DISCREPANCY IN STUDENT PEER AND SELF-EVALUATIONS: A PREDICTOR OF TEAM PERFORMANCE?

DISCREPANCIA ENTRE LAS EVALUACIONES DE LOS COMPAÑEROS Y LAS AUTOEVALUACIONES: ¿UN PREDICTOR DEL RENDIMIENTO DEL EQUIPO?

Matthew D. Marmet*
https://orcid.org/0000-0002-8743-8376 (ORCID iD)
Utica University, USA

ABSTRACT

Performance prediction and evaluation (rating) have been investigated by psychologists for years. One aspect of performance description that has increased in popularity is multisource feedback. This process typically involves a focal person receiving feedback on their performance from the perspectives of others. Additionally, these multisource feedback systems call on this focal person to rate their own performance, so an evaluation of the discrepancy between self and others' ratings can be made. The current study aims to assess the impact of self-other rating congruence in an academic setting. Specifically, can team performance be predicted by the level of agreement between self and others' ratings? The magnitude of the discrepancy between self and others' ratings on a student peer evaluation form was appraised, and the correlation between discrepancy magnitude and team performance on a final project was assessed. Initial data analysis yielded results contrary to the proposed hypothesis, but also called into question the overall utility of the evaluation process itself.

RESUMEN

Los psicólogos han investigado durante años la predicción y evaluación del rendimiento (calificación). Un aspecto de la descripción del desempeño que ha ganado popularidad es la retroalimentación de múltiples fuentes. Este proceso generalmente involucra a una persona focal que recibe retroalimentación sobre su desempeño desde las perspectivas de los demás. Además, estos sistemas de retroalimentación de múltiples fuentes requieren que esta persona central califique su propio desempeño, de modo que se pueda realizar una evaluación de la discrepancia entre las calificaciones propias y las de los demás. El estudio actual tiene como objetivo evaluar el impacto de la congruencia de calificación entre uno mismo y otros en un entorno académico. Específicamente, ¿se puede predecir el desempeño del equipo por el nivel de acuerdo entre las calificaciones propias y las de los demás? Se evaluó la magnitud de la discrepancia entre las calificaciones propias y las de los demás en un formulario de evaluación por pares de estudiantes, y se evaluó la correlación entre la magnitud de la discrepancia y el desempeño del equipo en un proyecto final. El análisis inicial de los datos arrojó resultados contrarios a la hipótesis propuesta, pero también puso en duda la utilidad general del proceso de evaluación en sí.

INTRODUCTION

The importance placed on team-based work in industry has been accompanied by a rise in team-based work in business school education to help prepare students for this reality (Andrade, Miller, & Ogden, 2020). Due to issues such as free riding and social loafing in such work, academics have introduced organizational tactics in an attempt to diminish their prevalence. One such tactic comes from the world of performance appraisal, and involves students rating their own performance in addition to receiving ratings from their peer teammates.

Prior research on the peer evaluation process has focused on teamwork effectiveness (Petkova, Domingo, & Lamm, 2021). The objective of this study, however, was to investigate whether agreement between these self and peer ratings impacts team success as measured by each team's final grade on a team-based project. It was hypothesized that greater agreement between these ratings would yield better final performance. Despite past literature indicating that such agreement tends to yield positive outcomes such as this, analysis of the data produced results to the contrary.

In trying to find meaning in these results, the author was forced to examine the utility of the peer evaluation process employed. In doing so, potential techniques for improving the evaluation process were uncovered. Of note are in-flow peer

review, instructor coaching, and performance evaluation training, as all could potentially increase student engagement in the process itself.

What follows is a review of the literature associated with the research endeavor at hand. From there, the study's methodology will be described, followed by the results of the inferential statistical analysis. Interpretation of these results is subsequent to this, including a discussion of the study's limitations and potential directions for future research efforts.

LITERATURE REVIEW

Performance appraisal

Performance appraisal has occupied the minds of psychologists for many years, and is a process that has seen exponential increases in complexity during that time. Fletcher (2001) states, "performance appraisal was a term once associated with a rather basic process involving a line manager completing an annual report on a subordinate's performance and (usually, but not always) discussing it with him or her in an appraisal interview" (p. 473). With the advent of methodologies such as multisource rating, and true one-on-one executive coaching, the increase in complexity is undeniable, and also speaks to the importance of the appraisal process itself.

