

Students' perceptions of fairness in peer assessment: evidence from a problem-based learning course

Ana Carvalho*

Economics and Management School, Department of Management, University of Minho, Campus de Gualtar, Braga 4710-057, Portugal

(Received 20 December 2011; final version received 30 September 2012)

This paper describes the experience of peer assessment in a problem-based learning course in a Portuguese university, and presents the results of a survey of 120 students who attended the course over seven semesters. Cluster analysis revealed that students group into two distinct clusters with significantly different experiences of peer assessment. Although the larger cluster (70%) expressed a positive experience, the other group perceived peer assessment results and final marks to be unfair, and reported incidents of friendship-marking and conflict arising from peer assessment. Results show that perceptions of fairness (both regarding peer assessment and final marks) are very much associated with those problems. We reflect on these findings to explore implications for practice as well as future research.

Keywords: peer assessment; fairness; assessment biases; student perceptions; cluster analysis

Introduction

Peer assessment has featured increasingly in higher education accompanying the growth in active learning and student-centred approaches, including problem-based and cooperative learning (Dochy, Segers, and Sluijsmans 1999; Lew, Alwis, and Schmidt 2010; Van den Berg, Admiraal, and Pilot 2006).

In peer assessment, individuals rate their peers' performance, usually within a teamwork setting (Dochy et al. 1999). In addition to addressing a number of concerns regarding the evaluation of group assignments (King and Behnke 2005), peer assessment is particularly congruent with active and self-regulated learning, emphasising students' active involvement and the development of behavioural, cognitive and meta-cognitive skills, some of which are professionally relevant (Boud, Cohen, and Sampson 1999; Van den Berg et al. 2006). Peer assessment is especially pertinent in problem-based learning (Segers and Dochy 2001).

Despite the advantages of peer assessment, it also presents implementation problems (Vickerman 2009; Lew et al. 2010). Moreover, students' perceptions of their learning environment, including assessment modes, influence their learning strategies and outcomes, thus also affecting the effectiveness of peer assessment (Segers and Dochy 2001; Struyven, Dochy, and Janssens 2005).

*Email: anac@eeg.uminho.pt

In this paper we describe the peer assessment model designed for a problem-based learning course, and analyse the perceptions of seven student cohorts regarding peer assessment. The study has two main objectives: (1) to learn about how students experience peer assessment and what affects their perceptions and (2) to investigate whether the model used is being successful in attaining its goals and preventing some of the problems associated with peer assessment.

Informing literature

Accompanying the increase in cooperative learning and teamwork (Cheng and Warren 2000; King and Behnke 2005), the use of peer assessment has also grown (Van den Berg et al. 2006; Lew et al. 2010). Students often complain about the evaluation of teamwork when all team members get the same mark (Conway et al. 1993; Freeman and McKenzie 2002). A major concern for students and instructors is ‘free-riding’ (Brooks and Ammons 2003), when some students contribute less than their fair share but still get the group mark. As individual contributions are hard to determine by instructors (Maiden and Perry 2011), and team members can be better positioned to assess each other, peer assessment is an interesting alternative to individualising group marks. It also has a number of other advantages. Vickerman (2009) lists the main benefits reported in the literature, which include improved motivation, sense of autonomy and responsibility, developing lifelong learning skills and encouraging deep learning. Other authors add that peer assessment (and self-assessment) increases students’ maturity and confidence (Cheng and Warren 2000), allows them to regulate and monitor their own learning (Vu and Dall’Alba 2007; Lew et al. 2010) and, because it ‘requires students to closely scrutinize their peers’ work, guided by criteria of desired performance’ (Vu and Dall’Alba 2007, 542), promotes critical and reflexive thinking (Boud et al. 1999; Cheng and Warren 2000; Taras 2010).

Peer assessment has, however, some potential drawbacks. The main concern is with students’ ability to reliably assess themselves and their peers, either for lack of proficiency in what they are evaluating (Dochy et al. 1999; Lew et al. 2010) or for any number of subjective biases (King and Behnke 2005). Dochy et al. (1999, 340) describe assessment biases typical of peer assessment, including ‘friendship-marking’ (students overmark their friends); ‘collusive-marking’ (students pre-arrange undifferentiated marks); ‘decibel-marking’ (dominant students get higher marks); and ‘parasite-marking’ (similar to free-riding).

Studies have presented mixed results about the impact of these biases, but the consensus is that students can make useful assessments (Orsmond 2004; Lew et al. 2010), especially if they are experienced and receive proper guidance (Orsmond 2004; Vickerman 2009; Vu and Dall’Alba 2007). Providing students with clear and specific criteria (Brooks and Ammons 2003; Miller 2003) and descriptive scales (Vickerman 2009) is also helpful.

