Delay and Distortion: Tacit Influences on Performance Appraisal Effectiveness

Michael E. Benedict and Edward L. Levine
University of South Florida

A laboratory study of performance appraisal delay and distortion (a) investigated the effects of rater gender, performance level, and feedback requirement, and (b) defined feedback delay as the separate components of appraisal delay and feedback session scheduling delay. Subjects were 64 university business students who had at least 6 months of supervisory experience. Female raters delayed longer when conducting performance appraisals and when scheduling feedback sessions with subordinates, and more positively distorted ratings than did male raters. Moderately low performers were evaluated and scheduled for feedback sessions later, and were evaluated with more positive distortion than were moderately high performers. When required to provide feedback, raters delayed longer appraising performance and distorted ratings more positively. Significant interactions suggest that when feedback is required, female raters may delay appraisals, delay scheduling feedback sessions, and more positively distort their ratings of performance, particularly when rating low performers.

The communication of performance appraisal information to subordinates is an important function of organizational control systems. As organizational control information, feedback is an important factor in the enhancement of organizations' overall effectiveness. Accepting the assumption that supervisors are hesitant to convey negative feedback to subordinates (Meyer, Kay, & French, 1965), managers may use one of three strategies to affect the variability in the delivery of feedback: (a) avoidance, (b) delay, and (c) distortion. Evidence suggests that individuals attempt to avoid communicating negative or unpleasant information to others (Blumberg, 1972; Katz & Kahn, 1978; Tesser & Rosen, 1975). However, the current emphasis on employee performance appraisals by many American companies, as well as the emphasis on corporate requirements for annual employee reviews, seems to preclude the avoidance option for many supervisors. Therefore, this study focuses on only the last two strategies available to the supervisor, delay and distortion.

Studies suggest that delays in preparing and communicating feedback are widespread (Gruenfeld & Weisberg, 1966; Tesser, Rosen, & Tesser, 1971), particularly if that feedback is to be negative. However, delay is not always found. In a study of 168 male college students who acted as supervisors, Fisher (1979) found that when they were required to provide feedback to both below-average and above-average employees, subjects made performance appraisals of moderately poor performing subordinates more quickly (decreased delay), as opposed to moderately high performers. Fisher's results, therefore, are contrary to a general tendency to delay the transmission of negative performance information to subordinates. Why were negative performance appraisals made sooner than positive appraisals? Fisher's postexperimental questionnaire responses, obtained from the supervisors, suggest that the reversal can be explained by the supervisors' desire to allow subordinates time to improve their performances after the delivery of feedback, but before the end of the work session. Although Fisher's results may indicate a situational element that would operate to qualify a general tendency to delay conducting negative performance evaluations and providing feedback, other possible explanations for the inconsistency must be considered.

Results of Fisher's (1979) study may be considered equivocal because of a lack of consideration for the whole performance appraisal process regarding the experimental subjects. Central to the issue of delay in providing performance feedback is the supervisor's perception of the interaction with the rated subordinate. Fisher's supervisors were college students who were required to participate in the experiment for class credit. These subjects never experienced any personal contact with their subordinates. The subjects in the study neither developed the supervisor–employee relationships nor experienced any of the supervisor–subordinate interactions that are typical of organizational activities. We attempted to address these and related issues that arose when we reviewed Fisher's study.

Supervisors may resort to distorting the performance appraisal feedback that they must provide to subordinates, as well as the actual performance ratings. By suppressing unfavorable information during the performance appraisal feedback session, a supervisor may depict an employee's true performance

---

This article is based on the first author's doctoral dissertation under the direction of the second author.

We would like to thank the members of the dissertation committee (James Jenkins, Herbert Meyer, Francis Sistrunk, and Frederick Swierczez) for their advice and support. We would also like to extend our appreciation to David Stephens at the Department of Management, University of Texas at El Paso, for his assistance in facilitating part of the data collection, and to Marion Palmer for her support of the research and comments on an earlier version.

Correspondence concerning this article should be addressed to Michael E. Benedict, who is now at the U.S. Army Research Institute for the Behavioral and Social Sciences (ATTN: PERI-RG), 5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600.
in a more positive, and thus more satisfactory, manner. Studies of this phenomenon (Tesser, Rosen, & Batchelor, 1972; Tesser, Rosen, & Conlee, 1972) have reported that subjects will selectively transmit messages that contain bad news by transmitting only those portions that are neutral in content and avoiding the unpleasant portions. This effect, described as positive distortion, has been discussed as a factor in performance appraisals (Huttner & O’Malley, 1962; Oberg, 1972) and can be differentiated from such rating errors as central tendency, halo, recency, or bias (Cummings & Schwab, 1973).

