Journal of Education for Business

ISSN: 0883-2323 (Print) 1940-3356 (Online) Journal homepage: https://www.tandfonline.com/loi/vjeb20

The effectiveness of peer assessment and a proposal for its analysis using game theory

Gene Klein

To cite this article: Gene Klein (2018) The effectiveness of peer assessment and a proposal for its analysis using game theory, Journal of Education for Business, 93:8, 436-442, DOI: 10.1080/08832323.2018.1494531

To link to this article: https://doi.org/10.1080/08832323.2018.1494531

Published online: 13 Sep 2018.

JOURNAL OF EDUCATION FOR BUSINESS
2018, VOL. 93, NO. 8, 436–442
https://doi.org/10.1080/08832323.2018.1494531

Routledge
Taylor & Francis Group

PROFESSIONAL PERSPECTIVE

The effectiveness of peer assessment and a proposal for its analysis using game theory

Gene Klein

Mathematics, Keiser University, Fort Lauderdale, Florida, USA

ABSTRACT
A close look at the educational literature reveals both negative perceptions among peer assessment participants and evidence that peer assessment may not as effective as generally thought. An analysis of peer assessment using game theory is proposed and examined. This examination helps to explain the aforementioned lack of effectiveness and negative perceptions. The resulting model suggests interventions that may mitigate these negative perceptions and lead to an increase in peer assessment effectiveness.

KEYWORDS
Effectiveness; game theory; learner perceptions; peer assessment

There is much literature laying claim to the effectiveness of peer assessment as a teaching strategy, not only in connection with merely reducing instructor labor and time, but also providing benefit to the learners (Kaufman, & Schunn, 2011; Klucevsek, 2016; Nicol, Thomson, & Breslin, 2014). However, a closer reading presents a less rosy picture. This same literature noted student dissatisfaction with peer assessment, due in large part to their perception of their peers' lack of ability to provide an effective review and feedback. If learners distrust their instruction, then that instruction will hardly enjoy effective transmission.

Too often, authors arrived at conclusions regarding peer assessment effectiveness that may be unwarranted, due to their generality. For example, Klucevsek (2016) found evidence of peer assessment effectiveness by drawing on the improved results of students in a second-language classroom. In fact, most prior research on peer assessment effectiveness has focused on second-language learners (Patchan, Schunn, & Correnti, 2016). However, generalizing from second-language learners, where almost any practice and repetition can be expected to yield exponential improvements, to students experiencing instruction in their native language, is hardly appropriate.

These issues regarding the actual effectiveness of peer assessment may be amplified when individuals experiencing peer assessment lack complete information about factors germane to the experience. This lack of knowledge of various germane factors, such as the skill set of the reviewer, may be inherent to the peer assessment environment. Practitioners in other fields have used the tools of game theory to mitigate and even overcome issues of incomplete information. Here, I propose a game theory model to analyze the effectiveness of peer assessment. The model reveals and explains the less rosy picture just noted.

Literature review

Wilson, Diao, and Huang (2015) reviewed the use of peer assessment as a learning tool in institutions of higher education (IHEs). They argued that this use of peer assessment as a learning tool is an institutional reaction to the development of student-centered, collaborative, and online learning. They observed that assessing this peer assessment strategy is a "challenge" (Wilson et al., 2015, p. 15). One reason for this assessment challenge could be the ambiguity of the term *peer assessment*. Peer assessment, peer review, peer feedback (PF), and peer learning are all referred to as peer assessment (Wilson et al., 2015). Arguably, awareness on the part of the student reviewer that his or her review will have an impact on a fellow student's grade will materially impact the review experience. Accordingly, PF should result in a materially different learning experience, for both the reviewer and the reviewee, than a true peer assessment (TPA).