The systems involved in the performance appraisal process are designed to hold individual employees accountable in all sectors of industry, including public, private, and non-profit (Rubin & Edwards, 2020.) Performance rating also plays a key role in nearly all personnel decisions. Examples here include criteria necessary for training evaluation, indices of effectiveness necessary for administrative decision making such as promotion opportunities or merit-based pay increases, and finally, the performance-related information needed to provide developmental feedback and counseling to employees (Jefferson, 2010; Landy, Barnes, & Murphy, 1978). A large part of this developmental feedback discussion has become the aforementioned multisource feedback.

Multisource feedback

Multisource feedback typically involves an evaluation of a focal person's performance by others around them who have observed that performance. A far cry from Fletcher's (2001) depiction of traditional performance appraisal, multisource feedback adds layers of complexity to the process as an individual's performance is evaluated through the perspectives of multiple unique observers. Prior research has shown value in assessing a person's strengths and weaknesses through these unique lenses. For example, Church (2000) found that comparing multisource ratings with one's own self-ratings can improve self-awareness and behavioral change. Additionally, Sinha, Mesmer-Magnus, and Viswesvaran (2012) indicate that, compared with other evaluative techniques, multisource evaluation provides a more complete picture of individual job effectiveness than traditional methods of performance appraisal.

Self-other agreement in multisource ratings

As indicated in Church's (2000) findings above, part of the multisource rating process involves the focal person rating their own performance, which allows for an analysis of the agreement between those self-ratings, and the ratings provided to that focal person by others (Sala, 2003). In other words, is there alignment

between one's perception of their own performance, and how that same performance is perceived by the others around them?

Prior research has highlighted a multitude of factors that may account for a lack of alignment among self and others' ratings. For example, Sinha, Mesmer-Magnus, and Viswesvaran (2012) found that the personality characteristics of the raters involved in the appraisal process contributed to discrepancies among self and others' ratings. However, the performance appraisal literature also provides several indications for why agreement in self-other ratings is important. First, since it represents different perspectives on the same phenomena, "the overlap or degree of consensus or agreement is valuable information in itself" (Yammarino & Atwater, 1997, pp. 39-40). Second, agreement in self-other ratings is preferable as it indicates a mutual understanding between the individuals involved in the rating process. Disagreement, on the other hand, can be seen as dysfunctional in a managerial capacity as it may indicate ineffective or incomplete communications with subordinates with regard to performance (Smircich & Chesser, 1981). Finally, and perhaps most importantly, prior research has yielded support for predictions that self-other rating agreement is related to positive organizational outcomes. For example, Atwater and Yammarino (1992), in a study involving naval officers, found that self-other agreement was positively related to leadership performance and advancement. Additionally, Church (1997) indicated that high-performing managers displayed greater levels of congruence between their own ratings of performance and those made by their direct reports than did average-performing managers. Interestingly, this relationship held true across four independent datasets representing three different organizations and industries (Church, 1997).

Performance appraisal in the educational setting

With its known importance in industrial settings, one of the main questions this research effort attempted to address was whether or not the benefits yielded through self-other agreement in ratings can translate to an educational setting. This question is valid because, as indicated by Sherwood and DePaolo (2007), the rise of team-based work in organizations has led to a subsequent rise in business schools adopting such work. Although benefits of team-based work abound in both the global marketplace and academia (Sherwood & DePaolo, 2007), this type of learning presents challenges for instructors. "Areas of concern often include free rider/social loafing problems, challenges regarding how to effectively develop individual member team skills, and issues regarding how to assign grades to individuals for team projects" (p. 109). These same authors go on to present the use of student peer evaluations as a potential method for addressing these serious challenges. Additionally, Andrade, Miller, and Ogden (2020) reference a multitude of studies indicating that business and management faculty "have recognized the need for both peer and self-evaluations to encourage reflection on both individual contributions and team processes" (p. 5).

The use of student peer evaluation data

Prior research in the area of student peer evaluations has focused on the evaluation process in general, and whether or not it had a significant impact on teamwork effectiveness (Andrade, Miller, & Ogden, 2020; Petkova, Domingo, & Lamm, 2021; Politz et al., 2014). Additionally, the relationship between student peer/self-assessment and ratings made by these students' instructors has been

investigated (Suñol et al., 2016).   When evaluating oral presentations, these authors found a significant degree of discrepancy between the peer/self-ratings and the instructors' ratings. Interestingly, these peer evaluations in general, along with the results of the aforementioned research studies, also provide a glimpse into phenomena that are conceptually similar to those found in industry-based multisource feedback (please see Figure 1). Using a peer evaluation tool, students are charged with providing their perceptions of their own performance. In addition, their performance is rated by their peer team members. Consequently, as in multisource feedback, each student then becomes a focal person that is rated by others.   This methodology allows for an opportunity to assess self-other agreement in a purely educational setting, across all criteria contained within the evaluation instrument (described below).

