CAN WE TRUST PEER GRADING IN ORAL PRESENTATIONS? TOWARDS OPTIMIZING A CRITICAL RESOURCE NOWADAYS: TEACHER'S TIME

Nuno Gracias, Rafael Garcia

Computer Vision and Robotics Group, Department of Computer Architecture, University of Girona (SPAIN)
ngracias@silver.udg.edu, rafael.garcia@udg.edu

Abstract

This paper discusses the issue of peer grading in oral presentations, applied to engineering students. It represents an initial effort towards a longer term objective, of rationalizing the teacher's time by reducing or eventually eliminating the need for teacher grading in this activity. The paper reviews previous literature in the topic, and reports preliminary results of an assessment of fifth year students of Computer Engineering.

Keywords: peer assessment, oral presentation.

1 INTRODUCTION

A new education law was recently approved by the Spanish government [1]. This law involves an increase in the lecturing hours for most of the faculty, unless they could prove to have a very strong research record. This new situation has forced Spanish universities to look for strategies that optimize the time of faculty members. According to the principles of the Bologna Process, one ECTS corresponds to 25 hours of student effort. A wide majority of Spanish universities have established that this corresponds to 10 contact hours, while some understand this should correspond to 10 hours of professor effort. Our aim is to explore alternatives so that teacher time is optimized, while the quality and level of service provided to the students is maintained or even improved. Of course, this is not an easy task, but we feel that diminishing teacher time can be beneficial for the students if the professor devotes his time to high value activities.

In this paper we address the problem of how to grade oral presentations of students while trying to we use the available professor's time on those activities that provide more value to the students. Grading is necessary to provide feedback to the student, but as the class size increases, the further we are from a one-to-one tutoring strategy, and grading involves a very large amount of time. So we explore whether peer assessment is reliable enough to reduce or even eliminate professor grading in evaluating oral presentation skills.

Regarding professor time as a valuable resource, we can grade students in an efficient way by using multiple choice exercises, or online forms, although these methodologies have other drawbacks, such as possible ambiguity in the student's interpretation of the question. Moreover, higher-order reasoning or even problem-solving skills are difficult to assess using multiple choices, although some faculty still attempt to evaluate these skills through multiple-choice tests. However, in subjects for which there are no single correct answers, objective grading is often controversial, generating dissatisfaction both on the student and on the professor's side. This discomfort makes multiple-choice exams more appealing than assessment mechanisms based on essay answers or even oral presentations. In this article, we attempt to use peer assessment to grade students regarding their oral competence. We present an experience in which a peer-grading is compared to professor-grading. This experience seeks optimizing teachers' time, although some authors [2] state that there could be no saving of time in the short term, since good quality peer assessment requires time for organization, training, and monitoring the students' activities.

If we try to analyse the validity of peer assessment, i.e., how similar the grading is between students assessment and expert (professor) assessment, most studies reported in the literature find reliability and validity adequate [3, 4, 5], although some find them variable [6, 7]. When specifically dealing with oral presentation skills, Magin & Helmore [8] compared grades awarded by peers and several professors attending the same individual presentations. They concluded that although teacher assessments were more reliable than peer assessments, single teacher assessment of oral

presentation skills is inadequate as a reliable assessment measure, provided the variance of grades awarded by the different teachers. In their study, they also conclude that combining teacher assessment scores (from several teachers) with those produced by averaging multiple peer grading increases the mark reliability when compared to just single teacher ratings or averaged peer assessments. Moreover, they state that by involving the audience of students in the task of assessment we are fostering skills of professional judgment and, at the same time, assessment of oral presentations is attained with greater reliability than just averaging teacher assessment scores. In general, all studies seem to agree on the fact that grading their peers becomes another learning activity for the students, and they can actually learn from the experience [9, 10, 3, 11].

It should be noted that some authors (e.g. Freeman [12]) state that unless student grades can reliably reproduce professor marks, then peer assessment should have a very low weight, if any, in a student's final grade. This is in contradiction with what we aim, which is using peer assessment as the standard or reference point for the final grade. Other authors, on the contrary, state that the reliability and validity of peer assessment tend to be at least as high, and often higher, than teacher assessments [10, 7].

