PEER ASSESSMENT IN PROBLEM-BASED LEARNING: STUDENTS' VIEWS

Magdeleine D.N. Lew[1], W.A.M. Alwis[1], Henk G. Schmidt[2]
Republic Polytechnic, Singapore[1], [2]Erasmus University, Rotterdam, The Netherlands

ABSTRACT

The purpose of the present study was to describe the peer assessment experience from the viewpoint of the students themselves in a post-secondary educational institution which organizes its curriculum according to principles of problem-based learning. Here, assessing peers' learning within the team is a continuous (i.e., daily) activity. To that end, a questionnaire containing statements inquiring about their experiences with peer assessment was administered to 897 first-year students. Analyses of the data collected demonstrate that students were generally positive about the peer assessment process. Descriptive statistics showed that students agreed that peer assessment helped their peers in their learning (69% agreed); they found the peer assessment process a valuable learning experience (43% agreed), and contended that they judged their peers in a manner which was honest and unbiased (55% agreed). The outcomes of a correlational analysis further suggest that those students, who agreed that the peer assessment enabled them to aid their peers in their learning, regarded the peer assessment exercise a valuable learning experience, and tend not to let interpersonal relationships, influenced their judgments of peers. Furthermore, students, who did not consider the peer assessment process as a valuable learning experience, also claimed to be less honest and biased when judging their peers.

Keywords: Peer assessment, students' views, student learning, problem-based learning

INTRODUCTION

Peer assessment requires students to provide feedback or grades (or both) to their peers on a product, process, or performance, based on the criteria of excellence for that product or event which students may or may not be involved concomitantly in determining the criteria (Falchikov, 2005). Implicit in the design of peer assessment is the assumption that students will be accurate and fair when assessing their peers. This, it is claimed, encourages students to develop high levels of responsibilities and a sense of ownership for their peers' learning (Dochy, Segers & Sluijsmans, 1999). Nicol and Milligan (2006) further substantiate that as peer assessment engages students in the process of reflecting on and evaluating the performance of others, they can develop objectivity in relation to standards which can be then be transferred to their own work.

Beneficial effects of peer assessment on student learning have repeatedly been reported in the literature. For instance, Ballantyne, Hughes, and Mylonas (2002) reported that peer assessment enhanced the metacognition of learners and improved understanding of subject matter. Peer assessment was regarded as an awareness raising exercise which enabled students to consider their own work more closely, highlighted what they needed to know in the subject, helped them make a realistic assessment of their own abilities, and provided them with skills that would be valuable in the future. Furthermore, the peer assessment process also encouraged students to compare and reflect on their own work, which in turn is an important element of self-directed learning. More recently, Papinczak, Young and Groves (2007) conducted a qualitative study of first-year medical students' attitudes to, and perceptions of, peer assessment. Their study revealed that students felt a sense of increased responsibility towards their peers' learning. Students also reported that feedback from peers also assisted

them in identifying deficiencies in their understanding and skills that were not readily apparent, thereby enabling them to take steps to further improve (see also Falchikov, 2005; Nicol & Milligan, 2006).

Although the studies mentioned thus far seem to suggest support for peer assessment, there are, however, several problems and limitations that have repeatedly been associated with the process of assessing others. For example, evaluation of student progress in a student-centred curriculum like problem-based learning (PBL), however, has remained a challenge (Eva, Cunnington, Reiter, Keane & Norman, 2004). This is because teachers often develop assessment procedures that test content knowledge (e.g. end-of-course examinations), rather than on areas like problem solving, and skills as a group member. Eva et al. (2004) suggest that the assessment of student skills, processes, and attitudes in PBL-based schools will take place most appropriately within the tutorial setting. They go on to substantiate that since there are many opportunities to assess areas of student proficiency such as communication skills, teamwork, and respect for others (which are not readily evaluated by other forms of content knowledge-based tests) in the tutorial setting, PBL educators are strongly encouraged to adopt tutorial-based peer assessment in their classrooms.