Figure 1. The peer evaluation process
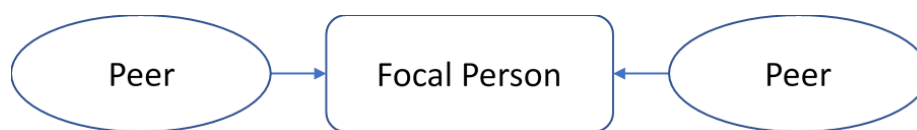
With the known positive impact of self-other agreement in organizational (industrial) settings, the following hypothesis was posited in this current study:

   H1: Greater levels of self-other rating discrepancy will have a significant negative impact on a team's final project grade. In other words, the more teammates agree on each other's performance, the better the team will do overall.

What follows in the sections below is a thorough explanation of the data and data analysis techniques employed to gain a better understanding of the relationship between self-other agreement among peer evaluation ratings and overall team performance on a final project. Included in this explanation are the results of correlational analysis, and a deeper dive into the data that left the author questioning the utility of the peer evaluation process altogether.

METHODOLOGY

Participants
   Data for this study were collected from 108 undergraduate students enrolled in several sections of a 300-level management of technology course at a small liberal arts university in the northeastern United States. The course is compulsory for all business majors, and is typically taken during the students' third or fourth year of the program. This sample had to be trimmed to 57 students, representing 17 teams, due to unusable data, which will be discussed in greater detail later. After this data trimming, the average team size was approximately three students. The sample was comprised of 40% women and 60% men.

Data collection: self-other agreement
   Performance ratings from students were collected over the course of three semesters. For the purpose of this study, the scores were taken from evaluations that were completed at the end of the semester, once the teams had submitted their final projects.   Participants were asked to rate their own performance, as

well as the performance of their team members, across the five dimensions of Listening Skills, Openness to Others' Ideas, Preparation, Contribution, and Leadership. Ratings were made on a 5-point scale ranging from *0 = missing (never shows up and never contributes)* to *5 = Excellent*. A behaviorally anchored rubric (please see Table 1; Altman, 2018) was provided to each student along with the peer evaluation instrument to help inform numerical ratings for each criterion. An example of the verbiage contained within these behavioral anchors can be seen below for the criterion of Listening Skills:

Excellent (5) = Routinely restates what others say before responding; frequently solicits others' contributions; sustains eye contact

Table 1. Peer evaluation behaviorally anchored rubric

Criterion	Excellent (5)	Good (4)	Fair (3)	Needs to Improve (2)	Unacceptable (1)	Missing (0)
Listening Skills	Routinely restates what others say before responding; rarely interrupts; frequently solicits others' contributions; sustains eye contact	Often restates what others say before responding; usually does not interrupt; often solicits others' contributions; makes eye contact	Sometimes restates what others say before responding; sometimes interrupts; sometimes asks for others' contributions; makes eye contact	Rarely restates what others say before responding; often interrupts; rarely solicits others' contributions; does not make eye contact; sometimes converses with others when another team member is speaking	Doesn't restate what others say when responding; often interrupts; doesn't ask for contributions from others; is readily distracted; often talks with others when a team member speaks	Never shows up and never contributes
Openness to others' ideas	Listens to others' ideas without interrupting; responds positively to ideas even if rejecting; asks questions about the ideas	Listens to others' ideas without interrupting; responds positively to the ideas even if rejecting	Sometimes listens to others' ideas without interrupting; generally responds to the ideas	Interrupts others' articulation of their ideas; does not comment on ideas	Interrupts others' articulation of ideas; makes deprecatory comments and/or gestures	Never shows up and never contributes
Preparation	Always completes assignments; always comes to team sessions with necessary documents and materials; does additional research, reading, writing, designing, implementing	Typically completes assignments; typically comes to team sessions with necessary documents and materials	Sometimes completes assignments; sometimes comes to team sessions with necessary documents and materials	Sometimes completes assignments; sometimes comes to team sessions without necessary documents and materials	Typically does not complete assignments; typically comes to team sessions without necessary documents and materials	Never shows up and never contributes
Contribution	Always contributes; quality of contributions is exceptional	Usually contributes; quality of contributions is solid	Sometimes contributes; quality of contributions is fair	Sometimes contributes; quality of contributions is inconsistent	Rarely contributes; contributions are often peripheral or irrelevant; frequently misses team sessions	Never shows up and never contributes
Leadership	Seeks opportunities to lead; in leading is attentive to each member of the team, articulates outcomes for each session and project, keeps team on schedule, foregrounds collaboration and integration of individual efforts	Is willing to lead; in leading is attentive to each member of the team, articulates general direction for each session and each project, attempts to keep team on schedule	Will take lead if group insists; not good at being attentive to each member of the team, sometimes articulates direction for sessions, has some trouble keeping team on schedule	Resists taking on leadership role; in leading allows uneven contribution from team members, is unclear about outcomes or direction, does not make plans for sessions or projects	May volunteer to lead but does not follow through; misses team sessions, does not address outcomes or direction for sessions or projects, team members become anarchical	Never shows up and never contributes

