Accountability Forces in Performance Appraisal

RICHARD KLI MOSKI AND LAWRENCE INKS

Ohio State University

This experiment investigated the effects of two factors felt to influence the quality of ratings: anticipated feedback sharing and knowledge of subordinates' self-assessment. One hundred and eighty subjects receiving either favorable, unfavorable, or no subordinate self-assessment information were led to anticipate either face-to-face, written, or no feedback sharing with a subordinate. Main effects were found both for type of feedback sharing and level of subordinate self-assessment. Partial support was also found for the interaction of feedback sharing and partial self-assessment effects. Supervisors anticipating sharing face-to-face feedback with a subordinate rated the latter's performance significantly more positively than did supervisors who received no self-assessment data; while supervisors receiving knowledge of an unfavorable self-assessment rated their subordinates significantly more negatively than those receiving no self-assessment information. Ancillary analyses support the contention that the impact of knowledge of self-assessment information is largely motivational, as opposed to informational, in nature. Despite their potential to influence ratings, it is suggested that face-to-face feedback sharing requirements and the use of subordinate self-assessment data are not necessarily detrimental, but rather that care should be taken to minimize their potential to reduce rating quality.

Performance appraisal is assumed to be a primary mechanism through which the level of individual performance is improved within an organization (Bernardin & Beatty, 1984; Wexley & Klimoski, 1984). In order to do this, it is essential that these ratings be accurate and unbiased. However, there is a significant body of literature which supports the proposition that the accuracy of a performance evaluation can be influenced by a large number of different factors, bringing about poor rating quality and the manifestation of such symptoms as leniency/severity and halo.

A subset of these factors is felt to operate on the motivation to rate accurately (Banks & Murphy, 1985; Wexley & Klimoski, 1984). Specifically, having to provide feedback can produce such an effect. At the most basic level, past research shows that people try to avoid sending negative or unpleasant messages to others (Blumberg, 1972; Tesser & Rosen, 1975). Taken in an organizational context, Stone (1973) found that while approximately 80% of managers and professionals polled enjoyed giving positive evaluations, the vast majority of them disliked giving negative evaluations. In addition, Stone also observed that ratees generally find receiving evaluations to be difficult only if the evaluation is negative. Positive evaluations seem to be accepted quite easily by the ratee. Thus, in spite of the obligations inherent in the supervisor's role to attempt to shape or improve the performance of subordinates and the existence of some reasonably powerful forces to do so (see Katz & Kahn, 1978; Feild & Holley, 1975), this often does not occur.

The reluctance of a rater to transmit negative feedback can impact on the rating and feedback process in several ways. The rater may, for example, refuse to give feedback at all. Similarly, the rater may delay the transmission of negative feedback for as long as possible. He or she might also choose to transmit very specific feedback to provide justification for the low ratings. Finally, the feedback may itself be distorted in such a way as to be less disturbing to the subordinate (see Larson, 1984 for a review).

An empirical test of the hypothesis that assessments made in anticipation of having to provide feedback will be distorted by raters was conducted by Fisher (1979). Her study revealed that performance ratings by supervisors who were to feed back their ratings to subordinates were significantly more positive than ratings that were not to be fed back when subordinate performance was poor. When subordinate performance was good, no such distortion effects were found. Ilgen and Knowlton (1980) replicated Fisher's findings with respect to distortion of feedback to poor performers. However, they do caution that they could not separate the effects of having to give feedback from the effects of anticipated face-to-face interaction with the ratee per se. Fisher (1979) also seemingly did not address this possible confound.

Weary (1979) maintains that distorting ratings in anticipation of feedback may be viewed as "a response bias or strategic self-presentation designed to maximize public esteem" (p. 1419). This idea of "self-serving biases" has been more carefully examined in a study by Mitchell and Klimoski (1984). This susceptibility to the expectations or wishes of others (in this case the subordinate) is defined in the Mitchell and Klimoski study as "accountability," and their research supports the contention that accountability forces do produce distortion in performance evaluations of poor performers.

