IJPCE
International Journal of Physics and Chemistry Education, 11(1), 7-11

www.ijpce.org
ISSN: 2589-8876

Peer and Self-assessment: A Mathematical Model to Improve Students' Accountability in Laboratory Stations Model

Ana Rita Mota*
Universidade do Porto, Portugal
ana.mota@fc.up.pt

João B. Lopes dos Santos
Universidade do Porto, Portugal
jlsantos@fc.up.pt

Paula Esperto
Escola Secundária Poeta Joaquim Serra, Portugal
paula.esperto@gmail.com

Isabel Coutinho
Escola Secundária Poeta Joaquim Serra, Portugal
i.coutinho@campus.fct.unl.pt

Abstract
There is an increasing interest in peer and self-assessment (P&SA) in science education due its importance in students' performance in collaborative work and in the development of their metacognitive skills. We developed and tested an algorithm to help teachers organize the P&SA process. Collected data relative to the 8th and 11th grades were analyzed to examine the viability of this algorithm and to determine the degree of agreement between a student's self-assessment and that of his/her peers. We found no significant differences between peer and self-assessment, which means that students have a fair perception of their contribution to the team. The data also suggest that the maturity of the students can be an important variable in the self-regulation learning and in metacognitive skills acquisition, since the correlation between self and peer assessment was more robust in the 11th grade.

Keywords
Collaborative learning, Metacognitive skills, Peer and self-assessment, Group assessment

◆Received 11 December 2016 ◆ Revised 16 March 2018 ◆ Accepted 20 November 2018

Introduction
As the emphasis in education is moving towards developing student-centered curricula, interest is growing in the use of self and peer assessment. In a collaborative learning scenario, assessment is a fundamental part of the learning process, since the manner in which students approach their learning is highly conditioned by the way they are assessed and by their social interactions (Michaelsen et al., 2008; November, 2012). In this sense, peer and self-assessment (P&SA) play an important role in achieving an active learning environment in a collaborative setting, providing a formative and summative data (Topping, 2009) to both instructor and student.

Regarding formative assessment, P&SA creates the opportunity to receive prompt feedback about students' performance since they can observe and learn how peers address the same problem during their work (Topping, 1998). In other words, they learn from the cognitive processes of their peers (Mazur, 1997). This fosters improvement in their performance, individual accountability, problem solving and team skills over time. Peer evaluation scores are also important in summative assessment, allowing teachers to ensure fairness in the grading process according to each member's contributions (Falchikov & Goldfinch, 2000; Topping., Smith, Swanson, & Elliot, 2000).

In addition, P&SA is helpful in gaging individual contributions to a group task and is essential to motivate each student's performance, promote professionalism, develop self-monitoring and self-regulation, and mitigate free riders in team activities (Conway, Kember, Sivan & Wu, 1993; Brooks & Ammons, 2003). Getting students to assess each other and themselves encourages them to take required to reflect on their knowledge, their strengths and their weaknesses, and compare them with their peers, they learn to target their learning accordingly and, in the process, develop important metacognitive skill.

Social team interactions, cognitive cohesion, team productivity, are also enhanced by P&SA, preparing students for the future. In the real world, individuals who do not pull their weight on teams eventually suffer consequences far worse than low grades (Kaufman, Felder & Fuller, 2000).

Despite these advantages, students often voice objections to P&SA. Many have a pessimistic idea of teamwork, probably based on their experience with previous groups that functioned poorly together (Brooks & Ammons, 2003). The free-rider problem, also known as social loafing, is the focus of many complaints voiced by students regarding unsatisfactory teamwork experiences. When one or more members of a group fail to do their fair share of the work or are unable discuss at an adequate level, they disconnect from the group. According to Freeman and McKenzie (Freeman & McKenzie, 2002), students consider group work assessment as inequitable if there is equal reward for unequal contributions. These negative experiences can lead to students

*Corresponding author
https://doi.org/10.12973/ijpce/16031

experiencing discontent and dissatisfaction with group work (Wilson, 2013). As a result, the positive outcomes that P&SA can bring, such as critical thinking, communication skills improvement, self-learning, increased understanding of subject matter, and respect for others, fail to be recognized by the students. On the plus side, students tend to appreciate the fairness of P&SA and the lack of instructor input in the P&SA process (Kaufman & Schunn, 2011).

which they see as a hindrance for a constructive peer feedback process Nevertheless, there seem to be a majority view that allowing students to grade offers logistical, pedagogical, metacognitive and affective advantages over teaching assessment (Sadler & Good, 2006).

The challenge for educators is to develop new strategies of self and peer assessment which promote student participation and satisfaction with group work, in other words, assessment schemes' that are recognized and accepted by students as ensuring fairness in group work assessment.

