primarily examples of critical incidents depicting an individual's behavior. (See Appendices A and B).

Purpose and significance of the study: It is the purpose of this study to concentrate on problems affecting the use of the performance appraisal in making promotion decisions in the Federal Personnel System. This aspect of performance evaluation is significant both to the organization and to the individual in that the selection from among the best qualified employees for specified vacancies is both a declared and bona fide objective of Government organizations. At the same time, a valid selection process for promotion is a paramount criterion in maintaining employee satisfaction and the integrity of the promotion selection system. Since performance ratings are the most frequently used means of obtaining information on how well employees perform, it is important to understand the limitations in using these ratings under specific organizational constraints.

It is my contention that psychologists should be trying to measure the validity of using performance appraisals in predicting success or failure in higher level positions. While peer ratings have been tested as to their predictive validity, (Hollander 1954, Korman 1968, Miner 1968, Lindzey and Byrne 1969) the testing of supervisory ratings in this regard has been overlooked. Furthermore, investigation of supervisory ratings have pointed out some serious difficulties which could influence the predictive validity of

the ratings. Lawler (1967), indicated that there existed severe difficulties in obtaining accurate ratings of performance because of rater errors and the general lack of discriminant validity in a multitrait-multirater analysis. Borman (1978), classified the problems with performance ratings into four categories: 1) the lack of opportunity of the raters to observe the behavior being rated, 2) the raters limited familiarity with the types of rater errors, 3) rating formats which do not allow raters to translate the behavior observed to a specific level on the rating scale and, 4) organization constraints on raters to provide ratings different from that which the rater believes is more correct. Such difficulties with supervisory ratings of subordinates casts doubt on the benefits claimed by management in using performance appraisals to predict success or failure in higher level positions. Federal managers, personnel specialists, and psychologists must address themselves to the questions raised by these limitations.

For example, we need to ask what organizational constraints are placed on raters by the laws and regulations governing performance appraisals? Do these constraints impact on our use of performance appraisals in promotions and if so how do they impact on the rating process? We need to understand the difficulties that rating scale and format design causes raters and how these difficulties can be alleviated. And finally we need to ask 1) can rater error be reduced? and 2) what effect does reduction of error have on the use of peer distinctions as a valid way of predicting success or failure in a higher level

position?

It is the intent of the present study to consider three aspects of performance appraisals with regard to their use in promotion selections. The first part of the study will concentrate on determining what organizational constraints Federal Civil Service laws and regulations (including the recently passed Civil Service Reform Act), and the Uniform Guidelines on Employee Selection Procedures have placed on the use of performance appraisals in selecting employees for promotion. The second part of this study will be to review and critically evaluate the current research literature to the extent it may be utilized in developing more valid performance appraisals for use in selecting employees for higher grade positions. In evaluating the literature the effects of training on performance appraisals will be considered. In addition time and cost factors will be considered as possible organizational constraints on management's use of the various performance appraisal research methods. The third part of this study will be to quantitatively explore the existence of contrast error (race, sex and age) in supervisory ratings used in a federal government organization. For this paper I am assuming that problems such as organizational constraints, rater error and rating scale design preclude valid use of performance appraisals in selecting employees for promotion.

REVIEW OF LAWS AND REGULATIONS

Federal Civil Service Systems: The first efforts to establish a performance appraisal system in the federal government appeared during the administration of President Tyler (1841-45) and continued to various forms under different administrations. However, it was not until 1923 when Congress enacted the Classification Act that legislation governing performance appraisal systems was also enacted. These performance ratings were, according to the law, to be used for personnel decisions affecting (1) within grade salary increases (2) retention in grade without advancement or reduction in salary (3) reduction in pay within the grade and (4) immediate demotion or dismissal. A graphic rating scale was added in 1924 to provide a quantitative measure for the ratings.

The graphic rating scale is a check list form listing various traits of performance or personality and providing a variety of categories from which the rater could choose. The rater chooses those categories (such as outstanding, excellent, average, below average, unacceptable) which best describe the rater's perception of the trait. A numerical value is provided for each category so that a quantitative measure is available.

The graphic rating system was quite unpopular. For one reason, the supervisor had no control over the final rating as the final rating was determined by higher echelon officials and

a board of review. The boards were allowed to adjust the ratings to conform to a predetermined pattern of distribution. For example, a rating board would determine the number of employees that would be allowed to be rated in each category of outstanding, excellent, average, below average, and unacceptable. This predetermined distribution of employees in each category would limit the kinds of ratings the supervisors could give to their employees and would not provide an accurate basis for rating. The supervisor found it difficult to explain the rating to employees or to work with them in improving their performance and their rating. Due to its unpopularity, the Civil Service Commission and the Federal Personnel Council revised the system in 1936 under the Uniform Efficiency Rating System.

The Uniform Efficiency Rating System was still basically the graphic rating system but with different organizational factors. The system provided for five ratings of (1) excellent, (2) very good, (3) good, (4) fair, and (5) unsatisfactory. The new system also had quantitative measurements with numerical values assigned to various performance factors. The organizational factors did differ, however, in that adjustments to obtain a predetermined pattern of distribution were not allowed. Also, a formal appeal system was introduced providing for each agency to establish an independent review board. The members of this independent review board were comprised of representatives from the agency, the Civil Service Commission and the employees. The Uniform Efficiency Rating System remained in effect until 1950 and was the basis for personnel decisions affecting

promotions, salary increases, reassignments, reductions-in-force, and removal.

As in the previous changes of each preceding performance appraisal system, dissatisfaction played a key note in the decision to change. Thus, in 1950 the Civil Service Commission reported to Congress its study of the Uniform Efficiency Rating System along with recommended changes. The existing rating system, reported the study, was too static, too little related to employee development, too closely tied to personnel decisions and too costly for conducting appeals. The report further stated its recommendations to:

- (1) let the agencies decide whether to have a rating plan,
- (2) eliminate the adjective summary rating,
- (3) restrict rating appeals to an intra-agency administrative review, and
- (4) determine personnel decisions on other factors in addition to the performance rating.

Hearings held by the Congress disclosed other dissatisfactions. Various employee organizations responded to the proposed legislations and maintained that employee dissatisfaction was primarily caused by (1) the variety of ratings obtainable, (2) the tendency toward personal favoritism and (3) the use of rating elements which were not valid for the jobs performed. Once again dissatisfaction with the existing performance appraisal system had led to cries for change and improvement. On the basis of the Civil Service Commission report and subsequent hearings,

Congress enacted the Performance Rating Act of 1950.

In 1978, riding on a cresting wave of public opinion and strongly supported by the broadcasting and newspaper media, by business and industry, and by other special interest groups, President Carter proposed to Congress a massive Civil Service Reform bill. Included in the reform proposals was legislation to revise the Performance Rating Act of 1950. It was proposed to:

- (1) authorize each agency to develop one or more performance appraisal systems,
- (2) provide for the establishment of performance standards and critical elements based on the requirements of the employee's position,
- (3) encourage participation of employees in establishing performance standards,
- (4) provide for an extensive method of documentation,
- (5) require the communication of such documentation with employee affected by the performance appraisal system, and
- (6) establish an appeal rights to a Merit Systems Protection Board.

Congress enacted the proposed legislature under the Civil Service Reform Act of 1978 to be effective January 11, 1979. Recent interim regulations have been published by the Office of Personnel Management (formerly the Civil Service Commission) to implement the establishment of the performance appraisal systems as required by Subchapter I of Chapter 43 of the Act.

The regulations provide for the use of the results of the performance appraisal systems as a basis for personnel decisions involving employee development, rewards, reassignments, promotions, demotions, retaining in grade and removals. Provisions have been made for evaluation of agency performance appraisal systems by the Office of Personnel Management and the Comptroller General.

Equal Employment Opportunity Applications:

Other legislation and regulations which may have significant impact on the future design of performance appraisal systems and the use of their results as a basis for personnel decisions is found in the area of equal employment opportunity.

When Congress passed the Civil Rights Act of 1964 one of the problems confronting them at the time was that the use of tests as an employee selection tool had the effect of denying employment to large groups of minorities. While Congress was reluctant to abolish the use of tests as an employee selection tool, they did wish to eliminate any possibility that tests or any other selection techniques would be used to discriminate against the protected groups. The definition of protected groups periodically changes and currently includes women as well as minorities such as blacks and hispanics. In an attempt to compromise on the issue of using various selection processes, such as tests and performance appraisals, Congress adopted language in Title VII of the Civil Rights Act of 1964 which allowed the use of tests provided the tests were not used in discriminating against the protected groups. In addition, the tests had to be professionally developed. The language of the Act was such that disagreements

arose between employers and the Equal Employment Opportunity Commission (EEOC) as to the interpretation of Title VII with regard to the use of employee selection techniques.