In addition to distorting the feedback about a subordinate’s performance, the supervisor may also distort the subordinate’s performance evaluation by inflating the actual ratings. By making the subordinate appear on paper to be performing at a more satisfactory level than is justified by actual job performance, the supervisor reduces the probability that the feedback will be negatively received.

Reasons why performance appraisal ratings and feedback may be distorted are suggested in the literature. For example, supervisors may feel that giving negative information is distasteful because of subordinates’ tendencies to overestimate their personal performance levels (Hanson, Morton, & Rothaus, 1963; Prien & Liske, 1962; Thornton, 1968).

Although the importance of accurate performance appraisal results seems obvious, few studies directly address the distortion effect. Stockford and Bissell (1949) studied 485 aircraft assembly plant supervisors in a company in which performance appraisal results historically were not made available to subordinates. Supervisors were asked to reevaluate their employees 2 weeks after they had completed regular quarterly performance ratings, with the intention of conducting feedback sessions with each employee. These ratings were then compared with the original ratings, and a significant increase in mean ratings was observed when managers anticipated that the results would be communicated to their subordinates. This provides strong evidence for the contention that performance ratings made for the purpose of providing employees with feedback are positively distorted.

Studying distortion in performance ratings made for subordinate feedback purposes of moderately low and moderately high performing subordinates, Fisher (1979) found that supervisors who were required to give feedback to below average performers tended to distort their performance appraisal ratings, judging their subordinates’ performances less negatively (positive distortion) than did supervisors who completed the same performance ratings but were not required to give feedback. She also found no significant difference in ratings of moderately high performers related to expectations of face-to-face feedback. The latter finding may have been due to certain features of Fisher’s study such as raters who lacked supervisory experience, raters who never experienced any personal contact with their subordinates, and raters who were not required to experience any of the supervisor-subordinate interactions that are typical of organizational activities.

Past studies of delay and distortion in performance appraisal ratings thus far have not systematically addressed the issue of the gender of the rater. Such an omission ignores that women are a growing presence in managerial roles (Brown, 1981; Reha, 1979). A number of studies have suggested that there may be systematic differences in the ratings completed by women as opposed to those of men. London and Poplawski (1976), in a simulated work setting, found that evaluations made by female college students were higher than those by male college students in appraisal and interview situations for individual performance dimensions. Hamner, Kim, Baird, and Bigoness (1974) also found that women gave higher ratings than did men when they evaluated performances in a simulated work setting, especially for high levels of subordinate performance. Although there are inconsistencies, available research in this area suggests that women generally evaluate others more positively than do men. As a result, this study incorporated gender of rater as an independent variable to probe its effect on distortion and delay.

Ratee Gender

Overall, a general tendency has been documented that men are more favorably evaluated than are women. Male job applicants tend to be selected more frequently than are equally qualified female job applicants for managerial, scientific, and semiskilled positions (Gutek & Stevens, 1979; Haeffner, 1977; Rosen & Jerdee, 1974a, 1974b). Male candidates are also rated more positively than are female candidates on dimensions such as acceptability and service potential (Gutek & Stevens, 1979; Rosen & Jerdee, & Prestwich, 1975). Rater-ratee gender further complicates the interpretation of previous findings. Pheterson, Kiesler, and Goldberg (1971) showed that women evaluate the accomplishments of women less favorably than they do those of men, unless clear evidence of outside recognition is present. In a related study, Goldberg (1968) showed that women evaluate the accomplishments of comparably performing male subjects more favorably than do female subjects, when performance criteria are less clearly defined. Even when objective measures of performance were defined, Bigoness (1976) and Hamner et al., (1974) showed that sex stereotypes influenced performance ratings. Although the sex of ratee effect seems to be an important consideration, the present study held ratee gender constant to reduce the variability previously introduced by this variable.

To clarify further the dynamics of performance appraisal delay and distortion, the present research incorporated factors that appear to be potential contributors to delay and distortion, namely, rater gender, level of subordinate performance (moderately high or moderately low), and whether face-to-face feedback is provided to subordinates. The research, although completed in the laboratory, was carefully designed in an attempt to attain fidelity with key aspects of performance appraisal processes in organizations.