An examination of the literature indicates a big range of different methods to conduct peer evaluations and, therefore, some different findings, but it seems that learners who accepted this process believed that the quality of their work improved as a result of the peer assessment, in spite of a possible socio-emotional discomfort (Michaelsen et al., 2008; Topping et al., 2000). The correlation between students' assessment grades (self and peer) and teacher assessment has been inconclusive. For instance, Sadler and Good (Sadler and Good, 2006) found that when grading themselves, lower performing students tended to inflate their own low scores. However, many studies show a significant correlation between peer and self-assessment scores (Alias et al., 2015; Mehrdad, et al. 2012). On the other hand, some international studies found that a well-organized and systematic P&SA can provide reliability and validity of the assessment process (Falchikov, 2000; Sande & Llorent, 2014).

Some peer and self-assessment schemes have been reported in the literature (Chen and Lou, 2004), all wrapping an organized system where students rate their own and peers' contributions and the average ratings are consequently used to moderate team marks to reflect individual differences. The schemes differ from each other and emphasize different goals. However, in almost all of them a mathematical formula is used to calculate mark-weighting factors.

The purpose of this paper is to report on an action research approach to the development and evaluation of a self and peer assessment strategy, based on an algorithm, which was designed to promote student participation in a student-centered curriculum and enhance fairness in group work assessment. This grading system ensures individual student accountability to both the teacher and peers since it has peer evaluation component and a team performance component.

The Algorithm

Evaluation in group tasks require fair grading systems that not reward and encourage social loafing.Thus, is important to develop an individual score, after the group grade, focused on not unfairly penalizing hard-working students (who may have a lower grade because they were randomly assigned to a weak performing group) and not benefiting the weak students (who may be carried along by members with high conceptual and team skills).

There are several desirable requirements to guide the development of the algorithm:
(a) Student's assessment should combine with the teacher's assessment of the group work, in such a way that their grades are constrained to an interval around the group grade awarded by the teacher;
(b) The teacher controls the range of variation of the student's grades around the group grade;
(c) Students' assessment should not be allowed to inflate the grades of all members of the group; it should discriminate the different contributions of each member to the group work.

We developed a procedure to help teachers organize the P&SA process and their students with self and peer monitoring, respecting these principles. Each student is required to complete a form assessing his/her own contribution and those of their team members to the effectiveness of their team, by awarding a grade to each of the team members (including his/herself).

Each score ranges from - 1.0 (deficient) to + 1.0 (exceptional), when compared with the other team members. The sum of all relative contributions (self and peer) must be zero. Thus, each student can choose -1.0, -0.9, - 0.8, ... 0... + 0.7, 0.8, 0.9, 1.0. The highest score that each student can have is n (the number of elements of the group) and the minimum grade is - n. These scores are normalized by the total number of the group members and each student has a score between - 1.0 and + 1.0. Because each student is required to award grades with zero sum, the average of the scores of the members of a group is zero.

In accordance with requirement (b) above, the teacher decides the value β that determines the maximum variation with respect to the average group score (GS). In Portuguese middle schools, the score can range from 0 to 100%. In the high schools, the score ranges from 0 to 20 points. Thus, in the middle school, β = 20% means that each student's score can fluctuate ± 20% of group score. In the high schools, β = 2 means a fluctuation among ± 2 points.

The students' final grade (FG) is calculated as:

$$FG=GS+APA \times \beta \quad (1)$$

FG - Student's final Grade (0-20 or 0-100%)
GS - Group Score
APA- Average of the Peer + Self-Assessment
β - Factor determining the variation with respect to the average group score.

Since APA averages zero among the members of a group, the average of FG in a group is equal to GS, thus ensuring requirement (c) above.

Let's give an example: Peter, Paul and Mary are high school students and belong to the same group. **Table 1** shows the self/peer-assessment and the final grade of each student. The group score in the task, assessed by the instructor, was 16/20 and the teacher decided that the FG (each student final grade) could fluctuate ± 3 points.

As we can see in **Table 1**, the average of students' grades in the group equals the group score. (17.5+13.5+17.0)/3 =16.0. The algorithm is flexible enough to adapt to different school levels and different collaborative assignments. It accommodates instructor specification of the weight of P&SA in the student's final grade in the choice of the β parameter.

Table 1. Student's final grade calculation

			Evaluated		
Evaluator	GS	X	Peter	Paul	Mary
Peter			0	-1.0	1.0
Paul	16	3	1	- 0.5	- 0.5
Mary			0.5	-1.0	0.5
APA			0.50	- 0.83	0.33
FG			17.5	13.5	17.0

Methodology

This model was implemented in Portugal, in five middle and high school classes. Once a week, students performed a lab class working in teams, through the Lab Stations Model (Mota and Lopes dos Santos, 2013). The class was divided into four/five completely independent experiment stations, each with different apparatuses/materials. Students traveled from station to station and, in addition to conducting experimental activities, they answered conceptual and quantitative questions on a worksheet. All the stations lasted the same time to allow an orderly rotation.