The Government's position was that even though the tests were professionally developed and were not intentionally used to discriminate, any adverse impact on employment opportunities of protected groups could be considered discrimination. To clarify the situation, the EEOC in 1966 and the Department of Labor in 1968 issued guidelines setting forth their position and requiring private business and industrial firms to adhere to those guidelines. The guidelines required employers to justify the selection practices by demonstrating that the selection practice fairly measured or predicted actual performance on the job. While the early emphasis of the guidelines tended to concentrate on preemployment selection tests, later efforts were made by both EEOC and the Department of Labor to strengthen the guidelines for all selection procedures including the selection of employees for promotion through the use of performance appraisal systems. The conflicting interpretations between Government and private enterprise resulted in a number of court cases which supported the Government's position.

Probably the single most significant case regarding selection procedures was the Griggs vs. Duke Power Co., 401 U.S. 424 (1971). In Griggs vs. Duke Power the Supreme Court announced that employer selection practices which had an adverse impact on minorities and which could not be justified by "business necessity" constituted discrimination under Title VII. While the conflict

continued on a lesser scale in other court actions, the Government for all practical purposes had won its argument. Subsequent actions by the courts and Congress solidified the Government's position and refined the principles set forth in the guidelines. The Griggs vs. Duke Power decision became the turning point in the Government's efforts to regulate employee selection procedures by providing legal support for the principles that had been emphasized in the regulations.

The effect of the court actions and the EEOC and Labor Department regulations were not immediately applied to Federal selection procedures. There were, however, several occurrences in 1978 which brought the Federal selection procedures under the EEOC regulations. With the passage of the Civil Service Reform Act of 1978, the EEOC assumed jurisdiction over Federal employee equal employment opportunity matters. Also affecting the Federal personnel operations was the decision by the four Federal agencies overseeing equal employment opportunity matters for the Government to agree on a single guideline for all four agencies (the Equal Employment Opportunity Commission, the Department of Labor, the Office of Personnel Management, and the Office of Revenue Sharing, Treasury Department). This agreed upon guideline was published on August 25, 1978. Thus these two occurrences brought the full weight of the Griggs vs. Duke Power Co. decision and subsequent court actions and EEOC regulations to impact on federal employee selection procedures.

There are two basic principles included in the guidelines and in the Griggs vs. Duke Power decision in determining

discrimination. The first principle is that there must be an adverse impact on protected groups. The second principle is that the selection practice such as performance appraisals is not justified by "business necessity." Both of these situations must be present for discrimination to exist legally.

Adverse impact is defined in the Federal Guidelines on Employee Selection Procedures (1978) as a "substantially different rate of selection in hiring, promotion or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group." The substantially different rate is an arbitrary figure adopted by the Government to provide some point at which the attention of enforcing agencies would become necessary. This rule of thumb is generally known as the "4/5ths" or "80 percent" rule. Under this rule any selection rate for protected groups which is less than four-fifths of the selection rate for the group with the highest selection rate (principally white males) is considered substantially different. While the Government reserves the right to pursue any case of discrimination, from a practical viewpoint, the Government has declared its intention to concentrate on those organizations which are identified by the 80 percent rule of thumb.

If evidence of adverse impact can be shown then the employer must show that there is a business necessity, the user of the selection process must show that there is a valid relationship to the job; that is, the procedure is necessary for the safe or efficient operation of the business. The Government guidelines thereby require professionally validated studies be conducted

and the relationship between the selection procedure and the job duties and responsibilities be determined in accordance with professional standards. The Government recognizes the same three validity strategies recognized by the American Psychological Association: 1) criterion-related validity, 2) content validity, and 3) construct validity. The three validity strategies are defined in the Federal Register, Vol. 44, No. 43 in the following manner:

"Criterion-related validity -- a statistical demonstration of a relationship between scores on a selection procedure and job performance of a sample of workers.

Content validity -- a demonstration that the content of a selection procedure is representative of important aspects of performance on the job.

Construct validity -- a demonstration that (a) a selection procedure measures a construct (something believed to be an underlying human trait or characteristic, such as honesty), and (b) the construct is important for successful job performance."

For each strategy, there is the necessity to conduct a job analysis to determine measures of work behavior(s) or performance that are relevant to the job or jobs under study. There is also a need to analysis, so as to identify critical or important job elements, work behavior, and work outcomes or constructs necessary to perform the job elements. Work behavior and outcomes must be observable in order to be identifiable and measurable.

In summary, the intent of the regulations is to provide an acceptable guide for determining adverse impact on protected groups in both the private and public sectors. If the adverse impact is caused by selection practices then the employing

organization has the responsibility of eliminating the adverse impact or validating the use of the selection practice. The guidelines have been designed to be consistent with the accepted professional standards in validating selection practices and are patterned after the standards described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, The American Educational Research Association, and the National Council on Measurement in Education (American Psychological Association; Washington, D.C. 1974).

The full impact of the Civil Service Reform Act of 1978 and the Uniform Guidelines for Employee Selection Procedures of 1978 is yet to be felt in Government personnel decisions. One outcome will be heavier reliance on the use of psychologists in advising managers and personnel administrators on the development of performance appraisal systems and on their use in making personnel decisions. With this in mind an understanding of the psychological research on performance appraisals becomes important if we are to comply with constraints placed upon the federal personnel process.

REVIEW OF RESEARCH LITERATURE

In my review and critique of the research literature I attempted to concentrate my efforts on five areas of the performance evaluation studies which appeared to have the most relevance to peer comparison by performance appraisals. These areas were (1) studies regarding the reliability and validity of performance ratings, (2) behavior based ratings, (3) rater errors and the impact of rater training on rater errors, (4) prediction validity of performance ratings, and (5) studies of racial and sexual bias.

Area #1: Reliability and Validity of Performance Ratings

The Borman (1975) and Borman (1978) reliability and validity studies were used for my first area of concentration. In the first study, Borman investigated the impact of short training sessions designed to reduce halo error on the reliability and validity of performance evaluation ratings.

To offset the common problem of not having external criteria against which to evaluate performance ratings, Borman prepared six vignettes about individual hypothetical first-line supervisors. Each vignette contained known performance scores or "true scores" for six performance dimensions against which the ratings given by the raters could be compared. Both before and after a short training session the raters were randomly assigned rates

for evaluation. The result of this test indicated that short training sessions introducing halo effect and then discouraging the committing of the halo error may be successful in reducing halo error. However, the reduction of halo error could reduce the interrater reliability. The results also indicated that the raters were somewhat more accurate in their rating of individual rates after the training than they had been prior to their training.

Borman's later study examined the upper limits of reliability and validity of performance ratings. Borman assumed, for the purposes of the study that raters were capable of making more valid performance evaluations. The questions to be answered were (1) how much improvement in the validity of performance evaluations could be made? and (2) was there a leveling off or ceiling due to some kind of inherent limitations in the precision with which raters could judge the performance of others? The laboratory experiment attempted to create an ideal condition with regard to four problem areas identified by Borman. These four problem areas dealt with (1) the rater's opportunity to observe relevant job related behavior, (2) the rater's knowledge of rater errors, (3) the rating format, and (4) organizational constraints. Results indicated that there may be a "ceiling" for interrater agreement as well as limitations on how much convergent and discriminant validity of ratings psychologists could obtain. Borman also found that under these almost ideal conditions most rating errors were reduced sharply and that discriminant validity was considerably higher than that typically found in applied

rating studies.

Area #2: Behaviorally Based Ratings

Behaviorally anchored ratings have their beginning in such methodologies as critical incidents Flanagan (1949 and 1954) and Smith and Kendall's (1963) use of expectations based on observed behavior.

Smith and Kendall (1963) considered the use of critical incidents as "extremely desirable" in anchoring rating scales but determined that variations in observation opportunities by raters would preclude the use of critical incidents as a basis for rating. Smith and Kendall's (1963) behaviorally based scaling methodology was designed to improve rater comparability. Smith and Kendall felt that the interpretation of ratings must not vary too widely between raters or from one occasion to another in the evaluation or in the job dimension (trait, job requirements, job behavior, etc.). In order to utilize ratings with any confidence there must be a consistent interpretation or agreement between raters as to what is meant by a particular job dimension. Furthermore this agreement must be maintained by each rater during the different rating periods or rating occurrences. Their scaling methodology appears to have good potential for overcoming or reducing many of the rater errors encountered in job performance rating systems.

