How to Choose an Appropriate Reviewer Assignment Strategy in Peer Assessment System? Considering Fairness and Incentive

Yanqing WANG[1,a,*] and Fuquan SUN[2,b]

[1]School of Management, Harbin Institute of Technology, Harbin 150001, China

[2]School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Hebei 066004, China

[a]yanqing@hit.edu.cn, [b]sunfq@neuq.edu.cn

*Corresponding author

Abstract. As a prior process of peer assessment, reviewer assignment plays a critical role and, eventually, determines the performance of peer assessment. Nowadays, so many reviewer assignment strategies exist that relevant researchers need a big picture of reviewer assignment strategies. In this study, we propose a taxonomy of reviewer assignment strategies, which consists of four dimensions including stability, anonymity, number ratio (reviewer to author), and assigner. We analyze the fairness and incentive effect of each reviewer assignment strategy. Then we share our experiences and lessons from implementation of an online peer assessment system in educational context, which could be references for relevant scholars and instructors.

Introduction

Peer assessment, also called peer evaluation, peer feedback or peer review, is defined in Topping's highly cited review as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance by other equal-status learners" [1]. The worth of peer assessment is widely illustrated especially in educational context [2-5]. In peer assessment, every author's work should be assigned to one or more reviewers before the assessment process starts. As a critical step of peer assessment, reviewer assignment strategy has a huge impact on attitudes of reviewers and authors, review efficiency, and review quality. There are several successful online peer assessment systems, such as Calibrated Peer Review [6], Expertiza[4], SWoRD[3], and EduPCR[5,7,8].

EduPCR (peer code review in educational context) is an online peer assessment system for programming language learning. It is suitable for undergraduates, postgraduates, and junior college students who need to learn programming skills. In EduPCR, every learner is required to participate some phases, such as completing coding task, reviewing peer's work, revising own program, doing back-evaluation, etc. Instructor is responsible for setting tasks and summarizing final scores of students. Computer assigns reviewers and manages the schedule [9].

Since 2004, EduPCR has been updated for several versions. We applied it in the evaluation process of three courses, including *C++*, *C Programing*, and *Object-Oriented Programming in Java*, at two schools of Harbin Institute of Technology. With this open-type learning approach, students improved greatly in their high-order capabilities, such as analyzing, expressing in writing, critical thinking, and innovative thinking.

In fact, the optimum strategy that suits all conditions does not exist. Practically, instructor should weigh the pros and cons of every aspect and choose appropriate strategy. This study puts forward a taxonomy of reviewer assignment and gives a brief description of each strategy. Then an analysis is conducted from fairness and incentive aspects to illustrate the choosing of reviewer assessment strategy.