An example of this different learning experience may be inferred from the following observation of Wilson et al. (2015): that universities "quietly

CONTACT Gene Klein gklein@keiseruniversity.edu Mathematics, Keiser University, 1500 NW 49th St., Fort Lauderdale, FL 33309-7203, USA.

encourage an individualistic and competitive atmosphere" (p. 16). Wilson et al. observed that such an atmosphere limits the value and effectiveness of peer assessment to the students. Arguably, this limited effectiveness should be differentially experienced. While the strategies of both PF and TPA should be negatively affected by an individualistic and competitive atmosphere, this atmosphere should have a different impact on students experiencing PF than students experiencing TPA. For example, a competitive TPA reviewer might give the reviewee a lower-than-deserved grade, whereas a competitive PF reviewer might give the reviewee unhelpful feedback. Importantly, Wilson et al. observed that learner unhappiness with peer assessment has been deemphasized and underexplored in the literature. In fact, these theoretical inaccurate and suboptimal reviewer results could be a significant contributing factor to this underexplored learner unhappiness.

An example of this underexplored learner dissatisfaction with peer assessment may be inferred in a study of online instruction best practices by Batts (2008). He reviewed both learner and instructor perceptions of seven instructional strategies. The second instructional strategy was student cooperation. Batts reported that the mean scores for student cooperation, for both student and instructor, were "dramatically lower" (p. 482) than the other six instructional practices. Although Batts did not specify exactly which practices make up student cooperation, it is reasonable to conclude that at least some of these practices include peer assessment. In fact, this vagueness regarding the definition of student cooperation may be indicative of what Wilson et al. (2015) meant by underexplored.

Wilson et al. (2015) questioned the appropriateness of entrusting feedback and review to learners, asking if these learners have either the necessary competencies or motivation to successfully provide such review. Interestingly, this issue of competencies, or lack thereof, may have been unknowingly addressed by Nicol et al. (2014). Nicol et al. observed that learners identify PF as "more understandable" (p. 103) than instructor feedback. However, this increased understandability could actually reflect difficult concepts being improperly filtered, to the point of being diluted, through learners who have not yet received the scaffolding necessary for genuine understanding.

In an attempt to further calibrate and understand this dissatisfaction, Kaufman and Schunn (2011) studied 250 undergraduates and graduate students. These students were enrolled in 10 different courses and matriculated at six different universities. They found that students' perception of peer assessment unfairness correlated with their perception of the peer's lack of ability to engage in an accurate and meaningful review. Significantly, students' perceptions of peer assessment fairness significantly dropped after their peer assessment experience. This suggests that they did not have preconceived ideas of their peer's lack of ability, but rather discovered them during the experience. If this discovery process was valid, and can be confirmed, then it further calls into question the effectiveness of peer assessment.

Peer assessment and participants with incomplete information

There appears to be a certain amount of incomplete information that is inevitably part of important dimensions of the peer assessment experience, and in particular the peer assessment experience in an online environment. Most students tend to view their peer-learners as less knowledgeable than their instructor. Moreover, in a required general education (GE) class, one's peers are pursing many different and specific academic subjects. These subjects may be very different than the particular GE class that the diverse cohorts are learning together. Accordingly, learners may not know, or question, if their potentially distant peer assessor is capable or knowledgeable enough to effectively review their work. Indeed, in a study on peer assessment, Patton (2012) noted that one student had this concern about her peer assessor in every class. The peer assessment system itself may be set up anonymously (Kaufman & Schunn, 2011), meaning that the reviewee cannot request, even informally, a desired reviewer. Patton noted a further example of incomplete information when English as a second language students were being reviewed. The reviewer could be unaware that he or she was reviewing an English as a second language student, and thus could potentially give the reviewee a lower-than-appropriate grade.

Other factors, possibly external, of which both the reviewee and the reviewer may or may not be aware, can significantly impact the peer assessment experience. For example, grade inflation may exist at the university or individual class level. Reviewers who believe that they operate under this grade inflation regime (GIR) may logically conclude that even if the reviewee's work is less than adequate, it is adequate enough given the GIR. Accordingly, they may deliver inferior feedback to their reviewee.