Smith and Kendall believed that most rating errors were not due to deliberate faking; therefore, they did not attempt to design a rating scale which would be fake proof. Their firm belief was that a rater who deliberately wished to distort the

scores on a rating scale could readily do so.

Four groups of head nurses in medical surgical nursing from different hospitals participated in conferences to identify examples or expectations with which to anchor such job characteristics or dimensions as knowledge and judgment, conscientiousness, organizational ability and skill in human relationships. In addition two other head nurse groups were used to gather data by mail. These two groups were from the same hospitals as two of the four groups participating in conference. The examples chosen to anchor the job dimensions were based on inferences or predictions made from observations of behavior. The expectations agreed upon by raters thereby served as a mutual frame of reference. Smith and Kendall found, moreover, that this mutual frame of reference provided the opportunity to supplement predictions with recordings of actually observed behavior upon which the predictions are based.

The procedure seemed promising, with excellent discrimination and high scale reliability. An advantage of this method was the extensive participation of the raters who obtained greater understanding of the job dimensions and behavioral anchors.

Other investigators have studied this approach with varying results. Campbell, et al. (1973) developed an experiment using a modification of Smith and Kendall's scaled expectations to assess department manager performance in terms of concrete job functions. The purpose of the study was to determine if such scales would reduce leniency and halo errors more than rating methods that were not behaviorally anchored and to determine

if they would exhibit significant convergent and discriminate validity. Their experiment supports the behaviorally anchored methodology approach. They concluded there was less leniency and halo error.

Arvey and Hoyle, (1974) used a scalogram analysis in developing behaviorally based rating scales for systems analysts and programmer/analysts. Their study showed that such developed scales exhibit reasonably good convergent and discriminate validities.

Borman and Dunnette (1975) conducted an empirical study on Navy officer raters using three methods of rating: (a) behaviorally anchored scales, (b) identical scales without behavioral anchors, and (c) a series of scales using trait oriented dimensions, also not behaviorally anchored. They found that while the behaviorally anchored rating scale was superior in all counts the superiority was a modest one. They concluded that the greatest value is in the wealth of information about job requirements and job performance which is yielded by the technique.

Researchers tend to agree on several issues with regard to the behaviorally anchored scales. First, they are time consuming and expensive to develop. Second, they do yield useful side benefits such as the obtaining of a wealth of information on job requirements and job performance. Third, they provide a definite basis for training. However, because of the time and cost of developing such scales it is questionable whether these scales should be used solely for performance ratings. A wide range of uses would be necessary to provide an adequate basis for

the amount of time and effort required. The potential for providing such a basis is available if the wealth of information on job requirements and job performance obtained during the development of behaviorally anchored scales is used for applicant selection and employee development as well as performance appraisal. Furthermore, while the concept seems to have potential for better differentiation between persons rated, the studies to date are equivocal on this point.

Area #3: Rater Errors and Rater Training

Rater training to reduce rater errors holds the most promising avenue for improving the quality of evaluations. Rater errors occur frequently in performance appraisals and selection interviews, namely in the form of contrast errors, halo effect, similarity, first impressions and leniency errors (See Appendix A). Borman and Dunette, (1975), concluded that the one line of research that could lead to improvements in evaluations is to train the raters in conducting error-free evaluations and in greater systematic work-related behavior observations.

Wexley, et al. (1973) were among those who recognized the possibilities of reducing rater errors through training. Their laboratory efforts were aimed at the elimination of contrast errors by employment interviewers through the use of various training approaches. Four experiments were conducted with raters: (1) warning about such errors, (2) providing interviewers with absolute standards, (3) combining the first two experiments, and (4) providing a 2-hour workshop on contrast errors. While the first three experiments showed no improvement, the workshop on

contrast errors had the effect of substantially reducing contrast errors.

Borman (1975) attempted to determine the effect of instructions to avoid halo error on the reliability and validity of performance evaluations. By developing a laboratory experiment Borman was able to identify "true scores" which were used to determine validity. The training consisted of short sessions (5 to 6 minutes) used to introduce the concept of halo error to the raters. The results of the study suggests that such short training sessions may have the effect of reducing halo error without affecting the validity of the ratings. Interrater reliability, however, was somewhat lower.

Latham, Wexley, and Pursell (1975) laboratory testing the effects of the workshop (9 hours) versus group discussion (6 hours) approach on reducing contrast, halo, similarity and first impressions found the workshop (while more costly) a slightly more effective way of reducing such rater errors. While the workshop was the most effective way of reducing rater errors, group discussions should not be ignored, particularly where a choice between the two methods is based on cost effectiveness. It would appear from these studies that training is an effective way of improving the usefulness of performance evaluations.

Area #4: Predictive Validity

Despite the enormous amount of research on performance appraisals, little work has been done on the predictive validity of the ratings. In addition, no studies appear to have been performed which investigated the validity of the common practice

of using the performance appraisal to distinguish between candidates for filling a position.

If managers and personnel specialist are to continue to use performance appraisals as a selection technique we need to ask the question, "Is there merit in considering that most employees who are rated high in a performance appraisal system will do well in a similar job at a higher grade level in the same organization or in a different organization at the same or higher grade?" Conversely, we need to ask a similar question, "Is there merit in assuming that employees who do not do well on a performance appraisal rating will not do well in a similar job at a higher grade level in the same organization or in a different organization at the same or higher grade level?"

One study which lends some support to the validity of using performance ratings to predict future performance was conducted by Lewin and Zwany (1976). Lewing and Zwany performed research on peer nominations in an effort to develop a model of the peer nomination process which could be used in predicting future performance. Other studies performed by Hollander (1954) and Korman (1968) support the concept that peer performance ratings are valid predictors of future performance. Hollander (1954) concluded that buddy ratings do predict success or failure in flight training and leadership ability. Korman (1968) in turn concluded that "peer ratings are better predictors than psychometric procedures" Similar determinations were made by Lindzey and Byrne (1969) and Miner (1968).

While these studies show support that performance appraisals

provide some acceptable degree of predictive validity we must use caution in concluding that supervisory ratings would also provide similar predictive validity. The variables in peer ratings may be quite different than the variables in supervisory ratings which would affect the degree of predictive validity. For example do peers have a greater opportunity to observe behavior than the supervisor, thereby aiding in more accurate ratings? Are there unknown variables of leadership which are sensed by peer raters through frequent contact? Do unknown or known variables create a "peeking" order which in turn is translated into a rating of future success or failure? The author has been unable to locate any studies directly related to how predictive supervisory ratings are when used for promotion selections.

Area #5: Racial and Sexual Bias

Schmidt and Johnson (1973) examined the effect of race on peer ratings used in an industrial setting. The sample used in the study was approximately 50% black and had recently been exposed to training in human relations. The Schmidt and Johnson (1973) study contrasted with previous studies conducted by Cox and Krumboltz (1958) and DeJung and Kaplan (1962) in that no race effect was found. The latter two studies found that raters rate members of their own race higher than members of the other race and that this effect is more marked for black than for white raters. Schmidt and Johnson (1973) contributed their reverse outcome to their relatively large proportion of blacks (46.2%) in these peer groups and/or to the human relations training received.

Possible sex discrimination in hiring practices was re-

searched by Kryger and Shikiar (1978) by examining the use of letters of recommendation. Past studies by Fidell (1970), Rosen and Jerdee (1974) and Dipboye, et al. (1974) indicated that sex discrimination at various stages of the hiring procedure still exists. Kryger and Shikiar (1978) investigated the screening and evaluation of letters of recommendation by selecting officials. They hypothesized that "both the sex of the person writing the letter of recommendation and the sex of the applicant would affect the applicant's likelihood of obtaining the job." Eight different letters of recommendation were tested for applicants applying for a management (trainee) position. Although the investigators expected to find discrimination against female applicants, their findings indicated a preference for female applicants. Kryger and Shikiar (1978) reasoned that the explanation for this "reverse discrimination" may have been based (1) upon publicity caused by the feminist movement changing the public's behavior toward females as well as (2) the impact of affirmative action regulations which has sensitized managers to give greater consideration to qualified female applicants than has been given in the past.

