

CEO

Established 1979

Center for Effective Organizations



CEO Working Paper Series

Aligning Research and the Current Practice of Performance Management

CEO Publication: G16-04 (671)

Gerry Ledford
Senior Research Scientist
Center for Effective Organizations
Marshall School of Business
University of Southern California

George Benson
Associate Professor
Associate Professor of Management
University of Texas at Arlington

Ed Lawler
Director
Center for Effective Organizations
Marshall School of Business
University of Southern California

No society can provide its members with a high quality of life unless it has effective organizations.

Edward Lawler III

February 2016

Background Note

Issues of the journal of the Society for Industrial and Organizational Psychology (SIOP), Industrial and Organizational Psychology, are organized around two focal articles and commentaries about those articles. This paper is a commentary. The focal article that it concerns is a debate about the scientific and practical merits of ratingless performance appraisal. Readers interested in ratingless performance reviews may find the entire issue of the journal to be of interest; it will appear later in 2016.

The debate over eliminating performance ratings addresses many important theoretical and practical issues. However, the academic debate on the topic is disconnected from the concerns of practitioners. Knowledge gained from theory-driven research is not leading practice on the use of performance ratings, despite the large volume of potentially relevant research findings. Many organizations are charging ahead with performance management solutions that seem sensible to them. They may be interested in academic research, but they not waiting for it. We will argue that academic researchers who hope to influence practice need to better understand the concerns of practitioners and the research opportunities that are presented by contemporary practice.

The Evolution of Performance Management Practices

There has been a major disconnect between research and practice in performance management for decades. Table 1 outlines three major waves of practice in performance management (Ledford, Benson, & Lawler, in press). Here we will focus on the rating approaches used, while noting that these are embedded in a set of complementary practices.

Conventional performance management practices were implemented in the 1950s and 1960s. This was the era of complex rating scales, in which more was better – more subscales, each using at least five and often more points on the rating scale. Such complex schemes reflected the highly bureaucratic, hierarchical organizational designs of the era.

Transitional performance management began to be used in the 1990s and is often considered to represent best practice. The rating approach involves a radical simplification, consistent with the attempt to run organizations with leaner, less cumbersome processes of all kinds. Often there are only three points on the scale: a small group top performers (10% to 25%, depending on the company), typical performers (70% to 85%), and poor performers (usually only a few percent of the distribution). Stacked ranking and forced distributions were common at the start of this era but largely disappeared by the end of it.

Cutting edge performance management is only a few years old and its use is still limited, but is rapidly gaining ground. The rating approach associated with this phase is text-based appraisal with no scoring or rating. Companies do not cease to manage performance, provide feedback, or differentiate consequences; they simply do not provide the employee with rating numbers, letters, or categories that summarize performance.

Insert Figure 1 about here

Misunderstandings of the Phenomena of Study

The Adler et al. debate was spirited and thoughtful, but we think that most of the debate concerns performance ratings as they were done long ago, not how they are done now and will be done in the future. As such, we suspect that most practitioners would find the discussion interesting but not very relevant to the issues with which they are grappling. Specifically:

1. Most of the debate is about a type of rating that is waning in practice. The very complex rating schemes of an earlier era invited a great deal of concern about rating accuracy, interrater reliability, rater training, and so on. These problems are automatically less important if the complexity of the rating task is reduced using either simplified ratings or no ratings.

Consider three organizations, each representing one of the historical phases of performance management practices. An aerospace company uses a 150-point rating scale (15 factors that each use 10 point scales). Executives insist that an employee with 125 points is a better performer than one with 124 points. An insurance company uses a three-point scale: up to 25% of the population is in the top performer category, about 2% are in the nonperformer category, and the rest are in the “good performer” category. Finally, an entertainment company has no ratings. Its performance distribution prior to going ratingless was very similar to the insurance company’s three-point scale, and indeed the ratingless company found that the distribution of rewards did not change after it moved to a ratingless system. The company with the 150-point rating scale automatically has a

much greater challenge with its system of fine-grained measurement; all of the problems associated with ratings accuracy and equity are magnified. Which company is most likely to have disputes about the level of employee performance? When the number of discriminations is vastly reduced, the cognitive task is simpler and the likelihood of a valid and reproducible outcome increases.

We do not know of a single company that has made its performance rating scheme more complex and detailed during the past five years. Organizations are choosing between a very simple rating and a ratingless system that looks, in practice, very much like a simple rating system without a rating score. Most of the considerable research literature on performance appraisal has limited applicability to these systems. Research that addresses the accuracy and validity of the newer systems is badly needed.