2 METHODOLOGY

Participants were students in the 5th year of the Computer Engineering studies (last promotion of the old programme, to disappear next year in favour of the new degree on Computer Engineering following the Bologna process). They were taking a course on Computer Vision, in which one of the assignments was to work on a project in teams of 3 students. Students were given a video and were asked to propose and execute projects which involved the use of *invariant features*, i.e. algorithms to detect and describe local features in images. Within this project-based context, students were encouraged to search for information resources that would accompany their project ideas. Every student team was supervised by a tutoring instructor and periodic meetings were scheduled during the execution of the project to follow their progress. The assignment was designed to be carried out in six weeks and the evaluation was based on their 20-minute oral presentation on week 6. This assignment had a weight of 20% of the final grade of the course.

For the peer assessment trial, a rubric was designed (see Appendix A). This rubric was made available to students some weeks before the peer assessment was carried out, and was discussed in the class to make sure all students understood every aspect to be evaluated. Specifically, the rubric is requesting feedback in relation to the following aspects of the presentation:

- **Execution/ Structure**: Have the speakers correctly introduced the topic and the presentation follows a coherent structure?
- **Content of the message delivered**: Did they prove a complete understanding of the topic and deliver it efficiently? Did they present the information logically?
- **Non-verbal and gestural communication**: Did they maintain a natural and appropriate posture? Were they facing the audience and keeping eye contact to be able to capture the audience reaction?
- **Para-verbal Aspects**: Did the speakers speak loudly enough? Did they have clear diction?
- **Enthusiasm**: Are facial expression and body language generating strong interest on the topic?
- **Time**: have they adjusted to the allocated time?
- **Group**: Did all team members participate in the presentation? Did they choose adequately their roles to decide who explains what? Did they look coordinated?
- **Slides**: Have they designed a creative an adequate audio-visual support? Did they use visual aids effectively?
- **Proactivity during questions**: Did they handle the questions well?

Performance was rated against these criteria using marking one of the options in the rubric of Appendix A, thus based on a 4-point scale.

Inspired by Tseng & Tsai [13], who proposed some methods for computerizing peer assessment, the rubric was introduced in Google forms [14] to collect the data automatically, and both students and staff were requested to bring their own web devices to complete the form online once the presentation was over. A different form was prepared for every group, and although completing the same rubric,

professors and students were completing different forms to make data extraction easier. During the evaluation, every group had to present his own project, and all other students had to complete a rubric for that group.

The forms were completed by 14 students and 3 professors. Five groups were presenting their project, so every student completed 4 forms (they were not self-evaluating their own project). Once all the presentations were finished, the links to the raw assessment data were forwarded to every group, so that they obtain feedback from the evaluation process. This results in every group of students receiving 3 feedback forms from faculty and an average of about eleven anonymous assessments from their peers.

It should be noted that the form included a field in which the student had to fill in his student number (although this was only accessible by the faculty acting as coordinator of the activity). Moreover, the students knew that their grades will be averaged with those of the academic staff to obtain the final grade. It has been proved that knowing that peer assessments will "count" is likely to promote a greater seriousness and commitment from students [15]. Knowing that their assessment counts on the final grade and that it can be traced back may also promote their level of responsibility when answering. We should also note the level of maturity of the students taking part in this study as being quite high, since the students are in the last year before becoming engineers.

3 RESULTS

Table 1 shows the means and standard deviations of the global faculty and student assessments for every group. It also indicates the sample Pearson correlation coefficient, which was computed using the assessment means.

Table 1. Comparison of Faculty and Peer assessment: mean, standard deviation and correlation values.