A significant number of students do not actually read their class syllabus (Chapman, 2017). Moreover, undergraduates who are new to heavy college workloads arguably lack the skills of strategic academic time management. This means that the peer assessment's stakes may be another unknown quantity to these students. Moreover, instructors are free to grade the reviewer on the quality of his or her peer assessment. Accordingly, if the reviewer believes he or she is not being graded on the review, and that the stakes for the reviewee are low anyway, arguably he or she has very little incentive to deliver anything more than cursory feedback. Importantly, when the reviewee's belief of either of these factors is not aligned with the reviewer's skills or motivations, the results may not only be ineffective, they may actually be damaging. For example, the reviewee could incorrectly think that the assignment under review is high stakes, while the reviewer correctly believes it is low stakes. This could result in the reviewer doing an inferior job (as he or she believes it is not important), but with the reviewee unaware that the feedback is inferior and accordingly uses it. This could mean that students internalize false messages about course topics, the adoption of bad academic writing behavior, or various other counterproductive outcomes.

Game theory

Scenarios such as the foregoing may be analyzed using game theory. Game theory is less about games in the traditional sense and more of an "exact science of strategy" (Poundstone, 2008, p. 31). In fact, any situation involving at least two individuals who are seeking to either gain a prize or avoid a punishment is considered a game (Pitt, 2000). Game theory examines how "rational adversaries" make decisions, "knowing that their opponents are trying to second-guess or double-cross them" (Poundstone, 2008, p. 31). In other words, the players' decisions are not only molded by a desire to gain a prize or avoid punishment but they also must take into account the expected decisions of the other player or players. Importantly, these other players' decisions, and indeed the basis for their decisions, may not be entirely known. This decision making based on incomplete information forwards the case to use game theory to analyze peer assessment.

The standard tool used in the analysis of the player's decisions is the payoff matrix (Ruby, 2003). This payoff matrix displays the universe of both players' results—or payoffs—given the choices available to both players. The probable and optimum strategies are revealed through a comparative analysis of the payoffs. If the point of PF as a teaching strategy is for students to achieve increase increased learning through positive curriculum-related messages from their peers, then effective PF is a goal worth students' time and effort to pursue. Thus, equating effective PF with the reviewee's desired payoff, and the reviewer as the reviewee's opponent, game theory becomes a productive lens through which to view peer assessment.

Absence of game theory in the peer assessment literature

The initial article on game theory was published almost 100 ago, in 1928, by John von Neumann (Poundstone, 2008). In 1944, John von Neumann and Oskar Morgenstern published the first article on the application of game theory to economics. By 2012, The Massachusetts Institute of Technology was delivering an online course which discussed the applications of game theory to economics, politics, law, biology, and computer science (Massachusetts Institute of Technology, 2017). However, a review of the educational literature reveals hardly any mention of applying game theory to assessment in education and no mention of game theory's application to peer assessment in particular. In fact, the only related study found was by Pitt (2000), who studied the expected behavior and actual results of learners involved in group work using the lens of games theory. Indeed, Pitt noted with surprise the rarity of applying game theory to assessment in education.

Analyzing peer assessment effectiveness using game theory

In IHEs, students arguably want to avoid at least the punishment of a failing grade, and many students will desire the reward of a high grade. Even when peer assessment only functions as PF and not TPA, students will tend to value their peer's feedback only if they think it will improve their work and accordingly the grade they earn on the work. However, as noted, some seasoned learners have experienced less-than-adequate peer assessment and could reasonably expect this pattern to continue. Less seasoned learners might not know what to expect regarding the effectiveness of peer review. When low-grade avoidance and the desire for a high grade is coupled with the aforementioned lack of complete information regarding the reviewer's abilities, the idea of viewing peer

assessment participants as players in a game becomes clearer. Importantly, it is possible that the reviewer might not take the view that he or she is striving against the reviewee for any payoff. However, the reviewee's prize—feedback—is still dependent on the reviewer. This means that the notion of viewing peer assessment through the lens of game theory becomes relevant.

