

Rater Characteristics Affecting the Validity of Performance Appraisals

Arthur G. Bedeian, *Auburn University*

The task of developing effective performance appraisal systems is one of the most pressing contemporary problems of personnel administration. An abundance of literature is available detailing the problems and difficulties inherent in subordinate appraisals. [29] Numerous studies have made suggestions for rating format and content changes. [2] Some have even suggested the elimination of appraisal. [17] More recently, an identifiable body of knowledge which seriously casts doubt on the use of supervisors' judgment in evaluating employee performance has begun to emerge. The purpose of this paper is to explore this emerging body of knowledge and to examine its ramifications for performance appraisal.

Influences on the Ability to Judge

The answer to the question of whether or not supervisors' ratings actually measure the performance they are meant to is extremely important. It is a core concern for those interested in accurately evaluating performance and certainly of great importance to all those being judged. While it has long been recognized that people differ in their standards of judgment, [8, 9] it is not generally recognized that people differ in their ability to judge. Gordon W. Allport suggests that the ability to judge people is much like artistic ability — it is not entirely widespread nor entirely specific to particular individuals. He writes: "It would be unreasonable . . . to expect a judge of people to be uniformly successful in estimating every quality of every person It seems more of an error, however, to consider the ability entirely specific than to consider it entirely general." [1, p. 512]

In a review of studies on the relationship between various factors and the ability to judge people, psychologist Robert Taft [35] formed numerous conclusions that bear on the issue being examined. He concluded, for example, that a positive relationship exists between intelligence and the ability to judge and that the ability to judge increases with emotional adjustment. Significantly related to this point is the fact that poor judges are more likely to be poorly adjusted and are more likely to be affected by personal biases in their judgments.

Thus, it seems that Taft's conclusions raise some very serious questions pertinent to accurate performance appraisal. In postulating a positive association between intelligence, emotional adjustment and the ability to judge others, Taft has touched upon a seldom researched area. Regarding the relationship between intelligence and the ability to judge, there seems to be little direct industrial data available. An early study by Stockford and Bissell [34] found a correlation of .33 between the intelligence of 30 department heads and the reliability of the ratings they gave their subordinates. They also found that in assessing subordinate performance those department heads lower in intelligence exhibited a tendency to be more influenced by rater popularity than department heads of higher intelligence.

Related to this point, Sales has expressed the belief "that democratically-oriented supervisors can be expected to be more intelligent than are autocratically-oriented supervisors." [32, p. 283] He bases this contention on the fact that intelligence scores usually correlate negatively with measured (F-scale) authoritarianism [38] which in turn generally correlates highly with authoritarian behavior. [28] Thus, a combination of several research findings encourages the hypothesis that democratically-oriented supervisors are both more intelligent and better judges of people than their autocratically-oriented colleagues.

In considering the relationship between emotional adjustment and the ability to judge people, several additional significant relationships emerge. As Massarik and Wechsler [27] have observed, a person's emotional adjustment is closely associated with his self-concept, i.e., "how he sees himself and how he feels about himself." They further note that a person's self-concept establishes his "psychological state of mind" which affects all those with whom he may come in contact — e.g., family, friends, superiors, and subordinates. Further, an individual who has satisfactorily attained a level of self-acceptance is better able to develop what is commonly termed a "healthy personality." In relating personality dynamics and the ability to judge to the topic of evaluation, Jerome Bruner has provided considerable evidence to indicate, especially in "ambiguous" situations, that the personality characteristics of a judge (subject) are important influences in the formulation of his evaluations. [7, pp. 132-135]

Performance evaluations unquestionably qualify as "ambiguous." [40] Experimental evidence to demonstrate the powerful influence of personality characteristics is available in a study performed by Kallejian, Brown, and Wechsler. [19] Conducted in a research laboratory setting, a clinically skilled interviewer interviewed 32 subordinates concerning their relationships with fellow workers and superiors. Following these interviews, the interviewer conducted a personality evaluation of the supervisors responsible for rating the employees interviewed. Based solely on these personality evaluations and the information gathered in the interviews he had conducted earlier with their subordinates, the interviewer attempted to predict how the supervisors actually rated their subordinates' performance. The interviewer's predictions were made without any knowledge of the technical work performed in the laboratory. When the interviewer's predictive ratings were compared with the actual ratings of the supervisors on "overall effectiveness of performance," they correlated at the .03 level of confidence. When the interviewer's predictive ratings were compared with the

actual ratings of the supervisors on 17 specific characteristics of performance, they correlated at the .05 level of confidence.