Data collection: team performance

Team performance was assessed using a final team project designed to evaluate the students' grasp of course content. This project involved student teams creating a 12- to 15-minute video that tasked them with relating course content to a real-world scenario presented in a news article. The scenario, in this case, was various Major League Baseball (MLB) teams implementing things like improved Wi-Fi access and augmented reality to enhance the fans' experience at their games. This final project was worth 30% of the students' final grade in the course. Division of labor during the execution of the project was entirely up to the teams, but smaller, incremental deliverables were assigned to each team to help prepare them for delivery of the final product. Final grades on the project ranged from 0 to 100. Each project was evaluated only by the course instructor, using a separate behaviorally anchored rubric that the students also had at their disposal.

Data analysis

Previous research has measured the discrepancy between self and others' ratings in two primary ways. The first is the direction of the discrepancy. This refers to whether the focal person (self) tends to overestimate or underestimate their performance compared to the ratings made by others. The second is the magnitude of the discrepancy. Using this methodology, a researcher assesses how large the difference is between self and others' ratings (Atwater & Yammarino, 1992; Church, 1997).

The current study measured only the magnitude of the discrepancy in order to assess its impact on the final project grade. Both self and others' ratings across the five dimensions included on the peer evaluation form were averaged. The differences between these averages were then calculated to arrive at the discrepancy between them.

RESULTS

Assessing self-other agreement

Across the 17 teams included in the study, the average magnitude of discrepancy was 0.49. These discrepancies ranged from 0 to 1.2 ($SD = .38$). Table 2 provides information on the mean differences between the self and others' ratings. It is important to note here that this sample was obtained as a result of necessary data trimming. This was due in part because some students simply did not submit the evaluation, and others failed to follow directions and did not rate themselves as part of the evaluation. These cases were eliminated due to the inability to assess rating congruence. Those teams whose participants rated themselves and their peers with 5's across all criteria were maintained in this study.

Table 2. Average difference in disagreement between self and others ratings

| Team Number	Avg. Self Rating	Avg. Peer Ratings	Avg. Level of Disagreement
1	5.00	4.50	0.50
2	5.00	4.40	0.60
3	4.87	3.97	1.17
4	5.00	5.00	0.00
5	4.80	3.60	1.20
6	4.95	4.98	0.08
7	4.65	4.77	0.42
8	4.53	4.33	0.67
9	5.00	5.00	0.00
10	5.00	5.00	0.00
11	4.75	4.30	0.45
12	4.68	4.46	0.52
13	4.73	4.23	0.57
14	4.20	3.88	0.32
15	4.76	4.75	0.21
16	4.90	4.20	0.90
17	5.00	4.20	0.80

Final project grades

Final team project grades were assessed on a standard scale from 0-100. Grades across the various course sections ranged from 73.59 to 95.00. The mean final project grade was 86.15 ($SD = 6.20$). It is worth noting again here that these projects were evaluated only by the course instructor, who taught all sections of the course across the three semesters in question.

Correlational analysis

In order to assess the relationship between self-other agreement and team performance, as measured by these final team project grades, a Pearson's correlation coefficient was computed using IBM SPSS 24. The analysis yielded a weak, positive relationship ($r = .14, p = .59$) between the two variables, lending no support for the researcher's hypothesis. Please see Figure 2 for a visual representation of the relationship between self-other rating discrepancy and team performance.