While feeding back ratings almost certainly promotes a considerable amount of supervisor accountability to the subordinate (Mitchell & Klimoski, 1984), it is important to determine how this comes about. Is it the case that a performance rating made for any feedback purposes will promote accountability, or is it the anticipation of face-to-face contact with the subordinate that is important?
A review of related literature in social psychology (Darley & Berscheid, 1967; Bond, 1979; Bond & Dutton, 1973; Shapiro, 1975) suggests that anticipated face-to-face contact should make a difference. For example, Shapiro (1975) demonstrated that expected future interaction caused subjects to divide a reward equally rather than based on contributions to a joint effort (equitably). Work in decision making is also suggestive. Beach and Mitchell (1978) view accountability as strongly affected by the individual’s belief that he or she will have to share a decision’s results with others who have a vested interest in such outcomes (e.g., a subordinate). They demonstrated that accountability can have effects on the quality of decisions (McAllister, Mitchell, & Beach, 1979). Actually, it appears that having to justify a decision to others affects both the decision process and decision outcomes (Ford & Weldon, 1981; Tetlock, 1983a, 1983b, 1985a, 1985b).

It is interesting to note that in several studies accountability was found to improve the care with which decisions were made and the complexity with which information was processed. For example, Ford and Weldon (1981) report that their subjects in a high-accountability condition took longer to make a decision than low-accountability subjects. Rozelle and Baxter (1981) led subjects to believe that they would have to participate in a postdecision discussion with knowledgeable others. This caused subjects to produce higher quality reports when compared to control conditions.

However, other research indicates that accountability forces can also detract from quality. Tetlock (1983b) had subjects in his study justify their decisions to others with known or unknown views on the decision topic area. His results show that when a subject was accountable to an individual with known views, he or she shifted opinions toward the direction of the other person. He coined the term “preemptive self-criticism” for this tendency to anticipate or prepare for a justification session and to accommodate the other party’s perspective.

While the notion of accountability implies personal responsibility for decisions and judgments, it clearly rests heavily on the dynamics of the social context in which the decision or judgment must be made. It is not just that the individual or rater is charged with the task. The decision outcomes usually impact on and therefore have relevance to others. And we appear quite sensitive to this. In fact, efforts to separate the forces stemming just from the obligation (responsibility) for making a decision and those created by the requirement to justify a decision have often not been successful (Whipple, 1988). All this would lead us to believe that having to present a decision to someone else should have greater impact on subjects than merely having to make it.

Accountability has been examined in only a limited way in a performance appraisal context. However, applied to performance feedback situations, the available literature suggests that situations in which a supervisor anticipates face-to-face feedback sharing with a subordinate will promote even greater accountability to that subordinate (and therefore the potential for greater rating distortion) than situations in which performance feedback is not to be transmitted in a face-to-face meeting, although the latter should still promote some distortion. Moreover, according to accountability theory (Mitchell & Klimoski, 1984) and consistent with the findings of Tetlock (1983b), the direction of the distortion itself should be toward what the rater perceives the ratee’s wishes or expectations to be. This, we would argue, in most cases, will be in an upward (more positive) direction.

Another potential source of bias involves knowledge of a subordinate’s self-assessment of his or her performance. More specifically, knowledge of the nature of a subordinate’s task-relevant self-image is likely to operate on a rater in two different ways. First, in the case where raters are relatively unfamiliar with the subordinate’s task, the self-assessments might act as an informational anchor for the raters’ assessments of performance. Second, knowledge of a subordinate’s self-assessment is likely to promote anticipation of particular reactions of that subordinate to performance feedback. In the case of the poor performer, the effects of this knowledge are problematic. Thus, when the supervisor is aware that a subordinate’s self-assessment is high (but, in fact, unwarranted), we would expect a fair amount of apprehension on the part of the former, in essence, he or she will have to convey an assessment that is neither desired nor expected.

One way to minimize the potential for conflict would be to modify the evaluation that is to be given in the direction of a self-assessment. This should result in positive distortion in supervisor ratings of subordinate performance in most cases. However, we would expect little, if any, distortion in ratings when the supervisor is aware that the subordinate’s self-assessment is actually low (and ostensibly more accurate). In fact, under these circumstances we might predict greater candor and accuracy on the part of the supervisor.