STUDY OF RACIAL, SEXUAL AND AGE
BIAS IN PERFORMANCE RATINGS

The Civil Service Reform Act and the Federal Guidelines on Employee Selection Procedures placed constraints on the use of performance appraisals in response to discrimination complaints. While the use of performance appraisals may continue to provide the basis for personnel decisions they must not be allowed to impact adversely on the racial and sexual composition of an organization's workforce.

These constraints along with stronger emphasis on affirmative action programs have contributed to increased numbers of racial minorities, women and older workers in the workforce. Schmidt and Johnson (1973) found that racial bias does not inevitably occur and Kryger and Shikiar (1978) found that some preference toward female applicants was evidenced in their study. These studies ran contrary to results of earlier investigations (Cox and Krumboltz, 1975, DeJung and Kaplan, 1962, Fidell, 1970, Rosen and Jerdee, 1974).

A study of racial, sexual, and age bias occurrence in a government agency (Washington Office, Agriculture Stabilization and Conservation Service) was conducted to determine if the findings would support Schmidt and Johnson's determination that racial bias does not inevitably occur and to see if these findings are similar to other types of group differences such

as sex and age.

Methodology: For the purposes of this study I used data which was available in existing personnel records and on computer tapes. Consequently the study was based on data available prior to the implementation of the Civil Service Reform Act. The use of data which was already on record was necessary to compensate for the shortage of time and lack of resources (such as funds and staffing) required to develop original and perhaps more suitable evaluations. Most of the records were located in the personnel folders of the employees within the agency or else were on computer tape. The primary document in the personnel folder was the form (see attached AD-434 in Appendix C) used in evaluating the employee for the preceding year.

The Performance Rating document (AD-434) contained all but three of the data (age, sex and race) necessary to conduct the study. The employee identifying information included name, pay plan, grade, salary, step, title, job number, organizational location and rating period. The rating periods, while varying from a few months to a little over a year do not go back beyond October 1977. The reason for this time limitation was due in part to the establishment of a new form and in part to a major reorganization of the Washington Office. The AD-434 in addition to the employee and rater identification information, contains such information as 1) major activities, 2) percent (%) of time for each major activity 3) the rating of each major activity on the factors of quantity and quality, and 4) on overall rating of either unsatisfactory, satisfactory, or outstanding.

The major activities are identified by each rater for each of the individuals rated and are governed by U.S.D.A. regulation. The parameters established by regulation limit the rate to identifying at least one major activity and not more than five major activities. The description of each major activity is limited to 40 characters and spaces. The restriction of forty characters and spaces is dictated by the need to put this information on a computer. The major activities identified must cover at least 80% of the employee's time. For each major activity a percentage of time is given to the nearest 5 or 10 percent. The quantity and quality of each major activity is rated on the basis of a five point rating scale ranging from 1 for unsatisfactory to 5 for distinguished with 3 the midpoint for proficient. In addition to the title each numerical rating has a brief definition.

Major activities varied so greatly from ratee to rater that the individual's ratings for each major activity could not be compared. For example, major activities for secretaries ranged from such activities as "prepares correspondence," "places telephone calls," "files, types and takes dictation" to activities such as "manages office" and "exercises extensive knowledge of organizational programs." While there were a total of twenty-three different activities listed for secretarial type activities each secretarial employee had only 3 to 5 of these activities listed on their ratings forms. Because of the wide variation in major activities, an average rating for each employee was determined. One average for the quantity of work and a separate average for the quality of work was obtained and rounded off to the nearest

whole number.

The total population for the Washington Office, Agricultural Stabilization and Conservation Service consisted of 545 permanent full time employees. The total population was screened to obtain nonrandom samples of raters with five or more ratees. Groups of five or more ratees under individual supervisors was determined necessary to provide statistical validity. All groups of five or more ratees were retained for the population samples. A total of 19 rating groups were identified. All of the 19 raters were white males. Twenty-nine job series out of a total of 77 for the agency were represented. It is recognized that this limited population is too small for real significance. (See Appendix D).

For this statistical analysis I used four null hypotheses. The variables were identified as 1) race, i.e., black, white and other minorities, 2) sex and 3) age. The first three hypotheses were tested with The Chi Square test (goodness-of-fit). The .05 level of significance was used to evaluate the null hypotheses. Each hypothesis was tested twice, once for the differences in ratings on the quantity of work and once for the differences in ratings on the quality of work.

The fourth hypothesis was tested by using the Pearson product-moment coefficient of correlation. The $\alpha = .4$ level of significance was used to evaluate the null hypothesis. Correlations were determined for both the quantity and quality performance ratings.

Hypothesis 1: There is no significant difference (at the .05 level) between performance ratings given to black employees and performance ratings given to white employees by white supervisors.

There were 130 rates in the sample used for this hypothesis. This sample was similar to the sample used in the second and fourth hypothesis except that two employees, neither white or black were excluded. Of these 130 employees, 41 were black and 899 were white for an approximate ration of 1 black for every 2 whites. Most of the blacks (36) were concentrated in grade levels 7 and below. Of the remaining five blacks, 4 were at the GS-9 level and 1 was GS-12. Whereas 85 percent of the blacks were at the grade level 7 or below, 20 percentage (18) of the whites were grade level 7 or below, and 49 percent (44) were grades 13, 14, and 15. The grades ranged from grade 2 through 15.

The Chi Square Test of the differences between the ratings of black employees and the ratings of white employees given by white supervisors resulted in a calculated value of 1.13 for the ratings on quantity of work and 2.39 for the ratings on the quality of work. The ratio of 1.13 for the quantity rating was greater than $p=.05$. The calculations failed to reject the first null hypothesis -- that there are no significant differences (at the .05 level) between performance ratings given to black employees and performance ratings given to white employees by white supervisors.

Hypothesis 2: There is no significant difference (at the .05 level) between performance ratings given to female employees and performance ratings given to male employees by male supervisors.

The second hypothesis was tested on a sample of 132 rates.

There were slightly more males (67) than females (65) in this sample. Most of the females (47) were concentrated in grade levels 7 and below. The remaining 18 females were found at grade levels from 8 through 13. Whereas 72 percent of the females were found at the grade level 7 and below, only 10 percent of the males were found in grades 7 and below, 87 percent in grades 11 and above and 62 percent in grades 13, 14, and 15. The grade levels ranged from grade 2 through 15.

The test of differences between performance ratings given to female employees and performance ratings given to male employees by male supervisors resulted in a calculated value of 2.12 for the ratings on quantity of work and 1.12 for the ratings on the quality of work. Both values were greater than $p=.05$. The calculations failed to reject the second null hypothesis -- that there are no significant differences between performance ratings given to female employees and performance ratings given to male employees by male supervisors.

Hypothesis 3: There is no significant difference (at the .05 level) between performance ratings given to black secretaries and performance ratings given to white secretaries by white supervisors. A total of 21 employees were identified for use in testing the third hypothesis. Of these 21 secretaries, 17 were classified in the secretarial field (0318 series) while 4 were classified as clerk stenographers (0312 series). The clerk stenographer series was included as part of the population distribution for the test because it is the custom of the agency to use the clerk stenographer series for lower level (GS-4 and GS-5 grades) employees who are training in the secretarial field.

Of those 17 employees classified in the 0318 series 14 were graded at the GS-6 level and one each were graded at the GS-5, 7, and 8 levels. Ages ranged from 63 years of age to 24 years of age with an average age of 37 years and a standard deviation of 12. While all of the secretarial population were female, the racial ratio was 9 blacks to 12 whites.

The Chi Square Test of the differences between the ratings of black secretaries and the ratings of white secretaries given by white supervisors resulted in a calculated value of .269 for the ratings on the quantity of work and 1.43 for the ratings on the quality of work. The ratio of 2.69 for the quantity rating was greater than $p=.05$. Therefore, the tests failed to reject the third null hypothesis -- that there are no significant differences between performance ratings given to black secretaries and performance ratings given to white secretaries by white supervisors.

Hypothesis 4: There is no significant correlation (Pearson $=.4$) between the age of employees and the performance ratings they receive.

The sample used for the fourth hypothesis was the same as the 132 employee sample used for the second hypothesis. The ages for this group ranged from 20 years of age to 63 years of age. The average age for this group was 42 years of age with a standard deviation of 11.