Figure 2. The relationship between self-other rating discrepancy and team performance.
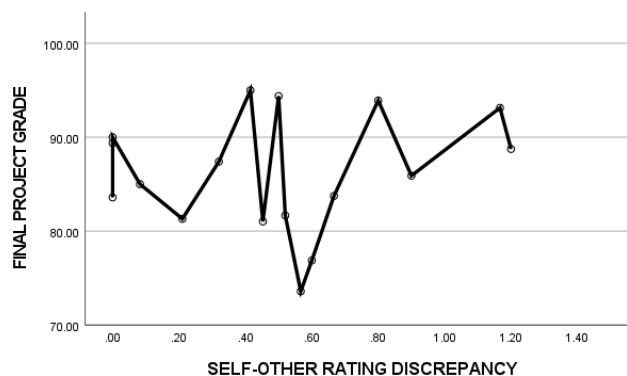
DISCUSSION

Conclusion

It was hypothesized that increased discrepancy between self-other ratings on a peer evaluation instrument would lead to lower scores on a final team project. However, the opposite was found. Additionally, the weak relationship between the two variables investigated in this study ($r = .14$) indicates that, despite direction, the impact of self-other rating agreement on team performance is not substantial. From here, the question for this researcher became: Can any meaning be derived from these findings?

A closer look at the data

Further investigation into the data itself proved fruitful in the attempt to derive such meaning. However, the results of this further investigation brought focus away from the relationship between self-other rating agreement and team performance, and centered it on the actual utility of this peer evaluation process overall.

Recall that the study sample was cut from 108 observations down to 57 due to unusable data. This was necessary because students either failed to follow directions and did not rate themselves on the evaluation form, or they simply did not submit a peer evaluation at all. In either case, a comparison of self-other ratings became an impossibility. Additionally, the researcher noticed an abundance of evaluation forms where students rated themselves and their teammates with 5's across every criterion. This was present in approximately 30% of the usable observations, and would obviously have an impact on the ability to assess true agreement or discrepancy in ratings.

Performance evaluation: a goal-directed behavior

In light of this information gathered from further investigation into the data, and to gain a better understanding of what could potentially be happening in this scenario, a more macro-level view of performance evaluation was assessed. In particular, a review was conducted of Murphy and Cleveland's (1995) four-component model of performance appraisal. This model posits that performance evaluation is a goal-directed behavior on the part of raters (please see Figure 3 for a visual depiction of the model). Based on this, could the students who filled out the peer evaluation forms have been doing so with certain goals in mind?

According to the model, contextual variables can be broken up into two levels. The first are proximal variables, and can be thought of as the environment within the organization. These are very salient to the rater, as they operate within the proximal setting of the organization itself. The second are distal contextual variables, and refer to the actual external environment the organization is in. This includes things such as the economic and cultural climates surrounding the organization (Murphy & Cleveland, 1995).

Perhaps both of these were at play in this peer evaluation scenario. In terms of proximal context, although it did count as part of the students' grade, the peer evaluation component was a small portion, which could equate to a lack of incentive to take it seriously. Also, these evaluations were conducted at the end of the semester, when student focus is hardly at its peak. Additionally, the impact of COVID-19 cannot be ignored as a potential distal contextual factor. Anecdotally, this researcher has had numerous conversations with students who

indicated that "learning during a pandemic" is extremely difficult, and takes a toll on motivation levels.

With these two factors in mind, it may not be surprising that 30% of students completed their peer evaluations in the most efficient way available to them. Instead of taking the time to think critically about their own performance and that of their peers, they simply assigned themselves and their team members the highest scores possible across all dimensions. In other words, their rating behavior may have been directed by the goal of simply completing the form rather than accurately depicting performance, be it their own or of their peers.

Figure 3. Murphy and Cleveland's (1995) model of performance appraisal
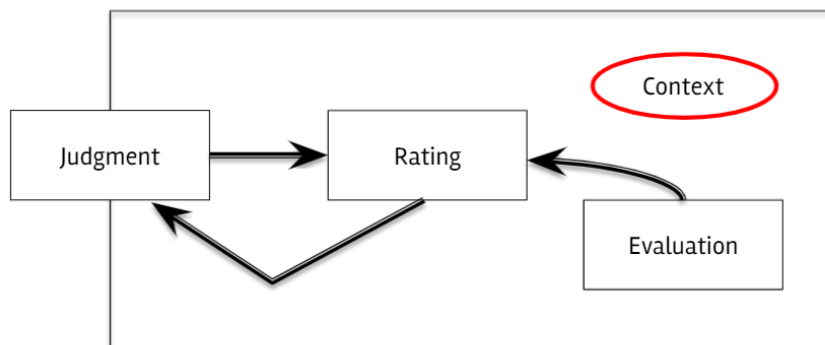
Note. The contextual variable is called out for emphasis due to its importance to the current study.