As students work, the instructor goes round and monitors their work, looking at what is being discussed and having a quiet discussion if any group seems to be stuck. Each worksheet is assessed by the teacher and has a score (GS) common to all the team members. The students worked on the same team during 3 months. After this period, students completed a self/peer assessment and a reflection form. In addition to P&SA, the reflection form also included some questions related to the team performance:

- Q1: What went well in the team?
- Q2: What were the main problems of the team?
- Q3: Please select the best option that characterizes your team's performance during this term:

My team was amazing.
My team worked well.
My team worked satisfactorily.
My team didn't work satisfactorily.
My team was a disaster.

- Q4: In the next term, do you want to belong to the same team?

Ninety-eight students were involved in this study; 49 from 8th grade and 49 from 11th grade. All of the students belong to the same school, distributed in five classes, as we can see in **Table 2**. The instructor, without any specific criteria, randomly chose the team members. Each team usually had three or four members.

Table 2. Distribution of students (by grade and class)

		Class				
		A	B	D	E	Total
Level	8th		12	25	12	49
	11th	22	27			49
	Total					98

Results and Discussion

Team performance

In the reflection form, students had four questions about their team. Q1 and Q2 are open-ended answers and were analyzed only by the instructors to assess their work and the team's performance. In this paper we only discuss the multiple-choice question Q3 and the closed-ended question Q4.

Students showed satisfaction with their teams since 82% of the students in 8th grade and 86% of the students in 11th grade wanted to work again with their teammates in the next term (Q4). This opinion was corroborated with their answers about their team. In a scale from 1, meaning amazing, to 5, meaning a disaster, students had to select how they evaluate the effectiveness of their team. As we can see in **Figure 1**, students think that the teams worked satisfactorily/well. In fact, 9.2% of the students considered that their team was a success, 48.0% of the students considered that their team worked well and 35.7% worked satisfactorily. We see no statistical difference among grade and gender. The results were very close. Both groups seem to experience the same (good) feelings about their work through this model.

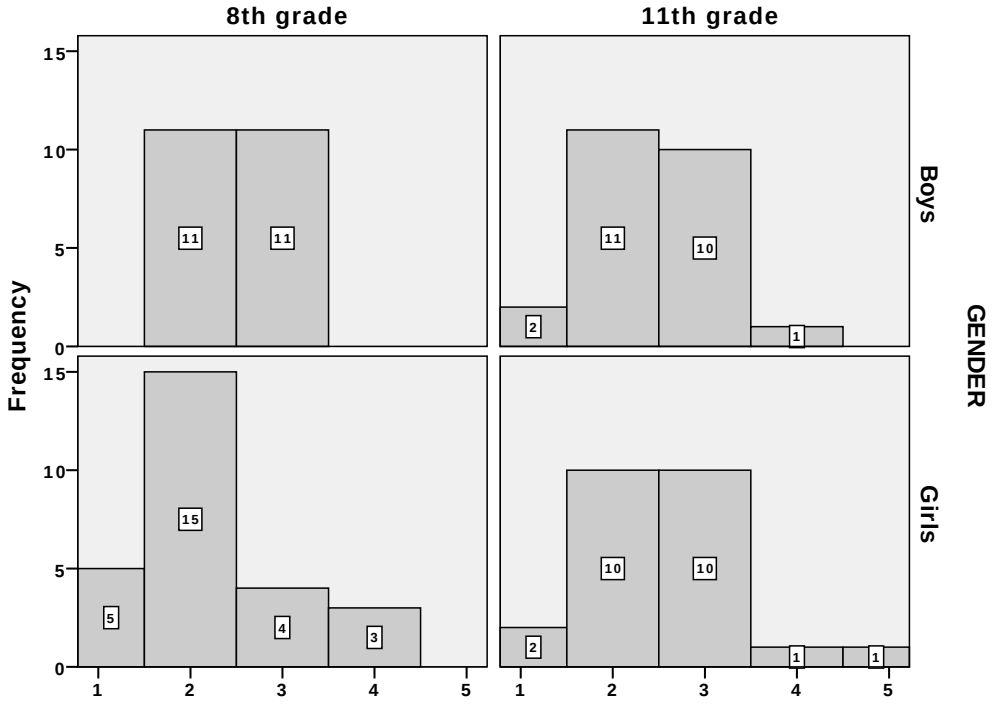

Figure 1: Team's effectiveness; (1) My team was amazing; (2) My team worked well; (3) My team worked satisfactorily; (4) My team didn't work satisfactorily; (5) My team was a disaster.

Self and peer Assessment mathematical model

Descriptive (mean and standard deviation) and inferential statistics (paired t-test and Pearson Correlation Coefficient) were used to analyze the data. To compare the results of evaluations (self and peer), and to show correlation of the scores, paired t-test, and Pearson Correlation Coefficient were used, respectively. SPSS software (version 17) was used for data analysis.