Reflecting on these findings, the question naturally arises as to which personality characteristics most influence performance ratings. The answer to this question, however, is not simple. As acknowledged by Krech, Crutchfield, and Livson, "the multiplicity of personality determinants and the infinite nuances in the ways they may interact make the prediction and understanding of personality characteristics one of the most difficult tasks within the province of psychology." [24, p. 729] The development of an individual's personality is generally considered to be a continuous process, ever changing in response to a myriad of influences. Such influences include hereditary determinants, physical characteristics, environmental influences, and even childhood experiences. [23, pp. 729-741] It is generally recognized, however, that an individual's attitudes reflect his/her personality. [23, p. 199] For this reason, it is interesting to postulate the existence of a relationship between a supervisor's attitude toward performance appraisal and his/her personality.

Additional Factors Influencing The Accuracy of Subordinate Appraisal

The realization that accurate performance appraisal is positively associated with intelligence, emotional adjustment, and the ability to judge others implicitly carries with it the recognition that additional factors may also play a part in employee evaluation. Factors such as similarity of background, supervisory ability, organizational level, and leniency or severity of raters are all additional areas of possible influence.

Similarity of Background

Numerous sociometric studies [e.g., 12, 13, 39] have indicated that the degree of similarity between two persons relative to characteristics such as age, sex, intelligence, race, religion, and factors such as common attitudes, interests, values, and abilities influence interpersonal attraction and relationships. However, relatively few personnel studies have investigated the influence of rater-ratee background similarity as a source of error in performance evaluation. Do various degrees of background similarity between superior and subordinate positively bias evaluations of performance?

While the findings of three research studies in this area are available for reference, the answer to this question is clouded. Ruda, [31] after analyzing the backgrounds of 279 superior-subordinate pairs, concluded in favor of the similarity hypothesis. His investigation was based on the factor analysis of a 247 item personal history questionnaire and its relationship to the performance ranking of subordinates by their immediate superiors. Quinn, [30] utilizing the personnel files and performance ratings of 21,518 male captains, evaluated by male captains through colonels, found little support for the similarity hypothesis. Each rater/ratee pair studied was matched on 14 major background characteristics — e.g., age, marital status, education, race, religion, etc. A positive relationship between ratings and background characteristics was found only in the case of race and marital status. A third

study by Huber [22] of 30 management consultants comprising three separate levels of the same organization investigated not only biographical variables, but personality variables as well. Neither set of variables, however, was found to be related to performance evaluation. Huber did note that such similarity was associated with increased subordinate job satisfaction.

Supervisory Ability

Research is available that suggests performance appraisals conducted by more-effective supervisors are more valid than those conducted by less-effective supervisors. Both Kirchner and Reisberg [20] and Levy and Stone [25] have found that more-effective supervisors discriminate more in rating subordinates, showing more "spread" and variance in ratings. Less-effective supervisors demonstrate a tendency to rate all subordinates the same. It has also been reported [20] that more-effective supervisors tend to place a greater emphasis on such actions as persistence, initiative, and planning ahead, while less-effective supervisors tend to value such things as loyalty, "getting along," and not "rocking the boat." The implication of this finding is far-reaching. It suggests that subordinates demonstrating such qualities as persistence and initiative may be penalized by a less-effective supervisor. Conversely, the same subordinates, if working for a more-effective supervisor, would be encouraged to continue their performance and would be rated higher. It would seem, then, that less-effective supervisors are more likely to nurture and reward poorer performance than their more-effective counterparts.

Closely related to the Kirchner and Reisberg findings are three studies that indicate a further relationship between supervisory characteristics and performance appraisal. The first deals with supervisor attitudes. Gruenfeld and Weissenberg [16] in a study of 72 male civil service supervisors found supervisory effectiveness to be correlated with the attitudes of supervisors toward performance appraisal. In a manner similar to Kirchner and Reisberg, Gruenfeld and Weissenberg strongly suggest that supervisors high in supervisory quality, self-assurance, initiative, structure and interpersonal trust tend to be more inclined toward the development of subordinates than supervisors low in these characteristics.