The testing of the correlation between the age of employees and the performance ratings they receive resulted in $Y=.76$ for the ratings received on the quantity of work and $Y=.44$ for the

ratings received on the quality of work. Both correlations showed a significant level of differences and therefore we can consider the null hypothesis rejected.

Discussion: The results of the statistical tests on the differences between ratings of black employees and the ratings of white employees given by male supervisors would support the finding found in the Schmidt and Johnson (1973) study that racial bias effect is not an automatic occurrence in ratings. Their study is further supported by the statistical tests conducted on the differences between the ratings received by black secretaries and the ratings received by white secretaries from white supervisors. Schmidt and Johnson (1973) indicated that it was quite possible the absence of racial bias effect in peer ratings may have been affected by the approximately equal proportion of minority and majority group members in peer groups as well as the human relations training associated with its nonoccurrence. It should be noted that while the groups selected in this study were not equally proportioned between white and black employees the ratio of blacks to white employees was approximately one black to every two whites. The secretarial groups, however, were more closely equal in size -- 9 black secretaries to 12 white secretaries. Another parallel between the sample of employees studied and the Schmidt and Johnson (1973) sample is that human relations training had been given to the USDA supervisors, although this was several years ago. It is impossible to say what effect this training had at the time it was given and whether the effect could continue over an extended

period of several years.

The results of the statistical tests on the differences between ratings of female employees and the ratings of male employees given by male supervisors would support the finding of Kryger and Shikliar (1978) in that no discrimination was found. While this study did not deal with hiring practices as did the Kryger and Shikliar (1978) study the same explanations given for the lack of evidence of discrimination may apply. Kryger and Shikliar (1978) felt that there were two plausible explanations for their not finding discrimination. The first explanation was the impact of the feminist movement may have changed management's behavior. The second explanation was that affirmative action regulations may have sensitized managers to give greater consideration to qualified female applicants. A third explanation may be in order and that is that contrast bias may not exist once the female applicant is employed and has an opportunity to demonstrate her abilities. Finally, there is the possibility that bias may never have existed in the performance appraisal process, but may exist only in the selection process.

The statistical study on correlations of age and ratings received was not tied into any previous studies. The high correlations which rejected the null hypothesis that there is not significant correlations ($r=.4$) between the age of employees and the performance ratings they receive may have several plausible explanations. First, the additional years of experience that often goes with older employees may provide an advantage to older employees in that it allows them to bring greater expertise to the job. Second, the agency benefits from a stable workforce and

stable programs. A stable workforce allows the older employee to develop stronger friendships both among peers and superiors and to develop greater influence which I believe is more readily reflected in global ratings.

CONCLUSION

Performance appraisals provide a basis for many types of personnel decisions such as promotions, appointments, training, rewards and punishments. Further, performance appraisals are the means most frequently used to gather performance information on employees and often are the only means to provide criterion scores against which psychologists can validate these decisions.

Researchers and management officials have recognized difficulties in developing and using performance ratings for personnel decisions. Borman, (1978), categorized these difficulties under four headings involving rater errors, observation opportunities, rating formats and organizational constraints. Concern by personnel specialists has been voiced as to the affect of these difficulties on the usefulness of ratings in distinguishing between peers when making selections for promotion.

A review of research literature indicates that the affect of reduction of rater errors can be to increase the validity of the performance appraisal. Various studies (Borman, 1975; Borman and Dunnett, 1975; Latha, Wexley and Parsel, 1975; and Wexley, et al. 1973) have supported the use of training in reducing rater errors such as halo effect, contrast, similarity, first impressions.

Schmidt and Johnson (1973) also concluded that the non-occurrence of racial bias was in some measure attributable to human relations training. Training is a method easily available

to federal agencies and by comparison with other methods (development of behavioral rating scales and validation studies) is the most cost effective means of improving the usefulness of performance evaluations. From management's point of view it would appear to be the most acceptable approach to increasing the validity of performance appraisals as a selection method.

However caution should be exercised before embarking on training programs to reduce rater error. It would be prudent to determine in advance what type of rater error, if any, is prevalent in an organization before conducting training programs designed to reduce rater errors. The Schmidt and Johnson (1973) study indicated a non-occurrence of contrast error, while Kryger and Shikliar (1978) found a minimum of contrast error in favor of women applicants. Their findings were supported by the author's study of rater error with regard to black and white employees rater by white raters and female and male employees rated by male raters. In addition, Borman (1978) concluded that there may be a ceiling on how much validity can be achieved. Based on these studies, training should be undertaken only after some evidence is available to support the undertaking and to justify the time, cost, and effort involved.

Other difficulties in the use of performance appraisals stem from the design of rating formats and the lack of opportunity to observe the behavior being evaluated. Recognizing that the evaluator does not always have the opportunity to observe the behavior being rated, Smith and Kendall, using techniques of critical incidents developed by Flanagan (1949 and 1954),

designed their rating format around examples of expected behavior. The use of expectations based on observed similar behavior has the advantage of offsetting the lack of opportunity for rater observation. Further studies of the behaviorally anchored rating scales (Borman and Dunnette (1975), Irvey and Hoyle (1974), and Campbell, et al. (1973)) support their use as a means of increasing the validity of performance ratings as well as increasing the manager's knowledge of job requirements and job performance. The disadvantage of this type of format design is that a large amount of time is required and the development of the behaviorally anchored rating scales is expensive. From a practical viewpoint the manager would have to consider whether the modest amount of increased validity offset the time and cost involved. Only, large numbers of jobs of similar job requirements would make the use of behaviorally anchored rating scales worthwhile in terms of the increased validity and the confidence gained in the use of such scales.

What constraints are placed on an organization such as ASCS by the new Civil Service Reform Act and the Uniform Guidelines on Employee Selection Procedures? Through much trial and error the Office of Personnel Management (previously the Civil Service Commission) has adopted regulations which appear to be based on the more recent behavioral studies. The regulations would provide wide latitude for a federal organization to adopt such approaches as behaviorally anchored rating scales and experiment with rating scales designed specifically for promotion selections. Further, the Equal Employment Opportunity Commission

requirements have adopted the validity strategies recognized by the American Psychological Association, thereby encouraging agencies to work closely with behavioral scientists. The time appears favorable for federal agencies such as ASCS to pursue needed studies in the use of performance appraisals to predict success or failure in promotion selections. Certainly, the organizational constraints placed on agencies by the Office of Personnel Management and the Equal Employment Opportunity Commission are limited in scope.

In summary, what do these findings mean with respect to my assumption that problems such as organizational constraints, rater error and rating scale design preclude valid use of performance appraisals in selecting employees for promotions. Studies have indicated that greater reliability and validity can be obtained through training programs to reduce rater error, and through better designed rating formats. However, the cost and time involved may from management's point of view offset the validity to be gained. Management may decide to provide organizational limitations on such improvements through constraints on available expenditure and manpower.

Caution then should be exercised by agencies before undertaking expensive and time consuming efforts by determining ahead of time whether such undertaking will prove beneficial. There is some evidence to indicate there may well be a point of diminishing return for the effort put forth (Borman, 1978).

Organizational constraints on experimentation, previously

required by the Office of Personnel Management (OPM) have been lifted. Both the Equal Employment Opportunity Commission and the OPM requirements reinforce the need for more extensive validation studies in the use of performance ratings as a basis for personnel decisions. Such ratings must have a strong foundation in job requirements and show a relationship between these requirements and success on the job.

I have prepared a model of the variables which I perceive influence the use of performance appraisals for promotion decisions within ASCA (Appendix E). The evaluator's rating is strongly dependent on procedural constraints format design his observation opportunities, his perception of the relationship between the rater and the employee in achieving common objectives, and his knowledge of rater errors. To a far lesser extent the rating is influenced by the employee's perception of fairness of the system, the right of grievance and the right to know. The ASCS procedure requires a pass-fail determination be made. The standard for the pass-fail level is determined by each promotion panel for the specific vacancy being filled.

Very little is known about the use of performance appraisal at the panel and selecting official stages of decision making. The strong concern by personnel specialists and the EEOC and the lack of research on the use of performance appraisals in the selection for promotion process indicates to me that one research effort which would prove most useful is in the area of predictive validity studies on the use of supervisory ratings in selecting employees for promotion. The lack of such studies

is cause for a lack of confidence in the use of performance ratings to predict success or failure in promotion decisions.

APPENDIX A

Documentation for Quality Increase (ASCS)
(Demonstrating incidents of high level performance)

APPENDIX A

Documentation for Quality Increase (ASCS)
(Demonstrating incidents of high level performance)

I recommend _____ for a quality increase based on her consistent outstanding performance during the past year.

