RATINGLESS PERFORMANCE REVIEWS: A QUASI-EXPERIMENT

Gerald E. Ledford, Jr.
Affilitated Research Scientist
Center for Effective Organizations

George Benson
Associate Professor, College of Business
University of Texas at Arlington

CEO CENTER FOR EFFECTIVE ORGANIZATIONS

Ratingless Performance Reviews: A Quasi-Experiment

Executive Summary

We conducted a study of ratingless reviews in an organization that had an innovative and effective performance management process prior to adopting ratingless reviews. We collected data from the pilot and comparison units before and after the change to ratingless reviews. Results indicated some positive changes and no clear negative changes in the pilot units compared to the comparison units. Analysis suggests that positive results are primarily the result of more frequent feedback that is more oriented toward employee development rather than the ratingless reviews, per se.

Ratingless Performance Reviews: A Quasi-Experiment

Performance management is often called the human resources process most hated by employees and managers alike (Ledford, Benson, and Lawler 2016a). The history of research and practice on performance management is, most of all, the search for new techniques that will finally turn this vexing HR process into one that is effective at improving performance and developing employees, while offering a less frustrating experience to both employees and managers. The ratingless review is the latest in a long line of techniques that have been advocated for these purposes.

Ratingless reviews offer the employee feedback on performance and/or development but offer no summary numerical or letter grade. Ratingless reviews have received tremendous attention in popular media and the business press in recent years. They have been adopted by prominent companies such as Accenture, Cargill, Gap, Eli Lilly, IBM, Microsoft, New York Life and Zappos. A survey of 430 organizations found that 12% had abandoned performance ratings and 25% were considering doing so (Human Capital Institute [HCI] 2015). A Corporate Executive Board (CEB) survey of 296 organizations found that 6% had abandoned ratings and another 15% planned to do so soon (CEB Corporate Leadership Council 2016). Whatever the exact percentage, it's clear that a number of organizations have adopted ratingless reviews and there appears to be a trend in that direction.

However, there is relatively little research about the effects and effectiveness of the ratingless approach. There is a large volume of evidence that quantitative rating scales have many problems that different rating schemes and rater training have been unable to overcome (Adler et al. 2016). But this does not necessarily mean that ratingless reviews are

effective. Indeed, there is so little academic research on ratingless reviews that even recent reviews of the performance management literature typically do not even consider ratingless reviews. There is limited evidence available on two key questions: What are the effects of ratingless reviews on employees and the organization and what causes those effects?

There is more data about why organizations adopt ratingless reviews and what kind of effects they hope to obtain than about the effects of ratingless reviews. A large-scale survey by the Center for Effective Organizations that was sponsored by WorldatWork (Ledford, Benson, and Lawler 2016b; Ledford, Benson and Lawler 2016c) examined 244 organizations that had adopted one or more of three cutting-edge practices: ratingless reviews, ongoing feedback (feedback four or more times per year) and crowd-sourced feedback (the use of social media to collect feedback from peers and others). The three primary reasons that companies adopted cutting-edge practices (including ratingless reviews) were:

1. To improve the effectiveness of the process (e.g., better experience for employees and manager, more useful feedback, more time spent on coaching).
2. To improve strategic alignment (e.g., better support for the business strategy, increased organizational performance).
3. To help meet traditional reward objectives (rewarding, motivating, attracting).

Process effectiveness and strategic alignment were rated as more important than reward objectives and companies indicated that cutting-edge practices had a slight positive

impact on process effectiveness and strategic alignment, but little effect on rewards objectives. It was impossible to identify the separate effects of ratingless reviews in this study because only seven of 244 organizations used ratingless reviews without the other two practices; nearly all used ongoing feedback. However, ratingless reviews plus the other practices were more effective than ongoing feedback alone.

Other survey studies have found similar reasons for adoption and also suggest that it is difficult to disentangle the effects of ratingless reviews from other performance management innovations that are often adopted at the same time, especially ongoing feedback. The HCI study found that companies adopt ratingless reviews to:

- Improve process effectiveness

- Increase performance

- Reduce the time and resources devoted to ratings

- Help make the process more agile.

A study of 30 companies by Rock and Jones (2015a, 2015b) indicated that the reasons for adoption reflect the changing nature of work (annual reviews by one manager do not fit current work patterns), the need for better collaboration (because competition for ratings can prevent this), the need to attract and retain talent (to foster more frequent and developmentally oriented performance conversations) and the need for faster employee development. It is notable that most of these objectives could potentially be achieved without ratingless reviews, for example by implementing ongoing feedback.

