

AN EXERCISE TO TEACH PERFORMANCE REVIEW SKILLS AND AN ANALYSIS OF ITS EFFICACY

James W. Smither
La Salle University

ABSTRACT

From the perspectives of both the manager and the employee, a fair and helpful performance review requires the manager to possess two skill sets. First, managers need the skills to provide accurate (evidence-based) ratings. Second, they need the interpersonal skills to facilitate a constructive performance review conversation. But how can undergraduate students develop these skills before they are working in a management role? The multi-part (including a role-play of a performance review meeting) experiential exercise described in this article is designed to help upper-level students develop both of these skills. I used an experimental design to evaluate the efficacy of the exercise. The results show that the exercise enhances students' skills at facilitating a performance review meeting and provides detailed documentation to support their ratings of an employee.

Keywords: performance ratings, performance appraisal, performance management, performance reviews, performance feedback, experiential learning

INTRODUCTION

Undergraduate textbooks in human resource management often include a chapter about performance management but acquiring knowledge of performance management principles is very different from demonstrating the skills that supervisors need to conduct effective performance management reviews. This paper has two, closely related purposes. The first purpose is to present a multi-part experiential exercise (including a role-play) to help students develop two important skills that are required to conduct an effective performance review: completing accurate (evidence-based) ratings of performance and facilitating a constructive performance review conversation. The second purpose is to describe a research study, using an experimental design, that demonstrates the efficacy of the exercise. Presenting these results should enhance readers' confidence that the exercise can help their students develop important skills required of effective managers.

In recent years, there has been a debate about performance management and performance ratings (Adler, Campion, Colquitt, Grubb, Murphy, Ollander-Krane, & Pulakos, 2016). Some scholars have argued that performance management systems are broken and do not motivate employees (Bernardin, Hagan, Kane, & Villanova, 1998; Pulakos, & O'Leary, 2011). Grubb

(2007) described many problems associated with the use of traditional performance appraisals (e.g., short-term focus, creating rivalries, and negative effects on teamwork) and argued that their use should be discontinued. Others have argued that supervisors too often do not provide accurate feedback or useful ratings and that ratings should be decoupled from compensation (Colquitt, 2019). Practitioners note that, with or without ratings, performance is always evaluated in organizations (Effron, 2015). That is, “judgments are made and decisions about rewards and roles in the future are based on those judgments” (Adler et al., 2016, p. 234). Moreover, performance ratings are ubiquitous and inescapable. That is, performance ratings are the only feasible method for evaluating performance in most settings because ‘objective’ measures are nearly always deficient in some respects. Also, research has shown that strong performers tend to leave organizations where high performance is not rewarded (while weak performers are likely to remain in such organizations) (Adler et al., 2016). Despite the debate, there is little evidence that performance ratings are likely to disappear anytime soon (Smither, 2015).

The value of helpful, constructive feedback is widely accepted. And recommendations for guiding feedback delivery have been offered by many (Hauenstein, 1998; London, 2015). London (p.8) suggests that effective feedback is clear, focused on behavior (not personality), and frequent. He further suggests that constructive feedback avoids comparisons between people. It also offers suggestions to the employee about specific ways to improve performance.

In almost all organizations, performance ratings and feedback come together in the performance review. But managers are often apprehensive about conducting a performance review conversation and too often lack the training to do so (Aguinis, 2009).

Although several papers have focused on teaching students how to deliver effective feedback (e.g., Hess, 2007; Michaelsen & Watson, 1987; Seltzer, 1986), the management education literature has directed little attention to teaching students the skills associated with performance management. Bull-Schaefer (2018) described the use of a performance appraisal week during which the instructor provides mid-course feedback to each student in a manner that parallels mid-year performance reviews in business settings. Although the exercise does not teach students how to conduct a performance review themselves, it provides them with the opportunity to experience the emotions associated with receiving a mid-year review and helps them engage in a conversation that leads to planning for the rest of the course (year). An exercise developed by Nandedkar (2018) taught students how to evaluate the effectiveness of three different performance appraisal methods (graphic rating scale, behavioral observation scale; and essay method) and recommend which method would best fit the needs of a hypothetical organization. Gruys and Stewart (2007) describe an exercise where participants are asked to conduct an audit of an organization’s performance appraisal process, including decisions about salary and promotions. Students are asked to describe the strengths and limitations of the process and to provide recommendations for improvement. Hannay (2010) presented an exercise in which students, acting as supervisors, make recommendations about salary increases for eight employees (based on brief, three-sentence notes about each employee’s performance). Similarly, an exercise by Hyland (2017) asked students to review performance review forms concerning two employees and then decide (and subsequently discuss) which of those employees should receive a \$1,000 bonus. A paper by Rachman-Moore and Kenett (2006) focused on creating performance management systems but included no exercises to help develop students’ skills in actually conducting

performance reviews. While dealing with the broad topic of performance management, none of these six exercises teach students how to conduct a performance review. In sum, there has been almost no attention to helping students learn how to conduct performance reviews.

One notable exception is an excellent paper by Mumford (2009). The Mumford paper described an exercise where each student acted as a supervisor in a role-play and conducted a performance review with another student who played the role of a corrective maintenance mechanic. Before conducting the performance review, the student-supervisor first received information concerning the employee's performance, which included multisource ratings about the employee for the current and previous year and qualitative information in the form of brief notes about the employee's performance on four dates throughout the year (e.g., "8-15 - Went 25% over budget for the quarter ordering additional inventory"). Students completed the role-play twice: before classroom training about performance reviews and after the classroom training. During each role-play, the student-supervisor conducted a performance review by providing verbal feedback and written ratings concerning the employee's performance. The pre-training and post-training role-plays were video-recorded and scored by graduate students who were blind to whether a video was recorded pre-training or post-training. Mumford described very positive reactions from the students who participated in the exercise but there was no statistically significant difference between pre-training and post-training videos as rated by the graduate students. (The Mumford paper also included two other role-play exercises: providing a written warning to an employee and conducting an employee termination; both of these role-play exercises showed statistically significant improvement from pre-training to post-training as rated by graduate students.)