Taxonomy of Reviewer Assignment Strategy

From four dimensions, including *stability*, *anonymity*, *number ratio*, and *assigner*, taxonomy of reviewer assignment strategy is depicted, as in Figure 1.
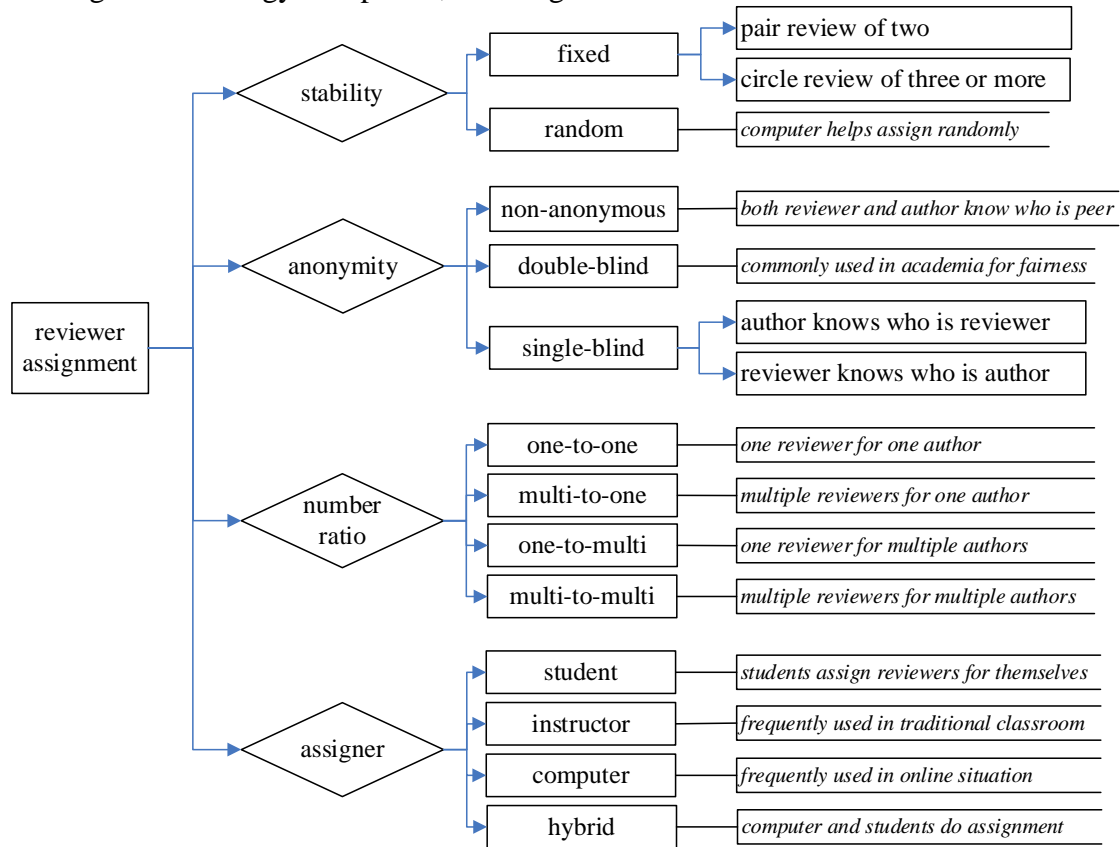
Figure 1. Reviewer assignment taxonomy in peer assessment

Stability

Fixed assignment means the review pairs are fixed for a relatively long period, e.g. course or a semester. Pair review of two and circle review of three were used by us. Every reviewer reviews the same author's programs and every author's work was reviewed by the same reviewer for a long time. This often brings on the phenomenon that student peers collaborate to finish a task (work together), and the problem of ghostwriting is exposed.

Random assignment denotes that computer generates designating rings by random. Topping discussed random assignment in their peer assessment systems [1]. Schunn also applied random assignment strategy in SWoRD [3]. This strategy enables each reviewer to review different author's work and each author's work is reviewed by a different reviewer for each time. In this strategy, different writing styles and review perspectives help students learn more.

Anonymity

Double-blinded method is an important part of the scientific method, used to prevent research outcomes from being "influenced" by the placebo effect or observer bias. Blank found that the reviewers are more critical when they are unaware of the author's identity [10]. In double-blinded strategy, a student does not know peer's identity. This strategy ensures the quality of the review by excluding the factors of human relationships. Moreover, the problem of ghostwriting could be inhibited.

Single-blinded can be defined by authors' blindness to reviewer's identity and reviewers' blindness to authors' identity. Knowing reviewer' name helps author shield against disturbance caused by

misunderstanding of reviewer's suggestions or mistakes made by reviewer. Getting authors' names allows reviewers shield against bewilderment of authors' advanced algorithm.

Peer assessment should be anonymous. Ballantyne et al. reported that anonymous peer assessment reduces the opportunity of collusion and biased marking [11]. *Double-blinded* strategy is also supported by interviews with our students. Some students said:

"Peer assessment - anonymous is a good idea, because if it's not anonymous, I incline to give bad marks to people I know."

"I don't want to do peer assessment exercise, if it is not anonymous."

Number ratio

From the aspect of number ratio, there are four cases of reviewer assignment:

 (a) one-to-one. One reviewer is assigned for one author's work

 (b) multi-to-one. Multiple reviewers are assigned for one author's work

 (c) one-to-multi. One reviewer is assigned for multiple authors' work

 (d) multi-to-multi. Assigning in the ways of case (b) and case (c)

The problem is to balance participants' workload and the reliability of their assessment.