It was predicted that female raters, moderately low subordinate performance levels, and the requirement of providing face-to-face feedback would each produce greater delay and distortion than their counterparts. These variables were also predicted to interact in such a fashion that delay and distortion would be disproportionately greatest when women rated moderately low performing subordinates with whom they would provide face-to-face feedback.
Method

Subjects

The sample consisted of 32 male and 32 female students recruited from business and management courses at three public colleges and universities in western Texas. The number of subjects was determined by means of a power analysis, with power at .9, a set at .05, and effect size estimates drawn from previous research. Subjects were invited to participate through the use of a verbal announcement read to them by their course instructors before lecture. Only White subjects were included in the sample. Overall sample ages ranged from 20 to 45 years, with an average of 30.4 years. The ages of male subjects ranged from 20 to 45 years, with an average of 29.9 years, and the ages of female subjects ranged from 20 to 43 years, with an average of 30.9 years.

It was required that subjects have a minimum of 6 months experience in a paid supervisory capacity in which they assigned work and evaluated performance. Male subjects had an average of 3.5 years of such experience, and females subjects had an average of 3.2 years.

Independent Variables

Rater gender. Equal numbers of male and female raters served as raters during the study. Male and female subjects were randomly assigned to one of the four combinations of observed subordinate performance level and feedback requirement conditions appropriate to their gender at the time their qualifications for participation were verified.

Feedback requirement. Subjects who were randomly assigned to the feedback condition were verbally instructed that they would have to participate in a face-to-face performance appraisal session, during which they would present their performance appraisals to the rated subordinate. Subjects who were assigned to the no-feedback condition were verbally informed that their ratings of performance would not be transmitted to the rated subordinate.

Subordinate performance level. The experimenter manipulated subordinates' performances as moderately high or moderately low by returning to the subject prearranged, completed answer sheet sets consisting of 10 variations of the Clerical Speed and Accuracy sections of the Differential Aptitude Test (DAT; Bennett, Seashore, & Wesman, 1972). The pattern of subordinate performance reflected average scores at either the 25th or 75th percentile as defined by the test standards. Individual answer sheet scores were randomly varied from the 20th to 30th, and from the 70th to 80th percentiles, respectively.

Dependent Variables

Performance appraisal rating delay. Delay was defined as the number of 3-min work sessions that subjects observed before completing a performance appraisal. Subjects who waited until the last sessions to rate the subordinate were viewed as delaying their ratings more than those who made their ratings after seeing only a few sessions.

Performance appraisal feedback session scheduling delay. Considering only those subjects in the feedback condition, delay was measured as the number of days between the experimental session and the date subjects scheduled a feedback review session with the subordinate, but counting only those days in the week subjects reported that they would be available.

Performance appraisal ratings distortion. The distortion measure was constructed to closely parallel that of Fisher (1979) by using three separate scales that could be combined into one summary scale. The first of three graphic rating scales was a straight line scale consisting of three anchors (excellent, average, and extremely poor performance) that was converted to a 100-point scale using a metric template. The second consisted of a 100-point horizontal scale marked at 5-point intervals with the same anchors. The last was a modified Likert-type scale with numerical anchors within each response category, representing how well the supervisor felt the subordinate could perform as a programmer. The use of three items as the measure of ratings distortion also provided for an estimate of performance measure reliability. The three performance scales were highly correlated (all rs > .90) with a .99 computed KR-20 for the summary scale. The actual level of subordinate performance (either 25% or 75%) was subtracted from the summary scales, and the remainder was used as the measure of performance appraisal rating distortion.

Simulation and Control Factors

In order to enhance the validity of this laboratory study, a number of precautionary steps were taken to minimize the effects of nonperformance influences on the performance appraisal process.

Subordinate gender and ethnicity. The potential for an interaction between rater and subordinate gender, as well as between rater and subordinate ethnicity, was avoided by using the same White male confederate to play the role of the computer science student for this study. The confederate chosen for the duration of the study was a 26-year-old mass communications student who was paid for his participation on a per experimental session basis. An examination of the subjects' postexperimental responses supported the contention that the confederate's appearance, speech pattern, and personal demeanor had little or no influence on the subjects' ratings.