The objectives of the experiential exercise described in this paper are for students to provide accurate (evidence-based) ratings of an employee's job performance, to provide written notes (based on the employee's behavior throughout the year) that support those ratings, and to facilitate a constructive and helpful performance review meeting with the employee. During the performance review meeting, students explain the purpose of the performance review process (e.g., to provide specific feedback, review the employee's strengths and any opportunities for improvement, and discuss whether the employee is interested in pursuing other opportunities at the company), encourage the employee to ask questions and participate throughout the conversation, tell the employee how he or she was rated on each performance dimension and provide specific examples of the employee's behavior to support each rating, summarize the employee's strengths as well as any opportunity for improvement, discuss the employee's career interests and offer suggestions about how the employee can pursue those interests and thank the employee for her or his contributions throughout the year.

The current experiential exercise differs from the Mumford (2009) exercise in several important respects. The focal job was a server in a restaurant rather than a corrective maintenance mechanic. The restaurant server job was chosen because, unlike the corrective maintenance mechanic job used in Mumford's exercise, many students have worked as servers and those who have not are familiar with the role (due to their experiences dining in restaurants). Student-supervisors were provided more behavioral incidents (see Appendix), 31 versus the four used in the Mumford exercise, concerning the server's job performance throughout the previous 12 months. Student-supervisors were not provided with multisource ratings about the server (as these

are not widely available in many applied settings). The instructor (rather than another student) played the role of the server. This provided a degree of uniformity (standardization) in the employee's behavior during the role-plays, thereby ensuring that differences between students (acting as supervisors) could not be attributed to differences in the behavior of those playing the role of the employee. Finally, the research design assigned students at random to one of two conditions. In the control condition, each student completed the performance review exercise twice (similar to Mumford, 2009); at the beginning of the semester *before* receiving any training on performance reviews, and toward the end of the semester, *after* receiving training on performance reviews. In the post-only condition, each participant completed the performance review exercise only once, toward the end of the semester, *after* receiving training on performance reviews. Like Mumford (2009), all role-plays were video-recorded and scored by judges (in this study, the judges were other faculty members rather than graduate students) who were blind to the student's research condition. The research design used here overcomes a key disadvantage of a simple pre-post only design where changes in behavior from pre to post can be due to practice effects, student learning, or a combination of the two.

Part 1: The Experiential Exercise

Overview. The exercise takes place over several class sessions. During the first class session (50 minutes), the instructor (a) provides a brief overview of the exercise for students, (b) delivers a lecture that includes an overview of performance management principles and guidelines, including the purpose of performance management, rating accuracy (and avoiding rating errors), and suggestions for facilitating a constructive performance management review. Students are also given notes that the supervisor has taken about the employee's behavior throughout the year and are given the company's rating form and dimensions. In separate sessions that last 15 to 20 minutes, each student participates in a role-play of a performance management review meeting (with the instructor playing the role of the employee). The role-play sessions ideally take place in a small office with a desk and two chairs. The student (who is playing the role of the supervisor) typically sits behind the desk and the employee (role-played by the instructor) sits on the other side of the desk. With a class of 20 students, the cumulative time devoted to the role-plays is 6.67 hours. I usually devote class time to conduct about half the role-plays (the class does not meet as a whole on these days) while the remaining role-plays are conducted before or after the usual class meeting time (or at times that are convenient for the student and instructor).

Step-by-step instructions. All handouts are provided in the Appendix. I place all the materials for the exercise on the course website.

1. Provide a brief overview by explaining the purpose of performance management in organizations. For example, I emphasize that performance management is more than performance appraisal. It includes setting goals and tracking progress, facilitating performance (by removing obstacles, providing adequate resources, and offering helpful feedback and coaching), and encouraging performance by providing valued rewards and recognition in a timely manner. I explain that effective managers provide timely and helpful feedback throughout the year. Finally, I note that performance reviews are often (but not always) used to influence decisions about compensation, promotions, training needs, and dismissals.

2. Explain that they will be learning how to complete performance ratings and facilitate a performance review discussion with an employee. Emphasize that nearly all of them will become a supervisor at some point (likely sooner than they think) and they will be required to rate the people who report to them and facilitate an annual performance review discussion with each of their employees. Tell them they will be playing the role of a manager in a restaurant and they will complete performance ratings for, and conduct a performance review session with, Chris Ramos, a server in the restaurant.
3. Distribute copies of notes about the performance of Chris Ramos throughout the year (see Appendix A, Performance documentation concerning Chris Ramos). I explain that good supervisors take notes throughout the year about employee performance and that those notes can include examples/incidents of good or poor performance. These notes ensure that the supervisor does not have to rely only on his or her memory months later. I ask the students to consider the notes about Chris Ramos' performance to be notes that *they* have taken about Chris' performance throughout the year. It is also important to explain that, consistent with best practices in performance management (Aguinis, 2019; Smither, 2012), the supervisor (i.e., the role they will be playing) provided ongoing feedback to Chris throughout the year. For example, in March, Chris and another employee got into a verbal argument and their remarks could be overheard by customers at several tables. In addition to making a note about this incident, the supervisor spoke with Chris shortly after it occurred to provide timely feedback.
4. Distribute the rating form that they will use to evaluate Chris' performance (see Appendix B). Note that each of Chris' behaviors can be assigned to one of the dimensions on the performance rating sheet. For example, being "10 minutes late on January 2" is relevant to the Punctuality dimension and a "large dining party complimented Chris' work ('very friendly and attentive')" is relevant to the Quality of Service and Interactions with Customers dimension. "I have observed Chris 'up-selling' more often" is relevant to the Knowledge of the Menu Offerings dimension. Explain that *each* of Chris' behaviors (as shown in Appendix A) should be assigned to the relevant dimension and the rating of Chris on each dimension should reflect those behaviors.
5. Provide a lecture that includes an overview of performance management principles and guidelines, including rating accuracy (and avoiding rating errors), and suggestions for facilitating a constructive performance management review (see Appendix C).
6. I distribute a sign-up sheet that lists 20-minute blocks on multiple days and ask students to put their name next to a time slot that is convenient for them. I remind students to dress business casual and to bring two copies of their typed, completed rating form (one for me and one for them to refer to during our meeting).
7. During the one-on-one performance review sessions, I (playing the role of the server, Chris Ramos) am careful to respond in similar ways to each student. This ensures that differences in the behavior of the students reflect differences in their skills and styles

