

Peer-Assessment and Learning Outcomes: Product Deficiency or Process Defectiveness?

Chadi Aoun
caoun@efs.mq.edu.au
61-2-9850-9178

Department of Accounting and Finance
Division of Economic and Financial Studies
Macquarie University – NSW 2109
Sydney – Australia

Abstract

Over the past decade, peer assessment has become a controversial concept in higher education. There is much research evidence highlighting the effectiveness of peer assessment and its' potential advantages including the promotion of learners' responsibility, meta-cognitive development, autonomy, and deep learning. On the other hand, significant research findings point to its ineffectiveness in improving students' marks and performance, poor quality feedback, and potential bias. This paper presents a conceptual overview of the literature on peer assessment with particular emphasis on prior empirical findings. A case study with practical evidence from the author's context is then presented, reflecting on the implementation of peer assessment in a postgraduate unit over a two-year period. This sets a preliminary means to investigate whether the problem is with product (ie. peer assessment is fundamentally and conceptually flawed) or process (ie. peer assessment is wrongly implemented). The paper concludes with recommendations for future research and practice.

Introduction

There are many conflicting opinions about the validity and usefulness of peer assessment in higher education. In his meta-analysis of pervious studies, Topping (1998, p. 267) concludes that "peer assessment seems equally likely to contribute to or not contribute to the assessee's final official grade." This seemingly paradoxical finding is reported in many studies that have found no difference in performance between two groups of students in the same unit, one adopting peer assessment, and the other not (Sluijsmans et al, 2004), with no obvious explanation. There was also no evidence of a relationship between students' capacity to mark their peers and their own performance on an assessment task (Bloxham and West, 2004) raising questions about the validity and reliability of the process. These contradictory views are rather intriguing, and have prompted our investigation of peer assessment, with the intent to examine whether the concept is inherently flawed, or whether it the problem emanates from improper implementation.

Definition

Many definitions exist about what constitutes peer assessment. Topping (1998, p. 250) defines peer assessment as “an agreement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status.” Similarly, Van Den Berg, Admiraal and Pilot (2006, p. 19) propose that “student assessment is understood to be an arrangement with students assessing the quality of their fellow students’ writings and giving feedback to each other”.

It is important to highlight that some of the recent hype about peer assessment is due, at least in some part, to the facilitation role of information technology as a major component in peer learning (Topping, 2005). Another fundamental concept proposed in the literature proposes that peer assessment is a skill that can, and should, be acquired through training (Sluijsmans et al, 2004), and as such many implementations could fail if the training component is neglected.

Advantages

Topping (2005, p. 640) asserts that “peer assessment can enhance self-assessment, and both can yield metacognitive gains.” Various other advantages to peer assessment are suggested in the literature, a summary of which is presented by Cassidy (2006, p. 509), and includes:

- “Increased student responsibility and autonomy;
- evaluative skill development;
- insight into assessment procedures and expectations for high quality work;
- students work harder with the knowledge that they will be assessed by their peers;
- potential for providing increased levels of feedback without increased demands on tutors; and
- encourages deep rather than surface learning.”

In addition, peer assessment is advocated as a means of improving employability skills, especially those related to non-technical (generic) aspects including oral and written communication, reading, learning skills and strategies, problem solving, decision making, dependability, and responsibility (Cassidy, 2006). It is therefore often recommended to be routinely implemented in the context of higher education as one of the ultimate objectives of university students is to obtain suitable employment. This view is well supported by similar studies advocating that peer assessment “promotes the acquisition of life-long skills due to the active involvement of students in the assessment experience” (Ballantyne et al, 2002, p. 428). This is because formative feedback from peers, especially detailed and open-ended feedback on assessment tasks is linked to improvements in students’ confidence, presentation, and skills appraisal (Topping, 1998).

Disadvantages

Conversely, concerns were also raised about the validity and reliability of the marks given by students due to lack of expertise, potential bias, discomfort with the extra responsibility, lack of formal training, and the perception that it's the job of the tutor to do the assessing (Cassidy, 2006). All of these issues could adversely influence assessment marking, and lead to a central tendency in granting marks namely, giving 'average' rating for peers (Topping, 2005), regardless of the quality of their work.

The most significant disadvantage emphasised in the literature is potential bias. While peer assessment seems to be more reliable than self assessment (Topping, 1998), it still suffers from a perceived lack of objectivity (Brindley and Scoffield, 1998). Bushell (2006) affirms that it is essential to moderate the peer assessment results if they are to be used for summative tasks, in order to ensure equity due to possible conscious or unconscious bias of peer assessors. This bias is often gender specific, favouring females, especially in peer assessment of group work, but that was not particularly due to same or opposite gender raters favouring similar or different genders to their own (Flachikov and Magin, 1997).

Students' Perceptions

Almost all studies indicate that the majority of students' exposed or involved in peer assessment held favourable views regarding the process. More than two-thirds of students were positive about peer marking in Bloxham and West's study (2004). Similarly, in a questionnaire survey (Van Den Berg et al, 2006), 80% of students attributed their improvements in writing to peer feedback. Students also felt personally motivated as a consequence of being involved in peer-assessment and saw it as an incentive to perform well (Brindley and Scoffield, 1998).

