Performance Ratings: Tainted or Sainted?

***Authors**: Lisa Harpe, PhD and Sarah Gilbert, M.S.*

Performance Ratings: Tainted or Sainted?

By Lisa Harpe, PhD and Sarah Gilbert, M.S.

DCI Consulting

Published in Law360 on July 18, 2018.

Organizations want to recruit, hire and retain the best performers. Many companies use a performance management system to identify high performing employees that they want to retain through rewards such as merit increases, bonuses, or promotions, and to identify low performing employees to target for improvement or separation.

The inherent association between performance ratings (e.g. performance ratings, objective metrics) and employment outcomes (e.g., merit pay or termination) often puts the performance management system under legal scrutiny. Whereas an evaluation of performance alone is not an employment decision, performance ratings that influence employment decisions (like pay, promotion, or reduction-in-force) are subject to the same anti-discrimination regulations and laws as other employment practices. As such, employers may want to proactively evaluate the job-relatedness of the performance ratings and whether there are subgroup differences in performance on any of the ratings.

In particular, the societal focus on pay equity and employers' pay practices has heightened the need for employers to consider EEO outcomes in their performance ratings systems. In the following article, we (1) provide some background on performance ratings, (2) discuss the legal environment related to performance ratings, and (3) describe proactive steps to examine performance ratings directly and indirectly in the context of evaluating pay equity.

Why evaluate performance?

Performance evaluation systems are an important organizational tool that can help companies make meaningful, job-related distinctions in performance within an employee (by identifying within-person strengths and weaknesses) and/or between employees (by identifying the strongest or weakest employees in their workforce). When done properly, having a formal performance evaluation system is a valuable human resource practice that can serve many purposes. For example, employers may use the resulting performance evaluation to:

- Communicate expectations to employees,
- Motivate employees to perform to their fullest potential,
- Identify strengths and weaknesses within an employee to help them focus on areas that need improvement or training,
- Establish the relationship between selection procedures and business outcomes through criterion-related validation studies, and
- Make personnel decisions such as who to promote, give raises or bonuses, or terminate.

Over the years, the specific mechanisms for evaluating performance have changed, primarily in response to employee dissatisfaction and concerns about the effectiveness of the systems. Companies have tried different rating scales (e.g., numerical, categorical, 3-point, 5-point) and different performance dimensions to rate (e.g., objectives, competencies, company values). Methods to improve the job-relatedness and accuracy of performance ratings have included rater training, use of behaviorally anchored rating scales, and 360-degree feedback. Recently, some organizations have joined a movement away from formal ratings—so called 'ratingless performance management.' In the place of a rating, these companies emphasize regular performance discussions between supervisors and incumbents. The value of this movement away from formal ratings has received extensive debate among Industrial Organizational Psychologists[i]. For example, though rating-less systems may provide the opportunity for managers to provide more timely feedback to their employees, it may be harder for them to justify employment decisions (e.g., raises, terminations) without formal documentation of employee performance.

Pay for Performance in a Rating-less Organization

Many organizations claim that they 'pay-for-performance' and insist that performance will explain pay differences among their employees. The formality of this relationship may be dictated via merit matrices with specific merit increases corresponding to specific performance ratings (e.g., 2% for "Meets Expectations", 3% for "Exceeds Expectations"). In these situations, the relationship between performance ratings and pay seems relatively straightforward.

So, how does performance influence pay in a ratingless system within a company that 'pays for performance'? Some organizations have shifted away from ratings to performance snapshots based on supervisor responses to certain questions or statements about employees (e.g., based on their performance, I would give this person the highest possible compensation; I would want this person on my team).[ii] Other companies have switched from ratings to adjectives such as 'industrious' or 'prolific'.[iii] A supervisor may then use the answers to these questions or the adjectives assigned to an employee to allocate merit pay or bonuses. A 2016 study by the Center for Effective Workplaces examined 244 companies with ratingless performance management systems and found the top two methods used to allocate merit increases were manager discretion and calibration sessions. [iv] In those cases, the merit increase itself becomes a proxy for performance.

It is important for organizations to be cognizant of the laws and regulations under which performance evaluation systems are monitored and regulated. This is because, when used to make employment decisions (e.g., merit increases, bonuses, promotions, terminations) performance evaluations, regardless of their form, represent a selection procedure from a legal perspective, and, thus, in some situations are subject to the same legal scrutiny and anti-discrimination laws and regulations as any other selection procedure. As such, having no direct measures of performance could put organizations in a difficult situation should employment decisions related to performance be challenged in the EEO context.