rather than any idiosyncratic behavior on my part. For example, the supervisor usually states that I was late 7 times between January and mid-July, but things have improved since mid-July (e.g., Chris was late only once between mid-July and the end of the year). I might respond by saying, “I live about 20 minutes from the restaurant, so I usually left my home about 20 minutes before my shift started. Of course, if there was a lot of traffic or other problems, I was late. But after you (the supervisor) spoke with me in July, I started leaving home about 35 minutes before my shift starts, and this has really reduced the number of times I have been late.” The supervisor usually mentions the verbal argument that I (Chris) had with a coworker in March. I might explain that “I really like the people here and that guy is the only person I just don’t get along with. Whether it’s sports, movies, books, or politics, we never agree. I was embarrassed by the argument and especially embarrassed that some customers overheard it (and tipped less). Since then, I have realized that I don’t have to like everybody I work with, so I avoid any casual conversation with him. We’ve worked the same shift numerous times since then, and there has never been another problem.” Instructors might create their own ways of responding during the performance review meetings, but it is important that the instructor (employee) adopts a consistent approach with different students (supervisors).

To debrief the exercise, I first provide individual feedback to each student about his or her performance during the role-play. I also return my comments on the performance appraisal rating form that the student completed. Next, with the class as a whole, I discuss the performance appraisal ratings. I ask them what rating Chris deserved for each performance dimension (Appendix D contains an explanation of a justifiable, evidence-based rating for each dimension). I also emphasize that the best appraisals contained detailed examples to support each rating (e.g., the specific dates when Chris was late).

In my experience, students (supervisors) sometimes struggle with giving negative feedback to Chris. For example, some don’t mention the argument that Chris had with a coworker in March. Almost all students mention that Chris was late too often. But some soften this feedback by saying something such as, “I know that sometimes things happen that make us late.” I explain that such statements diminish the seriousness of the problem and Chris’ responsibility to arrive on time every day. Others fail to mention the consequences of Chris being late (e.g., sets a bad example for coworkers and causes resentment, forces other servers to stay after their shift should end, causes delays in seating and serving customers).

Part 2: Evaluating the Efficacy of the Experiential Exercise

In 2015 and 2016, I asked students in my class to provide ratings on a 0 to 10 scale of (a) how comfortable they think they would be conducting a performance review meeting with an employee who reports to them and (b) their level of knowledge about how to conduct an effective performance review. Students answered these questions about five weeks *before* participating in the performance review exercise and completed them again several days *after* participating in the performance review exercise (but before receiving feedback on their effectiveness in the exercise). (See Appendix E).

In 2015 (N = 35), ratings for the ‘comfortable’ item increased from the pre-test (M = 4.99, SD = 2.26) to the post-test (M = 8.50, SD = 1.09; $d = 2.09$; $t = 10.22$, $df = 34$, $p < .001$). Ratings for the ‘level of knowledge’ item increased from the pre-test (M = 4.43, SD = 2.88) to the post-test (M = 8.81, SD = 1.10; $d = 2.20$; $t = 9.47$, $df = 34$, $p < .001$).

In 2016 (N = 25), ratings for the ‘comfortable’ item increased from the pre-test (M = 3.80, SD = 2.53) to the post-test (M = 8.32, SD = 1.11; $d = 2.48$; $t = 9.93$, $df = 24$, $p < .001$). Ratings for the ‘level of knowledge’ item increased from the pre-test (M = 3.72, SD = 2.59) to the post-test (M = 8.88, SD = 0.82; $d = 3.03$; $t = 10.00$, $df = 24$, $p < .01$).

In both years, the effect sizes for the increases on both items (expressed as Cohen’s d) are very large (Cohen, 1988; *Statistics for Psychology: Null Hypothesis Testing and Effect Sizes*, 2019). In sum, student reactions indicate that they became more comfortable and believed they were more knowledgeable after participating in the exercise described above. Stated differently, students’ self-efficacy (Gist, 1987) increased substantially.

In addition to student reactions, researchers have developed several methods to determine whether specific teaching methods are useful (Kosslyn, 2018). For example, we can use an experimental design in which students are assigned at random to either receive instruction (the treatment group) or not receive instruction (a control group). Both groups can then be tested on the same dependent variables. In this study, I used an experimental design to determine the efficacy of the exercise described above. I hypothesized that, relative to students who had not yet received training about performance reviews, students who had received training would provide more accurate (evidence-based) ratings of an employee’s job performance, provide more detailed written notes (based on the employee’s behavior throughout the year) to support their ratings, and would be more effective at facilitating a constructive and helpful performance review meeting with the employee. The research design and analyses are described below.

METHOD

Background

The study described below took place in an undergraduate course in 2017 that focuses on the development of skills related to human resource management, such as conducting employment interviews, creating job descriptions, providing on-the-job training, and conducting performance reviews. The course focuses on skills that will be valuable for line managers (i.e., its focus is not on creating human resource managers). The course is required for all Management and Leadership majors at the university. Most students take the course in the final semester of their senior year, but some take it during the second semester of their junior (third) year. The performance review exercise contributed 20% to the student’s course grade.