Nonetheless, many students are apprehensive about their capability to assess their peers and the responsibility associated with such assessment (Cassidy, 2006), leading to anxiety, stress, and discomfort. Moreover, most students found the peer-assessment process time consuming, intellectually challenging, and socially uncomfortable, but effective in improving the quality of their individual work and developing their skills (Topping et al, 2000).

Case Study

The author was involved in the administration of peer-assessment activities in a foundational postgraduate subject offered by the discipline of Information Systems, over the period of two years. The details below are based on the experience and perceptions of the author after maintaining a reflective journal over that period.

The unit in which peer-assessment was implemented is compulsory for all Information Systems and Accounting Masters Degree students, with many other students from various majors doing it as an elective to get a business focus or to advance their Information Systems knowledge and skills.

The subject included four summative assessment activities, two of which were peer-assessed quizzes, each worth 5% of the total assessment mark. Following are brief details of the rationale, preparation, and process of peer assessment as applied in the context of this unit.

Peer Assessment Process

About five (5) quiz questions were prepared for every quiz. Most of the questions were analytical and open ended. A marking guide was also prepared with sample answers, allocating a binary mark for each question. Students can only tick (1) if the answer is appropriate as per the marking guide; tick (0) if the answer is incorrect; or raise their hand and ask for the tutors' feedback if they thought the answer was reasonable, but is not accommodated for in the sample answers. No other formative or textual feedback or justification is given by markers. An initial formative practice quiz was administered early in the semester to familiarise students with the process.

The quizzes were closed-book and were administered during lab time. A marking guide was given to peer markers, immediately after quiz completion along with another student's quiz paper, distributed at random. The quiz paper only included the student's number and signature, to help preserve anonymity, to a practical extent. The students would then work on their marking in pairs (2 students assessing 2 papers), with guidance from the lab demonstrator. Finally students would put their details at the bottom of the marking sheets and return both the quiz and marking sheet to the tutor.

Tutors would then remark all papers to ensure an equitable mark has been granted. This is due to students in previous semesters perceiving the process as flawed and biased, citing that a student marking another's paper could have no knowledge of the concepts, and may not adequately understand the sample answers or request tutor's assistance.

The quizzes were usually returned the following week. Students would get back both their quizzes and the marking sheet. Even though all papers have been quality assured by tutors, about a quarter of the student cohort would appeal their marks and approach the lecturer for assistance. All issues were then resolved on the spot with about half of the appeals found valid, requiring a change in marks.

This process was refined over several semesters. The main issue is that there seems to be no connection between the peer marking process, and the dramatic increases in performance and grades suggested in some of the literature when compared to previous semesters that did not incorporate peer assessment, eg. the claim by Gibbs (1998, cited in Biggs, 2003) which reports on the outcomes of the introduction of peer-assessment in an engineering unit indicating that the average mark in the final examination had increased from 45% to 75%, with no failures.

Analysis and Recommendations

Ballantyne, Hughes, and Mylonas (2002) identified 10 conditions that are critical to peer assessment success which we assessed our case findings against in table 1 below:

Criteria for Critical Success in Peer Marking	IS Case
1. Prior students' experience with university assessment tasks.	Yes (Postgraduates)
2. Stand-alone item with significant length and complexity.	No (5 short answer Question)
3. Having all tutors 'on-side' and positive about peer marking.	Yes (Tutors in favour of peer Marking)
4. Staff-generated criteria sheet and a central form for coordination and dissemination of information (eg. Lecture).	Yes (Both)
5. Consider awarding marks for the feedback the peers provide.	No (No marks for assessors)
6. Incorporate practice sessions to familiarise students with the process.	Yes (Practice Quiz)
7. Use codes to preserve anonymity.	Yes (Student ID used)
8. Introduce a tutor moderation system for a sample of the assessments.	Yes (All quizzes checked and moderated)
9. Allow remarking by tutors if some students' are dissatisfied, provided the request for remarking is made rigorous (eg. Formal application with a written justification)	No (Remarking allowed, but non-formal request)
10. Do not overuse.	Yes (Only two quizzes)

*Table 1: Criteria for Critical Success in Peer Marking
(adapted from Ballantyne et al, 2002)*

Consequently, the peer assessment process applied in this unit met seven (7) of the ten (10) criteria specified above. Particularly, the missed criteria were: the assessment task was not of significant length and complexity (suggested at 6-8 pages by Ballantyne et al, 2002); no marks awarded for assessors; and informality of the remarking process. It could then probably be argued that if these criteria were met, potential improvements to the students' knowledge and therefore exam performance and marks could be achieved as suggested by Gibbs (1998, cited in Biggs, 2003).

Moreover, several studies recommend further advice which could contribute to the success of peer-marking:

Van Den Berg, Admiraal and Pilot (2006) recommend a combination of written and oral peer feedback, for peer assessments, as written feedback concentrated on the assessment of the text, while oral feedback included an explanation of the students' judgement and suggestions for improvement. This leads to the elimination of anonymity and could be problematic given the large class size (about 400 students divided into four streams). This approach is however supported by Paquet and Des Marchais (1998) who found that students preferred a non-

anonymous format in the peer-marking process, and appreciated the formative aspects of the assessment more so than the summative aspects.