Legal Landscape

In the early 2000s, several companies—such as Ford Motor, Capitol One, and Goodyear[v]—faced allegations of discrimination in performance ratings from employees. During this time, many employers adopted performance appraisal systems that required managers to distribute ratings into pre-specified performance distributions, known as forced distribution rating systems. In each of these cases, plaintiffs claimed that the employer discriminated against protected class members by unfairly giving them lower ratings, which ultimately had a negative consequence on their career or livelihood (e.g., via lower pay or termination).

In addition to direct claims of a discriminatory rating system, employers should be aware that a performance rating system could play a key role in other types of employment discrimination cases. For example, in Walmart vs. Dukes (2011) [vi]—one of the largest class action lawsuits filed against an employer—a class of female, store-level managers claimed sex discrimination in promotion and pay, despite having higher performance ratings than their male comparators. In Velez et al v. Novartis Pharmaceuticals Corp (2004)[vii], another high profile class action lawsuit, a class of female sales workers claimed discrimination in pay and promotions, among other things. Unlike in the Walmart case, the female employees in this case, on average, had lower performance ratings than males, but claimed that supervisors purposefully gave them lower ratings as a pretext to discriminate.

Today, perhaps the likeliest scrutiny of performance ratings is in the realm of its inclusion in pay equity studies. For example, where performance ratings explain pay disparities among protected classes, a plaintiff or government agency may claim that the performance system is tainted, and is a pretext for discrimination. In other cases, if performance ratings suggest that protected class members should receive higher pay than their comparators, plaintiffs can claim that the pay system is discriminatory as well if, for example, despite having higher performance ratings, receive lower pay. In either case, employers may want to proactively understand how performance ratings vary along protected classes. For companies who pay for performance but do not use ratings, the question becomes how to distinguish performance levels among employees in a pay study.

Proactive Steps

Employers concerned about how their performance management system affects their vulnerability to legal scrutiny can consider taking two steps:

1. Examine performance ratings directly for evidence of race or gender disparities[viii]
2. Examine the relationship between performance ratings and other employment decisions such as pay and whether ratings explain race or gender disparities in those decisions[ix]

<u>Setting up a study of performance ratings</u>. Prior to the actual analysis of ratings along race, ethnicity or sex, it can be helpful for employers to consider the following[x].

First, it is important to gain an understanding of the process. Policies and procedures, forms, and training materials provide invaluable information regarding how the evaluation system works. Interviews with incumbents, human resources, and managers can supplement these materials.

Second, organizations may identify the types of performance ratings to study. Many performance management systems include ratings on multiple performance categories, for example, based on the accomplishment of specific objectives or based on organization-wide competencies. Often, supervisors then assign employees an overall rating. Organizations may choose to study one, multiple, or all of these ratings.

Next, an employer may consider the proper grouping of employees for comparison. Typically, the rater (e.g., immediate supervisor) may be included as a grouping factor as well as any organizational levels where performance ratings are reviewed and possibly changed (e.g., department calibration meetings). Employers may want to examine ratings distributions at other levels such as by job or unit or location.

Finally, employers should identify the protected groups of interest. Whereas race, ethnicity, sex, and age are commonly studied protected groups, employers can examine any group for which they have data. The previous decisions will identify the required data that should then be collected for analysis.

Types of Performance Rating studies

Prior research provides some indication of expected differences in performance ratings by race, gender, and age. Meta-analyses and other research suggests that, particularly when looking at overall ratings, females tend to have higher ratings than males, and Whites tend to have higher ratings than African Americans[xi]. These findings suggest that companies would be well advised to conduct their own proactive studies on their own ratings.

After making decisions about the study parameters and collecting the data, several types of studies will inform an employer about vulnerability to claims of discrimination. For example, an employer could examine the entire distribution of ratings to see whether females, minorities or older workers receive disproportionately lower ratings than their peers. Various statistical significance tests (e.g., t tests, rank sum tests) and practical significance measures (e.g., d statistics, various ratios) are available to evaluate the rating differences.

Where high ratings afford employment opportunities such as merit increases, bonuses or promotions, a study can be conducted to examine the assignment of high ratings to see whether females, minorities, or older workers receive disproportionately fewer high ratings (such as “Greatly Exceeds Expectations”). On the flip side, where low ratings lead to negative employment decisions such as disciplinary action, demotions, or dismissal, a study can be conducted to examine the assignment of low ratings to see whether females, minorities, or older workers receive disproportionately more low ratings (such as “Needs Improvement”). Employers can conduct these studies on preliminary or final ratings.

Still, another type of study could look at group differences in the actual merit increase percentages received by employees receiving the same performance rating. That is, do men and women who received a rating of “Above Expectations” receive similar merit increase percentages? Should an employer determine that a study is feasible and potentially of value, there are a variety of approaches to consider and preferred approaches will likely depend on specific context.