The 2016 CEB survey found that organizations without ratings hope that employees will become more engaged, managers will spend more time coaching, the quality of performance conversations will improve and managers will better differentiate pay.

However, the CEB survey of almost 10,000 employees in these organizations found that employees in organizations using ratingless reviews report slightly less engagement and perceived fairness than employees who received ratings and spend fewer hours per year in performance conversations. Notably, the CEB study is the only one we have found that indicates negative effects for ratingless reviews. Because the data were collected at one point in time, it may be that ratingless reviews led to better outcomes, but adopters were still behind non-adopters. For example, companies that were poor at encouraging development may have adopted ratingless reviews for this reason — and may have improved as a result, just not to the level of others in the study. The study makes clear that other changes that are adopted at the same time as ratingless reviews (ongoing feedback, forward-looking reviews and peer input) determine whether the ratingless approach will be successful.

Some of the best evidence about the effects of ratingless reviews has been internal studies conducted by managers of companies that have them. These often include before-and-after data and even quasi-experimental designs testing ratingless reviews against the traditional system and even sometimes other interventions. Case reports include studies at Adobe Systems (Morris 2016); Cardinal Health (George and Holbein 2016); Microsoft (Ritchie 2016); Sears Holdings Corp. (Engler and Mason 2016); and DIRECTV (Jaffe et al. 2015). Common reasons for adopting ratingless reviews in these cases were an interest in implementing a less painful and more effective performance management process, a desire to increase feedback and coaching from managers and a desire to reinforce the organization's business strategy and/or culture. All the cases reported positive effects and no negative effects from the adoption of ratingless reviews.

The most sophisticated of these studies is the DIRECTV case, in which the organization studied three pilot groups that implemented biweekly one-on-one conversations with employees, separate conversations about pay and performance and increased training and resources for performance management. The three pilots had variations in ratings, using the organization's five-point scale, a simplified three-point scale or no ratings. The company tracked results over three years, including a baseline year, a pilot year and a year after the company adopted ratingless reviews companywide. The results indicated that there was higher employee satisfaction with performance management and a more equitable pay distribution under ratingless reviews and the trends were positive over time. Managers felt more empowered to make pay decisions and differentiate pay and they coached more. The organization also assessed 1,100 managerial comments from the reviews and found that the commentary became longer and more effective over the three years.

A difficult problem in prior research is disentangling the effects of ratingless reviews from other changes in performance management — especially ongoing feedback — that typically are adopted at the same time in organizations. Research has long established that performance management effectiveness and employee satisfaction with the process are tied to the quality of feedback, the frequency of feedback and to the degree to which feedback is oriented toward employee development (Aguinis 2014; DeNisi and Smith 2014; Schleicher et al. 2018). Some argue that ratingless reviews promote these outcomes by decreasing defensiveness, facilitating more frequent conversations and increasing a sense of fairness (Rock 2008; Rock and Jones 2015a). However, do the effects of ratingless reviews add significant incremental power to ongoing feedback and other

interventions that are designed to promote effective performance and development conversations? There is little prior research on this point.

Our discussion of the literature suggests several research questions that we investigated with a company that was adopting ratingless reviews in part of the organization. These questions are:

1. Do managers provide more frequent feedback to employees in a ratingless system than in a ratings-based system?

2. Do managers encourage more employee development in a ratingless system than in a rating-based system, as indicated by higher levels of coaching, more developmental feedback and more developmental goals?

3. Do managers provide higher quality feedback in a ratingless system?

4. Are employees more satisfied with a ratingless process than one with ratings?

5. Is a ratingless process associated with more favorable outcomes for employees, namely thriving and organizational commitment?

6. Does a ratingless process lead to more favorable outcomes for the organization, namely lower intentions to quit and higher organizational agility?

CEO RATINGLESS REVIEW STUDY

This study is noteworthy for several reasons. It is the most rigorous and comprehensive study of ratingless reviews to date. The study is quasi-experimental, with before-and-after data from both pilot and comparison groups. We are able to distinguish the effects of continuous feedback vs. ratingless reviews, which is difficult to do in studies in which both are adopted at the same time. Moreover, rather than using a small number of employee survey items to evaluate ratingless reviews, as in most prior case studies, we

7

collected data on a wide range of variables. Finally, we examine several hypotheses about why ratingless reviews may be effective. Some prior case studies have examined the effectiveness of ratingless reviews, but we are aware of none that have analyzed causal factors that may explain success or failure.