A second study by Schneider and Bayroff [33] and a third by Mandell [26] support the existence of a relationship between supervisory ability and the accuracy of performance appraisals. Both studies identified a correlation between the validity of supervisory ratings and characteristics such as aptitude and achievement associated with effectiveness.

In summary, findings strongly suggest the existence of a relationship between supervisory ability and both the development of subordinates and the validity of performance appraisal measures.

Organization Level

The supervisor-subordinate relationship inherently involves different hierarchical positions. Research [20] has indicated that supervisors, subordinates and peers each appraise performance from a different viewpoint. A logical question that often arises is whether or not a ratee's organizational level has any effect on his performance rating. A study of the distribution of

performance ratings of Naval officers in World War II uncovered a widespread tendency to rate personnel on the basis of rank rather than performance. [14, p. 187] Glickman [15] contends that similar examples may easily be found in both civil service and industry. Research conducted by Klores [22] in the research division of a medium sized chemical corporation verifies this contention. In a four-year study involving 100 subordinates and 21 different raters, Klores found that overall performance rating was "strongly affected" by the level of the ratee, i.e., a tendency existed for higher level ratees to receive higher overall appraisals. Furthermore, a positive correlation was found to exist between a rater's job level and the mean job level rated by the rater, indicating that high level ratees are generally appraised by high level raters. Thus, it would seem that the influence of a man's job and the fact that the higher level ratees tend to be appraised by high level raters who hold similar organizational positions and viewpoints may often be an unrealized source of bias in performance appraisal.

Leniency or Severity of Raters

Differences in raters' standards has long remained one of the more difficult problems in performance appraisal. [34] Numerous attempts have been made to develop rating methods to minimize such errors. [3, 4, 22, 36, 37] Some supervisors are more lenient than others in rating subordinates, while others rate all of their subordinates the same. Mandell [26] in a study of 43 "tough" raters and 25 "lenient" raters found that while both groups were equal in performance, they differed in other respects. The tough raters tended to be young, lacked self-confidence, was "sour towards people," and had held his current position for three years or less. In contrast, the lenient rater tended to be old, confident, and of long service, but without a "driving" ambition.

Work performed by Klores, [22] Wilfong, [41] and Bayroff, Haggarty, and Rundquist [4] further relates to this issue.

In attempting to identify sources of rater bias, Klores [22, p. 420] noted that raters who place a greater importance on demonstrating subordinate consideration than on supplying structure in the work place are much more lenient in their appraisals than raters who display an opposite tendency. Interestingly, this is consistent with the finding that effective supervisors are task-oriented.

Wilfong, in a ten-year analysis of the performance appraisals of over 110,000 employees of the Air Force Logistics Command, found that "chronological leniency" predominated the appraisal distributions studied and that even the linear transformation of endorsing officials' raw score rating distributions did not counterbalance the influence of leniency. In addition, he found that the differences between the severity or leniency of raters in appraising ratee performance accounted for greater rating variance than differences between occupational specific appraisal formats.

Does a tendency toward leniency or severity in appraising performance affect the validity of ratings? Are ratings likely to be more lenient if they are to be shown to the ratee? The answer to the first question appears to be no. Using a forced choice rating form in a study involving a group of Army officers who had been identified as hard, average, and easy raters, the findings of Bayroff, Haggerty, and Rundquist [4] indicate only a slight

difference in the validity of ratings by hard raters and easy raters. It is significant to note, however, that while the raters in all three groups placed the ratees in roughly the same order, separate portions of the same scales were used.

Regarding the second question, whether non-confidentiality of ratings increases their leniency, available research findings conflict. The earlier reported research of Stockford and Bissell [34] indicated that raters are likely to be more lenient when performance ratings are shared with ratees. This has also been the experience of the U. S. Army. [6] A more recent study, however, of the performance evaluations of 217 professional health employees found that nonconfidential ratings were no less lenient than confidential ratings. [10]

Summary

As indicated at the beginning of this paper, an identifiable body of knowledge which seriously casts doubt on the use of supervisors' judgment in evaluating employee performance has begun to emerge. The standards employed in judging, as well as the ability to judge people, differ from individual to individual. In addition, a positive correlation has been found to exist between the ability to judge, intelligence, and emotional adjustment. Poor judges (supervisors) are more likely to be poorly adjusted and to be affected by personal biases in their judgments. A relationship may exist between intelligence, supervisor orientation (democratic vs. autocratic) and the validity of performance appraisals. A relationship may also exist between a supervisor's attitude toward performance appraisal and the nature of his/her personality. Additional factors that seemingly play a role in determining the validity of performance appraisals include similarity of rater/ratee background, rater supervisory ability, ratee organizational level and the leniency and severity of raters.