In light of the notions of accountability, the principal thrust of the present research is to investigate the relative effects of anticipated face-to-face (versus anonymously written) feedback sharing under different conditions of supervisor knowledge of a subordinate’s self-assessment of performance.

By way of summary, this study examines the impact of two factors thought to affect the accuracy of ratings through their impact on rater accountability and motivation. Based on the limited research evidence available, it is hypothesized that accountability forces will be stronger
when raters expect face-to-face feedback sharing with their subordinates than when either anonymous feedback or no feedback is expected to be given. The nature of the accountability forces produced by the feedback sharing requirement are in this case expected to be primarily motivational in nature.

In addition, it is expected that ratings in high (inflated) self-assessment conditions will be more positive than ratings in either the no self-assessment or low self-assessment conditions. In the present experimental context, these ratings are likely to reflect the use of self-assessments as a source of information as well as a motivational force.

Finally, the effects of anticipated feedback sharing and self-assessment knowledge are expected to interact with one another.

Based on the limited research available, when no feedback sharing requirement is anticipated, there should be relatively little motivational impact on ratings. Instead, feedback providers are expected to rely on the self-assessment information for additional guidance in their judgments of performance. In contrast, it is expected that ratings made when face-to-face feedback sharing is anticipated will be influenced by motivational factors. Under these circumstances, subjects in both the high self-assessment and no self-assessment conditions are expected to produce inflated ratings. Moreover, we would expect little or no distortion for those in a face-to-face feedback condition who anticipate meeting with a low self-assessment recipient. Here, there is little incentive not to be candid (and hence more accurate).

METHOD

Subjects

Subjects were 180 individuals from a pool of introductory psychology students. To help reduce any possible rater bias effects due to gender, all the subjects were male. In addition, they were told that they were evaluating the performance of other males.

Procedure and Design

When each subject arrived at the laboratory, he was told that he was going to take part in a study examining the dynamics and effectiveness of various types of work simulations for research. Each subject was asked to evaluate the performance of another subject working on a clerical task. All were told that the latter had already started on his task. In actuality, there was no other subject; the clerical work that the subjects evaluated was programmed. All work to be evaluated reflected moderately poor performance.

The scenario involved was similar to the one used by Ilgen, Mitchell, and Frederickson (1981). The subjects (acting as supervisors) were told that each of their respective subordinates had been given 40 five-digit code numbers which represented pieces of merchandise in a catalog. The subordinate’s task was then to look up each of these code numbers in the catalog, and write down the catalog number and suggested retail price (both of which would be listed) for that specific piece of merchandise. In addition, the subordinate was to calculate (using a calculator) a 15%-off sale price for each piece of merchandise, and then write down that sale price as well.

The subjects (as supervisors) were told to mark as incorrect any response on the subordinate answer sheet which did not correspond to the answers listed on the grading key sheet. After determining how many incorrect answers their individual subordinate gave, each supervisor was to evaluate him using the given rating form. To familiarize the supervisors with their grading task, each was given a short practice session. Participants were randomly assigned to one of the experimental conditions.

The independent variables in the study included anticipated feedback sharing and knowledge of subordinate self-assessment. There were three levels of feedback sharing. Subjects in the written feedback sharing conditions were told that while they would not meet face-to-face to discuss their performance ratings with their subordinate, but that their written performance ratings would still be transmitted to the subordinates by the experimenters. Subjects in the face-to-face feedback sharing conditions were told that after completion of the performance ratings, they would meet with their subordinates individually and feed back their ratings in a face-to-face manner. Subjects in the no-feedback conditions were told that their ratings would not be seen by their subordinate, only by the experimenters.

There were also three conditions of self-assessment knowledge. Subjects receiving self-assessment information were told that this was being done in an attempt to simulate as many organizational forces as possible in the given setting. Subjects in the high self-assessment conditions were given a self-assessment rating form (similar to the rating form which was to be completed by the supervisor) purportedly filled out by the subordinate in a manner reflecting high performance, while subjects in the low self-assessment conditions received a completed self-assessment that described low performance. Subjects in no self-assessment conditions were neither told about nor given any self-assessment information.