Method

Organizational Context

The organization for this study was a health insurer and provider of health-care services. The insurance company serves more than 2 million customers in one U.S. region. Subsidiary companies, covering more than 70 million customers nationwide, provide services that include technology that makes health care more economic and personalized.

The most important characteristic of the organization for purposes of the study is that it had a history of excellence in performance management practice and a strong performance feedback culture that required managers to provide regular, effective performance feedback (see Ledford and Schneider 2018). The goal of ratingless appraisals was not to fix a broken system, but to make an already excellent system better. This is very different than the motivation for some other adopters, who see ratingless systems as a solution to a broken performance management process (e.g., Morris 2016). Management hoped that the ratingless approach would complement its overall performance management approach.

The organization implemented several major performance management techniques prior to adopting ratingless reviews in pilot units. The organization used a simplified, three-point performance rating scale. One of those points was used only for the limited number of employees who were on performance improvement plans (PIPs). Therefore, the

ratings primarily served to identify top performers vs. others. Managers provided ongoing feedback at least four times per year for all employees and ongoing feedback was consistently practiced. The organization heavily emphasized coaching and development in communication and training on performance management. Quarterly performance conversations considered both operational and development goals and progress. The organization did not use crowdsourced feedback (feedback from peers and others using social media) but made available an optional 360-degree review process that covered about 40% of the workforce. In short, the organization had implemented a number of performance management innovations prior to adopting ratingless reviews in the pilots.

The organization reinforced its strong performance feedback culture in several ways before and after the pilots. The CEO and executive team modeled providing regular, open performance feedback and used their process with subordinate teams. There was extensive communication about the importance of performance management and the role of managers and employees in feedback sessions. All managers were offered training in providing effective feedback. The organization monitored the quality of manager feedback with questions on the employee survey and through spot audits of manager feedback. The organization also provided recognition to managers who exemplified the company's values through their behavior, including their performance feedback. Finally, processes for hiring and promoting managers took candidates' communication styles into account.

The organization implemented ratingless reviews in a pilot group that included members of the IT and HR departments. A comparison group was drawn from several different departments, including legal, marketing, development and public relations. Both groups were constituted primarily of exempt professional employees.

Study Data

The research team administered online surveys at two points in time to all managers and employees in the pilot and comparison units. Time 1 was three months before the introduction of ratingless reviews in the pilot units. Time 2 followed the completion of the next round of annual reviews. For the ratingless pilot units, this was six months after Time 1. The comparison units were on a later cycle and Time 2 in those units was completed 11 months after Time 1.

We collected survey data from 406 employees and managers in the pilot units at Time 1 and 427 at Time 2. From these overall samples, we were able to analyze data from a matched sample of 274 who completed both surveys. We collected survey data from 161 people at Time 1 in the comparison groups and 166 at Time 2, with a matched sample of 74 people who completed both surveys. Response rates were 45% at both points in time.

The pilot and comparison groups were closely matched on most dimensions. The workforce was primarily professional in both groups. The pilot group was 80% exempt employees and 15% managers, while the comparison group was 80% exempt and 16% managers. Average age was 47 in the pilot group and 44 in the comparison group. The pilot group was 80% Caucasian vs. 83% in the pilot group. The only significance between the two groups was gender: 42% of the pilot group was female, while 73% of the comparison group was female. Since we found that gender differences were not associated with differences on key study variables, this was not a problem for the analysis.

Measures

We collected many types of data and measured a number of variables from employee surveys and organizational records. Most of the study variables are well

established in the research literature. All multi-item survey variables reported in this study had very good to excellent measurement properties, such as factor structure and reliability. More information about the measures is available from the authors.

Demographics. We obtained a number of demographic variables from company records that we were able to match to survey responses. These included title, hierarchical level, gender, tenure and performance level in the year prior to the pilot.

Performance management process*.* We measured a number of aspects of the performance management process in both surveys. These included:

1. Number of meetings with the respondent's manager to discuss performance and/or development in the prior 12 months.
2. Emphasis of the feedback on evaluation vs. development (1 item).
3. Performance review satisfaction (4 items).