The performance review exercise had been used in the course for nine years. As part of the exercise, students first received in-class instruction about how to conduct an effective performance review (see details below). They were told that they would be playing the role of a manager in a restaurant who is conducting annual performance reviews with servers who report to the manager. The students were then given notes about the performance of one server (Chris

Ramos) during the past 12 months. The notes were purportedly taken by the manager over the past 12 months and included 31 observations (several for each month) concerning the server's effective and ineffective job performance. Examples include: an elderly couple who brought their young grandchildren to dinner told the hostess that Chris was a great server and really made their grandkids feel welcome, comfortable, and special; late 25 minutes on February 14 (Valentine's Day) – it was a very busy shift and this caused a delay in seating customers; volunteered to work two extra shifts to cover for another waitress who was on her honeymoon; doing a great job of 'up-selling' several customers during the week of March 12 (e.g., ordering bottled water not tap, ordering more expensive wine, adding appetizers).

Based on these 31 notes, each student was asked to rate the server's performance on five performance dimensions (attendance, punctuality, cooperativeness with management and other employees, knowledge of menu offerings, quality of service and interactions with customers) and to provide an overall rating of the server's job performance. All ratings were completed on a 5-point rating scale with anchors ranging from very poor to very good. Students were also asked to complete brief narratives that summarized the employee's strengths and areas for improvement.

Next, each student conducted a performance review meeting with the server. The role of the server (Chris Ramos) was played by the course instructor. On average, the performance review meetings lasted 15 to 20 minutes. After the exercise was completed, I (the instructor) graded the performance ratings provided by the student (in terms of rating accuracy and thoroughness of supporting notes) and the student's effectiveness during the performance review meeting.

Training on Performance Reviews

The training began by describing the distinction between performance management (which includes goal setting, providing helpful feedback and coaching throughout the year, recognizing good performance in a timely manner, etc.) and performance appraisal. Next, students learned that performance reviews can be used to influence decisions about employees such as compensation, promotions, and dismissals. Ten legal guidelines (e.g., focusing on behaviors instead of traits, ensuring that performance dimensions are based on a job analysis) were reviewed. Students also learned about the variety of methods that can be used including simple rankings, behavior observation scales, graphic rating scales, forced distributions, behaviorally anchored rating scales, and management by objectives. Typical appraisal errors were also illustrated including halo, central tendency, primacy, recency, and leniency.

Guidelines for providing negative feedback were described. These included (1) objectively describing the event, behavior, or situation, avoiding accusations or inferences about the person's motives or personal characteristics, presenting any data or evidence, (2) describing your reactions (or feelings) and the objective consequences that have resulted or will result, (3) focusing on solutions, avoiding discussing who's right or wrong, suggesting an acceptable alternative, and (4) summarizing areas of agreement, next steps, and when and how follow-up will occur. Students then applied these guidelines to a brief case study.

Next, students practiced assigning examples of behavior to specific performance dimensions. For example, being "10 minutes late" was relevant to Punctuality but not Attendance.

And a customer complaint that “Chris was inattentive and got their orders wrong” was relevant to Quality of Service and Interactions with Customers but not to Cooperativeness with Management and Other Employees.

Finally, considerable time was spent discussing guidelines for conducting effective performance review meetings. These guidelines are listed in Appendix C.

Participants

Twenty-six students were enrolled in the course during the semester when these data were collected. All students understood that the performance review assignment (described above) was a required part of the course. They were also invited to participate in this study; participation would mean that they would be assigned at random to one of two experimental conditions (described below) and that their performance review meetings would be video-recorded and scored (for research purposes only) by two other faculty members (not the course instructor). They understood that if they decided not to participate in the research study, their performance review meetings would not be video-recorded or viewed by any other faculty member. All students were told that there would be no extra-credit for participating in the study and there would be no penalty (effect on their grade) for declining to participate.

Twenty-two students volunteered to participate and signed an informed written consent form that had been approved by the university’s institutional review board (IRB). Participants included 14 females and 8 males; 5 were black and 16 were white.

Study Design

Participants were assigned at random to one of two conditions. In the control condition ($n = 11$), each student completed the three dependent variables (described below) at the beginning of the semester *before* receiving any classroom training about performance reviews. At that time, these students received no feedback about their performance. (These students did eventually receive classroom training on performance reviews later in the semester, and they then completed all three dependent variables again.) In the treatment condition ($n = 11$), each participant completed the dependent variables toward the end of the semester, *after* receiving classroom training on performance reviews. All the performance reviews completed by students (before and after classroom training) were video-recorded. Each of the video-recordings was assigned a random number. The research design described here closely parallels the design used by Latham and Saari (1979).

Two faculty members from the Management and Leadership department were familiarized with the performance review exercise and all the materials associated with the exercise. Both faculty members were tenured Professors with extensive backgrounds in organizational behavior. Each of these two faculty members viewed each of the video-recordings. The video-recordings were labeled only with the random numbers assigned to them and no information was provided to the faculty members about the condition (treatment versus control) in which the video-recording occurred. The two faculty members also viewed all the video-recordings in random order. That is,

the faculty members were blind to the experimental condition associated with each video-recording.

Although students in the control condition completed the performance review meeting twice (before receiving classroom training and then, weeks later, after receiving classroom training), this could, of course, enable the faculty members to deduce that those students must have been in the control condition. But, because all video-recordings from all conditions were viewed in random order, the faculty members had no way of determining whether a video from a student in the control condition had been recorded before or after the student eventually received the classroom training.