Mrs. _____ has performed the duties and responsibilities of her position in a manner which consistently meets the objectives of branch functions, as well as that of the Personnel Division. Mrs. _____ is one of five Personnel Staffing & Classification Specialists assigned to the Classification & Staffing Branch No. 1. Assigned responsibilities include the overall classification and employment program for county offices and the classification and employment responsibilities for the Export Marketing Service. The workload in the assigned areas is shared with another personnel specialist. Mrs. _____ gives prompt attention to work assignments which results in meeting or beating deadlines, many of which are dictated by the priority nature of the work. Her assignments are voluminous in nature and necessitate a high degree of productivity. She is extremely cooperative and provides, on her own initiative, assistance as needed to meet excessive branch workloads and assignments. She undertakes each and every task willingly, establishes her priorities, researches work thoroughly, and concludes with a finished product indicative of her outstanding ability and professional attributes.

Mrs. _____ thoroughness of work is apparent in the unusually high degree of accuracy which she maintains. Her job knowledge and expertise as a classifier has necessarily required her assisting the other specialists within the branch. For example, as a result of merging the functions of classification and employment, all specialists have been in the process of learning additional duties and responsibilities for which they had no prior experience. Inasmuch as Mrs. _____ is the only specialist with classification and organization background, she has provided the technical assistance and guidance pertinent to other specialists' training as it relates to classification and organization. The additional workload has placed unusual demands on her time. This involves training of specialists in desk audits, position evaluation, job descriptions, organizational review, etc. Her ability to handle the additional responsibilities, plus her regular work with consistency and on a timely basis. The continuity of work operations indicates the demonstrated ability and effectiveness in which she conveys written and oral communication to her co-workers.

APPENDIX B

Excerpt from Removal Action (ASCS)
(Demonstrating incidents of poor performance)

APPENDIX B

Excerpt from Removal Action (ASCS)
(Demonstrating incidents of poor performance)

Failure to Exercise Sound Judgment in Preparation of 8-PM

One of your major assignments as a GS-9 Employee Relations Specialist was the preparation of a revised 8-PM, the ASCS Safety and Health Handbook. As a GS-9 you are expected to work without close supervision and generally exhibit sound judgment. Lack of good judgment requires close supervision and impacts on the efficiency of the Branch. The record reflects the following instances where your judgment in completing your 8-PM assignment was not sound:

Specification 1: Incorporation of Notice PM-877 titled, "Revised Form AD-278, Supervisor's Report of Accident." You failed to properly incorporate this notice into 8-PM. You did incorporate this notice in paragraph 16 of 8-PM, "Supervisor's Report of Accident" (AD-278). While the heading of the notice and the paragraph are similar, the contents of the notice deals with matters not properly discussed in paragraph 16. Paragraph 16 is primarily a referral paragraph for four subjects: (1) responsibility for completion of form, (2) instructions for completion of form -- refers reader to exhibit 2, (3) distribution of copies -- refers reader to exhibit 3, (4) assistance in investigation of accident -- refers reader to exhibit 4. It was necessary for you to read the substantive part of the notice in order to put information in its proper location in 8-PM. A superficial review results in the entire notice being placed in a paragraph with the same heading as the notice. The correct location of this information should have been:

<u>Notice Paragraph</u>	<u>Corresponding Paragraph -- 8-PM</u>
1A	14, Special Reporting (regarding deaths, property, damage, etc.)
1B	13, Reportable Accidents
2A	Exhibit 2 (instructions for preparation of AD-278)

Failure to incorporate the substance of PM-877 in the proper paragraphs of 8-PM denotes the lack of good judgment in completing this assignment.

Specification 2: Incorporation of Notice PM-930, titled "Submission of Form AD-278, Supervisor's Report of Accident." You failed to properly incorporate this notice into 8-PM. You did incorporate this notice in paragraph 16. You made the same error in incorporating Notice PM-930 in 8-PM that you made in incorporating Notice PM-837 in 8-PM. That is while the heading of the notice and the paragraph are similar, this similarity was not the determining factor in deciding where the information should go. The substantive information contained in the body of the notice is the determining factor in deciding where the information goes. The notice deals with distribution of copies of AD-278. Paragraph 16D also deals with the distribution of copies of the AD-278 by referring to exhibit 3. Yet, you established another paragraph, "F," in paragraph 16 which is concerned with the distribution of the AD-278.

In addition, in the case of Notice PM-930, you included information concerning counties in 8-PM, paragraph 16, when, in fact, county offices are not on the distribution schedule for 8-PM. Handbook 22-PM (formerly 6-CA) is the appropriate handbook for county personnel matters.

Finally, State Office Action (paragraph 16G in your submission) is unnecessary because it was a one-time action item required by the notice and not a continuing requirement to be included as a permanent feature of the handbook.

Failure to incorporate the substance of Notice PM-930, titled "Submission of Form AD-278, Supervisor's Report of Accident," inclusion of county office references and inclusion of paragraph 16G denotes the lack of good judgment in completing this assignment.

Specification 3: Incorporation of Personnel Bulletin No. 810-16, Subject: Realignment of Office of Workers' Compensation Program District Office. You failed to properly update exhibit 6 by introducing addresses which did not meet commonly accepted addressing practices and by introducing superfluous information. In the first instance the USDA Personnel Bulletin No. 810-16 instructs you to use the address below for the Dallas, Texas Office of Workers' Compensation Programs.

U.S. Department of Labor
Employment Standards Administration
Office of Workers' Compensation Programs
555 Griffin Square Building
Griffin and Young Streets
Dallas, Texas 75201

You failed (1) to use this address as the USDA bulletin instructed, or (2) to at least identify OWCP as part of the Department of Labor as was the common practice of the exhibit 6 which you were updating, or (3) if nothing more, spell out what OWCP means. Instead you used only the acronym OWCP without any further identifying information. While it may be argued that the information regarding street addresses was correct and would have gotten the claims forms to the correct address, it is far more proper and certainly safer to spell out the full identity of OWCP or if not that, then to identify OWCP as part of the Department of Labor. The reason it is safer and more proper to do so is that there is no room number or box number on some of the addresses. Also, the room number or box number may change and since there may be more than one federal organization in the building, the envelope could be delayed in reaching its designation or lost.

APPENDIX C

United States Department of Agriculture
Performance Rating

APPENDIX C

UNITED STATES DEPARTMENT OF AGRICULTURE

PERFORMANCE RATING

1 NAME (Last, First, Middle)		2 SOCIAL SECURITY NUMBER	3 AGEN. CODE	4 PAY PLAN	5 OCCUPATNL SERIES	6 GRADE/STEP		
7 OFFICIAL TITLE OF POSITION		8 PERS. POSITION NUMBER	9 STANDARD JOB NUMBER	10 RATING PERIOD FROM		TO		
11 DUTY STATION CODE	12 ORGANIZATIONAL STRUCTURE CODE	13 INT. DAYS IN PAY STATUS NO. DAYS AS OF (Date)		FOR NFC USE ONLY				
14 MAJOR ACTIVITY RATING SCALE								
<p>(1) UNSATISFACTORY – Fails to meet position requirements even after a written warning at least 90 days prior to being rated officially.</p> <p>(2) PASSABLE – Meets position requirements, but does so at a level lower than normally expected for the position.</p> <p>(3) PROFICIENT – Meets position requirements and does so at the level expected of a fully qualified incumbent.</p> <p>(4) SUPERIOR – Exceeds position requirements at a level high enough to identify as different.</p> <p>(5) DISTINGUISHED – Exceeds position requirements at a level that is high enough to deserve special commendation.</p>								
15 MAJOR ACTIVITY RATING (In the words and phrases you normally use to plan, assign, and discuss work with your subordinate, list up to five major activities in which this employee works. Each major activity description must be limited to 40 consecutive printed characters. The activities listed should account for at least 80% of the employee's time. For each activity, show the amount of time as a percentage of total duty time in the rating period. This should be to the nearest 5% or 10%. For each activity, rate the quantity and quality of achievement compared to performance for the position. Use the five-point rating scale listed in block 14.)								
MAJOR ACTIVITY (PRINT ONLY)					% OF TIME	QUAN- TITY	QUAL- ITY	
1								
2								
3								
4								
5								
16 RATING OF SUPERVISORY PERFORMANCE (In a brief narrative, describe how and how well this employee directs, motivates, evaluates, and develops subordinates. Include responsibility in furthering Equal Employment Opportunity and responsibility in occupational health and safety.)					16A SUBORDINATES (List the occupations this employee supervises and give the number in each.)			
					Occupational Series	Number of Subordinates		
17 POSITION ACCURACY					YES	NO		
a. Are the major activities and supervisory responsibilities rated above covered by the position description?								
b. Are all the major duties and responsibilities in the position description reflected in Items 15 and 16 above?								
18 PERFORMANCE ADJECTIVE RATING								
<input type="checkbox"/> Unsatisfactory <input type="checkbox"/> Satisfactory <input type="checkbox"/> Outstanding								
19 RATER'S SIGNATURE					DATE			
20 REVIEWER'S SIGNATURE					DATE			
NOTICE: An employee desiring to appeal this rating should follow instructions on the back of the employee copy of this form.								
21 I have read Appendix I of the Department's Employee Handbook, Employee Responsibilities and Conduct, and have had my questions answered to my satisfaction.					YES	NO		
22 My supervisor has discussed this rating with me.								
23 My position description is accurate.								
24 EMPLOYEE'S SIGNATURE					DATE			
25 REASON (If employee did not sign)								

