Industrial and Organizational Psychology, **1** (2008), 190–193.
Copyright © 2008 Society for Industrial and Organizational Psychology. 1754-9426/08

Inaccurate Performance Ratings Are a Reflection of Larger Organizational Issues

MICHAEL M. HARRIS
University of Missouri-St. Louis

DAN ISPAS AND GREG F. SCHMIDT
University of South Florida

Murphy (2008) suggests that there is generally only a "weak" relationship between job performance and ratings of job performance, arguing that supervisory performance ratings get little respect, and he questions whether the benefits of performance appraisal even outweigh the costs. Despite these doubts, most organizations continue to collect performance ratings and to conduct performance appraisals with their employees. We contend that industrial and organizational (I–O) psychologists should conduct more research on this issue before discarding the notion that performance ratings are practically useless. We extend Murphy's comments by elaborating on rater motivation (mentioned in passing by Murphy) and rater accountability (not included in Murphy's discussion).

Our informal, anecdotal conversations with managers suggest to us that indeed, line managers do not often think highly of their performance management system. The question of whether ratings formed under different instructional sets (e.g., research vs. administrative purposes) leads to major differences in practice is, however, not so clear. For example, Harris, Smith, and Champagne (1995) found a statistically significant mean difference between the performance ratings made for administrative purposes and the performance ratings made for research purposes (i.e., a validation study). However, they also found that performance ratings made for administrative purposes correlated .58 with performance ratings made for research purposes. Finally, they did report that an overall assessment center rating exhibited a nonsignificant correlation with the performance ratings made for administrative purposes ($r = .11$), whereas it exhibited a statistically significant correlation with the performance ratings made for research purposes ($r = .17$), but the difference between these two correlations was not statistically significant. Although the linkage between job performance and job performance ratings may not be as strong as I–O psychologists would like, and as Murphy notes, there are inaccuracies; the notion that the linkage is so weak as to make performance ratings unsuitable criteria for validation studies, training program assessment,

Correspondence concerning this article should be addressed to Michael M. Harris. E-mail: mharris@umsl.edu

Address: College of Business Administration, University of Missouri-St. Louis, 1 University Boulevard, St. Louis, MO 63121

Michael M. Harris, Department of Management and Center for International Studies, College of Business Administration, University of Missouri-St. Louis; Dan Ispas and Greg F. Schmidt, Department of Psychology, University of South Florida.

https://doi.org/10.1111/j.1754-9434.2008.00037.x Published online by Cambridge University Press

and so forth is based on very little data, to say the least.

Multiple Measures of Accuracy

There are different ways to measure rater accuracy. Typically, I–O psychologists have emphasized correlation coefficients or analytic tools based on correlation coefficients (e.g., factor analysis). Social psychologists, however, have developed several different analytic models to analyze rater accuracy, including the Cronbach (1955) model and the Social Relations Model (Levesque & Kenny, 1993). It is noteworthy that Jussim (1995), in his chapter on accuracy of social judgments, observed that "some biases can coexist with some degree of accuracy, and some biases enhance accuracy" (p. 16).

Both multifactor and mediated models of performance appraisal suggest that contextual factors (e.g., rater recall, purpose of the ratings for the organization, rating scales, rater motivations, etc.) interfere with the accurate measurement of performance. Murphy maintains that it is deliberate distortion that appears to account for most of the problems with performance rating inaccuracy. We will argue that nondeliberate sources of distortion are equally important.

Rater Motivation

Rater motivation can be another source of inaccuracy (Harris, 1994). Harris (1994) proposed several determinants of rater motivation: rewards, negative consequences, and impression management. Raters are usually not rewarded for engaging in performance management activities. It may be difficult to identify which managers are doing a good job with performance management tasks, and many organizations we have observed do not highly value this activity.

In terms of the perceived negative consequences, five related categories can be distinguished: damage to subordinate–supervisor relationship, damage to subordinate morale, criticism from the rater's subordinate, criticism from the rater's supervisor, and interference with other, more important, tasks. All these perceived negative consequences may be more imagined than real. Impression management concerns can be divided into self-impression management and management of others' impressions. Any of these factors could account for variations in rater motivation and subsequent rating accuracy.

Increasing the rater's motivation is likely to lead to more careful processing of the information and therefore to more accurate ratings. When producing ratings, raters are theorized to form schemas that categorize ratees. The term "cognitive miser" refers to a widely cited schema function in social psychology. To reduce the overall processing load, people use "shortcuts" to conserve mental resources when they are trying to make sense of other people (Fiske & Taylor, 1984). When making performance appraisals, raters have a considerable cognitive load: they usually have to consider multiple subordinates and multiple situations. Motivation can also impact the type of information processing used by the supervisor. Dual processing theories (e.g., Strack & Deutsch, 2004) propose two types of processing: implicit (fast and less flexible) and explicit (more flexible and requiring greater resources). In the absence of motivation, raters are more likely to use an implicit processing model that can lead to less accurate ratings. Unmotivated raters may, therefore, simply take the "easy" way out, by giving all their subordinates an average rating or using another, similar tactic, to avoid investing effort and time into performance management activities. We suspect that, depending on the organization, a relatively large number of managers fall into the nondeliberate (i.e., "lazy") category of inaccurate raters.