We also measured several survey variables only at Time 2 to assess the experience of ratingless reviews in pilot vs. comparison groups. These variables included preference for ratingless reviews (2 items), fairness of the pay administration process (4 items) and fairness of pay increases (4 items).

Manager feedback behavior. Survey respondents rated the behavior of their manager in their most recent performance management discussion. We measured feedback (5 items), coaching (5 items), providing consequences with recognition and rewards (3 items), goal setting (5 items) and communication (4 items). Furthermore, we analyzed performance review records downloaded from the HR information system to analyze the number of goals that were set in each performance review and classified the goals as performance-oriented or development-oriented.

Outcomes. Employee outcomes included employee thriving, a concept that has received considerable attention in recent years. It measures an employee's energy and engagement at work. We measured two dimensions (learning and vitality), each with 4 survey items. We also measured organizational commitment (6 items), which includes emotional attachment to the organization; it is similar to many operationalizations of employee engagement. We measured two organizational outcomes: intention to quit (4 items) and organizational agility (19 items). Turnover is a costly employee behavior that most organizations seek to minimize. Prior research has shown that survey responses on the turnover-intentions measure are highly predictive of whether an employee will actually leave the organization. Organizational agility indicates the degree to which the organization has built a culture that is able to act innovatively and manage change effectively. This variable was especially relevant to the organization we studied because senior leadership had been devoted for several years to multiple initiatives intended to increase organizational agility (Worley, Williams, and Lawler 2014), and the leadership team viewed ratingless reviews as a way to make the key HR process of performance management leaner, more adaptive to organizational needs and more agile.

Analyses

In addition to examining descriptive statistics (means, standard deviations, frequencies, etc.), we conducted several different analyses. We conducted T-tests to examine changes from Time 1 to Time 2. These examined whether there was a significant change in mean responses between the pilot and comparison groups at each point in time and whether each group changed from Time 1 to Time 2. Because we were able to track individual respondents, we were also able to do the same T-Tests using matched samples.

Using matched samples helps ensure that any differences are not due to different respondents completing the survey at different points in time. However, the matched samples are smaller because of the number of people who responded at only one point in time. In general, the results for the analyses were very similar for the full samples and the matched samples. Therefore, we will report results for the full samples here and note any important differences for the matched samples in our commentary.

We also conducted multiple regressions to look at the importance of the ratingless pilot vs. other variables, such as frequency of feedback and the degree to which the feedback was developmental, in explaining performance review satisfaction and other outcomes.

Results

Tables 1–4 report the means for the comparison and pilot groups on each measure and indicate whether the difference between Time 1 and Time 2 was statistically significant within each group (for example, within the pilot group from Time 1 to Time 2) and whether the difference was statistically significant across groups — that is, whether the comparison and pilot groups were different at Time 1 and Time 2. Because almost all the prior literature indicates that ratingless reviews have either positive or neutral effects, we used a one-tailed test of significance.

Feedback Frequency

Table 1 indicates that members of both groups received a great deal of feedback by contemporary standards. For the entire sample, 66% received feedback at least quarterly and 21% received feedback monthly. On average, members of the comparison group had more than five conversations about performance and/or development with their manager

in the twelve months prior to both Time 1 and Time 2. On average, members of the pilots had more than seven conversations during these times. The changes were not significant for either group. That indicates that the pilot did not lead to an increased amount of conversations. This number was already high compared to other organizations for both groups.

Emphasis on Development

Table 1 provides evidence that ratingless reviews are associated with an increase in developmental conversations. The "evaluation vs. development" measure uses a seven-point scale, where "1" indicates the conversation is entirely about evaluation of performance, "7" indicates that the conversation is entirely about employee development and "4" indicates an equal emphasis on both. For both groups, there was a mix of performance and development in the conversation, with slightly more emphasis on performance. The pilot group was not significantly different from the comparison group at either Time 1 or Time 2, but the pilot group showed a significant increase in developmental emphasis from one time to the next. This was true for both the overall sample (shown) and the matched sample. The difference within the comparison group was not significant.

For the pilot group only, we were able to obtain the actual text of goals and annual reviews. We found that for this group, the number of goals decreased to 5.2 from 6.3, on average, which is statistically significant. More important, the ratio of developmental to operational goals increased to 30% from 25%, a highly significant result. That indicates that managers were not simply adding developmental goals but were changing the mix of goals they set with employees.