	Professor Assessment		Peer Assessment		
Group number	Mean	s.d.	Mean	s.d.	Pearson r correlation
1	3.41	0.52	3.43	0.39	0.53
2	2.96	0.54	3.37	0.31	0.62
3	2.37	0.54	2.89	0.69	0.80
4	3.11	0.91	3.17	0.70	0.85
5	3.00	0.33	3.06	0.28	0.37
Total	2.97	0.67	3.19	0.53	0.73

From the table it can be seen that the standard deviations of the faculty assessments and the peer assessments are reasonably similar, which would indicate a similar level of internal agreement among faculty and among students. The peer averages are noticeably higher for groups 2 and 3 than those of the faculty, but the difference is still within one standard deviation. This fact could hint a bias towards higher student gradings, which could be in part explained by the student's lack of peer grading experience or personal affinity. The sample Pearson correlation coefficient varies considerably.

The means and standard deviations are illustrated in Fig. 1. These values were computed by averaging the professor and the student assessments for each item in the rubric, for each group. The mean and standard deviations were then extracted.

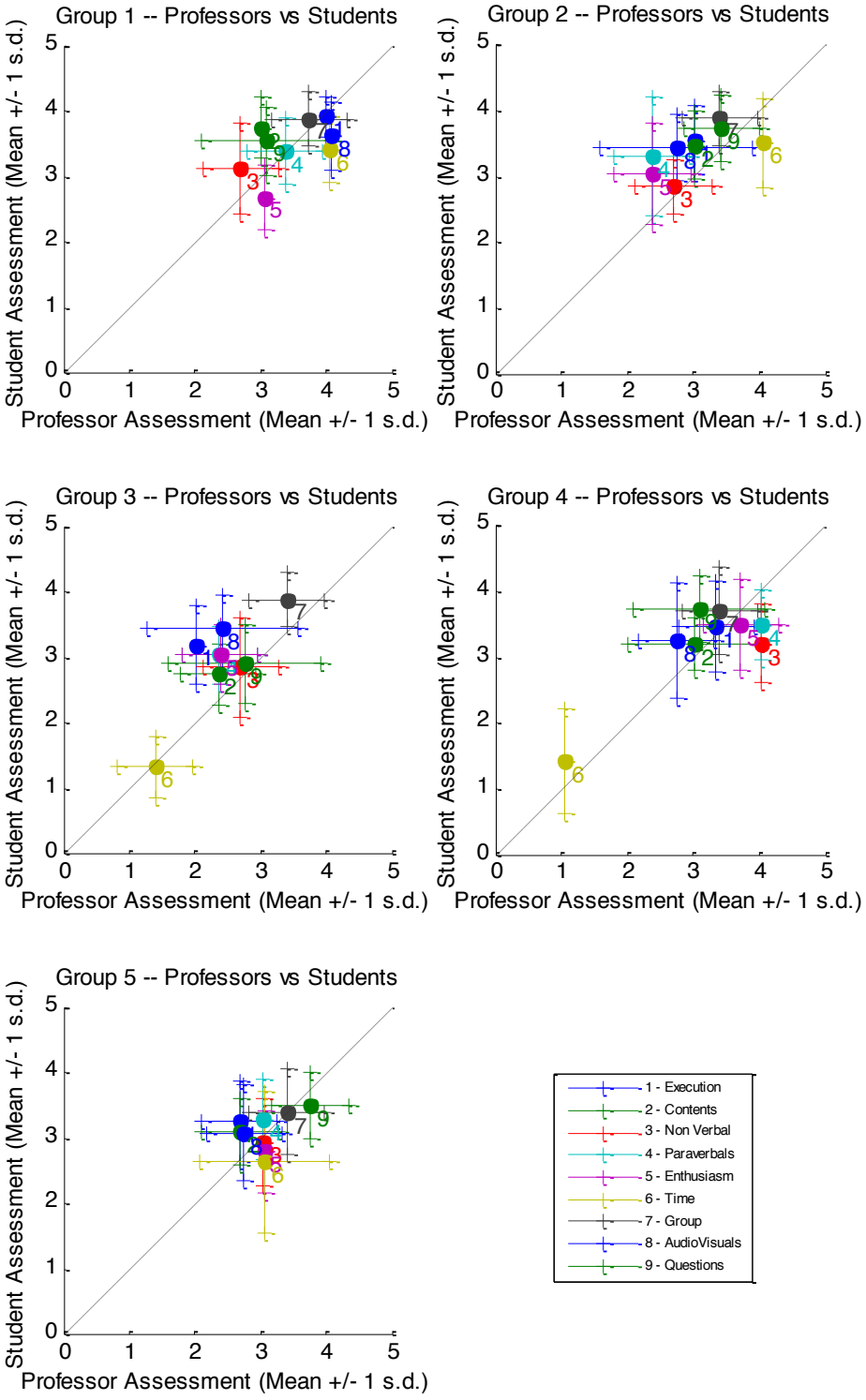

Figure 1 - Peer and faculty assessment per group.

Some authors have reported peer assessment to tend towards the middle values of the faculty assessment [6, 10]. Although being a small data set, our data is commensurate with that idea. An exception is the evaluation of the presentation time where the faculty assessment agreement is considerably higher (for 3 of the groups) than the peer assessment. As the presentation time is a very objective criterion to be evaluated, one may speculate that the variability of the peer assessment could be related with inadequate time logging by some of the students (for example, by forgetting to write down the starting and ending times of the presentations).The assessment for groups 2 and 3 shows

how most students are evaluating all items (except time) with higher grades than those awarded by the academic staff.

The students reported that having the rubric in advance helped them in deciding how to organise their presentations.

4 CONCLUSIONS

We have reported an experience that evaluates the reliability of peer and faculty summative assessments of oral presentation skills of engineering students in a fifth year subject. We have attempted peer assessment to see if we can reduce the teacher's time requirements, but results are not completely satisfactory since peer evaluation seems to be slightly higher than that awarded by the faculty.