A theoretical examination of how a rational reviewee should proceed, given his or her perceptions of the reviewer's skills, the reviewer's competitiveness, or other factors, regardless of these perceptions' validity, is illuminating. This examination may explain and even predict some of the negative perceptions that learners have about PF. This theoretical model assumes the reviewee and reviewer lack relevant information or evidence concerning characteristics that might have an impact on the review. Under these circumstances of uncertainness, the reviewee can only make guesses concerning this relevant information. These guesses may be entirely uninformed guesses or they may hinge on the reviewee's risk–reward assessment of the potential value of the reviewer's feedback. The model predicts that actual effectiveness, or lack thereof, of the PF is dependent on how well the reviewee's assessment of either the reviewer characteristics or external factors aligns with these actual characteristics or factors.

Using the model to analyze if the reviewee should value reviewer feedback

Tables 1 and 2 display payoff matrices using the proposed model to analyze peer assessment effectiveness. The tables display the possible reviewee payoffs given a particular reviewer characteristic. Tables 1 and 2 display the theoretical predictions of the reviewee's possible actions, if he or she is unaware of respectively, the reviewer's abilities or competitiveness. Each table displays how a rational reviewee should value his or her reviewer's feedback, given his or her perceptions of the reviewer's characteristic. The model assumes that a rational reviewee should value the reviewer's feedback only if he or she believes that the reviewer is either competent or noncompetitive. Solely in the case that the reviewee is correct in his or her assessment of the reviewer, the PF will at least be free from any potential noneducational agenda of the reviewer, and could possibly be useful. In all other cases, the model predicts that PF will be unreliable. Moreover, the model predicts PF to be actually harmful when the reviewee has incorrectly assessed the reviewer's characteristic and accordingly accepts either incompetent or purposely misleading feedback.

Table 1 displays the prediction that PF is effective only in one out of four cases—the upper left cell, which displays an alignment between the reviewee's belief in the reviewer's competence and the reviewer's actual competence. In all three cases in which there is a nonalignment between reviewee belief and reviewer actuality, the model predicts ineffective PF. In one of the four cases (upper right cell), the model predicts actual harm to the reviewee. In two of four cases no harm is predicted, but in one case (bottom left cell) the model predicts that the reviewee will reject feedback that he or she should have accepted.

If reviewer competitiveness is the sole variable that the reviewee is considering, then, arguably, he or she cannot automatically assume that the feedback from a noncompetitive reviewer will be effective. The feedback may be agenda-free, but this agenda-free feedback may come from an incompetent reviewer. Accordingly, in this case the model predicts that the reviewee should not necessarily accept, but merely consider, feedback when his or her belief is aligned with reviewer reality.

Although Table 2 appears to be isomorphic to Table 1, with noncompetitiveness replacing competence, there is no a priori reason to assume that a reviewee's estimation of a reviewer's competence should be the same as the reviewee's estimation of that reviewer's noncompetitiveness. In fact, an individual's competence was preselected, via the admissions process, by the IHE that he or she is attending. On the other hand, an individual's competitiveness, or lack thereof, is part of his or her personality and is

Table 1. Payoff matrix for reviewee given reviewer competence.

	Reviewer is competent	Reviewer is incompetent
Reviewee believes reviewer is competent	Reviewee accepts feedback and should accept it	Reviewee accepts feedback but should reject it
Reviewee believes reviewer is incompetent	Reviewee rejects feedback but should accept it	Reviewee rejects feedback and should reject it

Table 2. Reviewee's payoff matrix given reviewer competitiveness.