Dependent Variables

Faculty Member Ratings of the Student's Effectiveness during the Performance Review. After the end of the semester, the two faculty members viewed and rated all video-recordings independently. After viewing each video-recording, the faculty member rated the participant's performance on a 0 (very poor) to 100 (excellent) scale. Table 1 includes the instructions provided to the faculty members to guide their ratings. The inter-rater reliability of the ratings provided by the two faculty members was .62. This level of interrater reliability is consistent with levels of interrater reliability found for real-world performance ratings. For example, Rothstein (1990) examined the interrater reliabilities of ratings of 9,975 ratees from 79 organizations. She found that asymptotic levels of reliability were about .60, even for raters with years of exposure to the ratee. In a meta-analysis, Viswesyan, Ones, and Schmidt (1996) found the mean interrater reliability of supervisory ratings of job performance was .52. A more recent meta-analysis (Salgado & Moscoso, 2019) found that interrater reliabilities averaged .61 for ratings collected for research purposes and .45 for ratings collected for administrative purposes. Based on the above, I used the average of the two faculty members' ratings as the dependent variable.

Students' Notes about the Server's (Chris Ramos') Performance. Each student also provided performance ratings concerning Chris Ramos on each of the five performance dimensions along with notes/comments to support those ratings. The two faculty members independently rated the level of detail in the written notes that each student provided to support his or her performance ratings. The student's performance ratings and accompanying written notes were labeled only with the same random number associated with the student's performance review video-recording. That is, the two faculty members were blind to the experimental condition associated with the student's performance ratings and notes. For each of the five performance dimensions, each faculty member indicated whether there were detailed, behavioral comments that reflected all or nearly all information about the server's behavior over 12 months (2 points), some behavioral comments but did not reflect all or nearly all information about the server's behavior (1 point), or no comments or comments that were vague and lacked behavioral detail (0 points). The faculty member also evaluated the typing, spelling, and grammar in these written performance ratings (2 points for no spelling or grammar errors, 1 point for 1 or 2 spelling or grammar errors, and 0 points for notes not typed or more than 2 spelling or grammar errors). These points were summed to create a total score for the quality of the notes that accompanied each student's performance ratings. The inter-rater reliability of the two faculty members' ratings of student notes was .84. Based on the above, I used the average of the two faculty members' ratings as the dependent variable.

Accuracy of Students' Performance Ratings of the Server (Chris Ramos). The accuracy of students' performance ratings of the server (Chris Ramos) was scored by me (the instructor). Unlike the students' performance during the performance review meeting and the quality of the students' notes (which are inherently subjective and hence needed to be scored by faculty members who were blind to the student's experimental condition), the accuracy of performance ratings could be scored objectively (see Appendix D for details). For example, the server's attendance throughout the year was excellent with only one absence due to illness. Hence, the 'true' rating for the Attendance dimension was 5. Students who provided a rating of 5 received two points, students who provided a rating of 4 (because attendance was not perfect) received 1 point. Students who provided a rating of 1, 2, or 3 received zero points. There were six positive examples that illustrated the server's Cooperativeness with Management and Other Employees, but there was one example where Chris was not cooperative ("Chris and another employee got into a verbal argument and their remarks could be overheard by customers at several tables"). Hence, the 'true' rating for this dimension was 4 (i.e., a rating of 5 would be unreasonable given this negative incident). Students who provided a rating of 4 received 2 points, whereas students who provided a rating of 1, 2, 3, or 5 received zero points. A similar approach was used to determine the 'true' score for the remaining dimensions. The student's score for rating accuracy was the sum of the points the student earned across the five performance dimensions.

RESULTS

I compared the performance of participants in the control group (who completed the exercise at the beginning of the semester *before* receiving any classroom training about performance reviews) with the performance of participants in the treatment condition (who completed the exercise later in the semester *after* receiving classroom training about performance reviews).

Independent group t-tests found that effectiveness during the performance review meeting was higher for participants who had received training ($M = 84.73$, $SD = 6.90$) than for participants who had not yet received training ($M = 72.05$, $SD = 10.89$; $t = 3.26$, $df = 20$, $p < .01$, $d = 1.42$). Also, the level of detail in the notes/comments provided by participants who had completed training ($M = 9.57$, $SD = 2.97$) was greater than that provided by participants who had not yet completed training ($M = 7.18$, $SD = 1.96$; $t = 2.22$, $df = 20$, $p < .05$, $d = 0.97$). For the accuracy of performance ratings, although the means were in the predicted direction, the difference did not reach significance (participants that had received training $M = 10.36$, $SD = 1.21$; participants that had not yet received training $M = 9.82$, $SD = 1.99$; $t = 0.78$, $df = 20$, $p = .45$, $d = 0.34$). Taken together, these results indicate that participants who had received training performed more effectively on two of the three dependent variables than participants who had not yet received training.

As noted above, the 11 students in the control group did eventually (later in the semester) receive classroom training about completing performance reviews. After receiving the classroom training, they again completed the three dependent variables (described above). Paired t-tests found that the effectiveness of these students during the performance review meeting (based on ratings of the video recordings) was higher after training ($M = 85.09$, $SD = 9.12$) than before training ($M = 72.05$, $SD = 10.89$; $t = 3.23$, $df = 10$, $p < .01$, $d = 1.30$). Also, the level of detail in the notes/comments that supported the performance ratings was higher after training ($M = 10.20$,

SD = 1.86) than before training (M = 7.18, SD = 1.96; $t = 4.33$, $df = 10$, $p < .01$, $d = 1.58$). For the accuracy of performance ratings, although the means were in the predicted direction, the difference did not reach significance (after training M = 10.45, SD = 2.21; before training M = 9.82, SD = 1.99; $t = 1.55$, $df = 10$, $p = .15$, $d = 0.30$). These results indicate that participants' performance improved on two of the three dependent variables from pre-training to post-training.