Another ramification arising from the findings of research on peer assessment is that students often lack confidence in their own and their peers' abilities as assessors. Students frequently report feeling "uncomfortable" in carrying out peer assessment, often because they feel unqualified to make these judgments (Orsmond, Merry & Reiling, 1997). In another study, Sluijsmans, Moerrkerke and MerrKnboer (2001) confirmed the existence of bias in peer marking due to interpersonal relationships between students. Furthermore, Papinczak et al. (2007) highlighted that students were overt in their skepticism towards the peer assessment exercise. Students' verbatim responses such as "not taken too seriously" and "not much thought goes into the marking" reflected their casual attitudes towards the peer assessment process (see also Lew & Schmidt, 2006).

Most of the studies found in the literature on peer assessment focuses on the evaluation of individual contributions to group assignments or the validity and reliability of peer assessment. Although student perceptions of the peer assessment experience have been studied extensively in higher education, few studies have been concerned with evaluating students' views in a PBL tutorial setting (Eva et al., 2004). To add on, many of the studies reportedly used peer assessment for summative purposes to judge the product of collaborative work (e.g., a poster or report), and is mainly administered towards the end of a predefined period for judging the quality of peers' works (Ballantyne et al., 2002; Falchikov, 2005). The specific purposes of the present study are two-fold: first, to evaluate students' experiences with peer assessment in a PBL tutorial setting and second, to document the findings as a basis for future deliberation on peer assessment and student learning in PBL.

METHOD

Subjects

Participants were 897 students in their first-year of studies at the polytechnic in the 2007-2008 academic year. This group consisted of 467 (52.0 %) female students and 430 (48.0%) male students. The mean age of the participants was 18.32 years (SD = 1.53). The Grade Point Average (GPA) is calculated based on students' classroom performance grades as awarded by their tutors, and their grades on knowledge acquisition tests. The GPA values

which range from "A" to "F" were first converted to scaled numerical values on a 4-point scale. The mean GPA value of the participants at the end of the first semester of the academic year was 2.79 (*SD* = .56). Participants were representative of the entire cohort of 3588 first-year students: 1843 (51 %) were females and 1745 (49%) were males, mean age was 18.23 years (*SD* = 1.44), and mean GPA was 2.82 (*SD* = .47) (*z*-values less than 1.96 at the .05 significant level).

Educational Context

One-Day, One-Problem™ PBL. The research was carried out at a polytechnic that organizes its curriculum according to principles of PBL. Students work collaboratively in teams of 4-5, with learning centred on problems relevant to their domain of study. They work each day on one problem. The problem is initially discussed in the morning, followed by ample study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place nor is there any form of direct instruction (Alwis, 2007). One tutor supervises the student teams in a larger classroom. His or her role is to facilitate student learning. All students enrolled attend a different module every day guided by a different tutor during a five-day work week, and take five modules per semester. There are two semesters per academic year, with each semester lasting sixteen weeks. All the courses offered are three-year courses.

Assessment in the curriculum. The daily assessment approach consists of a reflection journal to be written by each student, a self-assessment exercise, a peer assessment exercise, and a judgment by the tutor of how well a student has performed during a day.

The reflection journal consists of a short essay created by the student, that documents his/her personal reflections to learning and development. The self-assessment consists of 8 items inquiring about the quality of students' performance within their teams. The peer assessment consists of 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. Students are asked to respond to these items on a five-point Likert scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree".

The tutor judgment is an opinion formed after deliberating on and considering students' learning *process*, and their responses on the reflection journal, self- and peer assessment exercises. It consists mainly of tutors' *casual* observations of students' performance in the classroom during a day. Tutors also provide written feedback to students on various aspects of their learning on a daily basis.

Students also have to take four knowledge acquisition tests per module, at timed intervals during a semester. Each test, consisting of at least three structured questions and 30 minutes in duration, and are conducted in a supervised manner akin to that of end-of-course examinations.

Instrument

A questionnaire, which had been validated, and containing eleven statements inquiring about students' experiences with the peer assessment exercise was administered to all participants. The questionnaire contains three underlying factors: Peer learning, Objectivity and Task worth. The factor descriptors are contained in Table 1.

Students were asked to respond to these items on a five-point Likert scale ranging from "strongly disagree," via "disagree," "neutral" and "agree" to "strongly agree."

Table 1. Descriptive information for questionnaire factors

Factor	Description
Peer learning	The extent to which students believe that peer assessment aids peers in their learning.
Objectivity	The extent to which students let interpersonal relationships with peers affect their assessments of peers.
Task worth	The extent to which peer assessment is regarded by students as a valuable learning experience.