Another concern is with the consequences on social relations among peers. Vu and Dall’Alba (2007, 542–543) mention that ‘peer assessment may cause friction among peers, including feelings of hurt or betrayal resulting from comments or unexpected marks’. For example, dominant students can impose their views on the group (King and Behnke 2005), and students may withhold honest reports if they fear disrupting their relationships with their colleagues (Dochy et al. 1999), which

might result in the subjective biases mentioned above. In such cases, combining peer assessment with self- and instructor-assessment (Dochy et al. 1999), as well as confidentiality (Lejk and Wyvill 2001; Freeman and McKenzie 2002; Maiden and Perry 2011), can be beneficial.

These concerns raise the question of peer assessment fairness itself. Dochy et al. (1999) identify fairness as one of the main lines of research in peer assessment. Still, most studies seem to concentrate on questions of reliability and validity as pre-requisites for fairness. Although various authors find fairness to be a major concern for students (Freeman and McKenzie 2002; Vu and Dall'Alba 2007; Maiden and Perry 2011), few explore their perceptions of fairness towards peer assessment and what might determine or influence them.

The 'Business Case' (BC)

The BC is a capstone course of the undergraduate programme in Management of the University of Minho (Portugal), which students take in their final semester. Following the Bologna restructuring of the programme, more emphasis was put on active learning, on the development of autonomous learning skills and on the opportunity for students to apply what they learn in real-life management problems. The BC was set up with an innovative design in the final semester of the programme as a problem-based course. Its goal is to have students, working in teams, apply the knowledge, tools and skills acquired throughout the programme to a given management problem (or 'Case'). This Case is a complex problem presented by a real company, of strategic reach and with implications on the various functional areas of management. Student teams work on sustained solutions and recommendations for the company to implement.

Students work in teams of 5–8 students, the composition of which is defined by the course coordinator based on a set of criteria (e.g. gender, language and computer skills, etc.) intended to ensure heterogeneity within teams and homogeneity across teams, as well as concurrent meeting times. The choice for instructor-selected rather than student-selected teams intends to avoid a number of problems associated with the latter (Oakley et al. 2004), including the assessment biases identified by Dochy et al. (1999). Teams work autonomously, being responsible for defining their own schedule and member roles, organising work and managing all communication and interaction with the company. They are, however, assigned a tutor (a member of faculty) with whom they meet weekly and whose job is to support and stimulate the team.

The evaluation in this course comprises three parts: a written report (40% of the global mark), an oral presentation (20%) and the evaluation of the team's work along the semester by the assigned tutor (40%). There are no individual evaluation components and all these marks are assigned to the team. A peer assessment method was therefore designed to individualise the marks.

Peer assessment in the BC

The literature highlights the importance of experience and training in assessment skills (Orsmond 2004; Van den Berg et al. 2006; Lew et al. 2010). When students attend the BC, they have had experience with peer assessment in three previous

courses, so no training is provided. Students are, however, given relevant literature, and tutors discuss the matter as required. Following Oakley et al. (2004), teams are encouraged to discuss their expectations as team members and instructed to draw out a Team Contract, establishing member roles, work procedures and rules to deal with decision-making, problems and conflicts.

In the BC each student is required to assess their own and their teammates' contribution on what can be termed 'process' (Miller 2003) criteria. The unpredictable and variable nature of the work in each Case would make 'content' criteria difficult to implement. A developmental purpose is added to the evaluative purpose of peer assessment (Brooks and Ammons 2003) by having three assessment moments (typically on weeks 5, 10 and 16). This seeks both to collect student input along the process, not just at the end, and to allow students to adjust their behaviour in response to the interim evaluations. The time-consuming computation difficulties for teaching staff are minimised by using an online open source application (iPeer) developed by the University of British Columbia (<http://www.elearning.ubc.ca/toolkit/ipeer>), as increasingly recommended (Freeman and McKenzie 2002).