Manager Feedback Behavior Overall

Table 1 indicates that performance conversations in the pilot group were more likely than those of the comparison group to provide good feedback overall, to provide consequences such as recognition and rewards, to engage in effective goal setting and to use good communication techniques at both Time 1 and Time 2. For feedback overall and coaching, the changes in the pilot group improved marginally, but these differences were not significant in the matched sample. Other measures of the perceived quality of feedback were not significantly different within each group between Time 1 and Time 2. Because the differences existed before ratingless reviews were adopted, we cannot conclude that they were due to the intervention.

Satisfaction with Performance Reviews

The pilot group was more satisfied with their performance reviews than the comparison group at both Time 1 and Time 2. However, the differences within groups were not significant at Time 2; the pilot group did not become more positive over time on this measure.

At Time 2 only, we asked respondents to indicate their preference for ratingless reviews. The pilot group was significantly more positive on both measures than the comparison group. (See Table 2.) In response to the question, "I prefer a performance review in which the manager does not use a performance rating label," 42% of the pilot group agreed (vs. 35% of the comparison group), 30% were neutral and 28% disagreed. On the question, "A performance review without a performance rating label leads to a better conversation between managers and employees than does a review with a rating," 46% of the pilot group agreed (vs. 39% of the comparison group), 27% were neutral and 27% disagreed. These results indicate that opinions were varied, but those in the pilot were

much more likely to endorse ratingless reviews and that more respondents were positive about ratingless reviews in the pilot group. We did not ask the questions at Time 1 because the questions would not have been relevant before the organization had experienced ratingless reviews. Therefore, we cannot know whether a ratingless process changed these responses.

Finally, we asked questions at Time 2 only about the fairness of the pay process and about the fairness of pay increases. On both of these measures, the pilot group was significantly more favorable, with a majority responding positively on these measures. Because we did not ask these questions at Time 1, we cannot know if the ratingless intervention resulted in more favorable responses. However, clearly the experience of a ratingless review did not lead pilot employees to believe that they received unfair compensation treatment as a result of those reviews.

Outcomes for Employees

We measured two types of employee thriving as well as organizational commitment (see Table 3). These three measures are related to, but distinct from, employee engagement. *Thriving-learning* measured whether employees feel as though they are continually learning and developing at work (e.g., "I find myself learning often"). This was very high for both groups at both points in time, with the pilot group significantly higher at both times. There was a marginally significant change for the pilot group from Time 1 to Time 2, with no change for the comparison group. However, the change in the pilot group was not detected in the matched sample. There were no significant changes within groups from Time 1 to Time 2. *Thriving-vitality* (e.g., "I feel energetic and vital at work") was significantly higher for the pilot group at both points in time, although the changes within

groups over time were not significant. *Organizational commitment* measures employee identification with and attachment to the organization (e.g., "This organization has a great deal of personal meaning for me"). As for many other measures, the pilot group was more positive than the comparison group at both points in time, but neither group became more or less positive over time. However, in the matched sample only (not shown), there was a significant improvement in the comparison group but not in the pilot group. Overall, we cannot conclude that ratingless reviews led to more favorable employee outcomes on any of the three measures.

Outcomes for the Organization

Table 3 also shows the results for two organizational outcomes. Intention to quit was low in both groups but showed a complex pattern of results. Turnover intentions were marginally lower for the full sample at Time 1 in the pilot group and were significantly lower in the pilot group at Time 2 for the full sample. In the matched sample, however, turnover intentions were significantly lower for the pilot group at Time 1 but not at Time 2, but significantly increased for the pilot group over time. This pattern is difficult to interpret but suggests a possible increase in turnover intentions in the pilot group.

Organizational agility was significantly greater at Time 2 for the pilot group compared to the comparison group for the full sample but not the matched sample. For both samples, the level of agility increased significantly in the pilot group and did not significantly change in the comparison group. Thus, it appears that there was a positive impact on a strategic outcome that was very much of interest to senior management.

Do Ratingless Reviews Predict PM Effectiveness?

We conducted multiple regression analyses to investigate the relative importance of ratingless reviews compared to frequency of feedback, coaching and the degree to which feedback is developmental to various outcomes. Table 4 presents results for regressions using performance review satisfaction, thriving (combining both learning and energy subscales) and organization commitment. The numbers in the table are standardized regression coefficients (beta), a measure of association ranging from 0 to 1. Betas are analogous to correlation coefficients. We controlled for a number of variables including gender, organization level, department, years of service and level of performance in a first step of the regression. None of these were significant for any of the three outcomes. This analysis allowed us to directly compare the effects of feedback frequency, coaching behavior and perceptions of developmental feedback on employee attitudes and performance management satisfaction while controlling for whether the employee received a performance rating. We found that feedback frequency and developmental feedback were important predictors of these outcomes, while being in the ratingless condition was not. Coaching behavior by the manager was by far the most important predictor. Together, these predictors accounted for 45%, 21% and 20% of the variance in the outcomes, respectively, indicating very strong relationships.