Finally, I compared the post-training scores of participants in the control condition (*after* they had received classroom training about performance reviews) with the post-training scores of participants in the treatment condition (*after* they had received classroom training about performance reviews). Because both groups had, at this point, completed training, there was no reason to expect any significant difference between the two groups of participants on any of the dependent variables. Independent group t-tests indicated that there were no significant differences in any of the three dependent variables (all $p > .50$) between the two groups after all participants had completed training.

DISCUSSION

The performance review process is complex; managers need to provide ratings that are supported by job-related evidence and they need the skills to facilitate a constructive conversation. The experiential exercise described here helps students develop those skills. At some time in their future, the majority of business majors will become managers and will need to apply these skills in a real-world setting.

One alternative to the approach described above would be to use other instructors or graduate students to play the role of the employee (Chris Ramos) during the performance review meeting. As noted above, whoever plays the role of the employee might create their own ways of responding during the performance review meetings, but it is important that all people playing the role of the employee adopt a consistent approach with different students (supervisors). Also, the training described above could perhaps be enhanced by showing videos that illustrate effective and ineffective performance review episodes.

In terms of impact, the evidence from the analyses described above shows (in contrast to Mumford's findings, 2009) that the classroom training enhances students' skills at facilitating a performance review meeting and providing documentation that supports their ratings. It is also noteworthy that effect sizes for all significant findings were large (approaching or exceeding one standard deviation unit in magnitude). The sample size for this research study was small. A small sample size limits the statistical power to detect an effect if there is an effect in the population of interest. Nonetheless, the statistically significant results and the large effect sizes that I observed indicate that statistical power was ultimately not a problem in this research.

The research design used here can be easily applied to many kinds of classroom exercises. When the outcomes of the training cannot be evaluated using some unambiguous, objective indicator, then it is important to have evidence from the exercise (e.g., video recordings, written reports) that can be evaluated by judges who are blind to the participants' experimental condition.

Although the experiential exercise described here helps students develop several important skills, there are other skills that managers need to develop as part of the broader performance management cycle. These include being able to set effective performance goals with employees, providing ongoing feedback ‘in the moment’ (rather than delaying feedback for weeks or months), providing encouragement and recognition when employees make progress, and (in most organizations) linking performance ratings to appropriate levels of financial awards (e.g., merit pay).

A search of the PsychINFO and ProQuest databases using the keywords ‘performance management,’ ‘performance ratings,’ ‘performance review’, and ‘experiential exercise’ yielded no other experiential exercises designed to develop the skills described above, with the exception of Mumford (2009) described above. Perhaps this exercise, and the research design used to evaluate its efficacy, will encourage others to create and rigorously evaluate other exercises related to the performance management process.

James W. Smither is a professor of management and leadership at La Salle University. He holds a Ph.D. from the Stevens Institute of Technology and M.A. degrees from Montclair State University and Seton Hall University. His teaching and research interests are in human resource management and leadership development. He is a Consulting Editor for the *Journal of Applied Psychology* and has served on several other editorial boards. Contact: smither@lasalle.edu.

REFERENCES

- Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E.D. (2016). Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology*, 9, 219-252.
- Aguinis, H. (2009). An expanded view of performance management. In J.W. Smither & M. London (Eds.), *Performance management: Putting research into action* (pp. 1-43). San Francisco: Jossey-Bass.
- Aguinis, H. (2019). *Performance management* (4th edition). Chicago Business Press.
- Bernardin, H. J., Hagan, C. M., Kane, J. S., & Villanova, P. (1998). Effective performance management: A focus on precision, customers, and situational constraints. In J.W. Smither (Ed.), *Performance appraisal: State-of-the-art in practice* (pp. 537 - 547). San Francisco: Jossey-Bass.
- Bull-Schaefer, R. A. (2018). Conducting midterm performance reviews: An exercise for teaching performance management, *Management Teaching Review*, 3 (1), 7-19.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Colquitt, A. L. (2019). Performance management 1.0 is a game and we should stop playing it. https://www.youtube.com/watch?v=-ieAthxn_e4

- DeGroot, T., Stambaugh, J., & Owen, J. (2009). An empirical examination of the use of a simulation in teaching human resource management. *Journal of Human Resources Education, 3* (3), 1-12.
- Effron, M. (2015). We love ratings. *Insights*, 1-5.
https://www.talentstrategygroup.com/application/third_party/ckfinder/userfiles/files/We%20Love%20Ratings.pdf
- Gist, M. E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review, 12*, 472-485.
- Grubb, T. (2007). Performance appraisal reappraised: It's not all positive. *Journal of Human Resources Education, 1* (1), 1-22.
- Gruys, M., & Stewart, S. (2007). Teaching human resource management concepts with experiential exercises. *Journal of Human Resources Education, 1* (1), 38-57.
- Hannay, M. (2010). Performance appraisal: Who knows best? *Journal of Human Resources Education, 4* (4), 15-25.
- Hauenstein, N. M. A. (1998). Training raters to increase the accuracy of appraisal and the usefulness of feedback. In J.W. Smither (Ed.), *Performance appraisal: State-of-the-art in practice* (pp. 537 - 547). San Francisco: Jossey-Bass.
- Hess, P. W. (2007). Enhancing leadership skill development by creating practice/feedback opportunities in the classroom. *Journal of Management Education, 31*(2), 195–213.
- Hyland, M. (2017). Who gets the money? An exercise to critically evaluate the annual performance review. *Journal of Human Resources Education, 11* (2/3), 15-24.
- Kosslyn, S. M. (2018, October). Minding education. *APS Observer*,
<https://www.psychologicalscience.org/observer/minding-education>.
- Latham, G. P., & Saari, L. M. (1979). Application of social-learning theory to training supervisors through behavioral modeling. *Journal of Applied Psychology, 64*, 239-246.
- London, M. (2015). *The power of feedback.: Giving, seeking, and using feedback for performance improvement*. Routledge: New York.
- Michaelsen, L. K., & Watson, W. (1987). Teaching through an experience with giving and receiving feedback. *Organizational Behavior Teaching Review, 11*(2), 111–113.
- Mumford, T. V. (2009). Developing performance management competence: An exercise leveraging video technology and multisource feedback. *Journal of Management Education, 33*, 617-644.