2. The elimination of ratings is almost always tied to other practices. Companies do not eliminate ratings while keeping all else constant. Our forthcoming study (Ledford et al., in press) examines 244 organizations that have adopted ratingless reviews, ongoing feedback (typically monthly or quarterly feedback meetings), and/or crowd-sourced feedback using social media. We found that 37% used ongoing feedback only; 34% used ongoing feedback plus ratingless reviews; 15% used ongoing feedback, ratingless reviews, and crowd-sourced feedback with social media, and the rest used other combinations of practices. Only 7 cases (3%) used ratingless reviews alone.

While the Adler, Campion, and Grubb team is correct that ratingless appraisals do not necessarily demand a greater emphasis on frequent performance conversations or a greater emphasis on development, in the world of organizations these practices almost always go together. Moreover, ratingless reviews are used with older performance practices as well. A majority of those adopting ratingless appraisal also used cascaded goals, calibration meetings, assessment of employee competencies, and 360 feedback, for example. This means that most research on ratingless reviews actually will be about a cluster of practices that include ratingless reviews,

inevitably making the analysis of results complex and requiring meta-analyses to disentangle effect. The cluster of practices used in each study needs to be identified to make this possible.

Theoretical Issues Relevant to Ratingless Reviews

There are many theoretical reasons bearing on the effectiveness of ratingless reviews, but these are largely unexplored. Here we outline the theoretical considerations that we consider to be the most intriguing.

Rewards distribution. The effect of ratingless reviews on the distribution of rewards is an important practical issue. Our cross-sectional survey indicates that organizations do not see a change in their overall reward costs with ratingless appraisals. Even so, the distribution of rewards may change considerably. Without the crutch of a rating number, how do managers allocate salary increases, bonuses, promotions, and other rewards? Is the distribution of rewards different under ratingless processes and older systems, and if so, why? The effects of the ratingless approach are likely to be mediated by the method of reward allocation that the organization chooses. Many organizations simply leave reward allocation to the first-line supervisor; others create intensive text-oriented calibration processes; others use “shadow” ratings that they do not disclose to employees; and so on. Each of these options probably has different effects. We are most concerned about potentially negative effects such as ratings bias when all decisions are left up to the supervisor, without the checks and balances of calibration sessions, for example.

Development focus. Performance management has always involved a balance of performance assessment and development, but ratingless appraisals shift the needle more in the direction of development. Does greater employee development occur in processes that use ratingless reviews? If so, is this due to ratingless reviews or related practices such as ongoing feedback? What is the impact on employee motivation and performance if management’s attention is shifted toward development?

Desensitization of performance discussions. The combination of regular feedback meetings (e.g., monthly or quarterly) plus ratingless appraisal may have a significant effect on performance discussions. Companies adopting these practices hope that regular performance discussions will become an ingrained habit rather than a torturous annual exercise. Perhaps shorter, more frequent, less formal sessions reduce the emotional baggage associated with annual performance discussions and produce better results. If so, this change could produce far higher satisfaction with the performance management process on the part of both employees and subordinates, as well as higher satisfaction with supervision.

Crowd-sourced feedback effects. When ratingless appraisal is combined with crowd-sourced feedback, the supervisor has access to all of the feedback provided by peers and others during the performance period. This may help overcome the recency effect, a serious rating problem, by reminding the supervisor of the accomplishments of the employee throughout the performance period. In this way, crowd-sourced feedback may provide a substitute for the supervisor diary that Denisi and Peters (1996) have shown to be an effective technique for overcoming the recency effect. An alternative is needed, because supervisors tend to view diaries as requiring unrealistic time demands.

Sensemaking. Another interesting issue is how employees and supervisors make sense of crowd-sourced feedback when it is combined with ratingless appraisal, an issue that has not been explored as far as we know. Crowd-sourced data are a psychometric train wreck. In the typical process, any person in the organization can provide feedback about anyone else using social media, and the results are public rather than anonymous. Feedback is free form and thus does not use any consistent format, and it usually is only positive. Yet our survey results indicate that the addition of crowd-sourced feedback to ratingless appraisal and ongoing feedback increases effectiveness on several measures. Numerous questions remain unanswered. If an employee receives a stream of positive feedback from others and a balanced review from the supervisor, how does the employee

make sense of the difference? Does the crowd-sourced feedback affect the supervisor review, changing the supervisor's understanding of who is a high or low performer, or is it simply disregarded? Do supervisor and crowd ratings converge over time or are they orthogonal?