Student perspectives on peer assessment
Finally, to gain a better understanding of the results of the current study, it is important to analyze the students' perspectives on the peer evaluation process. Prior research does indicate that students prefer teachers to peers when their performance is being evaluated (Kwok, 2008). Additionally, and related to the goal-directed behavior described above, it has been noted that the validity and reliability of student peer assessments can be called into question due to such social factors as friendship bias. In a review of the literature, Panadero (2016) cites several studies in which students reported inflating peer-evaluation scores in hopes of enhancing their relationships with peers. Results such as this shed light on the complexity of peer assessment, and the need to take into account the potential social factors and goal-directed behaviors at play.

Implications
The results of this research effort could lead one to question whether or not it is worth conducting these types of peer evaluations at all. Collecting the data from these instruments is time-consuming for the instructor, and perhaps even prohibitive in larger classes (Sherwood & DePaolo, 2007). Additionally, more recent research on the effectiveness of student peer evaluation processes in combating things like free riding has yielded mixed results (Pierson, 2016). Finally, the results of the current study suggest that students may not take the process seriously.

It seems that instructors who utilize such peer evaluation processes, and experience similar results, have several options. First, they can simply accept

the phenomena and continue to run their peer evaluations as is. This researcher would argue that this is not the best route to take given the time and effort that would be wasted conducting a faulty process. Second, the process could be eliminated altogether. Although this seems like the most effortless solution, it likely would come with increased time spent on using other measures to help deter things like free riding (e.g., an increase in time spent meeting with/checking-in on teams to assess holistic and individual performance). Third, changes to the peer evaluation process could be made. Perhaps an evaluation could be done at midterm, and another at semester end, with intervention after the first evaluation round if the "all 5's" rating behavior is noticed. This could potentially address the social and goal-directed behaviors impacting peer assessment described above. Finally, student motivation to provide meaningful ratings could be addressed. Friedman, Cox, and Maher (2008) define meaningful ratings as those in which "the student has taken the rating process seriously and spent time considering each rating" (p. 581). The authors utilized an expectancy theory approach (Vroom, 1964) and cite prior research (Chen & Lou, 2004) indicating that the attractiveness of outcomes, such as reducing uneven work distribution and enhancing group productivity, may play a role in students' willingness to take the process seriously.

To expand on this third option of making changes to the evaluation process, it is important to consider that students may lack the skills and training necessary to assign the meaningful ratings described above (Jassawala, Shashittal, & Malshe, 2009). Perhaps dedicated training sessions early in the semester are necessary to ensure that students are interpreting the scales used in the peer evaluation instrument the same way (Andrade, Miller, & Ogden, 2020). In addition, the results of this study indicate that students may not have taken the appraisal process seriously. Prior research has shown such training efforts to increase engagement in performance evaluation processes (Rubin & Edwards, 2020) and increase satisfaction with the appraisal system (Taylor et al., 1995).

These preparatory efforts could also be coupled with an expansion of the performance evaluation process itself. Politz et al. (2014) present what they refer to as in-flow peer review (IFPR). Here, performance evaluation is done at multiple stages during a project rather than just at the end. This could be fruitful because "[p]eer evaluations are most beneficial at a time when there are still opportunities to utilize the evaluation feedback to improve. Furthermore, multiple peer evaluations conducted throughout a course allow students to see if actions based on the evaluations resulted in improvements" (Morales-Trujillo et al., 2021, p. 2). The key to this methodology is to have students complete assignments early enough to receive and subsequently act on feedback. To accomplish this, assignments that make up a project could be broken up into multiple stages throughout a semester (Politz et al., 2014).

Regardless of what one wants to gain from the peer evaluation process, whether it is an attempt at reducing free-riding, or providing your students with a lesson in the value of evaluating their peers' performance, this researcher suggests complete transparency with students surrounding the process. The students need to be aware of why you are making them spend time working on this deliverable. As an instructor, one should also truly reflect on the costs and benefits of the process overall. If your response to the question "What are my students getting out of this?" is "I'm not really sure," it may be time to re-evaluate the utility of the methods being used, or how they are being presented to students.