	Reviewer is noncompetitive	Reviewer is competitive
Reviewee believes reviewer is noncompetitive	Reviewee considers feedback and should consider it	Reviewee considers feedback but should reject it
Reviewee believes reviewer is competitive	Reviewee rejects feedback but should consider it	Reviewee rejects feedback and should reject it

generally not revealed in the admissions process. Accordingly, a reviewee would have good reason to estimate a reviewer's competence and competitiveness differently.

A possible exception to the model's predictions in this case could occur at an IHE that has institutionalized competition through grading curves. For example, economics instructors of all classes with more than 60 students at Rutgers University are required to grade according to a published curve (Rutgers, 2017). Arguably, the number of students who would be unaware of such an important policy would be quite low. Indeed, learner resistance to PF as a teaching strategy would probably be quite high in such an environment.

The differential estimation of competitiveness and competence means that these characteristics should be treated as independent variables in the proposed model. This not only makes the analysis more complex, but also the model reveals the situation to be worse for the reviewee. An analysis that looks at this independency is displayed in Table 3. In this case, the model predicts that the reviewee will and should accept the feedback only when the reviewee's beliefs about the reviewer's two separate characteristics are aligned with the reviewer's actual characteristics.

Table 3 predicts effective feedback in only 1 of 16 cases. In 3 of 16 cases, the reviewee is harmed, and in 3 of 16 cases valid feedback is rejected. In other words, the model predicts that a rational reviewee, who wishes to avoid much risk, should always reject reviewer feedback.

Table 3. Reviewee's payoff matrix given reviewer competitiveness and competence.

	Reviewer actually is competent and noncompetitive	Reviewer actually is competent but competitive	Reviewer actually is incompetent and competitive	Reviewer actually is incompetent but noncompetitive
Reviewee believes reviewer is competent and noncompetitive	Reviewee accepts feedback and should accept it	Reviewee accepts comments but should reject it	Reviewee accepts feedback but should reject it	Reviewee accepts feedback but should reject it
Reviewee believes reviewer is competent but competitive	Reviewee rejects feedback and should accept them	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and should reject it
Reviewee believes reviewer is incompetent and competitive	Reviewee rejects feedback but should accept it	Review rejects feedback and should reject it	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and should reject it
Reviewee believes reviewer is incompetent and noncompetitive	Reviewee rejects feedback but should accept it	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and should reject it

Using the model to analyze incomplete information regarding external factors

Both the reviewee and reviewer may have incomplete information about external factors, such as if the IHE in which they are matriculating or the class in which they are enrolled is subject to a GIR. If the reviewee does not know if there is a GIR, the model assumes that his or her safest bet is to always reject reviewer feedback. This is because the reviewer will tend to provide inferior feedback if there is a GIR. Importantly, the model does not take into account if there actually is a GIR. The model assumes that it is only the reviewer perception of a GIR that affects his or her feedback. Table 4 displays the predictions for this GIR situation. Similar to in the other three cases, feedback is predicted to not be misleading when reviewee belief is aligned with reviewer belief.

Table 4. Reviewee's payoff matrix given GIR.

	Reviewer believes that class has a GIR	Reviewer believes that class does not have a GIR
Reviewee believes that class has a GIR	Reviewee rejects feedback and should reject it	Reviewee rejects feedback and but should not reject it
Reviewee believes that class does not have a GIR	Reviewee accepts feedback but should reject it	Reviewee accepts feedback and should accept it

Note. GIR = grade inflation regime.

Model implications

Peer assessment participants may not necessarily view themselves as rational players in a strategic game. However, frequent participants may not be surprised by the model's prediction of a low probability of effective feedback. In fact, the model may validate the aforementioned results of Kaufman and Schunn (2011). Their results that seasoned participants view peer assessment as unfair could be related to the model's predictions of a high probability of reviewees' receiving ineffective feedback. It would appear that seasoned participants are learning from experience that feedback from their peers tends to be ineffective.