- Nandedkar, A. (2018). Solving the performance appraisal puzzle: A HRM class exercise. *Journal of Human Resources Education, 12* (4), 13-26.
- Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management broken? *Industrial and Organizational Psychology, 4*, 146-164.
- Rachman-Moore, D., & Kenett, R. S. (2006). The use of simulation to improve the effectiveness of training in performance management. *Journal of Management Education, 30*, 455-476.
- Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75*, 322-327.
- Salgado, J. F., & Moscoso, S. (2019). Meta-analysis of interrater reliability of supervisory performance ratings: Effects of appraisal purpose, scale type, and range restriction. *Frontiers in Psychology, 10*, <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02281/full>.
- Seltzer, J. (1986). On teaching feedback skills. *Organizational Behavior Teaching Review, 10*(2), 97-97.
- Smither, J. W. (2012). Performance management. In S.W.J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology*. New York: Oxford University Press.
- Smither, J. W. (2015). The fate of performance ratings: Don't write the obituary yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*, 77-80.
- Viswesyan, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.
- Viswesyan, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.

Table 1: Instructions for Faculty Members Rating Video-recordings

In evaluating the student's performance during the performance review meeting, please consider the following:

At the start of the meeting, did the student ...

- Greet the employee in a professional manner?
- Explain the purpose of the meeting?
- Put the employee at ease?
- Invite the employee to ask questions at any time?

Did the student ...

- Offer specific examples that supported the rating for each performance dimension?
- Explain the rationale for the rating of overall performance?
- Praise the employee's performance by citing examples of effective performance (such as excellent attendance with only 1 absence all year, volunteering to cover other's shifts, positive feedback from customers, winning the upselling contest)?
- Provide direct negative feedback about the employee's punctuality (while acknowledging when it has improved)?
- Provide direct negative feedback about the employee's argument with a coworker while discussing how to handle such situations in the future?
- Avoid excessively dwelling on the negative (by not talking about punctuality 3 or more times)?

Toward the end of the meeting, did the student ...

- Summarize the employee's strengths?
- Summarize areas where the employee can or needs to further improve?
- Ask whether the employee has an interest in growing within the company?

Throughout the meeting, did the student

- Make appropriate eye contact?
- Treat the employee with respect?
- Maintain a professional tone (not too strict, not too friendly)?

Based on the above, rate the student's effectiveness on a scale from 0 (very poor) to 100 (excellent)

APPENDIX A

Performance Documentation Concerning Chris Ramos

January

- 10 minutes late on January 2
- 20 minutes late on January 17 – it was a busy shift and other employees had to cover tables/customers that would have been assigned to Chris at the start of Chris' shift
- Volunteered to stay after assigned shift to cover for an employee who was arriving late
- Large dining party complimented Chris' work ('very friendly and attentive')
- Received 'fair' and 'good' ratings from mystery shopper on January 25 – I shared the ratings with Chris

February

- Volunteered to help a new employee 'learn the ropes' – was very helpful
- Late 25 minutes on February 14 (Valentine's Day) – it was a very busy shift, and this caused a delay in seating customers
- An elderly couple complained that Chris was inattentive and got their orders wrong

March

- I observed Chris doing a great job of 'up-selling' several customers during the week of March 12 (e.g., ordering bottled water not tap, ordering more expensive wine, adding appetizers)
- On March 24, Chris and another employee got into a verbal argument and their remarks could be overheard by customers at several tables

April

- 15 minutes late on April 1
- Chris volunteered to work two extra shifts to cover for another waitress who was on her honeymoon
- Hostess incorrectly seated too many customers (5 parties of 4 and 3 parties of 2) in Chris' area and not enough in another waiter's area (only 2 parties of 2), but Chris handled the extra work efficiently and graciously and didn't complain

May

- On May 7, a mystery shopper gave Chris 'very good' to 'excellent' ratings – I congratulated Chris on 'a job very well done'
- Next day, Chris was later 15 minutes

- On May 22, an elderly couple who brought their young grandchildren to dinner told the hostess that Chris was a great server and really made their grandkids feel welcome, comfortable, and special

June

- Chris referred a friend to apply for work here – we hired her, and she has become a very good member of our waitstaff
- Chris was late 40 minutes on June 21 – we were all worried that something had happened to Chris. Chris showed up and apologized but other waitstaff had to cover tables in Chris' area during a very busy time of day

July

- On July 4, Chris was 15 minutes late – I spoke to Chris about recurring problems with being late and that I expected Chris to 'on time all the time' – Chris assured me it would not happen again

August

- On 5 occasions, Chris worked extra shifts to cover for employees who were on vacation – I told Chris how appreciative I was
- A very large party (6 adults and 10 children for a birthday party) told me that Chris made the occasion really special and that all the kids really loved Chris (Chris has a really good sense of humor)

September

- During Labor Day weekend, we had a contest among all waitstaff to see who could 'up sell' the most – Chris won the contest – I told Chris, privately, that I would love to see Chris demonstrate this skill more often
- On September 16, late in Chris' shift, a man and woman seated in Chris' area got into a heated and loud argument. Chris handled the situation very effectively. He brought the matter to my attention (so I could calm the situation down). He then apologized to a few customers seated nearby who appeared really annoyed by the episode – Chris offered them complimentary coffee and dessert for the inconvenience.