Rater-subordinate interpersonal bonding. Previous laboratory research has sometimes neglected the interpersonal bonding that occurs when supervisors interact with their subordinates. This lack of interpersonal experience with the subordinate, which occurs in actual organizational settings, could reduce the supervisor's reluctance to transmit negative performance information. The present study created an opportunity for the prospective rater and the subordinate to meet and to discuss their backgrounds, as well as to provide information about the consequences that the performance ratings had for the subordinate.

Procedure

The procedure consisted of the following steps. First, suitably qualified subjects were placed in a waiting room with the confederate, who attempted to establish rapport. Next, the subjects' supervisory responsibilities were explained, along with the nature of the task. Subjects then reviewed up to 10 sessions of performance data, completed their ratings, and made an appointment to deliver feedback (if in the feedback condition). Finally, they completed a postexperimental questionnaire and were debriefed. A detailed description can be obtained.

Results

The results are presented in the following sections: (a) manipulation and control checks, (b) performance appraisal rating delay results, and (c) performance appraisal feedback scheduling delay results.

Manipulation and Control Checks

Prior personal contact with subordinate. One factor that could have biased the results was the possibility of prior personal contact with the confederate. Responses to a question included in the postexperiment questionnaire to check for this possibility indicated that none of the subjects had any personal knowledge of the individual they supervised prior to their meeting at the experimental site.

Subordinate consequences. An important consideration of
Table 1
Analysis of Variance Summary Table: Performance Appraisal Rating Delay

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater gender (A)</td>
<td>1</td>
<td>12.25</td>
<td>35.64**</td>
<td>.04</td>
</tr>
<tr>
<td>Subordinate performance level (B)</td>
<td>1</td>
<td>225.00</td>
<td>654.63**</td>
<td>.74</td>
</tr>
<tr>
<td>Feedback (C)</td>
<td>1</td>
<td>36.00</td>
<td>104.74**</td>
<td>.12</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>1.56</td>
<td>4.55**</td>
<td>.01</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>5.06</td>
<td>14.73**</td>
<td>.02</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>0.56</td>
<td>1.64</td>
<td>—</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>2.25</td>
<td>6.55*</td>
<td>.01</td>
</tr>
<tr>
<td>Error</td>
<td>56</td>
<td>0.34</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*p < .05, **p < .001.

The study was whether the raters were aware of the consequences to the subordinate. Two questions were included on the post-evaluation questionnaire to assess this.

The first question asked subjects whether they perceived that the subordinate would be interested in or concerned about their performance ratings. Of the subjects, 96% responded that they felt the subordinate would be interested or concerned. The second question specifically asked subjects whether they felt the subordinate tried his best on the work materials. This time 92% of the subjects responded that they felt the confederate had tried his best. Those negative responses to the previous questions exhibited no relation to experimental conditions. An examination of the length of contact time during the initial bonding sessions also revealed no relation to the experimental conditions.

Observed subordinate performance level and feedback requirement potency. The literature on the performance appraisal process suggested that supervisors would experience discomfort both when they appraised the performance of moderately low performing subordinates and when they were required to provide feedback to subordinates. This discomfort was expected to be particularly intense for female raters who rated moderately low performing subordinates when feedback was required. In order to ascertain if this emotional discomfort was active during the present study, subjects were asked if they would be willing to serve as raters for the test evaluation project at some future session. Analyses of subject responses by chi-square tests support the contention that both dependent variables were particularly effective in creating emotional discomfort. For example, significantly more subjects reported that they would not be willing to serve again as raters (66%) when feedback was required than when no-feedback was required (31%), $\chi^2(1, N = 32) = 7.57, p < .01$.

One additional check is important to note. Examinations of the score distributions on the dependent variables suggested that score transformations would not be needed.

Performance Appraisal Rating Delay

The results from a $2 \times 2 \times 2$ (Male vs. Female Rater x Moderately High vs. Moderately Low Observed Subordinate Performance Level x Feedback vs. No-Feedback Requirement) analysis of variance (ANOVA), with performance appraisal rating delay as the dependent measure are presented in Table 1. When significant interactions were obtained, individual comparisons were made using $F$ tests (Keppel, 1982). Table 2 displays the means and standard deviations for performance appraisal rating delay for each of the eight experimental treatment combinations.