Each subject was given the 40 code-number exercises of his “subordinate” to score using the master answer key. Upon completion of this, he was given self-assessment information (if appropriate), a short questionnaire to fill out, and the performance evaluation instrument itself. Ratings on performance were obtained in this way and constitute the
prime dependent variable of the study. After subjects completed the performance evaluation form, they were debriefed, pledged to secrecy, and thanked for participating.

**RESULTS**

**Manipulation Checks**

All subjects evaluated the same "subordinate" performance, and therefore were expected to see that performance as poor. Consistent with this notion, the average response to the manipulation check item "This is an example of good performance" indicated that subordinate performance was seen by all subjects as moderately poor ($M = 3.71/7.00$; all manipulation check ratings on a 7-point scale). In addition, there were no significant differences in the number of errors detected across the experimental conditions, indicating a consistent scoring pattern by subjects across conditions.

It was felt that attributions of the causes of performance could have an effect on ratings. To control for this, the task had been structured so that subordinate performance would be seen as due more to effort than to ability. Consistent with this, subjects saw performance as due more to effort ($M = 5.50$) than to ability ($M = 3.51$).

The feedback sharing manipulations appear to have been convincing. There was a significant main effect for feedback sharing, $F(1, 99) = 276.49$, $p < .001$, such that subjects in the face-to-face feedback sharing conditions ($M = 5.71$) expected to meet with their subordinates more so than did subjects in written feedback ($M = 0.25$) and no feedback ($M = 1.41$) conditions.

Consistent with accountability notions, it was hypothesized that self-presentation (motivationally-based) concerns should be strongest and distortion greatest when a supervisor expects to feed back ratings to a subordinate, especially when the feedback is to be presented in a face-to-face manner. Conversely, these self-presentation concerns should be weakest when the supervisor does not expect ratings to be fed back to the subordinate at all. A significant main effect was found for degree of feedback sharing, $F(2, 171) = 8.70$, $p < .001$, $\omega^2 = .06$ (see Table 1). Contrary to prediction, however, Dunn post-hoc tests revealed that supervisor ratings were not significantly different between conditions of anticipated face-to-face and anonymous feedback sharing, nor were they significantly different between conditions of no anticipated feedback and anonymous feedback sharing. But there were significant differences in these effects between conditions of anticipated face-to-face feedback sharing and no anticipated feedback sharing ($p < .05$) (see Table 1). That is, performance ratings of supervisors who expected to feed back their ratings in a face-

<table>
<thead>
<tr>
<th>Table 1: Mean Performance Rating across Experimental Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of feedback sharing</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>Anonymous</td>
</tr>
<tr>
<td>Face-to-face</td>
</tr>
<tr>
<td>(Marginal means)</td>
</tr>
</tbody>
</table>


to-face manner rated their subordinates significantly more positively than did supervisors who did not expect their ratings to be seen by the subordinates.

It was proposed that subjects with knowledge of a subordinate’s high self-assessment would evaluate him more positively than those receiving information that a low self-assessment was held by their subordinates. Ratings from subjects with no access to self-assessment information would fall between these two levels. A main effect for self-assessment was found on the performance ratings, $F(2, 171) = 33.83$, $p < .001$, $\omega^2 = .26$ (see Table 2). A Dunn post-hoc test was performed in order to determine the nature of the difference(s) between ratings made across the three conditions. Ratings in which high self-assessment information was available were the most positive, and ratings made when low self-assessment information was available were the most negative. In addition, ratings made when no self-assessment information was available fell in-between those made in the other two conditions (see Table 1).

This finding lends some support to the accountability notion discussed earlier, which suggests that accountability forces can promote both positive and negative distortion in ratings, depending on the nature of the presumed expectations of the agent to whom one is responsible. Supervisors who receive either a positive or negative subordinate self-

<table>
<thead>
<tr>
<th>Table 2: Analysis of Variance for Dependent Measure (Rating)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>Feedback sharing</td>
</tr>
<tr>
<td>Self-assessment</td>
</tr>
</tbody>
</table>

* Medium effect.  ** Large effect.
assessment should feel self-presentation concerns that motivate them to rate these subordinates in a manner consistent with the nature of the subordinate's expectations or wishes implied by the self-assessment. Because this effect did not occur to a greater extent in the context of having to personally provide feedback (see next section) it seems that knowledge of such self-assessments have an informational influence basis as well. Thus, it had a pervasive effect in this study.