As multiple and specific criteria (Brooks and Ammons 2003; Miller 2003) may have a higher formative value than global approaches, categorical criteria are used. In order to foster student participation in defining the criteria (Orsmond 2004), a designated set is proposed at the beginning of the semester, and students are invited to suggest alternatives. So far, the proposed criteria have always been accepted. They closely follow Brooks and Ammons' (2003):

- Attendance in meetings – effort in scheduling and attending team meetings
- Ideas for the work – input with useful ideas for progressing the work
- Task execution – participation in the execution of relevant tasks
- Submissions – contribution for the timely submission of reports and assignments
- Interpersonal relations – contribution to a good team environment and dealing with conflicts
- Team motivation – personal motivation and contribution to team motivation
- Team rules compliance – compliance with the Team Contract

The scale students are required to use follows Goldfinch (1994): students assess the contribution of each team member relative to the team average. The assessment is confidential (Lejk and Wyvill 2001) to obtain more honest and unconstrained reports. The self-assessment element is preserved, despite Lejk and Wyvill's (2001) advice, due to its inherent advantages (Lew et al. 2010; Taras 2010). The scale used is:

- (1.00) – above-average contribution (did more than other team members)
- (0.75) – average contribution (did as much as other team members)
- (0.5) – below average contribution (did less than other team members)
- (0.25) – no contribution (did not contribute)
- (0.00) – was an impediment to the teamwork

The computation of the final coefficient is similar to that proposed by Conway et al. (1993). It measures the distance of each member's rating to the mean of the team's ratings, with the advantage of being simple to compute and understood by students as it directly translates into a percentage. In detail, the sum of all team member

assessments for a given student is divided by the team's average. The peer assessment coefficient is the simple average of the three assessments gathered along the semester, which is multiplied by the team mark attributed to the written report component. Peer assessment thus affects 40% of the team mark to individualise marks. This restriction is intentional, and stems mainly from the lack of previous experience with courses where all evaluation is based on teamwork. Mathematically, this final coefficient can range from 0% (if all teammates rate a member as an impediment) to 500% or 800% (depending on team size¹), if a student consistently contributes more than their teammates. Fearing the unknown results of applying peer assessment to the whole mark, and believing that there should be some consistency regarding the marks of the students in the same team, tutors and course coordinator opted for this conservative stand.

The study

Over the seven semesters occurring between 2008 and 2011, a total of 271 students (organised into 40 teams) and 17 tutors participated in the course. Six companies, from both the manufacturing (textile and electronics) and service sectors (software development and retail), have presented their management problems.

The data collected on peer assessment and its impact on team and individual marks are analysed for the seven student cohorts. In addition, an online survey was designed (using the Qualtrics survey software – <http://www.qualtrics.com>) to assess students' perceptions on various aspects of this course. Apart from biographical information, students indicated their level of agreement with a number of statements regarding peer assessment, teamwork practices and skill development. A four-point scale (4 = 'Strongly agree', 1 = 'Strongly disagree') was used expressly to avoid a neutral middle point.

An invitation to take the survey was sent by email on September 2011 to the 271 students, and 120 responses were returned, corresponding to a response rate of 44.28%. Considering this is a final semester course, and some students had been away from university for three or four years, we were pleased with this response rate. A valid concern was that the earlier cohorts would be under-represented. Although the last cohort was clearly over-represented, the earlier ones were not under-portrayed. Table 1 shows relevant summary data for the seven cohorts.

Survey data allow us to address the two objectives stated earlier. In order to examine how students experience peer assessment, we consider: (1) whether they value peer assessment and perceive the specific model used to be fair and adequate; (2) if students associate the experience with learning and skills development; and (3) what teamwork practices are associated with positive assessment experiences. In addition, to gauge the success of the model used in attaining its goals and avoiding potential pitfalls, we analyse: (4) student acceptance of the peer assessment model and perceptions of fairness; (5) the incidence of problems such as difficulties in assessing, friendship-marking and team conflict; (6) teamwork practices associated with those problems.

The data collected were analysed using IBM's SPSS Statistics software package, version 19. Apart from descriptive statistics and frequency analysis, correlations among variables were examined, and a cluster analysis was conducted. Because most data collected in the survey are categorical or ranked-ordinal in nature, the

Table 1. Summary of relevant data.

Cohort	No. of students attending	No. of survey responses	Response rate	No. of teams	No. of tutors
2007/2008 – 2nd sem.	46	14	30.43%	6	5
2008/2009 – 1st sem.	29	14	48.28%	4	4
2008/2009 – 2nd sem.	35	15	42.86%	6	4
2009/2010 – 1st sem.	17	5	23.53%	3	3
2009/2010 – 2nd sem.	70	29	40.00%	9	5
2010/2011 – 1st sem.	21	9	42.86%	3	3
2010/2011 – 2nd sem.	53	34	64.15%	9	8

Spearman rank correlation test was preferred, and the non-parametric Mann–Whitney test was used to test for statistical differences (Conover 1980).