October

- I have observed Chris 'up-selling' more often. I congratulated Chris on the good work and suggested that Chris might show other employees how to do this effectively. Later that week, I saw Chris giving some advice to a couple of new employees (e.g., what you can say to customers and when to say it)
- Chris called in sick on October 28 – Chris' first day absent all year

November

- I complimented Chris on not being late for three months – it's nice to see that improvement
- November 18 was a weird shift because one customer (looked like a woman in her mid-40s) complained that Chris never checked back to see if her order had been prepared properly (the order had been overcooked), but about two hours later, a group of customers in their late teens told the hostess that Chris had been a terrific server and that they really liked the menu items that Chris recommended.
- A mystery shopper gave Chris 'very good' ratings

December

- On December 21, Chris arrived 10 minutes late
- The Christmas season often brings customers who are in a hurry (so they can get back to their holiday shopping). On several occasions, I observed Chris explain politely to customers (who were 'in a rush') how long it would likely take before their orders arrived (so the customers would have realistic expectations). On another occasion, I saw Chris suggest some menu items that could be served quickly, and the customers thanked Chris for being thoughtful.
- Chris did a great job working the New Year's Eve shift – handled a lot of customers very efficiently without becoming stressed out.

APPENDIX B

Rating Form

Employee Name _____ Supervisor Name _____

Attendance

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

Comments:

Punctuality

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

Comments:

Cooperativeness with Management and Other Employees

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

Comments:

Knowledge of Menu Offerings

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

Comments:

Quality of Service and Interactions with Customers

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

Comments:

OVERALL RATING

Very Poor	Poor	Average	Good	Very Good
1	2	3	4	5

What are this employee's strengths?

In what areas does this employee need to improve his or her performance? Provide specific expectations for behavior change, skill development, and/or performance improvement.

Please offer suggestions concerning skills this employee could acquire or further develop that would enable the employee to advance his or her career in the company.

APPENDIX C

Suggestions for Facilitating an Effective Performance Review Meeting

Don't delay feedback - feedback and recognition is a day-to-day activity (not an annual event)

- There should be no surprises during the performance review meeting– the employee should not be receiving negative feedback for the first time about performance/behavior that occurred weeks or months earlier
- Such feedback should have been given in a timely manner when the performance/behavior occurred

Take notes throughout the year about what the employee is doing well and areas needing improvement.

- The notes should be specific – describe behavior, not traits
- Review these notes to create your end-of-year performance evaluation

Your completed performance evaluation should be typed and include detailed notes to support each of your ratings.

- No spelling or grammatical mistakes
- The completed document should look very professional

Encourage employees to prepare for the meeting (e.g., by reviewing performance expectations, their work outcomes, both successes and problems).

Dress professionally.

Open the meeting by explaining its purpose.

- “A chance to briefly review what you are doing well and to discuss any areas where changes would be helpful”

Encourage participation – make it clear that the employee's comments and questions are welcome.

- This enhances the likelihood that the employee will perceive the process as fair

Review the employee's rating – and the examples/behaviors that support that rating – for each aspect of job performance.

- Focus on the employee's behavior and performance, not personality

Praise effective performance.

Offer candid and specific feedback with clear suggestions for the future

Listen actively (avoid interrupting, summarize what the employee has said, maintain eye contact, etc.).

Avoid destructive criticism (e.g., using sarcasm, attacking, being mean-spirited).

Toward the end of the meeting, briefly summarize the employee's strengths and thank the employee for his or her contributions.

- Also, briefly review areas where the employee should make improvements.
- Ask about the employee's career goals.
 - Provide helpful advice
- Set specific, challenging, mutually agreeable goals for the future.
- Close by thanking the employee again for her or his contributions to the organization.

APPENDIX D

Justifiable (evidence-based) Ratings for Each Performance Dimension

Attendance

The most realistic rating for this dimension is “5” because Chris was absent only once all year (and he called in to explain that he was sick). Occasionally, a student will rate Chris as a “4” on this dimension because attendance was not perfect. I ask the student if this rating is reasonable because it would be very rare for any employee to have perfect attendance (and not be sick) for an entire year.

Punctuality

For this dimension, the most realistic rating is “2” because of Chris’ frequent lateness in the first half of the year. A rating of “1” would seem too low given Chris’ substantial improvement on this dimension during the second half of the year.

Cooperativeness with Management and Other Employees

For this dimension, the most realistic rating is “4” due to the many ways that Chris was helpful throughout the year (e.g., volunteering to cover others’ shifts, referring a friend who was hired and became a very good employee, helping a new employee ‘learn the ropes,’ giving helpful advice to new employees), but a rating of “5” would seem too high given the argument that Chris had with another employee in March.

Knowledge of Menu Offerings

The most realistic rating on this dimension is “5”. Chris excels in this area. Examples include winning the upselling contest in September, explaining to “customers (who were ‘in a rush’) how long it would likely take before their orders arrived (so the customers would have realistic expectations),” and “suggesting some menu items that could be served quickly.”

Quality of Service and Interactions with Customers

For this dimension, the most realistic rating is “4” because of the many (9) compliments that Chris received throughout the year from customers and positive feedback from mystery shoppers. A “5” might be too high because of two customer complaints.

APPENDIX E

Pretest Survey: The 'Performance Evaluation and Review (Chris Ramos)' Exercise

Name: _____

_____ Using a 0 (*very low level of comfort*) to 10 (*very high level of comfort*) scale, rate how comfortable you think you would be conducting a performance review meeting with an employee who reports to you.

_____ Using a 0 (*very low level of knowledge*) to 10 (*very high level of knowledge*) scale, rate your level of knowledge about how to conduct an effective performance review.

APPENDIX F

Posttest Survey: The 'Performance Evaluation and Review (Chris Ramos)' Exercise

Name: _____

_____ Using a 0 (*very low level of comfort*) to 10 (*very high level of comfort*) scale, rate how comfortable you think you would be conducting a performance review meeting with an employee who reports to you.

_____ Using a 0 (*very low level of knowledge*) to 10 (*very high level of knowledge*) scale, rate your level of knowledge about how to conduct an effective performance review.