In addition to the simple main effects described above, predictions were made regarding an interaction between the effects of anticipated feedback sharing and knowledge of self-assessment. This interaction did not prove to be significant at the .05 level. Possible explanations for this finding will be presented later.

Ancillary Analyses

Ancillary analyses were also performed to examine subjects' perceptions and feelings associated with their role as evaluators of performance. It was expected that subjects would show less confidence in their ratings and greater anxiety in the feedback sharing conditions than in conditions in which no feedback sharing was anticipated. The means, however, were not significantly different across conditions. All means were toward the low end of the scale for anxiety and rating difficulty and toward the high end for confidence. To the extent that the questionnaire items were valid indicators of these factors, it seems likely that other forces were operating to produce felt accountability of the subjects.

DISCUSSION

Results of the study support the notion that accountability forces operating in performance appraisal contexts can have significant effects on ratings. It was found that when subjects anticipated face-to-face feedback sharing with a poor performing subordinate, they rated the latter in a significantly more positive manner than did both supervisors who expected only to share written feedback anonymously and those who did not anticipate feedback sharing at all. Besides partially replicating past findings (Fisher, 1979; Ilgen & Knowlton, 1980), it also stresses the importance of social forces and social pressure implied by future face-to-face interaction. These findings suggest that accountability may act to promote distortion in performance ratings in two ways: by purpose, whereby the amount of rating distortion is greater when evaluations for feedback purposes are to be face-to-face as opposed to written, and by direction, whereby the nature of distortion will be related to the expectations held by the party to whom you are to give feedback.

The analyses thus lent support to the potential biasing effects of subordinate self-assessment information. Knowledge of a subordinate's self-assessment caused subjects to distort performance ratings in a direction consistent with the nature of the self-assessment itself. It was found that subjects receiving high subordinate self-assessments rated the effectiveness of their subordinates significantly more positively than those receiving low self-assessments. In addition, when no self-assessment information was available, subjects rated the performance of their subordinates less positively than those receiving high self-assessment, but less negatively than others receiving low subordinate self-assessments. This suggests that the accountability construct might be useful to predict not only the existence but also the direction of distortion in ratings. In addition, findings supported the contention that the accountability forces operating in these situations are primarily motivational, as opposed to informational, in nature.

This use of self-assessment in rating contexts is not uncommon in organizations. Recent surveys have revealed its use in 5 to 14% of the performance appraisal programs sampled (Lazar & Wikstrom, 1977; Locker & Teel, 1977). Moreover, this use is expected to increase (Ferris, Yates, Gillmore, & Rowland, 1985). Thus, the finding that self-assessment information can be potentially biasing (especially in the direction of the self-assessment itself) has many implications.

One could argue that there will be no detrimental effects on rating accuracy as long as the self-assessment information is itself accurate, and this reasoning is probably correct. In fact, however, people often have a great deal of trouble evaluating themselves in an accurate manner (Meyer, 1978), especially when the consequences are serious. In many contexts it seems unlikely that the employees would be candid and, if warranted, rate their performance as below average. This would be especially true if they have purposely distorted their self-assessments in a positive direction in the past and have noticed a favorable effect upon the rater. It would appear then that self-assessment information should be gathered and shared with great care and perhaps under certain conditions. Fortunately, there appear to be some circumstances under which relatively accurate self-assessments can be obtained (Mabe & West, 1982; Ash, 1980). In particular, the role of the supervisor as a "coach" in some situations suggests that ratings in a developmental context might prompt more honesty (and more candid feedback) than in a performance appraisal context.