Seven semesters of peer assessment

The data accumulated over these seven semesters are interesting in itself, allowing an insight into how students use the peer assessment instrument and how that impacts final marks. Most teams use peer assessment to distinguish individual performance, with only 5 teams over the 7 cohorts choosing not to. On these instances, the final averaged peer assessment coefficient was 100% for every member of the team, and they all had the same individual mark. This tendency towards equal marks is more typical of the first peer assessment moment, when 15 teams rate themselves equally. As the work progresses egalitarianism fades, and 35 out of 40 teams (87.5%) differentiate final peer assessment marks. Although in theory peer assessment coefficients can range from 0 to 800%, in practice they have varied from 84.3% to 114.5%. This means students differentiate team member performance by no more than 30.2%. Applied to the written report mark only, this has translated into individual final mark variations of at most 2 points (out of 20) among members of the same team (for example, one member getting 14/20 while others get 16/20). Final mark differentiation occurred 13 times, out of 40 possible, corresponding to a 32.5% occurrence rate. This disparity between peer assessment and final mark variation happens because the peer assessment coefficient only affects 40% of the course evaluation.

Survey results

Of the 120 students responding to the survey, 68.3% are female, the majority (72.50%) is between 21 and 25 years old; 5% are 20 or less; 5.85% are over 30. During the semester they attended the course, 20% of the students had a job.

Table 2 presents the descriptive statistics for the statements concerning peer assessment, including the frequencies for each statement. These show how strongly and consistently students agree with the need to evaluate individual contribution and differentiate final marks accordingly. Most students also agree with the overall evaluation system used. However, some dissent occurs relative to the weights of each evaluation component, and regarding the specific peer assessment model used, with about a third of students (34.5%) disapproving.

Opinions are also divided in terms of the peer assessment exercise itself, and its impact on final marks. More students disagree with the results of peer assessment in their team, but more students strongly agree with those results. Similar responses were given relative to the fairness of final marks. This can be partly explained by the occurrence of problems and difficulties alluded to in the literature. Indeed, over a quarter (27%) of the students report having difficulty in assessing their peers, and friendship-marking affected nearly 30% of the students. Slightly more (33.60%) complained of conflicts arising from peer assessment.

Table 2. Descriptive statistics for perceptions about peer assessment.

Variables	<i>M</i>	<i>SD</i>	Response frequencies (%)			
			Strongly agree (4)	Agree (3)	Disagree (2)	Strongly disagree (1)
1. I think final marks should be different for team members with different contributions	3.54	0.55	56.76	40.54	2.70	0.00
2. I think it's important that each team member's contribution be evaluated	3.52	0.57	55.86	40.54	3.60	0.00
3. The three evaluation components (Semester Work + Report + Presentation) are appropriate	3.12	0.61	23.42	66.67	8.11	1.80
4. The relative weight attributed to each component (40% + 40% + 20%) is appropriate	2.95	.71	19.09	60.91	16.36	3.64
5. The peer assessment model used in this course is adequate	2.71	0.83	14.55	50.91	25.45	9.09
6. My team's final mark was fair relative to the work we presented	2.68	0.93	19.27	42.20	25.69	12.84
7. The peer assessment in my team was adjusted relative to what happened throughout the semester	2.63	0.92	17.27	40.91	29.09	12.73
8. Peer assessment led to conflict in my team	2.21	0.89	9.09	24.55	44.55	21.82
9. In my team, peer assessment was more reflective of friendships than actual member contribution	2.20	0.98	15.45	13.64	46.36	24.55
10. It was very hard for me to assess each team member's contribution	2.14	0.75	4.50	22.52	55.86	17.12

Although most students seem to have had a positive experience with peer assessment, it is clear that some of the potential pitfalls did occur to a significant proportion. In order to understand this better, we analyse how these perceptions are correlated.

Correlation analysis

Table 3 presents the correlation matrix for the peer assessment statements. Perceptions concerning the acceptance of peer assessment (statements 1 and 2) and the adequacy of the specific model used (statements 3–5) are closely associated. In particular, the idea that final marks should be differentiated and individuals' contribution should be evaluated is substantially correlated ($r = 0.57$; $p \leq 0.001$), and agreement with the three components evaluated and their respective weights are also strongly correlated ($r = 0.69$; $p \leq 0.001$). Perceptions of fairness (statements 6 and 7) are very much associated with each other ($r = .51$; $p \leq 0.001$), and also positively correlate with overall acceptance of peer assessment and the evaluation model. On the other hand, perceptions of fairness are negatively associated with the occurrence of problems and difficulties (statements 8–10). Specifically, agreeing with peer-assessment results is substantially and positively correlated with accepting the three evaluation components ($r = 0.42$; $p \leq 0.001$) and considering the peer assessment model adequate ($r = 0.63$; $p \leq 0.001$); but it is inversely correlated with perceptions of conflict in the team ($r = -0.46$; $p \leq 0.001$) and friendship-marking ($r = -0.44$; $p \leq 0.001$). The latter also negatively and strongly correlates with agreeing with the final mark ($r = -0.63$; $p \leq 0.001$), which is also moderately associated with finding it difficult to assess one's teammates ($r = -0.33$; $p \leq 0.001$). Problems and difficulties with peer assessment are very much correlated amongst themselves, with a substantial association between peer assessment related conflict and friendship-marking ($r = 0.53$; $p \leq 0.001$), and moderate correlations between these and difficulties in assessing teammates ($r = 0.25$; $p \leq 0.01$ and $r = .39$; $p \leq 0.001$, respectively).