Limitations

As with all research efforts, this study is not without its limitations. First, the assessment tool utilized needs to be evaluated for reliability and validity. Although this tool was taken from Altman (2018) for the purpose of evaluating student perceptions of their own and their peers' performance, this author is unaware of any exploratory or confirmatory analysis conducted on it. Going forward, tests to evaluate such things as inter-rater reliability should be conducted on this assessment to ensure that these ratings are not distorted. Second, the direction, or sign, of the discrepancy was not assessed in the current study. As mentioned above, along with magnitude, direction is another common way to measure disagreement in multisource feedback, and its assessment has proven useful in organizational settings (Atwater & Yammarino, 1992). Although it was beyond the scope of this study to do so, this type of analysis could prove fruitful in the future.

Directions for future research

Going forward, this researcher would like to extend these research efforts by assessing this phenomenon at the graduate level of education. The same peer evaluation instrument is utilized at the MBA-level within the institution's course on the principles and practices of leadership. Given the notable differences in degree of professionalism and work experience between most undergraduate and graduate students, this same peer evaluation process could potentially yield different outcomes at the graduate level. Further study here could also help address the issues of the reliability and validity of the assessment instrument mentioned in the limitations above.

Also, future research efforts are needed to drill down into this peer evaluation form to assess discrepancies in self-other rating agreement at the specific criterion level, rather than in the aggregate ratings. For example, one criterion within the form is "Contribution." Perhaps more significant discrepancies in self-other ratings exist here that could lead to better prediction of team performance. In accordance with this, it would be interesting to assess the impact of training students on the use of this peer evaluation tool. Prior research has shown instructor assessment to be more reliable than student peer assessment (Magin & Helmore, 2010). Additionally, Xiong, Hunter, Guo, and Tywoniw (2020) found that rater training had a positive impact on learning outcomes resulting from peer assessment. Currently, there is no training provided to students on utilizing the assessment. Perhaps including this in the curriculum could lead to more reliable ratings. Training here would include a more specific definition of the language contained within each criterion. For example, in relation to the "Contribution" criterion listed above, what is the difference between an "exceptional" versus a "solid" level of contribution? In addition, this could yield supplementary learning outcomes as students would gain greater exposure to the tool and the value of the evaluation process itself.

Additionally, as indicated in the limitations above, future research endeavors should ensure the ability to capture the direction of the discrepancy between self and others' ratings. This analysis is over and above the magnitude of the discrepancy presented in the current study. The importance of assessing discrepancy direction has been highlighted in prior research, as those who

overestimate their own performance display different organizational outcomes than those who tend to underestimate it (Atwater & Yammarino, 1992).

Finally, a change to the review process reflecting the in-flow peer review posited by Politz and colleagues (2014) could be administered. Efforts to decompose the final project assignment, as suggested by the authors, are already done in the course these evaluations are conducted in. Additional peer reviews could be added to the schedule of the course where students rate their own and their peers' performance on each of these smaller deliverables. Feedback received in these sessions could then be acted on by the students as they progress through the semester. Intermittent one-on-one coaching from the course instructor regarding the performance ratings may also pay dividends in this scenario.

REFERENCES

Altman, W.S. (2018). *Rubric for assessing group members' ability to participate effectively as part of a team.* Retrieved from: https://jillshultz.com/williamaltman/Bill_Projects/Rubrics.html

Andrade, M., Miller, R., & Ogden, M. (2020). Teamwork for business majors – The impact of peer evaluation. *E-Journal of Business Education & Scholarship of Teaching, 14*, 1-18.

Atwater, L.E. & Yammarino, F.J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.

Chen, Y. & Lou, H. (2004). Students' perception of peer evaluation: An expectancy perspective. *Journal of Education for Business, 79*, 275-282.

Church, A.H. (1997). Managerial self-awareness in high-performing individuals in organizations. *Journal of Applied Psychology, 82*, 281-292.