We have shown that the value of student assessment will depend on the many variables affecting learning in a specific course. Using a very clear rubric, which eliminates ambiguity, may help in obtaining reliable scoring of oral presentations, especially if it is analytic and complemented with exemplars, but it is far from being perfect. Some of the items in which we expected to have a perfect correlation between student and professor grading (such as how well the group adjusted to the available time) show clear inconsistencies, which questions the validity of the method. Therefore, our experience shows that if students are provided with a comprehensive rubric, there is just a certain correlation between student assigned grades and professor assigned grades, although it may not be enough to trust peer assignment alone, for the final grade of the students. However, even if peer grading is not a perfect strategy for providing reproducible grades, the used rubrics make the criteria explicit, which presents the benefits of facilitating feedback and student self-assessment when preparing their presentations. Moreover, slight "overmarking" in peer assessments is evident in the assessment of some groups, probably due to friendship. On the other hand, peer assessment seems to be really useful to provide constructive feedback to the students in a very short time.

Finally, it should be noted that our experience has only been tried in a small class in the fifth year of Computer Engineering, so it is not statistically significant, and may have biased our results due to outliers. However, it already shows a trend that we plan to scale for larger groups in the future to reach more reliable conclusions.

REFERENCES

[1] Real Decreto-ley 14/2012, de 20 de abril, de medidas urgentes de racionalización del gasto público en el ámbito educativo. Link: http://www.boe.es/boe/dias/2012/04/21/pdfs/BOE-A-2012-5337.pdf

[2] Falchikov, N. (2001). Learning together: Peer tutoring in higher education. London: RoutledgeFalmer.

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

[4] Kwan, K. & Leung, R. (1996) Tutor versus peer group assessment of student performance in a simulation training exercise. Assessment and Evaluation in Higher Education, 21(3), pp. 205–214.

[5] Stefani, L. (1998) Assessment in partnership with learners, Assessment and Evaluation in Higher Education, 23, pp. 339–350.

[6] 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.

[7] Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68, pp. 249–276.

[8] Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? Studies In Higher Education, 26, pp. 287– 298.

[9] Ngar-Fun, L. & Carless, D. (2006). Peer feedback: the learning element of peer assessment. Teaching in Higher Education 11(3), pp.279–290.

[10] Topping, K.J. (2009). Peer Assessment. Theory Into Practice, 48, pp. 20–27.