Several strategies can be used by organizations to increase rater motivation. For example, the raters can be trained in understanding the importance and consequences of providing accurate ratings for effective organizational functioning. Rewards can be provided to managers who expend the time and energy to engage in performance management activities. To encourage

https://doi.org/10.1111/j.1754-9434.2008.00037.x Published online by Cambridge University Press

accurate storage and retrieval of performance relevant information, raters could hold more frequent appraisal discussions with their subordinates and should be encouraged to keep accurate and up-to-date records.

Rater Accountability

Rater accountability is another important contextual variable that may have a large impact on accurate performance appraisals and should be included in Murphy's discussion of rater motivation and goals. As defined by Tetlock and Kim (1987), rater accountability concerns the social pressure to justify one's judgments to others.

Depending on the audience, raters may "simply adopt positions likely to gain favor of those to whom they feel accountable" (Tetlock, 1985, p. 310) or engage in more careful processing of performance, resulting in more accurate ratings (Tetlock). A number of factors influence whether the rater will attenuate or amplify performance ratings (e.g., rating heuristics, rater overconfidence, rater tolerance for ambiguity, etc.) (Lerner & Tetlock, 1999) and may, in addition to those issues already described by Murphy, help explain the weak relationship between job performance and ratings of job performance.

However, accountability as a social and cognitive psychological variable is not unidimensional and is anything but simplistic. According to Lerner and Tetlock (1999), when an audience is present (a) whose views are unknown, (b) who is concerned with rating accuracy, (c) who is also interested in how the rater arrived at their performance appraisal rather than simply the outcome, (d) who is well informed, and (e) who has a legitimate reason for learning more about the rater's decisions, the rater will engage in critical and effortful thought about accurately rating others. Any subset or combination of these audience characteristics can attenuate, amplify, or have no impact on the amount of bias present in performance measurement. This bias can take the form of any commonly referenced rating error—halo, leniency, central tendency, and so on. Consequently, rater accountability should also be included in any discussion of factors that affect performance rating quality.

Recent research has examined the role of accountability in ratings of job performance. Curtis, Harvey, and Ravden (2005) found that participants asked to justify their ratings to the experimenter (which can be thought of as the supervisor) provided significantly lower ratings than the participants asked to justify their ratings to the ratees. Mero, Guidice, and Brownlee (2007) found that raters accountable to higher status and mixed-status audiences were more likely to give accurate ratings.

In organizational settings, this discussion holds implications for how much effort and thought raters will expend to arrive at accurate ratings of performance. As Murphy mentions, more emphasis should be placed on building better climates for performance settings, taking into account the rater's perception of accountability for providing ratings. Holding raters accountable to their supervisors and providing incentives to motivate them seem to increase the accuracy of ratings. In other words, when building a performance appraisal system where the goal is to minimize bias and maximize accuracy, careful attention should be paid to who will hold raters accountable for their decisions. Nondeliberate sources of inaccuracy can be equally important and should be included in the discussion.

In conclusion, we agree that there is room for improvement in performance management systems. We have highlighted some additional sources of inaccuracy and emphasized the role of nondeliberate sources of rating distortion. Ultimately, however, we believe that performance management systems are much more a reflection of broader problems in an organization's human resource management systems. Principally, the effectiveness of the performance management system is a reflection of how important human resource systems are to the organization and whether their purpose is fully understood by the organization or whether the organization simply fails to

https://doi.org/10.1111/j.1754-9434.2008.00037.x Published online by Cambridge University Press

understand the importance of effective human resource management systems.

References

Cronbach, L. J. (1955). Processes affecting scores on "understanding others" and "assumed similarity." *Psychological Bulletin*, *52*, 177–193.

Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in performance appraisals. *Group and Organization Management*, *30*, 42–60.

Fiske, S., & Taylor, S. (1984). *Social cognition*. Reading, MA: Addison-Wesley.

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management*, *20*, 737–756.

Harris, M. M., Smith, D., & Champagne, D. (1995). A field study of performance appraisal purpose: Research- versus administrative-based ratings. *Personnel Psychology*, *48*, 151–160.

Jussim, L. (2005). Accuracy: Criticisms, controversies, criteria, components, and cognitive processes. *Advances in Experimental Social Psychology*, *37*, 1–93.

Lerner, J. S., Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, *65*, 1178–1187.

Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, *33*, 223–252.

Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 148–160.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Tetlock, P. E. (1985). Accountability: The neglected social context of judgment and choice. In M. B. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 7, pp. 297–332). Greenwich, CT: JAI Press.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, *52*, 700–709.

https://doi.org/10.1111/j.1754-9434.2008.00037.x Published online by Cambridge University Press