Church, A.H. (2000). Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology. *Consulting Psychology Journal: Practice and Research, 54,*166-172. doi: 10.1037/1061-4087.52.2.99

Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology, 74*, 473-487. doi: 10.1348/096317901167488

Friedman, B.A., Cox, P.L., & Maher, L.E. (2008). An expectancy theory motivation approach to peer assessment. *Journal of Management Education, 32*, 580-612. doi: 10.1177/1052562907310641

Jassawalla, A., Shashittal, H., & Malshe, A. (2009). Students' perceptions of social loafing: Its antecedents and consequences in undergraduate business classroom teams. *Academy of Management Learning and Education, 8*, 42-54. doi: 10.5465/AMLE.2009.37012178

Jefferson, A.L. (2010). Performance appraisal applied to leadership. *Educational Studies, 36*, 111-114. doi: 10.1080/03055690903148597

Kwok, L. (2008). Students' perception of peer evaluation and teachers' role in seminar discussions. *Electronic Journal of Foreign Language Teaching, 5*, 84-97.

Landy, F.J., Barnes, J., & Murphy, K. (1978). Correlates of perceived fairness and accuracy in performance appraisal. *Journal of Applied Psychology, 63*, 751-754.

Li, H., Xiong, Y., Hunter, C.V., Guo, X., & Tywoniw, R. (2020) Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education, 45*(2), 193-211.

Magin, D. & Helmore, P. (2010). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education, 26*, 37-41. 10.1080/03075070120076264

Morales-Trujillo, M.E., Galster, M., Gilson, F., & Mathews, M. (2021). A three-year study of peer evaluation in a software engineering project course. *IEEE Transactions on Education, early access*. doi: 10.1109/TE.2021.3123682

Murphy, K. & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-oriented perspectives.* Newbury Park, CA: Sage.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. *Handbook of social and human conditions in assessment* (pp. 247-266)*.* New York: Routledge.

Petkova, A.P., Domingo, M.A., & Lamm, E. (2021). Let's be frank: Individual and team-level predictors of improvement in student teamwork effectiveness following peer-evaluation feedback. *The International Journal of Management Education, 19*, 1-15. doi: 10.5465/ambpp.2016.90

Pierson, D.S. (2016). The influence of peer evaluation on individual contributions to group work in online graduate education (Publication No. 10140344) [Doctoral dissertation, Northcentral University]. Proquest Dissertations Publishing.

Politz, J.G., Patterson, D., Krishnamurthi, S., & Fisler, K. (2014). CaptainTeach: Multi-stage, in-flow peer review for programming assignments. *ITiCSE '14: Proceedings of the 2014 Conference on Innovation and Technology in Computer Science Education, June 2014*, 267-272.

Rubin, E.V. & Edwards, A. (2020). The performance of performance appraisal systems: Understanding the linkage between appraisal structure and appraisal discrimination complaints. *The International Journal of Human Resource Management, 31*, 1938-1957. doi: 10.1080/09585192.2018.1424015

Sala, F. (2003). Executive blind spots: Discrepancies between self- and other-ratings. *Consulting Psychology Journal: Practice and Research, 55,* 222-229. doi: 10.1037/1061-4087.55.4.222

Sherwood, A.L. & DePaolo, C.A. (2007). Student peer evaluations in business education: A web-based administration. *Academy of Educational Leadership Journal, 11*, 109-120.

Sinha, N., Mesmer-Magnus, J. & Viswesvaran, C. (2012). Personality antecedents of self-other rating discrepancy. *Journal of Work and Organizational Psychology, 28,* 3-14. doi: 10.5093/tr2012a1

Smircich, L. & Chesser, R. (1981). Superiors' and subordinates' perceptions of performance: Beyond disagreement. *Academy of Management Journal, 24*, 198-205. doi: 10.2307/255835

Suñol, J.J., Arbat, G., Pujol, J., Feliu, L., Fraguell, R.M., & Planas-Lladó, A. (2016). Peer and self-assessment applied to oral presentations from a multidisciplinary perspective. *Assessment and evaluation in Higher Education, 41*(4), 622-637. doi: 10.1080/02602938

Taylor, M.S., Tracy, K.B., Renard, M.K., Harrison, J.K., & Carroll, S.J. (1995). Due process in performance appraisal: A quasi-experiment in procedural justice. *Administrative Science Quarterly, 40*, 495-523. doi: 10.2307/2393795

Vroom, V.C. (1964). *Work and motivation.* New York: John Wiley & Sons.

Yammarino, F.J. & Atwater, L.E. (1997). Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management. *Organizational Dynamics, 25*, 35-44.

DECLARATION OF CONFLICTING INTERESTS

FUNDING

Cite