[11] Gopinath, C. (1999). Alternatives to Instructor Assessment of Class Participation. Journal of Education for Business, 75 (1), pp. 12-14.

[12] Freeman, M. (1995). Peer assessment by groups of group work. Assessment and Evaluation in Higher Education, 20, pp. 289–299.

[13] Tseng, S. C., & Tsai, C. C. (2007). Online peer assessment and the role of the peer feedback: A study of high school computer course. Computers and Education, 49, 1161–1174.

[14] Google Forms. http://spreadsheets.google.com

[15] Lejk, M., Wyvill, M. & Farrow, S. (1999). Group assessment in systems analysis and design: a comparison of the performance of streamed and mixed-ability groups, Assessment and Evaluation in Higher Education, 24, pp. 5–14.

APPENDIX - A

ORAL PRESENTATION ASSESSMENT FORM

SUBJECT: Computer Vision

SKILLS: Create and deliver an appropriate and creative oral presentation, with appropriate audiovisual support, following a script and displaying command of nonverbal communication strategies.

MARK	*4*	*3*	*2*	*1*
1 - EXECUTION /STRUCTURE	*Introduced the subject well and have followed a coherent structure.*	*The presentation is good but without advanced details, or without mention on how the work was distributed.*	*The presentation has some execution problems, no details are presented and no information on work distribution.*	***The topic was not introduced and the structure of the presentation is not consistent.***
2 - CONTENTS OF THE DELIVERED MESSAGE	*Demonstrated a complete understanding of the subject, and summarized the main ideas at the end.*	*Clear effort in making the explanation understandable but some concepts were not clear.*	*Understanding of the explanation was irregular. Some parts were difficult to follow.*	***The explanation could not be understood at a basic level.***
3 - NON VERBAL AND GESTURAL COMMUNICATION	*Maintained a natural and appropriate posture. Kept contact with the audience to capture their reactions.*	*Tried to maintain a natural and appropriate posture, but failed occasionally to keep eye contact*	*Were mainly unconnected with the audience and were gesticulating too much. Frequently lost eye contact.*	***Had poor posture and / or did not engage in eye contact during the presentation.***
4 - PARAVERBAL COMPONENTS	*Appropriate articulation and vocal delivery. The tone, pace and volume were suitable.*	*Correct tone, pace and volume, but with a few articulation flaws.*	*Irregular tone, pace or volume. Monotonous speech delivery at times.*	***Inadequate tone, pace and volume. Poor articulation impacting the understanding of the words.***
5 - ENTHUSIASM	*Facial expressions and body language generated a strong interest and enthusiasm among the audience about the topic.*	*Facial expressions and body language sometimes generated a strong interest and enthusiasm about the topic.*	*Facial expressions and body language are used to try to generate enthusiasm, but seem to be too forced and simulated.*	***Very sparse use of facial expressions and body language. Did not generate much enthusiasm, in the way it was presented.***
6 - PRESENTATION TIME	*Presentation was precisely adjusted and delivered to the established time (15 minutes)*	*The presentation was not adjusted to the prescribed time by a small margin (+/-2 minutes).*	*The presentation departed from the set time (2 to 4 minutes)*	***Did not follow the set time (over 4 minutes).***
7 - GROUP	*All students participated in the presentation. Good choice of presenting roles and coordination*	*All participated in the presentation. Good choice of presenting roles but some lack of coordination*	*All students participated in the presentation. Inadequate choice of presenting roles.*	***Some students did not participate in the presentation***
8 - AUDIO-VISUAL RESOURCES	*Have designed creative and very appropriate audiovisual support materials, which were used to complement the verbal presentation.*	*Designed correct audiovisual support material and used it to complement the verbal presentation.*	*Designed basic and predictable audiovisual support material. The material was read instead of delivered from memory.*	***No audiovisual or graphic materials were used; or were used at wrong timing or in excess***
9 - PROACTIVITY IN ANSWERING QUESTIONS	*Replied with interest to all questions posed by the teacher.*	*Replied with interest to part of the questions posed by the teacher.*	*Replied with little engagement to the questions formulated by the teacher.*	***Did not answer the questions.***