At first, multi-peer strategy is not encouraged. So long as multiple participants are involved in an algorithm, i.e. in case (b), (c) or (d), instructor should be careful because participants may spend so much time that they get tired and bored. That will in turn discount the outcome of the whole assessment process.

However, in order to enhance the reliability of students' assessment results, multi-peer strategy is necessary because more assessors may make more accurate assessment.

Practically, case (a) and (d) might be the most popular because the essence of "peer assessment" implies that every participant has equal right to be a reviewer and a reviewee.

Assigner

Who assigns reviewer to author's work is very important because it affects both quality control and participants' passion of learning. The assigner of a review task can be *student*, *instructor*, or *computer*.

Student as assigner (self-assignment or free-selection). It enables students to choose favorite reviewers. Li implemented self-assignment strategy: students chose their reviewers by themselves [2]. In this circumstance, capable students will be more popular and they will be chosen by more than one students. This will lead to problems such as unfair distribution.

Teacher as assigner. Li utilized a method in which reviewer assignment were made by teachers according to students' characters or randomly. This method helps to avoid the negative impact of "mutual admiration societies" [2].

Computer as assigner. In this way, reviewers are assigned by computers based on some algorithm. Thus, the quality assurance depends on the quality of assignment algorithm.

Mixed assigner. In recent years, we tried *mixed assigner* strategy in latest version of EduPCR. This strategy suits for multi-to-multi assignment. For example, in a 5-to-5 strategy, a student is allowed to choose *two* reviewers for self, and computer assigns *three* for this student [8].

Papadopoulos et al. compared and found that students following the free-selection protocol (students are allowed to freely explore and select peer work for review) demonstrate better domain learning outcomes and better reviewer skills, compared to the assigned-pair condition (the teacher assigns student works for peer's assessment).

Analysis of Reviewer Assignment Strategies

In peer assessment, fairness is critical because it correlates with education equality and affects the participants' mental feelings in peer assessment. Meanwhile, incentive mechanism may foster the learners' passion. Therefore, we analyze all strategies, as shown in Table 1.

Table 1. Fairness and incentive analysis of reviewer assignment policies

Category	Sub-category	Fairness	Incentive	+/-
stability	*fixed*	student assigned to a low-skilled peer may feel unfair	little incentive because peers are fixed	+
	random	everybody may feel fair	small incentive because student may often have a different peer	+
anonymity	*non-anonymous*	most low-skilled students may need more care (face saving)	most low-skilled students may feel uncomfortable	-
	double-blinded	everybody may feel fair	most students review peers' work without worries	+
	single-blinded	low-skilled students may need more care (face saving)	unhidden low-skilled student may feel uncomfortable	-
number ratio	*one-to-one*	everybody may feel fair	little incentive	0
	multi-to-one		author may feel overloaded	-
	one-to-multi		reviewer may feel tired	-
	multi-to-multi		little incentive	0
assigner	*student*	slow students may feel unfair because of contest; students having few friends may feel unfair	early student get high-skilled student or friend as reviewer	+
	instructor	some low-score students may query the fairness of instructor	little incentive except instructor design a motivation way	0
	computer	everybody may feel fair	little incentive except a good algorithm is designed	0
	mixed	slow students may feel unfair	early student get high-skilled student or friend as reviewer	+

Note: the last column is valence of incentive effect, being positive (+), negative (-), or none (0)

Experiences, Lessons, and Future Work

After over ten years' implementation of EduPCR, we share our ideas, as follows.

Experiences

Stability. In year of 2004 and 2005, we attempted *fixed* assignment strategy. The student number was small that time. However, with the increasing of student number in bigger classes, we discarded *fixed* and utilized *random* strategy instead (see Fairness analysis in Table 1).

Anonymity. In a small class, almost all students know each other. *Non-anonymity* has big challenge so we did not apply it. Most of time, we utilized *double-blinded* strategy, which did not cause much problems. Actually, we tested *single-blinded* strategy partially in *mixed-assigner* strategy (see following section about Assigner).