Cluster analysis

Correlation analysis revealed students' perceptions of fairness to be negatively and substantially associated with peer assessment problems and difficulties. Frequency analysis indicated these problems to affect a portion of students. In order to examine whether this is typical of a sub-group of students, a hierarchical cluster analysis was performed, using Ward's minimum variance method (Burns and Burns 2008). This indicated two distinct groups of students. Table 4 displays the descriptive statistics for the peer assessment perceptions in each cluster, indicating statistically significant differences between the two.

The two clusters significantly differ precisely in relation to perceptions of fairness and occurrence of peer assessment problems, although not in perceptions about the need for peer assessment. Cluster 1, which comprises roughly 30% of the students, is characterised by perceiving both peer assessment results and final marks to be unfair, considering the peer assessment model and the evaluation components less adequate, and experiencing friendship-marking and conflict arising from peer assessment. In Cluster 2, which includes the majority of the students, perceptions of peer assessment conflicts and friendship-marking are lower, while perceptions of fairness are stronger,

Table 3. Correlation matrix for perceptions about peer assessment.

Variables	1	2	3	4	5	6	7	8	9	10
1. I think final marks should be different for team members...	1.00									
2. I think it's important that... contribution be evaluated	0.57***	1.00								
3. The three evaluation components... are appropriate.	0.29***	0.40***	1.00							
4. The relative weight... is appropriate	0.16	0.37***	0.69***	1.00						
5. The peer assessment model used in this course is adequate	0.22*	0.15	0.31***	0.26**	1.00					
6. My team's final mark was fair...	0.22*	0.24*	0.38***	0.35***	0.38***	1.00				
7. The peer assessment in my team was adjusted...	0.16	0.15	0.42***	0.24*	0.63***	0.51***	1.00			
8. Peer assessment led to conflict in my team	-0.05	0.00	-0.11	-0.16	-0.38***	-0.46***	-0.26**	1.00		
9. ... Peer assessment was more reflective of friendships...	-0.11	-0.12	-0.29**	-0.19	-0.49***	-0.44***	-0.63***	0.53***	1.00	
10. It was very hard for me to assess...	-0.13	-0.01	-0.28**	-0.12	-0.27**	-0.16	-0.33***	0.25**	0.39***	1.00

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$, two tailed.

Table 4. Statistics for perceptions about peer assessment by cluster membership.

Variable	Cluster 1 (<i>N</i> * = 31)		Cluster 2 (<i>N</i> * = 73)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1. I think final marks should be different for team members...	3.52	0.63	3.55	0.53
2. I think it's important that...contribution be evaluated	3.42	0.62	3.55	0.55
3. The three evaluation components...are appropriate ^b	2.84	0.73	3.12	0.63
4. The relative weight...is appropriate	2.71	0.86	3.05	0.64
5. The peer assessment model used in this course is adequate ^a	2.13	0.72	2.96	0.77
6. My team's final mark was fair... ^a	1.90	0.75	2.98	0.84
7. The peer assessment in my team was adjusted... ^a	1.61	0.56	3.05	0.64
8. Peer assessment led to conflict in my team ^a	2.77	0.80	1.90	0.77
9. ...Peer assessment was more reflective of friendships... ^a	3.35	0.80	1.70	0.54
10. It was very hard for me to assess...	2.29	0.74	2.01	0.72

^{a,b}Statistically significant difference between clusters (a: $p \leq 0.001$; b: $p \leq 0.01$).

*Missing values were deleted listwise.

as is the acceptance of the evaluation model itself. In short, in Cluster 1 students had a negative experience with peer assessment, while students in Cluster 2 came out with a more favourable impression.

Statistical differences between the two clusters were also tested relative to perceptions of teamwork practices, teamwork difficulties and skill development. Table 5 shows these results, indicating the two groups of students differ in terms of teamwork practices and difficulties, but not significantly in their perceptions of learning and skill development. Specifically, students in Cluster 1 seem to have valued less the importance of teamwork rules and the teamwork contract, had less easygoing relationships with their teammates, perceived more conflicts and misunderstanding, and found teamwork (in a large team) a great difficulty. They were also less inclined to find the time they had for this course sufficient and to appreciate the feeling that they were learning.