An aggregate of research findings seriously questions both the accuracy and the ability of the "average" rater to adequately appraise the performance of his/her subordinates. The ramifications of this conclusion for personnel actions such as promotions, pay raises, and transfers are obvious. It is not possible, however, to completely assess its importance without a more thorough knowledge of: (a) the kinds of rater qualities, characteristics and circumstances which affect ones ability to judge; and (b) the degree of bias these qualities, characteristics, and circumstances induce.

References

1. Allport, Gordon W. *Personality: A Psychological Interpretation* (New York: Henry Holt, 1937).
2. Barrett, Richard S. *Performance Rating* (Chicago: Science Research Associates, 1966).
3. Bass, Bernard M. "Reducing Leniency in Merit Ratings," *Personnel Psychology*, Vol. 9 (Autumn, 1956), pp. 359-369.

4. Bayroff, A. G., H. R. Haggerty, and E. A. Rundquist. "Validity of Ratings as Related to Rating Techniques and Conditions," *Personnel Psychology*, Vol. 7 (Spring, 1954), pp. 93-113.
5. Blanz, Fredrich and Edwin E. Ghiselli. "The Mixed Standards Scale: A New Rating System," *Personnel Psychology*, Vol. 25 (Summer, 1972), pp. 185-199.
6. Brooks, W. W. *An Analysis and Evaluation of the Officer Performance Appraisal System in the United States Army*. M. S. Thesis, George Washington University, 1966. Cited in Reference (29).
7. Bruner, Jerome S. "Personality Dynamics and the Process of Perceiving," in Robert R. Blake and Glenn V. Ramsey (eds.), *Perception, An Approach to Personality* (New York: Ronald Press, 1951), pp. 121-147.
8. Conrad, Herbert S. "The Personal Equation In Ratings: I. An Experimental Determination." *The Pedagogical Seminary and Journal of Genetic Psychology*, Vol. 40 (December, 1932), pp. 267-292.
9. Conrad, Herbert S. "The Personal Equation in Ratings: II. A Systematic Evaluation." *Journal of Educational Psychology*, Vol. 204 (January, 1933), pp. 39-46.
10. Creswell, Marcia. "Effects of Confidentiality on Performance Ratings of Professional Health Personnel." *Personnel Psychology*, Vol. 16 (Winter, 1963), pp. 385-395.
11. Ferguson, Leonard W. "The Effect Upon Appraisal Scores of Individual Differences in the Ability of Supervisors to Appraise Subordinates," *Personnel Psychology*, Vol. 7 (Autumn, 1949), pp. 377-382.
12. Fiedler, Fred E., Willard G. Warrington and Francis J. Blaisdell. "Unconscious Attitudes as Correlates of Sociometric Choice in a Social Group," *Journal of Abnormal Social Psychology*, XLXII (October, 1952), pp. 790-796.
13. Fiedler, Fred E. "Assumed Similarity Measures as Predictors of Team Effectiveness," *Journal of Abnormal Social Psychology*, Vol. 49 (July, 1954), pp. 381-387.
14. Freeman, Graydon L. and Erwin K. Taylor. *How to Pick Leaders* (New York: Funk & Wagnalls, 1950).
15. Glickman, A. S. "Effects of Negatively Skewed Ratings On Motivations of the Rated." *Personnel Psychology*, Vol. 7 (1955), pp. 39-47.
16. Gruenfeld, Leopold W. and Peter Weissenberg. "Supervisory Characteristics and Attitudes Toward Performance Appraisals," *Personnel Psychology*, Vol. 19 (Summer, 1966), pp. 143-151.
17. Harper, Shirley F. and Thomas L. Whisler (Eds.). *Performance Appraisal, Research and Practice* (New York: Holt, Rinehart & Winston, 1962).
18. Huber, Norman A. "Superior-Subordinate Similarity, Performance, Evaluation, and Job Satisfaction," *Dissertation Abstracts International* (Vol. 31/7-B, 1971), p. 4380.