The model's predictions of a high probability of ineffective feedback stem from assumptions such as a lack of reviewer competence and an abundance of reviewer competitiveness. Instructors may mitigate competence issues by giving students instruction on how to conduct effective and positive PF. In fact, this would serve the dual purpose of increasing students' abilities while reducing reviewees' concerns of reviewer competency. Of course, an increased competency in conducting PF may not translate into competency in the content itself. It may be that the reviewer should limit him- or herself to feedback on how well the reviewee communicated the content.

Concerns about reviewer competitiveness can also be addressed by instructional intervention. Arguably, PF is an instructional activity for the reviewer as much as, if not more so, the reviewee. The class can be designed so that PF is an assessment for the reviewer only. In fact, this structure would allow PF to take on a discussion quality. All reviewer feedback could be made available for all learners to read and discuss. Indeed, upon reviewing input from other learners, the reviewer might even be allowed to revise his or her review. This structure could truly turn peer assessment into the aforementioned student-centered, collaborative, and online learning noted by Wilson et al. (2015).

Conclusion

Game theory has helped or allowed practitioners in numerous fields to analyze and find solutions to difficult problems. Yet, educational practitioners have denied themselves this useful tool. Game theory, and the proposed model, opens a door into a widely used instructional tool and helps to explain student dissatisfaction with it. Although peer assessment has been thought to be effective and welcomed as an instructional strategy by learners, a closer look at the literature displays this student dissatisfaction. The model's predictions appear to confirm that many learners have good reason to distrust feedback from their peers. This distrust may stem from both external factors and negative perceptions of reviewer characteristics. The model's assumptions may point to strategies that minimize this distrust and increase the overall effectiveness of peer assessment.

However, the model is currently a theoretical construct. Although its predictions appear consistent with findings in the educational literature, at present there is no primary research to support it. It is hoped that this introduction encourages the research community to further validate the proposed model.

References

Batts, D. (2008). Comparison of student and instructor perceptions of best practices in online technology courses. *MERLOT Journal of Online Learning and Teaching*, 4(4), 477–489.

Chapman, S. (2017). *Getting students to read the class syllabus*. Teaching @CSU, Retrieved from http://teaching.colostate.edu/tips/tip.cfm?tipid =50

Kaufman, J., & Schunn, C. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39(3), 387–406.

Klucevsek, K. (2016). Transferring skills from classroom to professional writing: Student-faculty peer assessment as an extension of cognitive apprenticeship. *Journal of the Scholarship of Teaching and Learning*, 16(6), 106–123.

Massachusetts Institute of Technology. (2017). *Economic applications of game theory*. MITopencourseware. Retrieved from https://ocw.mit.edu/courses/economics/14-12-economic-applications-of-game-theory-fall-2012/

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer assessment perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122.

Patchan, M., Schunn, C., & Correnti, R. (2016). The nature of feedback: How peer feedback features affect students'

implementation rate and quality of revisions. *Journal of Educational Psychology*, 108(8), 1098–1120.

Patton, C. (2012). 'Some kind of weird, evil experiment': Student perceptions of peer assessment. *Assessment & Evaluation in Higher Education*, 37(6), 719–731.

Pitt, M. (2000). The application of games theory to group project assessment. *Teaching in Higher Education*, 5(2), 233–241.

Poundstone, W. (2008). *Gaming the vote: Why elections aren't fair (and what we can do about it)*. New York, NY: Hill and Wang.

Ruby, D. (2003). *Game theory*. Digital Economist. Retrieved from http://www.digitaleconomist.org/game_theory.html

Rutgers, (2017). *Department grading policy*. www.ncas.rutgers.edu Retrieved from https://www.ncas.rutgers.edu/department-economics/department-grading-policy

Wilson, M., Diao, M., & Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': An investigation of an online peer-to-peer assessment workshop tool. *Assessment & Evaluation in Higher Education*, 40(1), 15–32.