The cluster analysis corroborates the notion that students' perceptions of fairness in peer assessment are tightly linked to the quality of their experience with peer assessment and teamwork. Students who perceive their peer assessment results and final marks to be unfair also experience peer assessment problems and teamwork difficulties.

Discussion

The analysis of the accumulated data on the peer assessment exercises allows us to conclude that peer assessment is successful in differentiating team member performance, with 87.5% of the teams exercising it. Limiting the impact of peer assessment to 40% of the final mark, although intentional, might be overly constraining, as it restricts differentiation to only about 30% of the teams. Results show that students themselves limit differentiation to no more than 16% above or below the team average. This suggests making the peer assessment coefficient affect a greater proportion of the final mark to reflect more closely the performance differentials expressed by students should be considered.

Table 5. Statistics for perceptions about teamwork practices, difficulties and skill development by cluster membership.

Variable	Cluster 1 (N = 31)		Cluster 2 (N = 73)	
	M	SD	M	SD
The Team Contract that we signed helped to clear expectations ^c	2.52	0.83	2.85	0.63
Defining roles . . . was important for our work ^a	2.84	0.69	3.37	0.63
We got along with each other from the beginning ^b	2.80	0.83	3.27	0.61
One of the things I most appreciated was . . . working in a team ^a	2.55	0.72	3.22	0.63
It was hard to deal with . . . misunderstandings in the team ^a	2.48	0.68	1.99	0.69
It was hard to work in such a large team ^b	2.67	0.75	2.17	0.72
The time we had to work on the Case was enough ^c	2.26	0.77	2.60	0.68
One of the greatest difficulties I felt was working in a team ^b	2.03	0.84	1.56	0.63
The BC allowed me to apply the knowledge and skills acquired	2.61	1.05	2.78	0.67
I most appreciated feeling that I was learning ^c	2.61	0.76	2.96	0.75
The BC made me more confident in my abilities	2.47	0.82	2.76	0.76

^{a,b,c}Statistically significant differences between clusters (a: $p \leq 0.001$; b: $p \leq 0.01$; c: $p \leq 0.05$).

Recalling the objectives of the study, the survey results help us recount students' experience with peer assessment. On the whole, we can surmise that most students are in favour of differentiating individual marks according to contribution. They also seem to favour peer assessment to achieve that, and generally approve of the evaluation system used. The specific peer assessment model is less consensual, with nearly 35% of students disapproving of it. Perceptions of fairness regarding the peer assessment exercise and final marks were lower, with almost 42% of students dissatisfied with the former and nearly 39% with the latter. Correlation analysis showed both acceptance of the peer assessment model and perceptions of fairness to be negatively related to occurrences of friendship-marking and team conflict. Indeed, the cluster analysis performed revealed two distinct groups of students who experience peer assessment in very different ways. Around 30% of students had a negative experience with peer assessment in this course, encountering instances of team conflict and friendship-marking, while also perceiving their peer assessment results and final marks to be unfair. They also found the peer assessment model less adequate than their colleagues, although they did not differ in their opinions about the appropriateness of differentiating final marks and evaluating individual contribution. These results stress the importance of acknowledging and dealing with those problems, which disturb the students' learning experience and seem to affect their perceptions of fairness.

As for the learning value of peer assessment, this is less evident. Students in each of the two clusters do not statistically differ in their perceptions of learning, other than those with more positive peer assessment experiences being more appreciative of feeling that they were learning. The fact that the model chosen for this course concentrates on 'process' rather than 'content' assessment criteria (Miller 2003) might explain this weak connection, especially if we consider students might be more easily aware of learning subject content and cognitive skills than interpersonal and meta-cognitive skills.

Some teamwork practices appear associated with a positive peer assessment experience. In particular, recognising the importance of defining team roles and having a Team Contract, as well as getting along with teammates and appreciating teamwork, is typical of the cluster of students who experienced peer assessment favourably and perceived it to be fair. Conversely, experiencing teamwork difficulties and conflict is typical of the student cluster with a negative peer assessment experience and higher perceptions of unfairness.