Number ratio. Considering the equality of learning opportunity, we deployed 1-to-1 and 5-to-5. They both were practical and successful. Not like in SWoRD [3], in which each student may finish any number of assigned reviews, every student had to finish all reviews in EduPCR.

Assigner. We had two modifications about assigner. First, we replaced *instructor-assigner* with *computer-assignor* to decrease instructor's workload and possible fairness complaints from some students. Second, we had a try on *mixed-assignor* strategy, which was interesting while the management process became complex.

Lessons

Fairness. Students are sensitive to fairness. If instructor allows some students to play merely the role of author, they are likely to feel discrimination because being a reviewer is regarded as an honor (like being a teacher). Thus, instructor should give all student equal opportunities.

Balancing teacher's workload and students' workload. In multi-to-multi strategy, the reliability of assessment is probably much higher than that in 1-to-1 strategy. However, students may complain about the workload. In 1-to-1 strategy, instructor has to do much work on scoring since students' assessment may be not reliable enough.

Self-assigning needs trade-off. Self-assigning ignites incentive, improves students' experience. However, it is likely to trigger collusion, in which some students give relatively higher scores to each other merely because they are friends or they admire each other.

Future Work

In *multi-to-multi* strategy, since many reviewers are doing an identical task, there may arise such issues as *non-consensus*, *radicalness*, and *bias from mutual admiration society*. That needs more concerns of related scholars.

If the shortage of fairness could be overcome, the motivation effect of *student-assigner* or *mixed-assigner* strategy seems like an interesting topic worthy of more investigation. Because *student-assigner* implies *non-anonymous*, the anonymity issue should be scrutinized as well.

Acknowledgement

References

[1] Topping, K. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research, 68*(3), 249–276.

[2] Li X. (2007). Incorporating a code review process into the assessment, In the *20th Annual Conference of the National Advisory Committee on Computing Qualifications*, Nelson, New Zealand, 125−131.

[3] Schunn, C. D. (2016). Writing to learn and learning to write through SWoRD. *Adaptive Educational Technologies for Literacy Instruction*, In S.A. Crossley & D. S. McNamara (Eds.), Adaptive Educational Technologies for Literacy Instruction. NY: Taylor & Francis, Routledge.

[4] Gehringer, E. F. (2010). Expertiza: Managing feedback in collaborative learning. Monitoring and assessment in online collaborative environments: Emergent Computational Technologies for E-learning Support, 75–96.

[5] Wang, Y., Li, H., Fend, Y., Jiang, Y., and Liu, Y. (2012). Assessment of Programming Language Learning Based on Peer Code Review Model: Implementation and Experience Report. *Computers & Education,* 59(2): 412−422.

[6] Russell, A. A. 2004. Calibrated Peer Review: A Writing and Critical-Thinking Instructional Tool. In Teaching Tips: Innovations in Undergraduate Science Instruction, edited by M. Druger, E. D. Siebert, and L. W. Crow, 54−55. Arlington: National Science Teachers Association.

[7] Wang, Y., Ai, W., Liang, Y., and Liu, Y. Towards Motivating Participants to Assess Peers' Work More Fairly: Taking Programming Language Learning as an Example. *Journal of Educational Computing Research, 2015*, 52(2): 180−198.

[8] Wang, Y., Jiang, Y., Chen, M., and Hao, X. (2013). E-learning-oriented incentive strategy: Taking EduPCR system as an example. *World Transactions on Engineering and Technology Education (WTE&TE), 11*(3): 174−179.

[9] Wang, Y., Liang, Y., Liu, L., & Liu, Y. (2016). A multi-peer assessment platform for programming language learning: Considering group non-consensus and personal radicalness. *Interactive Learning Environments, 24*(8), 2011−2031.

[10] Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *The American Economic Review*, 1041−1067.

[11] Ballantyne, R.,Hughes, K., and Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process, *Assessment & Evaluation in Higher Education, 27*(5):427−441.

[12] Papadopoulos, P. M., Lagkas, T. D., & Demetriadis, S. N. (2012). How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course. *Computers & Education, 59*(2), 182−195.