Main effects were obtained for each of the three experimental conditions on performance appraisal rating delay. Significant two-way interactions were obtained between the rater gender and observed subordinate performance level factors, as well as between the rater gender and feedback requirement factors. A significant three-way interaction was also obtained between the factors of rater gender, observed subordinate performance level, and feedback requirement.

As predicted, delays in the completion of performance ratings were greatest for female as opposed to male raters when feedback was required and when subordinate performance was moderately low. The magnitude of effect statistics reveal that subordinate performance level and the feedback requirement had substantial and moderate effects on delay, respectively. The effect of rater gender was slight.

The significant two-way interactions are attributable to the female raters. When faced with moderately low performing subordinates, women delayed disproportionately longer than did men. Likewise, when required to provide face-to-face feedback, women delayed disproportionately longer than did men.

The three-way interaction was also significant and its direction was exactly as predicted. As shown in Figure 1, female rats who were required to provide face-to-face feedback that was negative in content delayed disproportionately longer than did any other factor combination.

Performance Appraisal Feedback Session Scheduling Delay

The results from a $2 \times 2$ (Male vs. Female Rater x Moderately High vs. Moderately Low Observed Subordinate Performance Level) ANOVA, with performance appraisal feedback session scheduling delay as the dependent measure are presented in Table 3. Table 4 displays the means and standard deviations for scheduling delay for each of the four experimental conditions.

Table 2
Performance Appraisal Delay Cell Means and Standard Deviations

<table>
<thead>
<tr>
<th>Feedback requirement</th>
<th>Moderate high</th>
<th>Moderately low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>4.38</td>
<td>5.13</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.52</td>
<td>0.64</td>
</tr>
<tr>
<td>No feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>3.25</td>
<td>3.63</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.46</td>
<td>0.52</td>
</tr>
</tbody>
</table>

*Note. n = 8 for all cells.*
Main effects were obtained for each of the two experimental conditions on performance appraisal feedback session scheduling delay. A significant two-way interaction was obtained between the rater gender and observed subordinate performance level factors. As predicted, delays in the scheduling of performance appraisal feedback sessions were greatest for female as opposed to male raters and when subordinate performance was moderately low. The magnitude of effect statistics reveal that rater gender and subordinate performance level had moderate and substantial effects respectively on feedback scheduling delay.

Table 3
Analysis of Variance Summary Table: Performance Appraisal Feedback Session Scheduling Delay

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater gender (A)</td>
<td>1</td>
<td>357.78</td>
<td>30.48*</td>
<td>.20</td>
</tr>
<tr>
<td>Subordinate performance level (B)</td>
<td>1</td>
<td>913.78</td>
<td>77.86*</td>
<td>.53</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>87.78</td>
<td>7.48*</td>
<td>.05</td>
</tr>
<tr>
<td>Error</td>
<td>28</td>
<td>11.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .001$.

Table 4
Performance Appraisal Feedback Session Scheduling Delay Cell Means and Standard Deviation

<table>
<thead>
<tr>
<th>Performance</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderately high</td>
<td>3.00</td>
<td>6.38</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.27</td>
<td>2.13</td>
</tr>
<tr>
<td>Moderately low</td>
<td>10.38</td>
<td>20.38</td>
</tr>
<tr>
<td>$SD$</td>
<td>3.25</td>
<td>5.15</td>
</tr>
</tbody>
</table>

Note: $n = 8$ for all cells.

The significant two-way interaction is attributable to the female raters. When faced with providing face-to-face feedback to moderately low performing subordinates, women delayed scheduling the feedback sessions disproportionately longer than did men.

Performance Appraisal Ratings Distortion

The results from a $2 \times 2 \times 2$ (Male vs. Female Rater x Moderately High vs. Moderately Low Observed Subordinate Performance Level x Feedback vs. No-Feedback Requirement) ANOVA with performance appraisal rating distortion as the dependent measure are presented in Table 5. Table 6 displays the means and standard deviations for performance appraisal ratings distortion for each of the eight experimental conditions.

Main effects were obtained for each of the three experimental conditions on performance appraisal ratings distortion. A significant two-way interaction was obtained between the rater gender and observed subordinate performance-level factors. A significant three-way interaction was also obtained between the factors of rater gender, observed subordinate performance level, and feedback requirement.