Procedure

The questionnaire was administered online to the participants in the tenth week of the first semester in July 2007. Instructions for the questionnaire stated that there were no right or wrong answers to the items and that all answers were correct so long as they reflected students' opinions. No information was given regarding the factors underlying the questionnaire. Filling in the questionnaire took approximately a minute.

Analysis

Data collected was analysed by means of confirmatory factor analysis using SPSS AMOS™ software. Parameters for the model specified were generated using maximum likelihood. For evaluating the proposed questionnaire model, fit indices used in this study include: χ^2, accompanied by degrees of freedom (*df*), sample size and *p*-value; (2) the root mean square error of approximation (RMSEA); and (3) the comparative fit index (CFI).

Chi-square (χ^2) is used to test the goodness-of-fit between an observed and predicted covariance matrix. A χ^2/df value (i.e. ≤ 5) is an indication of good fit, although some researchers regard a smaller χ^2/df value of 3 indicative of a good model fit (Byrne, 2001). RMSEA appears to be sensitive to model specification, minimally influenced by sample size, and not overly influenced by estimation method and was therefore included. The lower the value of RMSEA, the better the fit, with a cut-off close to .06 (Hu & Bentler, 1999). CFI ranges from 0 to 1, with higher values indicating a better fit. Values greater than .90 are traditionally associated with well-fitting models, although more recently, values close to .95 are suggested (Byrne, 2001).

Coefficient *H* values for the factors were also computed to determine construct reliability. Unlike other measures of construct reliability (such as Cronbach alpha), coefficient *H* is unaffected by the sign of the standardised factor loadings, and a value of at least .80 is considered reliable for a construct. This is because the higher the values of coefficient *H* would mean stronger and more stable factor loadings which tend to fluctuate less from sample to sample (Hancock & Mueller, 2001).

Descriptive statistical analysis was performed in examining students' mean responses on the questionnaire scales. Correlational analysis was carried out to explore if any inter-relationships exist between the factors.

RESULTS AND DISCUSSION

The factor structure of the questionnaire model is contained in Appendix A. Analysis of the questionnaire model with four factors resulted in a CFI value of .94 and RMSEA value of .06. Results of the χ^2 analysis yields a χ^2/df value of 2.93, $p < .00$. These values suggest a reasonably good model fit (Byrne, 2001; Hu & Bentler, 1999). All the standardized factor loadings are significant at the .01 probability level, indicating that these items contribute significantly to their respective factors. The values of coefficient *H* for the constructs underlying the peer assessment questionnaire range from .86 to .88, reflecting good construct reliability (Hancock & Mueller, 2001). By and large, the fit indices and construct reliability values obtained are acceptable for further statistical analysis of the data collected.

In order to investigate if the mean response on each factor was significantly different from the Likert-scale score for "neutral" (i.e. a value of 3.00), one-sampled *t*-tests were performed on these scores and the results contained in Table 2.

Table 2. Factor descriptive statistics and one-sampled t-values

Factor	Mean	Standard deviation (*SD*)	*t*-value	Percentage of students who agreed (%)
Peer learning	4.09	0.59	55.34**	69
Objectivity	2.73	0.76	-10.49**	45
Task worth	3.26	0.77	9.89**	43

Note. $^{**}p < .01$ (2-tailed)

The results indicate that students agreed that the peer assessment exercise enabled them to help their peers improve on their learning, and the exercise was a valuable learning experience. By contrast, more than half of the students disagreed that their peer assessment process was corrupted by bias due to friendship marking or lack of honesty.

The correlations among the three questionnaire factors are contained in Table 3. The absolute values of the statistically significant correlations range from .36 to .42, indicating moderate inter-relationships among the factors. This implies that the factors could identify different domains of students' views about the peer assessment exercise.

Table 3. Correlations among questionnaire factors (N = 897)

Factor	Peer learning	Objectivity	Task Worth
Peer learning	--		
Objectivity	-.36**	--	
Task worth	.39**	-.42**	--

Note. $^{**}p < .01$(2-tailed)

The correlations suggest that students, who think that the peer assessment enables them to aid their peers' learning, also consider the peer assessment exercise a valuable learning experience. Such an observation is not surprising, since students who regard the peer assessment as beneficial to their learning and that of their peers, tend to treat it seriously, and strive to be unbiased and fair when judging their peers. Conversely, those students who are overtly skeptical towards the peer assessment exercise in improving learning tend to adopt

casual attitudes towards it, and are less likely to regard the peer assessment process as a valuable learning experience.