Turning to the second objective of this study, namely the success of the model used and its potential to avoid some typical peer assessment problems, the results are quite revealing. On the one hand, there is an acceptable level of agreement with peer assessment and the specific model used, and the majority of students (58.18%) perceive its results to be fair. The cluster of students with a favourable outlook on their peer assessment experience includes 70% of the students. However, the proportion of students who are (very) dissatisfied with peer assessment results is high enough (41.82%) to raise questions as to the success of the model. As reported above, these perceptions of unfairness are closely related to the incidence of certain problems and difficulties which, although not pervasive, are enough to warrant concern. Friendship-marking was admitted by 29.09% of the students, 33.64% reported conflict associated with peer assessment and 27.03% of students (strongly) agreed it was hard to assess teammates. Statistically significant differences were also found between the two clusters of students in terms of their perceptions of teamwork practices and teamwork difficulties, suggesting students' experience at this level to also be of consequence for how they experience peer assessment and perceive fairness in its outcomes.

One association that has not often been identified in the literature is between difficulties with time-management and peer assessment. Students with less time available for this course, or with poorer time-management skills, are likely to have less time to reflect on their own performance and that of their colleagues', or might have fewer occasions in which to observe and discuss each members' achievements, compromising peer assessment. So far, teamwork, time-management and peer assessment are areas where students are directed to written material but are not the subject of specific training. Although these students have previous experience, this may not guarantee they have developed the necessary skills, corroborating the results of Lew et al. (2010).

Conclusion

In this paper we describe the experience with peer assessment of seven cohorts of university students attending a problem-based learning course. We also report on a survey of their perceptions regarding their experience with peer assessment in this course. The two main objectives in this study were (1) to learn about how students experience peer assessment and what affects their perceptions of it and (2) to investigate whether the peer assessment model is successful in attaining its goals and preventing some of the problems associated with peer assessment in the literature.

On the whole, students appreciate peer assessment as a means of differentiating performance, and most have an overall positive experience. However, there is an appreciable incidence of perceptions of unfairness, and of problems such as

friendship-marking and conflict associated with peer assessment, which negatively correlate with perceptions of fairness. Cluster analysis revealed two distinct groups of students that have very different experiences with peer assessment in this course, one clearly unpleasant, and the other more favourable. These groups differ not only in their perceptions about peer assessment but also in their experiences with teamwork. Although no causal inferences can be made from our results, the problems encountered in peer assessment are very much interrelated, and also coincide with specific patterns of teamwork practices and incidence of teamwork difficulties. In particular, team conflict and friendship-marking go hand in hand, and these very likely affect students' perceptions of fair results. For this group of nearly a third of the students involved in this course, the success of the peer assessment model is highly questionable.

Implications can be drawn for practice, and for this course in particular, that reinforce previous recommendations. Peer assessment, teamwork and interpersonal skills should be the object of specific training, as previous experience alone seems to be insufficient to adequately prepare students (Lew et al. 2010). As O'Donovan, Price, and Rust (2004) point out, explicit criteria alone do not guarantee full and meaningful understanding and uniform interpretation of those criteria. Practices such as briefing by tutors, peer discussion, peer-marking exercises and the provision of exemplars can help students articulate assessment standards and criteria (O'Donovan et al. 2004). Also, if students discuss their expectations of peer assessment and learn to deal with interpersonal differences and conflict, some of the problems experienced might be mitigated and perceptions of fairness may improve. Discussing peer assessment results in the weekly meetings with tutors would also provide opportunities for oral feedback, which seems to be more formative than marks or written feedback alone (Van den Berg et al. 2006). Specific training on peer assessment might also make students more aware of how the assessment exercise influences their learning processes. Furthermore, larger teams should be avoided, since they appear related to the problems identified. And given the evidence of these seven cohorts, the peer assessment coefficient should be allowed to have a stronger impact on final marks by affecting a larger proportion of the evaluation.

This study highlights the negative relationship between perceptions of fairness and some of the potential pitfalls of peer assessment, namely friendship-marking and team conflict. Its many other benefits notwithstanding, peer assessment is most often used specifically to overcome the challenge of fairly evaluating teamwork. If its results are not perceived by students to be fair, peer assessment is arguably failing one of its critical objectives. Moreover, perceptions of unfairness may also to negatively affect students' learning, thus curtailing other peer assessment potential benefits and the advantages of a problem-based learning environment itself. Students' perceptions of fairness, its antecedents and consequences deserve therefore closer attention. In particular, the role of problems such as those encountered in this course, and others, merits further exploration.

These complex relations can only be glimpsed at through statistical analysis. Future research of qualitative nature should be used to study the students' point of view. Methods such as interviews or focus-groups might permit a deeper understanding of the complex relationships in action.

Acknowledgement

The author is grateful to an anonymous referee for their very helpful comments and suggestions.