An interesting issue relates to the role of attributions or explanations for poor performance as they directly or indirectly influence ratings. The literature suggests that attributions of lack of effort result in greater willingness of raters to record a low assessment (Ilgen & Knowlton, 1980), based on the assumption that effort is felt to be under the control (discretion) of the worker. Thus, he or she should get what is deserved. Moreover, Larson (1984) makes attributions a key factor in affecting the
behavior of those who are to provide feedback. It is felt that feedback to
individuals performing poorly for reasons beyond their control would be
different than when this performance is clearly due to a lack of effort.

In the present study, the cover story and task were designed to promote
effort attributions (rather than ability) in raters’ perceptions of the causes
of poor performance. This was done to increase the tendency toward
candor (valid and accurate assessments) in feedback. Ancillary data
indicated that this was done successfully. However, despite this, distortion
in ratings was observed. It might be expected, then, that where poor
performance is felt to be due to lack of ability, it might be even greater.
Raters might reasonably assume that clear and accurate (but in this case
negative) feedback would be resented.

As mentioned earlier, an interaction between the two factors manipu-
lated in the study was specified. However, support for this interaction
was not found. Looking at the pattern of means, when no feedback shar-
ing was anticipated, ratings made in high self-assessment conditions were
more positive than ratings made when no self-assessment information was
present. Similarly, ratings were more negative in low self-assessment
conditions than the no information baseline. This pattern of results in the
no-feedback condition is consistent with the general main effect found
earlier for the self-assessment factor, and most likely reflects the largely
informational, rather than motivational, forces operating to influence the
ratings made under these circumstances.

Contrary to prediction, under conditions of anticipated face-to-face
feedback sharing, ratings made in high self-assessment conditions were
only moderately more positive, than ratings in the baseline condition.
While ratings in low self-assessment conditions were lower. The combi-
nation of motivational forces operating when face-to-face feedback shar-
ing was anticipated and a high self-assessment provided were expected to
inflate ratings in this condition to a much greater extent than in any of the
others.

There are several possible reasons for the unanticipated pattern of re-
results. One possible explanation for this deals with the rater’s perception
of what rating was actually desired by subordinates. It could be the case
that, in the absence of any self-assessment data (but anticipating a feed-
back interview) raters in the baseline group were inclined to inflate ratings
more than was expected. If this were indeed the case, mean ratings made
when no self-assessment information was available would tend to more
closely resemble those made when the self-assessments were high, and a
significant difference between these two conditions would not occur. It is
important to recognize, however, that this possibility applies primarily to
ratings made when face-to-face feedback was anticipated. As noted ear-
lier, motivational forces are much more likely to be operating in these
conditions than when feedback is not expected to be given at all.

A second possibility stems from a related line of research which inves-
tigates feedback-sharing behavior in assessment contexts. It appears that
while people who expect to do well on tasks also welcome diagnostic
feedback and in some cases seek it out, those who do not expect to
perform well are much less likely to seek feedback. Instead the latter
show a tendency to avoid and even resist such diagnostic information as
this information will no doubt further threaten self-esteem (Trope, 1979;
Sach, 1982). Feedback providers in this study might justifiably think that
even well-intentioned honesty would create tension (even animosity) be-
tween them and the recipient and be unlikely to improve performance.
The motivational forces produced by accountability might thus com-
pletely overwhelm the effects of knowledge of subordinates’ self-as-
seassments. However, the relatively low levels of anxiety subjects re-
ported makes this dynamic seem less plausible.

A final explanation for the mean ratings found when face-to-face feed-
back was anticipated deals with a possible ceiling effect in the ratings
themselves. It could be the case that subjects perceiving a high self-
assessment from their “subordinate” were simply unwilling to give a
rating that was consistent with what the latter might desire or expect.
Actual performance levels, after all, were scripted to be quite low. Raters
might have been unable to justify giving a highly positive rating to their
subordinate, even when the alternative was potentially unpleasant.
Again, if this were indeed the case we would not expect a significant
difference between ratings made when high self-assessments and no self-
assessments were provided. Clearly, more work needs to be done here.