FORM AD - 434 (REV. 5/77)

ORIGINAL - NFC PROCESSING COPY (INSTRUCTIONS ON REVERSE)

APPENDIX D

Population Distribution -- Performance Appraisal Study
ASCS (Washington Office) -- 1978

APPENDIX D

Population Distribution -- Performance Appraisal Study
 ASCS (Washington Office) -- 1978

ID Code	Job Series	Grade	Age	Sex	Race	AV. RATING	
						Quantity	Quality
1	0335	08	58	2	2	5	5
2	0301	14	59	1	2	4	4
3	0318*	07	33	2	1	3	3
4	0301	12	52	1	2	3	4
5	0343	13	38	1	2	4	4
6	0560	13	41	1	2	3	3
7	0990	13	56	1	2	3	4
8	0305	04	58	2	1	4	3
9	1146	13	42	1	2	4	4
10	1071	11	55	1	2	3	3
11	0560	09	35	2	2	4	4
12	0334	13	31	1	2	4	4
13	0560	13	38	1	2	4	4
14	0201	07	33	2	2	4	4
15	1531	07	34	2	1	3	4
16	0334	13	35	1	2	4	4
17	0312*	05	24	2	2	4	4
18	0318*	06	53	2	2	4	3
19	0344	09	32	2	2	5	5
20	0301	13	57	1	2	4	4
21	2130	13	60	1	2	3	4
22	2130	12	48	1	2	3	3
23	1082	13	40	1	2	3	3
24	0343	13	36	2	2	4	4
25	1081	12	31	2	2	3	3
26	0322	04	21	2	1	3	3
27	0318*	06	34	2	1	3	4
28	0322	03	29	2	2	3	3
29	0322	04	26	2	2	3	3
30	0322	04	25	2	1	3	4
31	0201	12	35	1	2	3	3
32	0344	07	31	1	1	4	4
33	0343	09	49	2	2	4	4
34	0318*	08	39	2	2	3	4
35	0318*	06	58	2	2	4	4

Legend: Male - 1; Female - 2; Black - 1; White - 2;
 Other - 3; Secretarial - *.

APPENDIX D -- Continued

ID Code	Job Series	Grade	Age	Sex	Race	AV. RATING	
						Quantity	Quality
36	0318*	06	32	2	2	3	3
37	0318*	06	26	2	2	4	4
38	0343	13	42	1	3	4	3
39	1082	11	45	2	2	3	3
40	0318*	06	33	2	1	4	5
41	0510	12	52	1	2	4	4
42	2130	07	31	1	1	3	4
43	0312*	05	27	2	1	4	4
44	0344	06	44	2	1	3	3
45	2131	07	58	1	1	3	3
46	0301	06	55	1	1	3	3
47	1001	07	35	2	2	3	3
48	0312*	04	25	2	1	3	3
49	1145	13	54	2	2	3	3
50	0334	13	40	1	2	3	4
51	0334	07	31	2	1	5	4
52	0301	14	59	1	2	4	4
53	0318*	05	33	2	2	4	4
54	0301	06	32	2	1	3	3
55	0110	14	43	1	2	4	3
56	0560	09	37	2	2	3	4
57	0301	13	59	1	2	4	4
58	0560	14	63	1	2	4	3
59	0301	13	43	1	2	4	4
60	1001	07	36	2	2	3	3
61	0334	12	44	1	2	3	3
62	1081	14	55	1	2	3	3
63	0560	12	48	1	2	3	3
64	1146	11	44	1	2	4	4
65	0301	12	57	2	2	4	4
66	0201	12	57	2	2	4	4
67	1145	14	41	1	2	4	4
68	1145	14	46	1	2	4	4
69	0343	14	56	1	2	4	4
70	0343	13	48	1	2	3	4
71	1150	13	43	1	2	4	4
72	1150	13	59	1	2	4	4
73	1081	14	52	1	2	4	3
74	0343	13	41	1	2	4	4
75	0560	14	36	1	2	5	5

Legend: Male - 1; Female - 2; Black - 1; White - 2;
Other - 3; Secretarial - *.

APPENDIX D -- Continued

ID Code	Job Series	Grade	Age	Sex	Race	AV. RATING	
						Quantity	Quality
76	1145	07	41	2	2	4	4
77	1146	12	37	1	2	4	4
78	1146	14	46	1	2	4	4
79	0343	09	42	2	1	3	3
80	0388	04	62	2	1	3	3
81	0301	13	50	1	2	4	4
82	1145	13	48	1	2	4	4
83	0110	13	52	1	2	3	3
84	1145	09	39	1	2	3	3
85	1146	13	36	1	2	4	4
86	0110	14	51	1	2	4	4
87	0301	15	46	1	2	4	4
88	1145	13	40	1	2	3	3
89	1081	13	58	1	2	3	3
90	0334	13	52	1	2	4	4
91	1081	12	47	1	2	3	3
92	1145	13	44	1	2	4	4
93	1146	13	46	1	2	4	4
94	0560	12	58	1	2	3	3
95	1150	14	50	1	2	4	4
96	0560	12	51	1	2	3	3
97	0560	11	31	1	3	3	4
98	0301	09	55	2	2	4	4
99	0560	13	44	2	2	4	4
100	0344	07	38	2	1	4	4
101	0322	03	33	2	1	3	3
102	0343	09	32	2	1	4	4
103	0318*	06	31	2	1	4	4
104	0318*	06	30	2	1	4	4
105	0318*	06	30	2	2	3	3
106	0318*	06	30	2	1	4	4
107	0530	05	32	2	1	3	3
108	0301	04	27	1	1	3	3
109	0318*	06	27	2	2	4	4
110	0318*	06	54	2	2	4	4
111	0501	09	55	2	1	4	4
112	0318*	06	63	2	2	3	3
113	2130	11	50	1	2	3	3
114	1145	11	43	2	2	4	4
115	0312*	05	43	2	1	3	3

Legend: Male - 1; Female - 2; Black - 1; White - 2;
Other - 3; Secretarial - *.