As predicted, distortion in performance appraisals was greatest for female as opposed to male raters, when feedback was required, and when subordinate performance was moderately low. The magnitude of effect statistics reveal that rater gender and the feedback requirement had substantial and moderate

Table 5
Analysis of Variance Summary Table: Performance Appraisal Ratings Distortion

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater gender (A)</td>
<td>1</td>
<td>1482.25</td>
<td>578.44**</td>
<td>.62</td>
</tr>
<tr>
<td>Subordinate performance level (B)</td>
<td>1</td>
<td>10.56</td>
<td>4.12*</td>
<td>.003</td>
</tr>
<tr>
<td>Feedback requirement (C)</td>
<td>1</td>
<td>663.06</td>
<td>258.06**</td>
<td>.28</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>27.56</td>
<td>10.76**</td>
<td>.01</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>1.56</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>9.00</td>
<td>3.51</td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>56.25</td>
<td>21.95**</td>
<td>.02</td>
</tr>
<tr>
<td>Error</td>
<td>56</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .01$. 
effects on distortion, respectively. The effect of subordinate performance level was slight.

The significant two-way interaction is attributable to the female raters. When faced with moderately low performing subordinates, women distorted appraisals disproportionately more (in a positive direction) than did men.

The three-way interaction was also significant, and its direction was exactly as predicted. As shown in Figure 2, female raters who were required to provide face-to-face feedback to moderately low performing subordinates distorted appraisals disproportionately more (positive distortion) than for any other combination of factors.

Discussion

The results of this study support the contention that rater gender, level of subordinate performance, and feedback requirements affect the performance appraisals of subordinates. With respect to rater gender, women in this study delayed longer than did men both when making ratings of subordinate performance and when setting appointments to provide performance feedback. Female subjects also distorted performance ratings in a positive direction to a greater extent than did men. These outcomes may have resulted from lowered confidence when rating male subordinates (Mitchell & Liden, 1982). Alternatively, we conjecture that female raters may have exhibited more concern than did men with the accuracy of their ratings and the effect these ratings could have on their subordinates.

The pattern of interactions on all dependent variables is consistent with these propositions as well. Female raters’ tendencies to delay and distort ratings were exacerbated when subordinate performance was moderately low and face-to-face feedback was required. However, delay and distortion were not the sole province of women. Male raters also delayed and distorted performance ratings and delayed scheduling performance reviews.

These findings suggest that some action may be necessary on the part of employers to improve both their male and their female supervisors’ capacities to avoid distortion and deliver timely feedback across all subordinate performance levels. Future research should be directed at the further explication of the gender differences found in this study. In particular, the present

design should be extended to include female as well as male subordinates.

Generalizability of the Study Results

Given the steps taken to increase the external validity of the study, the generalizability would seem good. The subject pool included both men and women who possessed, on average, more than 3 years of supervisory experience. In addition, the stimulus materials for the coding task approximated an actual computer programming skill. Supervisors were given the opportunity (albeit brief) to interact with the subordinate and form a token interpersonal bond as well as to learn about the consequences for the subordinate. Raters possibly recognized that the subordinate’s employability was a concern and that their ratings might affect the subordinate’s career aspirations in the computer science field. Thus, it was assumed that raters would experience the concern or anxiety that often accompanies the rating process when ratings matter. Postexperimental questionnaire data and the accuracy of the predicted outcomes support this interpretation.

Despite these attempts to increase the generalizability of the

![Figure 2. Performance appraisal rating distortion: Interaction of rater gender, observed subordinate performance level, and feedback conditions.](image-url)
study, there are definite differences between the present study
and actual work organizations that affect its external validity.
The supervisors in the study did not have to consider any pro-
longed relationship with the rated subordinate, and they were
not subjected to the influences that could occur when observing
the subordinate's interpersonal relations with his peer group or
others on the job. The actual performance evaluation process
also did not contain such ramifications as performance ratings
affecting the promotions or job security of subordinates. Simi-
larly, there was no impact on the supervisors' job goals or any
consideration of the management philosophy of an actual cor-
porate environment. In addition, although the same White man
was observed by all supervisors in the study, individual differ-
ences among subordinates may affect supervisory ratings in the
workplace. The performance information of the study may also
have jeopardized the generalizability of the study by artificially
holding the moderately low and moderately high levels of ob-
erved performance relatively constant. Because the entire
range of possible subordinate performances was not included,
there may have been differences at the average performance
levels.