Topics Missing In Performance Management Research

The research on performance management is generally silent on several issues that have great theoretical and practical importance and are relevant to performance management in general, not just ratingless appraisal and other new practices. These include the following.

Need for customization. Researchers and practitioners alike often assume that a corporation should have one performance management system, and that research or “best practice” will lead an organization to the answer. For organizations, there are obvious administrative and oversight advantages if all units use the same process, the same calendar, and the same technology platform. However, different practices are likely to be appropriate for different organizations and even different units within large and diverse organizations. For example, frequent supervisory feedback is a straightforward task where supervisory spans of control are limited. However, we know of very lean, self-managed organizations that have spans of control of 75 to 125, making supervisor/ subordinate performance management untenable. In addition, individual contributors perform most of the work in some units, while the work of other units may be primarily team-based; 360 reviews and crowd-sourced feedback will have little value in units made up of individual contributors who lack familiarity with others' work. With regard to ratingless appraisal, it is likely that certain contingency variables will help explain effectiveness, including supervisor skill and organizational culture that reinforces good supervisor behavior.

Technology and job design issues. Employee behavior and performance is being electronically monitored and measured as never before. This has progressed from measuring the keystrokes of typists and the transaction time of call center operators to the intensive monitoring of the work behavior and performance of all types of employees, and the use of wearable devices, GPS,

and other technologies. This has profound implications for performance management because performance feedback is coming from intelligent software agents rather than from managers. Not nearly enough attention has been paid to how employees react to feedback from technology rather than supervisors, the relative advantages of and problems with machine versus human feedback, and the effectiveness of machine feedback in changing behavior and performance. We could imagine both positive and negative effects from these developments. Among the positive effects, it is possible that automated feedback may foster self-efficacy and self-control, giving employees ongoing performance feedback and suggesting expertise-based improvement strategies from a neutral and unemotional source. In any case, technology may render supervisor ratings increasingly irrelevant if it permits ongoing, real-time, direct performance feedback and evaluation.

Employee segmentation. It appears that far more companies are monitoring and conducting formal assessments of the impact of changes in performance management practices outcomes. The most common source of evaluation data is the annual employee opinion survey. Most companies conducting an evaluation ask whether employees in general prefer the new system to the old system, but usually there is limited analysis of the responses. One specific individual difference that is largely ignored in internal evaluations is segmentation by performance level (in part because many employee surveys are anonymous and performance level is unknown). This is potentially dangerous if high performers and typical performers have different preferences. In the case of ratingless appraisals, it is quite possible that those who receive top ratings place the most value on high ratings, since they are the “winners” in that system. If the ratingless approach alienates top performers, leading to dissatisfaction and turnover, the ratingless approach may be self-defeating even if the average employee prefers it.

In general, we would argue that there are now outstanding opportunities to study new issues of theoretical and practical importance by collaborating with practitioners to understand phenomena such as ratingless reviews. This will require researchers to understand the issues that

are important to practitioners, and it will require them to develop relationships with managers, as well as research frameworks and measures, that meet organizational needs as well as their own need for theory-based research. If they do it, they should be able to conduct research that contributes to both theory and practice.

References

- Denisi, A.S., & Peteres, LY. (1995). Organization of information in memory and the performance appraisal process: Evidence From the field. *Journal of Applied Psychology, 81:717-737.*
- Ledford, G., Benson, G., & Lawler, E. (in press). A study of cutting edge performance management practices: Ongoing feedback, ratingless reviews, and crowd-sourced feedback. *WorldatWork Journal, 25(2).*

Figure 1

The Evolution of Performance Management Practices

	Conventional PM	Transitional PM	Cutting Edge PM
Period of prominence	1950 – today	1995 – today	2010 – today
Timing of reviews	Annual	Annual, sometimes also midyear	Monthly or quarterly
Basis for performance appraisal	Goal attainment; traits	Cascaded goals; competencies	Cascaded goals; competencies
Appraisal scale	Complex ratings	Simplified ratings	Ratingless
Input from peers and others	None	360 appraisals	Crowd-sourced feedback
Reward allocation method	Supervisor determines using ratings	Calibration meetings and formulas	Varied
Method of differentiation	Varies: Open, stacked rank, forced distribution	Distribution guidelines	?
Appraisal target	Individuals	Mostly individual, some teams	Mostly individual, some teams
Balance of performance versus development	Balanced	Balanced	More development emphasis

Source: Ledford, Benson, & Lawler, (in press).