DISCUSSION

A consistent finding across many variables is that the pilot group was more positive than the comparison group — but it was also more positive at Time 1, before the ratingless reviews intervention. In most cases, the changes between Time 1 and Time 2 for both the pilot group and the comparison group were not statistically significant. Therefore, we cannot conclude that ratingless reviews led to more positive results in most cases.

Areas in which the ratingless pilot appears to have had a positive effect is in the increasing focus on development (on some indicators but not all), support for ratingless reviews and organizational agility. Ratingless reviews did not have any clear negative effects. There was not even any apparent negative effect on perceived fairness of the pay process or pay raises, despite concerns raised about how pay can be administered effectively without ratings (Ledford, Lawler, and Benson 2016).

This pattern shows the importance of measuring results before and after the change to a ratingless system in both a pilot and comparison group. If we only examined data from Time 2, we would be likely to conclude that there were systematic differences that were the result of the ratingless reviews, not realizing that the differences existed before the intervention.

The regression analyses suggest that most of the benefits claimed for ratingless reviews are not the result of that change, per se, but rather are the result of two important factors that prior research has shown to be related to performance management effectiveness: the frequency of feedback and especially the degree to which the feedback is developmental. Such benefits can be obtained through interventions other than ratingless reviews. The most notable alternative is the use of ongoing feedback, which in practice almost always accompanies the adoption of ratingless reviews. Many of the benefits claimed for ratingless reviews probably are due to ongoing feedback. We were in the unusual position of being able separate out the effects of these two interventions because ongoing feedback was already being used through the organization prior to Time 1.

Emphasizing development to a greater degree does not mean avoiding performance assessment in review meetings. Feedback for the average respondent was slightly more

oriented toward performance evaluation than development and very few managers

ignored evaluation. The emphasis on development is simply a matter of degree.

A limitation of our study is that it was in an organization that had an advanced, well-implemented performance management process prior to the adoption of ratingless reviews. This may not be typical of other organizations. Ratingless reviews were not part of a fundamental transformation of a broken process. Results might differ in organizations with poor performance management that use ratingless reviews and other changes to signal that management is resetting the entire process.

Finally, we did not assess the impact of ratingless reviews on corporate performance. Future studies are needed to determine the effects of ratingless reviews on performance. Removing ratings certainly can reduce administrative costs, as in Adobe's estimate of saving 80,000 hours of management time annually with a ratingless approach (Morris 2016).

ABOUT THE AUTHORS

Gerald E. Ledford, Jr., Ph.D. , gledford@usc.edu, is affiliated research scientist at the Center for Effective Organizations, Marshall School of Business, University of Southern California. Since 2004, he has been president of Ledford Consulting Network LLC. From 1998-2003, he was at Sibson Consulting, where he was senior vice president and practice leader, Employee Effectiveness. He received his Ph.D. in Psychology from the University of Michigan. He is author of 145 articles and 12 books and monographs. He was the winner of the 2012 WorldatWork Journal Author of the Year award. For more information, visit http://ceo.usc.edu.

George S. Benson, Ph.D., benson@uta.edu, is chairman of the Department of Management and an associate professor in the College of Business, University of Texas at Arlington. Benson earned his Ph.D. at the University of Southern California and is an affiliated researcher at the Center for Effective Organizations. He previously worked as a research analyst at the American Society for Training and Development. He also holds degrees from Washington and Lee University and Georgetown University. For more information, visit https://mentis.uta.edu/explore/profile/george-benson.

REFERENCES

Adler, Seymour, Michael Campion, Alan Colquitt, Amy Grubb, Kevin Murphy, Rob Ollander-
     Krane, and Elaine Pulakos. 2016. "Getting Rid of Performance Ratings: Genius or Folly?
     A Debate." *Industrial and Organizational Psychology* 9(2): 219-52.

Aguinis, H. 2014*. Performance Management,* 3rd ed. New York: Pearson Education.