The low rating provided by subjects in the face-to-face feedback con-
ditions and who also received low self-assessments does speak to certain
features of accountability dynamics. In retrospect, it appears consistent
with some of the ideas on accountability offered by Mitchell and Klimoski
(1984). That is to say, while the notion that people prefer and seek out
positive feedback will generally hold (Larson, 1984) under certain cir-
cumstances, the feedback given may be “allowed,” even expected, to be
candid in the case of a low evaluation. It may be that receiving an ac-
nowledged low self-assessment of performance from the person to
whom you are to give feedback is one of these. Thus, subjects in the
face-to-face feedback condition in this study were accountable to some-
one who was perceived to be candid and, (we are speculating) desire
c candor in return. If this effect were to be replicated, it would indeed
strengthen the notion that accountability can produce forces in the direc-
tion of promoting greater rather than lesser accuracy.
As with any laboratory study, the present research must limit its claims to generalizability. While the effect sizes for anticipated feedback sharing and self-assessment knowledge were of a reasonable order of magnitude in the present study, the reader should be cautioned about drawing population conclusions from fixed effort designs. Nonetheless, it is our feeling that many of the forces that promote both accuracy and distortion in performance ratings were not present in the experimental procedure. Moreover, factors such as a rater's constant and continuing interaction with a rater, the feedback sharing requirements and self-assessment knowledge are likely to be even greater in an organizational setting than in the laboratory (Ilgen & Favero, 1985). This should be especially true in the face-to-face feedback session (something that was not examined in this study). Thus, one would expect accountability forces to be even more pronounced in work settings. Future research, then, should be directed at measuring the strength of these effects in organizations, as well as the extent to which they promote distortion (or accuracy) in ratings of various types (ratings of performance vs ratings of potential, etc.), as well as in actual feedback. For example, while subjective, qualitative ratings are likely to show the effects of these accountability forces, more objective and quantitative rating criteria (such as sales figures) might inhibit accountability forces because they are verifiable.

Finally, the findings from the present study have several implications for performance-evaluation practices. They suggest that there are procedural and situational forces operating which serve to systematically promote distortion in ratings. This is not to say that these procedures should necessarily be abandoned; findings from the present study suggest instead that care be taken to balance any potentially biasing effects of these procedures.

REFERENCES


**Received: February 24, 1987**

---

**Task Complexity Effects on Non-task-related Movements: A Test of Activation Theory**

**DONALD G. GARDNER**

*University of Colorado at Colorado Springs*

The present experiment tested the activation theory-based predictions that variation in task complexity affects experienced activation level, which in turn affects performance, satisfaction, and the number of non-task-related movements made by task performers. Subjects performed both a low-complexity and a moderate-complexity task. It was predicted that subjects would make more non-task-related movements on the low-complexity task than on the moderate-complexity task, to increase experienced activation levels that were depressed by performance of the low-complexity task. It was also hypothesized that there would be inverted-U relationships between experienced activation level and number of non-task-related movements. Finally, it was predicted that the number of non-task-related movements would be inversely related to performance. Results provided modest support for the hypotheses. Implications for theory and work design are discussed. © 1990 Academic Press, Inc.

It has long been known that variation in task complexity has effects on affective, cognitive, physiological, and behavioral responses of task performers (e.g., Simonson & Weiser, 1976). For the purposes of this study, task complexity is defined as the degree to which a task provides a variety of stimulation to the task performer, in terms of number of distinguishable and dissimilar elements present in the task-based stimulation, as well as the degree to which the stimulation is variable or novel (cf. Berlyne, 1960; Scott & Erskine, 1980). There are many different explanations for the effects of task complexity on task performer responses (cf. Steers & Mowday, 1977). Activation theory represents one such explanation for these effects (cf. Fiske & Maddi, 1961; Gardner & Cummings, 1988; and Scott 1966, 1967, for extensive discussions).

Yet, only in recent years have activation theory predictions about effects of task complexity on task performers been rigorously examined in applied research (e.g., Gardner, 1986a; Huber, 1985). The purpose of the present study is to broaden this line of activation theory research by

The research reported here was funded by the University of Colorado—Colorado Springs Committee for Research and Creative Work. Send requests for reprints to Donald G. Gardner, College of Business and Administration, University of Colorado at Colorado Springs, Box 7150, Colorado Springs, CO 80933-7150.