The introduction of peer assessment should also emphasise the learning aspects associated with assessing peer work, rather than focusing on the allocation of marks and feedback (Cassidy, 2006). Our current assessment tasks (quizzes) may not be optimally conducive to this, due to its brief structure and non-complexity, and therefore considering other tasks could be more favourable (eg. content rich, complex assignments). To that end, Bloxham and West (2004) recommend the avoidance of over-specification of assessment criteria as it encourages instrumentalism and deters able students from extending themselves as they only focus on meeting the criteria. Therefore, adopting a binary approach is overly simplistic and fundamentally flawed. Furthermore, students' involvement in setting assessment criteria for peer-assessment tasks also contributes to their learning process (Sivan, 2000). Consequently, we need to involve students in developing such criteria and recommending sample answers, ideally for multiple peer-assessment tasks (Ballantyne et al, 2002).

Limitations:

The following limitations are inherent in this paper and could provide opportunity for further investigations:

Firstly, no empirical data were collected from students about their perceptions of the process, which could add a rich dimension to the analysis and suggestions presented above. Moreover, most academic literature on peer assessment addresses undergraduate programs. Our case study was on a postgraduate unit. This highlights the need for further empirical research on peer assessment and learning in a postgraduate studies context.

In addition, most academic literature on the subject targeted generic skills (eg. academic writing). More specific research to the Information Systems discipline could be beneficial in identifying peer assessment issues in this area, as Information Systems work is very collaborative and often relies on peer evaluation and feedback (Tu and Lu, 2005).

Finally, the question remains on the practicality of peer assessment in large classes incorporating students from various cultural backgrounds. The process is cognitively demanding on students and may require ongoing individual support and validation. This may put immense strains on the limited resources available to instructors of large classes.

Conclusion:

The benefits associated with peer marking seem to outweigh and disadvantages. However the mixed results regarding its validity in contributing to knowledge and performance, need to be considered within the context of how effectively it's been implemented. As our case study demonstrates, several amendments need to be undertaken before any critical judgements about the validity of peer-assessment in our context could be proven, or indeed refuted. Based on the evidence presented above, we are of the opinion that problems with peer assessment are due to process (ie. implementation) rather than product (ie. concept).

References

- Ballantyne, R., Hughes, K. & 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
- Biggs, J. 2003, *Teaching for Quality Learning at University*, Second Edition, SRHE & Open University Press, Buckingham
- Bloxham, S. & West, A. 2004, *Understanding the Rules of the Game: Marking Peer Assessment as a Medium for Developing Students' Conceptions of Assessment*, *Assessment & Evaluation in Higher Education*, 29, 6, 721-733
- Brindley, C. & Scoffield, S. 1998, *Peer Assessment in Undergraduate Programmes*, *Teaching in Higher Education*, 3, 1, 79-89
- Bushell, G. 2006, *Moderation of Peer Assessment in Group Projects*, *Assessment & Evaluation in Higher Education*, 31, 1, 91-108
- Cassidy, S. 2006, *Developing Employability Skills: Peer Assessment in Higher Education*, *Education + Training*, 48, 7, 508-517
- Magin, D.J. 2001, *A Novel Technique for Comparing the Reliability of Multiple Peer Assessments with that of a Single Teacher Assessment of Group Process Work*, *Assessment & Evaluation in Higher Education*, 26, 2, 139-152
- Paquet, M.R. & Des Marchais, J.E. 1998, *Students' Acceptance of Peer Marking*, *Education for Health*, 11, 1, 25-35
- Sivan, A. 2000, *The Implementation of Peer Assessment: an Action Research Approach*, *Assessment in Education*, 7. 2. 193-213
- Sluijsmans, D.M.A., Brand-Gruwel, S., Van Merriënboer, J.J.G & Martens, R.L. 2004, *Training Teachers in Peer Assessment Skills: Effects on Performance and Perceptions*, *Innovation in Education and Teaching International*, 41, 1, 59-78
- Topping, K. 1998, *Peer Assessment Between Students in Colleges and Universities*, *Review of Educational Research*, 68, 3, 249-276
- Topping, K.J. 2005, *Trends in Peer Learning*, *Education Psychology*, 25, 6, 631-645
- Topping, K.J., Smith, E.F., Swanson, I. & Elliot, A. 2000, *Formative Peer Assessment of Academic Writing Between Postgraduate Students*, *Assessment & Evaluation in Higher Education*, 25, 2, 149-166
- Tu, Y. & Lu, M. 2005, *Peer-and-Self Assessment to Reveal the Ranking of Each Individual's Contribution to a Group Project*, *Journal of Information Systems Education*, 16, 2, 197-205

- Van Den Berg, I., Admiraal, W. & Pilot, A. 2006, *Peer Assessment in University Teaching: Evaluating Seven Course Designs*, *Assessment & Evaluation in Higher Education*, 31, 1, 19-36