Cappelli, Peter and Anna Tavis. 2016. "The Performance Management Revolution," *Harvard
     Business Review*, October, 58-67.

CEB Corporate Executive Council. 2016. "The Real Impact of Eliminating Performance
     Ratings: Insights from Employees and Managers." Washington, DC: Corporate
     Executive Board. Viewed: Sept. 6, 2019.
     https://www.cebglobal.com/content/dam/cebglobal/us/EN/best-practices-
     decisionsupport/.

DeNisi, Angelo S., Kevin R. Murphy, and Gilad Chen. 2017. "Performance Appraisal and
     Performance Management: 100 Years of Progress?" *Journal of Applied Psychology*
     102(3): 421-433.

DeNisi, Angelo S. and Caitlin E. Smith. 2014. "Performance Appraisal, Performance
     Management and Firm-Level Performance." *Academy of Management Annals* 8: 127–
     179.

DeNisi, Angelo S. and Shirley Sonesh. 2011. "The Appraisal and Management of
     Performance at Work." In *APA Handbook of Industrial and Organizational Psychology*,
     Vol 2: Selecting and Developing Members for the Organization, ed. Sheldon Zedeck,
     255-279. Washington, DC: American Psychological Association Press.

Engler, Holly, and Chris Mason. 2016. "Enabling New Levels of Performance at Sears

   Holdings Corp." *WorldatWork Journal* 25(2): 35-46.

Human Capital Institute. 2015. *Talent Pulse: No Review, No Problem*. Cincinnati, OH: HCI.

Jaffe, Jennifer, Maren Dollwet, Traci Conner, Katelynn Duffel, and Matthew Lucy. 2015.

   "From Performance Management to Forward Focus — How DIRECTV Revolutionized

   the Process, Experience, and Culture of Feedback." Center for Effective Organizations

   Publication G15-20(667).

Ledford, Gerald E. Jr., George Benson, and Edward E. Lawler III. 2016a. "Aligning Research

   and the Current Practice of Performance Management." *Industrial and Organizational*

   Psychology 9(2): 253-260.

Ledford, Gerald E. Jr., George Benson, and Edward E. Lawler III. 2016b. "A Study of Cutting-

   Edge Performance Management Practices: Ongoing Feedback, Ratingless Reviews and

   Crowdsourced Feedback." *WorldatWork Journal* 25(2): 8-24.

Ledford, Gerald E. Jr., George Benson, and Edward E. Lawler III. 2016c. "Cutting-Edge

   Performance Management: 244 Organizations Report on Ongoing Feedback, Ratingless

   Reviews and Crowdsourced Feedback." Scottsdale, AZ: WorldatWork.

Ledford, Gerald E. Jr., Edward E. Lawler III, and George Benson. 2016. "Cutting-Edge

   Performance Management: What About Rewards?" *Workspan*, April, 12-15.

Morris, Donna. 2016. "Death to the Performance Review: How Adobe Reinvented

   Performance Management and Transformed Its Business." *WorldatWork Journal*

   25(2):25-34.

George, Lisa and Julie Holbein. 2016. "Engaging Employees to Transform Performance

   Management at Cardinal Health." *WorldatWork Journal* 25(2): 47-59.

Pulakos, Elaine D., Rose Muller-Hansen, and Sharon Arad. 2019. "The Evolution of

    Performance Management: Searching for Value." *Annual Review of Organizational*

    Psychology and Organization Behavior 6: 249–271.

Ritchie, J. 2016. "Transforming a Company: How Microsoft's New Employee Performance

    System Supports Its Business and Cultural Transformation." *WorldatWork Journal*

    25(2): 60-75.

Rock, David. 2008. "SCARF: A Brain-Based Model for Collaborating with and Influencing

    Others." *NeuroLeadership Journal*: 1(1).

Rock, David and Beth Jones. 2015(a). "Why More and More Companies Are Ditching

    Performance Ratings." *Harvard Business Review* (online), Sept. 10. Viewed: Sept. 4,

    2019. https://hbr.org/2015/09/why-more-and-more-companies-are-ditching-

    performance-ratings#.

Rock, David and Beth Jones. 2015(b). November 6. "What Really Happens when Companies

    Nix Performance Ratings." *Harvard Business Review* (online), Nov. 6. Viewed: Sept. 4,

    2019. https://hbr.org/2015/11/what-really-happens-when-companies-nix-

    performance-ratings.