Including Performance Ratings in the Studies of other Employment Outcomes, such as Pay

Performance may influence a number of employment outcomes. One of the theoretically most direct outcomes related to performance is pay, hence the designation as a ‘pay for performance’ organization. With the heightened concern with pay equity for legal and diversity reasons, many organizations have embarked on studies of their compensation practices. Most often, these studies focus on sex and/or race differences in current base salary. In ‘pay for performance’ companies, it is intuitive that such analyses may need to account for performance. However, it is important to examine how performance influences pay and to consider whether performance ratings could be challenged as a tainted variable.

Toward that end, organizations should think about the types of compensation directly influenced by performance ratings. For example, how is performance related to current pay? Current base pay does not reflect a single employment decision but is the result of multiple decisions (*e.g.,* starting pay, pay increases, promotional increases) over an employee's tenure with the company. Even the inclusion of multiple years of performance ratings may not accurately represent the influence of performance on current base salary because of other factors such as the merit pool of money available in any given year. It may be more fruitful to think about other pay decisions more directly influenced by performance ratings. For example, as mentioned above, many organizations utilize merit matrices that proscribe recommended merit increases based on specific performance ratings. Thus, performance ratings may be more relevant to an analysis of merit as opposed to base pay.

After controlling for performance ratings (and other explanatory variables), an employer can see whether there are statistically significant differences in merit pay along sex or race/ethnicity lines among similarly situated employees. However, employers may want to take an additional step. By conducting the analysis with and without performance ratings, employers can gauge their vulnerability to claims that performance ratings are tainted, that is, a pretext to discriminate. If a study of pay without performance ratings indicates significant racial/ethnic or sex differences, and these differences disappear with the inclusion of performance ratings, employers should be prepared to defend the job-relatedness of their performance management system. Industrial Organizational Psychologists often perform this type of evaluation.

Final Thoughts and Recommendations

- Employers may want to consider a review of their performance management systems on a regular basis.
- For employers considering ratingless performance management systems, think through how you will allocate merit pay or other performance related pay decisions. Will you use a proxy for a performance rating? Will managers decide? Consider how you will justify performance related pay differences without ratings.
- Proactive studies to evaluate sex, race, and age differences can help companies evaluate the extent to which their performance systems may be discriminatory.
- Employers should consider identifying the employment outcomes influenced by these ratings and conduct studies to examine their actual influence on these outcomes.
- Finally, the best defense may be a job-related performance management system implemented in a comprehensive and procedurally fair manner that will likely withstand legal scrutiny whether race, sex, or age differences exist or not

[i] Adler, S., Campion, M., Colquitt, A. & Grub, A. (2016). Getting Rid of Performance Ratings: Genius or Folly? A Debate. *Industrial and Organizational Psychology, 9*(2), 219-252. dio:10.1017/iop.2015.106

[ii] Buckingham, M & Goodall, A. (2015). *Reinventing Performance Management*. Harvard Business Review. April 2015 (pp.40–50). Boston, MA: Harvard Business Publishing.

[iii] Wilkie, D. (2016). *Instead of Rating Performance with Numbers, How About Adjectives?* Society for Human Resource Management. Retrieved from https://www.shrm.org/resourcesandtools/hr-topics/employee-relations/pages/instead-of-rating-performance-with-numbers,-how-about-adjectives.aspx

[iv] Miller, S. (2016). *Ratingless Reviews and Pay Practices*. Society for Human Resource Management. Retrieved from https://www.shrm.org/ResourcesAndTools/hr-topics/compensation/Pages/Ratingless-Reviews-Positively-Affect-Pay-Practices.aspx

[v] Siegel v. Ford Motor Co., No. 01-102583-CL (Wayne Co., Mich., Apr. 27, 2001). Feltman v. Capital One Services, Inc., No. C.A. 3:02cv894 (E.D. Va. Richmond Division, Dec. 9, 2002). Jones v. Goodyear Tire & Rubber Co., No. 02-00-5090 (Summit Co., Ohio, Sept. 12, 2002).

[vi] Wal-Mart v. Dukes, 564 U.S. 338 (2011).

[vii] Velez et al v. Novartis Pharmaceuticals Corp, U.S. District Court, Southern District of New York, No. 04-9194.

[viii] Similar studies can be set up to examine performance ratings along age lines. For simplicity's sake, this article focuses on race and gender, but the same concepts apply to other protected classes.

[ix] Similar types of studies can be set up to examine other employment decisions and other protected classes (e.g., whether performance ratings explain age disparities in termination or RIF decisions).

[x] For simplicity, the next few sections focus on organizations with performance rating systems.

[xi] Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance:
A new meta-analysis*. Journal of Applied Psychology, 88* (4), 694–706.
McKay, P.F., & McDaniel, M.A. (2006). A reexamination of black-white mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91*, 538-554.
Roth, P. L., Purvis, K. L., & Bobko, P. (2012). A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management,* 38 (2), 719-739.