To investigate the possibility of a bias between a student's self score and the score of his/her teammates, we looked for a difference between the means of the self and peer assessment. Although the means of self-assessment are slightly higher than those of peer assessment the difference is not statistically significant as shown in **Table 3**. Non-parametric tests were also conducted (Wilcoxon Signed Rank-test) and the conclusions were quite similar (p-value = 0.7 in both levels).

Table 3. Paired t-test

Level	Self-assessment		Peer-assessment		*p*-value
	Mean	Standard Deviation	Mean	Standard Deviation	
8th	0.0224	0.13270	- 0.0163	0.19639	0.273
11th	0.0224	0.19177	0.0010	0.18888	0.479

In the 11th grade there appears to be a correlation between each student's self-assessment and that of his/her peers (see **Table 4**). In the 8th grade, however, the self-score that not appear to be correlated with the score given by the teammates. We thus find better agreement between self and peer perception of each student's role in the group with older students.

Table 4. Pearson correlation

Level	Pearson Correlation Coefficient	*p*-value
8th	0.093	0.525
11th	0.390*	0.006

*Correlation is significant at the 0.01 level

Conclusion

Accuracy of self-assessment is often under suspicion. After all, the student has a vested interest in having a positive evaluation and that is sufficient reason for a bias in his/her own assessment. This study, however, found no significant differences between self and peer assessment, in other words, no evidence of the bias alluded above. We can conclude, therefore that the algorithm presented, that takes into account both types of assessment, is a good measure of the contribution of each teammate.

Still, our results show that the maturity of the students can be an important variable in the self-regulation learning and in the metacognitive skills, since the accuracy of the students' assessment was better in the 11th grade. These results are in line with the literature (Alias et al., 2015; Mehrdad, et al. 2012) in spite of the controversial results in international studies about this topic. We believe that the (long) time that each team worked together (3 months) was important to assess fairly each team mate and know better the role of each one.

We will use the results of this study to expand this algorithm to other collaborative learning environments and design future studies with larger sample sizes. On the other hand, it will be interesting to analyze if the difference between P&SA will decrease during the year. The benefits of collaborative learning and P&SA are widely recognized in academic literature. Group work can be used to encourage deeper learning, promote meta-cognitive skills improvement and student autonomy by transferring some of the responsibility for teaching and learning to students. In this sense, we suggest that even if we do not employ the results of P&SA in the student's final score, in order to increase precision and responsibility in students in the group work and to

expand critical thinking, communication skills improvement, self-learning and respect to others, self and peer-evaluation must to be considered as systematic activities, from early ages.

References

Alias, M., Masek, A., & Salleh, H. (2015). Self, peer and teacher assessments in problem based learning: Are they in agreements? *Procedia - Social and Behavioral Sciences*, 204, 309-317.

Brooks, C. M., & Ammons, J. L. (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.

Chen, Y., & Lou, H. (2004). Students' perceptions of peer evaluation: an expectancy perspective. *Journal of Education for Business, 79*(5), 275–282.

Conway, R., Kember, D., Sivan, A., & Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment and Evaluation in Higher Education, 18*(10), 45–46.

Falchikov N., & Goldfinch, J. (2000). Student peer assessment in higher education. A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287–322.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education, 20,* 289–300.

Freeman, M., & McKenzie, J. (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*, 551-569.

Kaufman, D. B., Felder, R. M., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education, 89*, 133–140.

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

Liu, N.-F., and Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education, 11*, 279–290.

Mazur, E. (1997). *Peer Instruction: A user's manual.* Prentice-hall, Upper Saddle River, NJ.

Michaelsen, L. K., Sweet, M., & Parmelee, D. X. (2008). *Team-Based Learning: small group learning's next big step*. New York: Wiley.

Mehrdad, N., Bigdeli, S., & Ebrahimi, H. (2012). A comparative study on self, peer and teacher evaluation to evaluate clinical skills of nursing students. *Procedia - Social and Behavioral Sciences, 47*, 1847-1852.

Mota, A. R., & Lopes dos Santos, J. M. B. (2013). Reflecting understanding: Using lab stations to teach image formation. *Science Scope, 37*(2), 20–29.

November , A. (2012). Who Owns the Learning? Solution Tree, United Kingdom.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*, 1–31.

Sande, J., & Llorent, G. (2014). Peer Assessment and Self-assessment: Effective Learning Tools in Higher Education. *International Journal of Engineering Education, 30,* 711–721.

Topping, K. J. (1998). Peer assessment between students in college and university. *Review of Educational Research, 68*, 249–276.

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,* 149-170.

Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*, 20-27.

Wilson, J. T. (2013). *Peer learning in higher education*. London: Routledge.