19. Kallejian, Verne, Paula Brown, and Irving R. Wechsler. "The Impact of Interpersonal Relations on Ratings of Performance," *Public Personnel Review*, Vol. 14 (October, 1953), pp. 295-302.
20. Kirchner, Wayne K. and Donald J. Reisberg. "Differences Between Better and Less-Effective Supervisors in Appraisal of Subordinates." *Personnel Psychology*, Vol. 15 (Autumn, 1962), pp. 295-302.
21. Klimoski, Richard J. and Manuel London. "Role of the Rater in Performance Appraisal." *Journal of Applied Psychology*, Vol. 59 (August, 1974) pp. 445-451.
22. Klores, Malcolm S. "Rater Bias in Forced-Distribution Performance Ratings," *Personnel Psychology*, Vol. 19 (Winter, 1966), pp. 411-421.
23. Krech, David, Richard S. Crutchfield and Egerton L. Ballachey. *Individual in Society* (New York: McGraw-Hill Book Company, 1962).
24. Krech, David, Richard S. Crutchfield and Norman Livson. *Elements of Psychology* (2nd Ed.; New York: Alfred A. Knopf, 1969).
25. Levy, S. and D. M. Stone. "Process and Content of Managerial Ratings of Subordinates." Paper presented at meeting of Eastern Psychological Association, New York, April, 1963. Cited in Bernard M. Bass and Gerald V. Barrett, *Man, Work and Organization* (Boston: Allyn and Bacon, Inc., 1972), pp. 233-234.
26. Mandell, Milton M. "Supervisory Characteristics and Ratings." *Personnel*, Vol. 32 (March, 1956), pp. 435-446.
27. Massarik, Fred and Irving Wechsler. "Empathy Revisited: The Process of Understanding People," *California Management Review*, Vol. 1 (Winter, 1959), pp. 36-46.
28. McGee, H. M. "Measurement of Authoritarianism and Its Relation to Teacher Classroom Behavior," *Genetic Psychology Monographs*, Vol. 61 (February, 1955), pp. 89-146.
29. Miner, John B. "Management Appraisal: A Capsule Review and Current References," *Business Horizons*, XI (October, 1968), pp. 83-96.
30. Quinn, James L. "Bias in Performance Appraisal," *Personnel Administration*, Vol. 32 (January-February, 1969), pp. 40-43.
31. Ruda, Edward S. "The Effect of Interpersonal Similarity on Management Performance," *Dissertation Abstracts International*, (Vol. 31/04-B, 1970), p. 2342.
32. Sales, Stephen M. "Supervisory Style and Productivity: Review and Theory," *Personnel Psychology*, Vol. 19 (Autumn, 1966), pp. 275-286.
33. Schneider, Dorothy E. and A. G. Bayroff. "The Relationship Between Rater Characteristics and Validity of Ratings," *Journal of Applied Psychology*, Vol. 37 (August, 1953), pp. 278-280.
34. Stockford, Lee and H. W. Bissell. "Factors Involved in Establishing a Merit Rating Scale," *Personnel*, Vol. 26 (September, 1949), pp. 94-116.
35. Taft, R. "The Ability to Judge People," *Psychological Bulletin*, Vol. 52 (January, 1955), pp. 1-23.
36. Taylor, Erwin K. and Robert J. Wherry. "A Study of Leniency in Two Rating Systems," *Personnel Psychology*, Vol. 4 (Spring, 1951), pp. 39-147.
37. Taylor, Erwin K., James W. Parker and Gerald L. Ford. "Rating Scale Content: IV. Predictability of Structured and Unstructured Scales," *Personnel Psychology*, Vol. 12 (Summer, 1959), pp. 247-266.
38. Titus, H. E. and E. P. Hollander. "The California F-scale in Psychological Research: 1950-1955," *Psychological Bulletin*, Vol. 54 (January, 1957), pp. 47-65.
39. Vroom, Victor H. "Projection, Negation and Self-Concept," *Human Relations*, Vol. 12, No. 4 (1959), pp. 335-344.
40. Wallace, S. Rains. "Criteria for What?" *American Psychologist*, Vol. 20 (1965), pp. 411-417.
41. Wilfong, Harry D. "Chronological Leniency and Linear Transformation of the Supervisory Appraisal Scale," *Dissertation Abstracts International*, (Vol. 33/8-B, 1973), pp. 4006-4007.