GENERAL DISCUSSION

The purpose the present study was to describe students' experience with peer assessment in a PBL curriculum in which judging peers' learning is a continuous (i.e., daily) activity. To that end, the validated form of a questionnaire containing eleven statements in three underlying factors was administered to 897 first-year students to inquire about their experiences with judging their peers' performances.

The data collected fitted the questionnaire model reasonably well, as indicated by the fit indices values. All the standardized factor loadings were significant at the .01 probability level, suggesting that the items contributed significantly to their respective factors. Factor reliability values (Coefficient *H* values) of the three questionnaire scales gave evidence of good reliability of internal consistency. These findings when taken together, suggest that students were able to identify the three factors underlying the questionnaire.

Descriptive statistics of students' mean responses on the questionnaire factors suggest that students agree that peer assessment aids their peers' learning, and found the peer assessment process a valuable learning experience. By contrast, more than half of the students contended that they were objective in making unbiased and honest peer judgments. Moderate correlations between questionnaire factors imply that the scales could identify different domains of students' views about the peer assessment exercise. The findings from the correlational analysis suggest that students, who think that the peer assessment enables them to aid their peers' learning, also consider the peer assessment exercise a valuable learning experience, but believed that they were fair in judging their peers; they do not let interpersonal relationships with peers influence their judgments of them. Furthermore, students who regard the peer assessment process as an invaluable learning experience are likely to be less serious towards the peer assessment process, and tend not to give much thought when judging their peers.

The findings from this study provide valuable insights on how students perceive peer assessment in a PBL curriculum where they have to make judgments of their peers' learning process on a continuous (i.e. daily) basis. A limitation of the current study is that the questionnaire should be administered to second- and third-year students to see if any significant differences between students' experiences with peer assessment in the polytechnic exist.

Three issues present themselves for future research. First, the questionnaire model should be tested into other independent groups comprising of students in their second or third year of studies. This will help us in investigating if the items operate equivalently across different groups (e.g.: age, gender and years of experience) with the peer assessment exercise. Second, the measurement stability of students' beliefs over time should be investigated. Third, further research should examine the predictive validity of the questionnaire in relation to student academic performance. If peer assessment really contributes to improving learning, then its influence should be reflected in student achievement.

REFERENCES

Alwis, W. A. M. (2007). Pedagogical Beliefs and Institutional Practices at Republic Polytechnic. *Keynote presented at the 2nd International Symposium on PBL: Reinventing PBL, Singapore*.

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.

Byrne, B. M. (2001). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. London: Lawrence Erlbaum Associates.

Dochy, F., Segers, M., & Sluijsmans, D. M. A. (1999). The use of self-, peer and co-assessment in higher education: A review *Studies in Educational Evaluation, 24*(3), 331-350.

Eva, K. W., Cunnington, J. P. W., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How Can I Know What I Don't Know? Poor Self Assessment in a Well-Defined Domain. *Advances in Health Sciences Education, 9*(3), 211-224.

Falchikov, N. (2005). *Improving Assessment Through Student Involvement: Practical Solutions For Aiding Learning in Higher and Further Education*: (London, Routledge).

Hancock, G. R., & Mueller, R. O. (2001). *Rethinking construct reliability within latent variable systems, in: R. Cudeck, S. du Toit, & D. Sörbom (Eds.), Structural Equation Modeling: Present and Future- A Festschrift in honor of Karl Jöreskog. Lincolnwood, IL: Scientific Software International, Inc.* .

Hu, L.-t., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Lew, M. D. N., & Schmidt, H. G. (2006). *Reflection upon learning between theory and practice: A focus-group study of tutors' and students' perceptions.* Erasmus University Rotterdam The Netherlands.

Nicol, D., & Milligan, C. (2006). Rethinking technology-supported assessment in terms of the seven principles of good feedback practice, in: Bryan, C. & Clegg, K. (Eds). In *Innovative Assessment in Higher Education*: (London, Routledge).

Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education, 22*(4), 357-369.

Papinczak, T., Young, L., & Groves, M. (2007). Peer assessment in problem-based learning: A qualitative study. *Advances in Health Sciences Education, 12*(2), 169-186.

Sluijsmans, D. M. A., Moerkerke, G., MerrKnboer, J. J. G. v., & Dochy, F. J. R. C. (2001). Peer Assessment in Problem based Learning. *Studies in Educational Evaluation, 27*(2), 153-173.

APPENDIX A

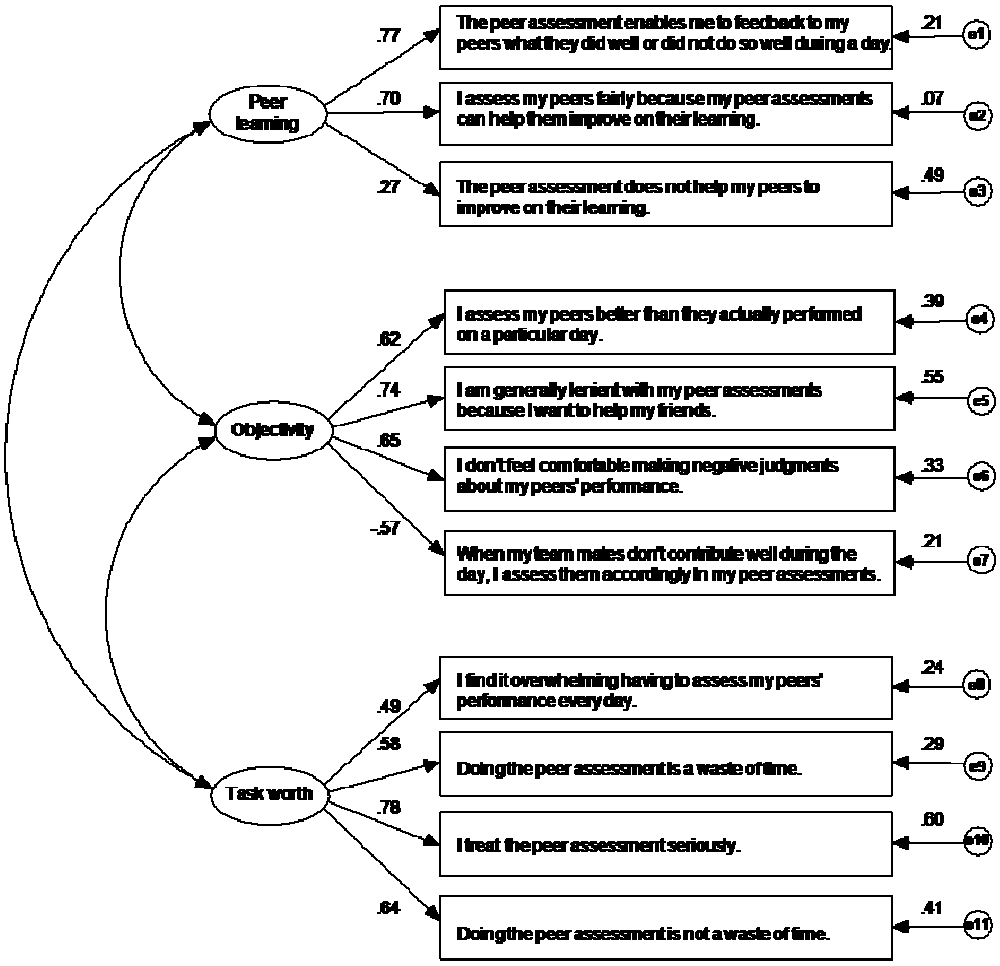

Figure 1: Factor structure of the peer assessment questionnaire model, showing standardized factor loadings

Acknowledgment

We gratefully acknowledge the participation of all students from Republic Polytechnic, Singapore who were involved in this research study. Special thanks go to Dr. W.A.M. Alwis, Deputy Principal (Academic Affairs) for his continuous support and encouragement throughout the course of this study.

Author contact details

Correspondence should be addressed to Magdeleine Lew, Office of Academic Affairs, Republic Polytechnic, Singapore, 9 Woodlands Avenue 9, Singapore 738964. Email: magdeleine_lew@rp.sg