One of the major threats to the internal validity of the study
was the possibility that subjects would discover the true aim of
the test development project deception. None of the subjects
reported that they thought the project was an experiment, and
the majority of subjects exhibited surprise when they were in-
formed at the end of the experimental session that they would
not be required to meet with the subordinate. Another major
threat to internal validity was the possibility that subjects would
learn about the experiment from other students. Three factors
may have minimized this possibility. First, any knowledge of
the performance task may have served as reinforcement that the
experiment was an actual test development project. Second, the
subjects for the study were drawn from a number of classes at
three different educational institutions. This reduced the pos-
\neli bility that discussions about the project would take place.
Third, the data collection for the study was accomplished in
less than 3 weeks, a relatively short time for many experiments,
reducing the total exposure time that subjects had to learn
about the actual purpose of the study.

Conclusions

Contrary to earlier findings (Fisher, 1979), moderately low
performers were rated later than were moderately high per-
formers. In addition, we found evidence for positive distortion
in the ratings of moderately high performers as well as moder-
ately low performers, although the amount of distortion was
considerably larger for moderately low performers. The differ-
ence in outcomes between this study and Fisher's may perhaps
be attributed to the higher degree of fidelity we achieved in our
attempt to simulate organizational performance appraisal pro-
cesses. Although other possible explanations for distortion and
delay cannot be unequivocally ruled out—explanations that
might include attributional processes—we view the distortion
delay found in this study as being due to discomfort or anxi-
ety associated with the appraiser's role that seems to be an im-
port ant element in actual performance appraisals.

In future studies, investigators should attempt to unearth fac-
tors that would effectively alleviate these tendencies to delay and
distort ratings. We would suggest that future studies maintain
the distinction between delay in actually completing perfor-
mance ratings and delay in providing feedback—a distinction
that has heretofore been ignored.

On the basis of our results, delay and distortion may be ex-
pected to occur when face-to-face feedback is to be provided
to subordinates by the supervisor who completes the ratings.
Organizations may avoid the problems that accompany delayed
and distorted performance appraisals by either of two potential
strategies. First, supervisors could and should be thoroughly
trained to carry out their performance appraisal and feedback
responsibilities, with particular emphasis directed at making
them aware of the potential for delays and distortion. Behav-
orial modeling training may be a promising method for ac-
complishing this (Goldstein & Sorcher, 1974). Second, sup-
\versisors could be relieved of the responsibility for rating their im-
mediate subordinates. Although resistance to this alternative
will in all probability be great, appraisals could be completed
by suitably constituted committees, thereby alleviating some of
the personal responsibility attributed to the supervisor by sub-
ordinates for the performance appraisal process.

As an organizational control mechanism, performance app-
raisal may fail when avoidance, delay, and distortion operate.
This study has shown that such phenomena as delay and distor-
tion do occur and that they affect the appraisal process most
under conditions in which performance is low and face-to-face
feedback is necessary. These are precisely the conditions un-
der which performance appraisal is intended to function best as a
remedial tool. It is clear that future research and improve-
ments in appraisal practices are needed to overcome this di-
lemma.

References

Aptitude Test: Clerical speed and accuracy (Pt I). San Antonio, TX: Psy-
chological Corporation.
Bignoness, W. J. (1976). Effect of applicant's sex, race, and performance
on employer's performance ratings: Some additional findings. Journal
of Applied Psychology, 61, 80–84.
Brown, L. K. (1981). The woman manager in the United States. Wash-
ington, DC: Business & Professional Women's Foundation.
Cummings, L. L., & Schwab, D. P. (1973). Performance In organiza-
tions: Determinants and appraisal. Glenview, IL: Scott, Foresman.
Fisher, C. (1979). Transmission of positive and negative feedback to sub-
ordinates: A laboratory investigation. Journal of Applied Psychology,
64, 533–540.
Goldberg, P. (1968). Are women prejudiced against women? Transac-
tion, 5, 28–30.
New York: Pergamon Press.
Grunfeld, I. W., & Weissenberg, P. (1966). Supervisor characteristics and
attitudes towards performance appraisals. Personnel Psychology,
19, 143–151.
and females to work situations which evoke sex-role stereotypes.
Journal of Vocational Behavior, 14, 23–32.
Haeftner, J. E. (1977). Race, age, sex, and competence as factors in em-


Received May 20, 1985
Revision received January 4, 1988
Accepted January 5, 1988