APPENDIX D -- Continued

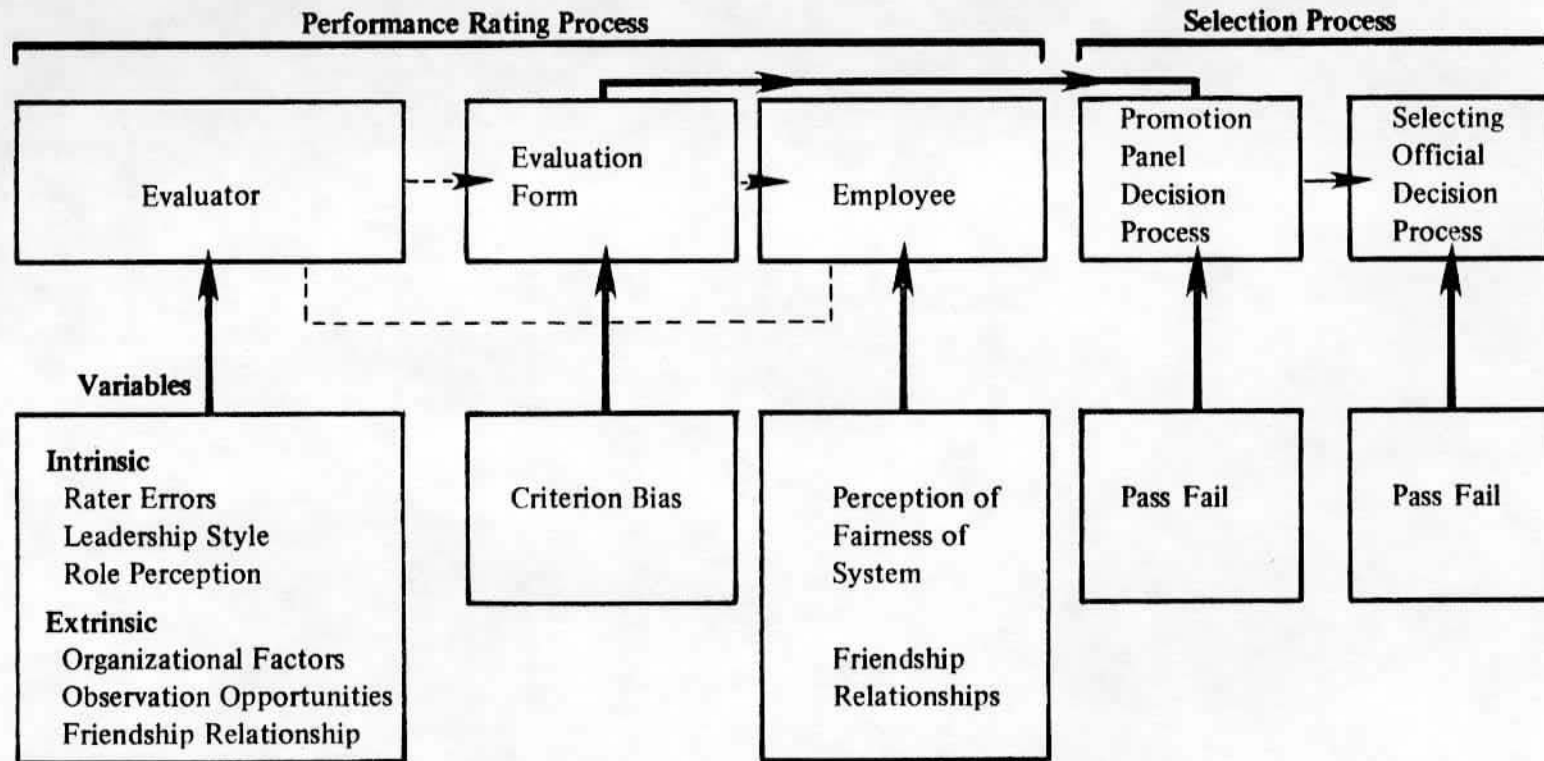
ID Code	Job Series	Grade	Age	Sex	Race	AV. RATING	
						Quantity	Quality
116	0344	07	38	2	1	4	3
117	1145	09	35	1	1	3	4
118	1531	05	28	1	1	4	4
119	0301	03	24	2	1	4	3
120	0322	02	22	2	1	3	3
121	0322	03	20	2	1	4	4
122	0318*	06	56	2	2	4	3
123	0305	05	55	1	1	4	3
124	0110	13	45	1	2	3	3
125	0305	04	54	2	1	4	3
126	1084	11	44	1	2	3	3
127	0322	04	34	2	2	3	3
128	0343	07	31	2	1	4	4
129	0501	07	31	2	1	4	3
130	0334	12	30	1	1	4	5
131	1087	05	25	2	1	3	3
132	0322	02	22	2	1	5	4

Legend: Male - 1; Female - 2; Black - 1; White - 2;
Other - 3; Secretarial - *.

APPENDIX E

Model of ASCS Use of Performance
Appraisals in Promotion Decisions

APPENDIX E
MODEL OF ASCS USE OF PERFORMANCE APPRAISALS IN PROMOTION DECISIONS



BIBLIOGRAPHY

BIBLIOGRAPHY

- Arvey, Richard D. & Hoyle, Joseph C.
A Guttman approach to the development of behaviorally based rating scales for systems analysts and programmer/analysts. Journal of Applied Psychology, 1974, Vol. 59 (1), 61-68.
- Borman, Walter C.
Effects of instructions to avoid halo error on reliability and validity of performance ratings. Journal of Applied Psychology, 1975, Vol. 60, (5), 556-560.
- Borman, Walter C.
Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 1978, Vol. 63 (2), 135-144.
- Borman, Walter C. & Dunnette, Marvin D.
Behavior-based versus trait oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, Vol. 60 (5), 561-565.
- Campbell, J.P., Dunnette, M.D., Arvey, R.D. & Hellervik, L.W.
The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, Vol. 57 (1), 15-22.
- Civil Service Reform Act
U.S. Code Vol. 5, Secs. 4301-5 (1978).
- Cox, J.A. & Krumboltz, J.D.
Racial bias in peer ratings of basic airman. Sociometry, 1958, Vol. 21, 292-299.
- DeJung, J.E. & Kaplan, H.
Some differential effects of race of rater and ratee on early peer ratings of combat attitude. Journal of Applied Psychology, 1962, Vol. 46, 370-374.
- Dipboye, R.L., Fromkin, H.L. & Wilback, K.
Relative importance of sex, attractiveness and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 1975, Vol. 60, 39-43.
- Equal Employment Opportunity Commission
Uniform Guidelines on Employee Selection Procedures (1978);
43 FR No. 166 (August 25, 1978); 29 CFR Part 1607.

- Fidell, L.S.
Empirical verification of sex discrimination in hiring practices in psychology. American Psychologist, 1970, Vol. 25, 1,094-1,098.
- Flanagan, John C.
A new approach to evaluating personnel. Personnel, 1949, Vol. 26, 35-42.
- Flanagan, John C.
The critical incident technique. Psychological Bulletin, 1954, Vol. 51 (4), 327-357.
- Guilford, J.P. and Fruchter, Benjamin
Fundamental statistics in psychology and education. Sixth Edition, New York: McGraw-Hill, 1978.
- Hollander, E.P.
Buddy ratings: Military research and industrial implications. Personnel Psychology, 1954, Vol. 7, 385-393.
- Huck, Schuylen W., Cormier, W.H. & Bounds, W.G.
Reading statistics and research. New York: Harper and Row, 1974.
- Kerlinger, Fred N.
Foundations of Behavioral Research. Second Edition. New York: Holt, Rinehart and Winston, Inc. 1973.
- Korman, A.K.
The prediction of managerial performance: A review. Personnel Psychology, 1968, Vol. 21, 295-322.
- Kryger, B.R. & Shikiar, R.
Sexual discrimination in the use of letters of recommendation: A case of reverse discrimination. Journal of Applied Psychology, 1978, Vol. 63 (3), 309-314.
- Latham, Gary P., Wexley, Kenneth N. & Pursell, Elliot D.
Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, Vol. 60 (5), 550-555.
- Lawler, E.E.
The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 1967, Vol. 51, 369-381.
- Lewin, Arie Y. and Zwany, Abram
Peer nominations: A model, literature critique and a paradigm for research. Personnel Psychology, 1976, Vol. 29, 423-447.

- Lindzey, G. and Byrne, D.
Measurement of social choice and interpersonal attractiveness. In G. Lindzey and E. Arson (Eds.) Handbook of Social Psychology (Vol. 2), Reading, Mass., Addison - Wexley, 1969.
- Miner, J.B.
The early identification of managerial talent. The Personnel and Guidance Journal, 1968, Vol. 46, 586-591.
- Rosen, R. & Jerdee, T.H.
Effects of applicant's sex and difficulty of job on evaluations of candidates for managerial positions. Journal of Applied Psychology, 1974, Vol. 59, 511-512.
- Schmidt, Frank L. & Johnson, Raymond H.
Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1973, Vol. 57 (3), 237-241.
- Sellitz, Claire, Johoda, Raymond H.
Research methods in social relations. Second Edition. New York: Holt, Rinehart & Winston, 1959.
- Smith, P.C. & Kendall, L.M.
Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, Vol. 47 (2), 149-155.
- U.S. Civil Service Commission
An Historical Survey of Performance Rating Systems in the Federal Service, U.S. Civil Service Commission, September 1974.
- Wexley, Kenneth N., Sanders, Raymond E., & Yukl, Gary A.
Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 1973, Vol. 57 (3), 233-236.