Note

1. The maximum coefficient a student can get is 100% times the number of team members (5–8).

References

- Boud, D., R. Cohen, and J. Sampson. 1999. "Peer Learning and Assessment." *Assessment & Evaluation in Higher Education* 24 (4): 413–426.
- Brooks, C. M., and J. L. Ammons. 2003. "Free Riding in Group Projects and the Effects of Timing, Frequency, and Specificity of Criteria in Peer assessments." *Journal of Education for Business* 78 (5): 268–272.
- Burns, R. B., and R. A. Burns. 2008. Cluster analysis. In *Business Research Methods and Statistics using SPSS (Chapter 23)* (accessed August 20, 2012) <http://www.uk.sagepub.com/burns/website/%20material/Chapter%2023%20-%20Cluster%20Analysis.pdf>.
- Cheng, W., and M. Warren. 2000. "Making a Difference: Using Peers to Assess Individual Students' Contributions to a Group Project." *Teaching in Higher Education* 5 (2): 243–255.
- Conover, W. J. 1980. *Practical Nonparametric Statistics*. 2nd ed. New York: John Wiley & Sons.
- Conway, R., D. Kember, A. Sivan, and M. Wu. 1993. "Peer assessment of an Individual's Contribution to a Group Project." *Assessment & Evaluation in Higher Education* 18 (1): 45–56.
- Dochy, F., M. Segers, and D. Sluijsmans. 1999. "The Use of Self-, Peer and Co-assessment in Higher Education: A Review." *Studies in Higher Education* 24 (3): 331–350.
- Freeman, M., and J. McKenzie. 2002. "SPARK, a Confidential Web-based Template for Self and Peer assessment of Student Teamwork: Benefits of Evaluating across Different Subjects." *British Journal of Educational Technology* 33 (5): 551–569.
- Goldfinch, J. 1994. "Further Developments in Peer assessment of Group Projects." *Assessment & Evaluation in Higher Education* 19 (1): 29–35.
- King, P. E., and R. R. Behnke. 2005. "Problems Associated with Evaluating Student Performance in Groups." *College Teaching* 53 (2): 57–61.
- Lejk, M., and M. Wyvill. 2001. "The Effect of the Inclusion of Self Assessment with Peer assessment of Contributions to a Group Project: A Quantitative Study of Secret and Agreed Assessments." *Assessment & Evaluation in Higher Education* 26 (6): 551–561.
- Lew, M. D. N., W. A. M. Alwis, and H. G. Schmidt. 2010. "Accuracy of students' Self-assessment and Their Beliefs about Its Utility." *Assessment & Evaluation in Higher Education* 35 (2): 135–156.
- Maiden, B., and B. Perry. 2011. "Dealing with Free-riders in Assessed Group Work: Results from a Study at a UK University." *Assessment & Evaluation in Higher Education* 36 (4): 451–464.
- Miller, P. J. 2003. "The Effect of Scoring Criteria Specificity on Peer and Self-assessment." *Assessment & Evaluation in Higher Education* 28 (4): 383–394.
- Oakley, B., R. Felder, R. Brent, and I. Elhajj. 2004. "Turning Student Groups into Effective Teams." *Journal of Student Centered Learning* 2 (1): 9–34.
- O'Donovan, B., M. Price, and C. Rust. 2004. "Know What I Mean? Enhancing Student Understanding of Assessment Standards and Criteria." *Teaching in Higher Education* 9 (3): 325–335.
- Orsmond, P. 2004. *Self and Peer assessment: Guidance on Practice in the Biosciences*. Leeds: Centre for Bioscience, The Higher Education Academy.
- Segers, M., and F. Dochy. 2001. "New Assessment Forms in Problem-based Learning: The Value-added of the Students' Perspective." *Studies in Higher Education* 26 (3): 327–343.

- Struyven, K., F. Dochy, and S. Janssens. 2005. "Students' Perceptions about Assessment in Higher Education: A Review." *Assessment & Evaluation in Higher Education* 30 (4): 331–347.
- Taras, M. 2010. "Student Self-assessment: Processes and Consequences." *Teaching in Higher Education* 15 (2): 199–209.
- Van den Berg, I., W. Admiraal, and A. Pilot. 2006. "Designing Student Peer assessment in Higher Education: Analysis of Written and Oral Peer Feedback." *Teaching in Higher Education* 11 (2): 135–147.
- Vickerman, P. 2009. "Student Perspectives on Formative Peer assessment: An Attempt to Deepen Learning." *Assessment & Evaluation in Higher Education* 34 (2): 221–230.
- Vu, T. T., and G. Dall'Alba. 2007. "Students' Experience of Peer assessment in a Professional Course." *Assessment & Evaluation in Higher Education* 32 (5): 541–556.