Schleicher, Deidra J., Heidi M. Baumann, David W. Sullivan, and Junhyok Yim. 2019.

    "Evaluating the Effectiveness of Performance Management: A 30-year Integrative

    Conceptual Review." *Journal of Applied Psychology* 104(7): 851-887.

Worley, Christopher G., Thomas Williams, and Edward E. Lawler III. 2014. *The Agility Factor*.

    San Francisco: Jossey-Bass.

TABLE 1 Performance Management Process Results: Group Differences

Variable	Time 1 Means		Time 2 Means		Differences Between Comparison and Pilot Groups		Differences Within Comparison Group	Differences Within Pilot Group
	Comparison	Pilot	Comparison	Pilot	Time 1	Time 2	Times 1 - 2	Times 1 - 2
Number of meetings – past 12 months	5.36	7.33	5.27	7.56	**	**	ns	ns
Evaluation vs. development emphasis	3.17	3.28	3.31	3.45	ns	ns	*	ns
Manager behavior:								
- Feedback quality	4.43	4.80	4.59	4.93	**	**	+	ns
- Coaching	4.50	4.66	4.53	4.82	ns	*	+	ns
- Provides consequences	4.27	4.56	4.30	4.61	*	*	ns	ns
- Goal setting	4.53	4.93	4.56	5.06	**	**	ns	ns
- Communication style	5.37	5.79	5.47	5.81	**	**	ns	ns
Performance review satisfaction	5.29	5.62	5.41	5.73	**	**	ns	ns

Notes:
1. N = 406 at T1 and 427 at T2 for pilot group, 161 at T1 and 166 at T2 for comparison group.
2. Analysis uses one-tailed T-tests.
2. Significance levels: ns = nonsignificant, + <.10, * <.05, **<.01

TABLE 2 Assessments at Time 2 Only

Variable	Time 2 Means		Difference Between Comparison and Pilot Groups
	Comparison	Pilot	Time 2
Performance reviews without label leads to a better conversation	4.14	4.43	*
I prefer a performance review without a label	4.04	4.35	*
Pay process fairness	4.41	4.70	*
Pay increase fairness	4.30	4.61	*

Notes:
1. N = 406 at T1 and 427 at T2 for pilot group, 161 at T1 and 166 at T2 for comparison group.
2. Analysis uses one-tailed T-tests.
2. Significance levels: ns = nonsignificant, + <.10, * <.05, **<.01

TABLE 3 Performance Management Outcomes for Employees and the Organization: Group Differences

Variable	Time 1 Means		Time 2 Means		Differences Between Comparison and Pilot Groups		Differences Within Comparison Group	Differences Within Pilot Group
	Comparison	Pilot	Comparison	Pilot	Time 1	Time 2	Times 1 - 2	Times 1 - 2
Thriving – Learning	5.87	5.87	4.72	4.92	ns	ns	ns	ns
Thriving – Energy	5.04	5.40	5.09	5.38	**	**	ns	ns
Organization commitment	4.63	4.93	4.72	4.92	**	+	ns	ns
Intention to quit	2.94	2.72	3.15	2.76	+	**	ns	ns
Organizational agility	2.78	2.77	2.76	2.85	ns	*	*	ns

Notes:

1. N = 406 at T1 and 427 at T2 for pilot group, 161 at T1 and 166 at T2 for comparison group.

2. Analysis uses one-tailed T-tests.

2. Significance levels: ns = nonsignificant, + <.10, * <.05, **<.01

TABLE 4 Regression Results: Ratingless Reviews Vs. Number of Reviews and Developmental Feedback

Predictors	Dependent Variables		
	Performance Review Satisfaction	**Thriving (Learning + Energy)**	**Organization Commitment**
	Beta / Significance	**Beta / Significance**	**Beta / Significance**
Number of reviews – prior 12 months	.12 **	.08 ns	.15 **
Manager behavior: Coaching	.56 ***	.35 ***	.33 ***
Evaluative vs. developmental feedback	.15 **	.19 **	.12 *
Ratingless reviews	-.03 ns	.03 ns	-.04 ns

Notes:
1. N =593 (all respondents at Time 2)
2. Numbers shown are standardized regression coefficients (betas); higher numbers indicate stronger associations.
3. Not shown are control variables entered in Step 1 of the regression; betas for all controls were nonsignificant.
2. Significance levels: ns = nonsignificant, + <.10, * <.05, ** <.01, ***<.001