

## SOME LIMITS IN PEER ASSESSMENT

**Joan Domingo<sup>1</sup>, Herminio Martínez<sup>2</sup>, Spartacus Gomariz<sup>2</sup>, Juan Gámiz<sup>1</sup>**<sup>1</sup>Department of ESAIL (Systems Engineering, Automation and Industry Informatics)<sup>2</sup>Department of ELL (Electronic Engineering)

Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB)

Universitat Politècnica de Catalunya (UPC)

[joan.doming@upc.edu](mailto:joan.doming@upc.edu), [herminio.martinez@upc.edu](mailto:herminio.martinez@upc.edu), [spartacus.gomariz@upc.edu](mailto:spartacus.gomariz@upc.edu), [juan.gamiz@upc.edu](mailto:juan.gamiz@upc.edu)*Received August 2013**Accepted January 2014***Abstract**

Nowadays, the educational methodology known as '*peer assessment*' constitutes one of the pillars of formative assessment at the different levels of the educational system, particularly at the University level. In fact, in recent years, it has been increasingly used to enhance students' meaningful learning, as it is considered to be an element of social learning, in which students benefit from the lessons learned by other classmates, and draw upon the ability to assess the quality of the learning, contrasting it with the level of knowledge that each has about the subject/course being evaluated, and using common evaluation criteria.

In this regard, this paper represents the experience of two groups of students. It allows us to determine how many peer assessments should be required of students in a particular course in order to constitute a serious, reliable activity. On the other hand, from the point of view of the student, the assessments are evaluated to the extent that they are seen as a required and mandatory exercise that must be carried out by students simply to pass the course. In the latter case, the activity can become extremely trivial and banal. Statistical analysis of the results indicates that three peer assessments per student appraised represents an adequate number. On the other hand, more than thirty peer assessments fail to contribute to learning, nor do they represent serious activities.

**Keywords** – assessment, peer assessment, assessment limits

-----

**1 INTRODUCTION**

Authentic, real learning always occurs as the result of reflection (Cowan, 2006), or in other words, awareness of learning and its implications for the personal structure of knowledge of each individual. In addition, the last objective of learning is frequently the ability to make good, correct decisions based on knowledge; i.e., the evaluation or assessment of a situation and in order to reach a decision. Accordingly, the failure to reflect on learning results in "low-quality" learning. As a consequence, evaluation should not be considered a simple act of classification or grading, as it has more and very important dimensions.

Perhaps the most important, critical, and judgmental of the different kind of assessments may be what is known as 'self-assessment', in which each student assesses himself/herself. This is, understood as one of several "reflections tool" that are available. Moreover, it can be also useful for adjusting the scope of learning (Boud, 1995; Andrade & Du, 2007).

In order to facilitate self-assessment, it is also very important to consider the tool known as 'peer assessment', which is understood as the exercise of value judgments regarding the learning of others, who are presumed to be cognitively equal, and in the most practical of cases, learning peers (classmates). When students reflect on



the product of the learning of peers (Keig & Waggoner, 1994), at the same time as they are also learning, this encourages an internal reflection on whether one's own learning is at the same, higher or lower level than that of others. Therefore, peer assessment posits the student as an observer and, at the same time, as an evaluator. Consequently, the student's own learning is, in turn, reinforced.

In terms of self-assessment, certain precise, external elements of control are required, according to which students can establish the authenticity of their knowledge, the understanding of concepts and, in general, of their learning. They provide reference models to the students, in order to compare the evolution of their own learning. If the learning is based on **concepts**, the references should be related to the students' ability to answer questions, make inferences, draw conclusions from situations, etc. These concepts should be presented in the texts or materials selected, or prepared by the course professors themselves. With regard to **procedures**, they should be oriented towards problems, situations, examples, etc. These should be also selected by the course professors. As far as **attitudes** (a much more complex competence to establish, since it is not restricted to a scientific approach or knowledge, rather it depends on the social and cultural needs of each student, among other things) are concerned, they are based on different factors, such as the attitudes of the professor and the educational center, the institution itself, appropriate readings, the proposal of situations, etc., trying to refrain from indoctrination.

Therefore, in self-assessment, there are many elements that promote authentic learning. It remains up to the professors to make proposals, giving the students the opportunity to engage in self-assessment so that they at least become aware of the learning that has taken place, of what remains to be learned, and the importance and status of such knowledge in the personal framework, under a "constructivist" approach to learning.

Without the ability to compare one's own learning to that of other classmates, the assessment process seriously lacks an element of reference. Furthermore, as Boud and Falchikov point out, "*peer assessment requires students to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining*" (2007, p.132). Actually, it is not only the comparison of one's own learning in terms of formal or scientific knowledge about concepts, procedures and attitudes. An element of reference would also be involved: the comparison of our own knowledge to that of our peers; i.e., the knowledge exhibited by other peers (usually fellow students or classmates). This allows the positioning of each student in relation to the rest of his classmates. Without the possibility to assess the knowledge of others (peer-assessment), the assessment triangle, formed by hetero-assessment (that carried out by the professor on his students) and self-assessment (that conducted by the student on his own performance), becomes faulty and weak. It would consist of an individual student who is presented formal knowledge, but without the support of peers, and thus, the support and assistance provided by social learning.

There are many aspects of social learning. The clearest is that two students learn more and faster than when working alone (conventional wisdom has summarized this in the old adage "two heads are better than one"). It is also true that nowadays, in most areas in our daily life, social learning occurs more frequently than individual learning. People are continually asked how to do certain things, perform different activities, etc. (for example, sending an e-mail, a fax, how a smartphone or PC application works, which TV or radio channel broadcasts a certain program, the time you need to be somewhere, etc.). This is a dimension of human relationships in which the social learning component is evident in our daily lives. This daily occurrence is also very common in academic learning: How do you calculate...?, How do you program...?, How do you say ...?, How do I mix ...?, What have you done with...?, etc.

One aspect of social learning that is favored by belonging to a certain group or collective (Pigott, Fantuzzo & Clement, 1986) is learning **from** others and **with** others and from the academic products of others, derived from a particular learning process. This means that the professor's educational goal is for his students to achieve a certain goal as the result of this training. The instructor establishes a learning procedure in which a teaching and educational methodology are chosen (narrative or cooperative, based on projects, problems or cases, portfolios, etc.), and students are required to provide a result of this learning in the form of a product that can be analyzed in terms of quality, having previously established quality criteria for said product. When students engage in the learning process established by the professor simply for the sake of carrying it out, they have already learned certain things (Cowan, 2006). However, if they do not reflect on what they have done, what they have learned, the learning may contribute little to the building of knowledge. This is the situation that occurs in the scaffolding of the concepts, procedures and attitudes inherent in the course assignment.

Previous ideas have been described by Topping (1998), Race (2001) and Nilson (2003), who outline how professors can use peer feedback as an alternative method of evaluation, in order to help students to acquire important life skills. Thus, convinced of its benefit, and the advantages suggested above, the experience presented in this article is derived from this approach. During the course of this experience, the authors have attempted to analyze some of the limits of peer assessment, an aspect which has been neglected in the previously consulted literature, and in particular the amount of work peer evaluation represents for the students. It would stand to reason that it is not the same thing to assess products made by only two classmates as it is to evaluate the activities of five - or even fifty - peers. The reader will probably perceive that fifty is a large number, but the questions remain: How many is it too many? and How many activity assessments are enough? Lin (2001), for example, describes a relatively good experience with six reviewers, in an effort to decrease bias due to assessments with fewer participants. On the other hand, it is worth examining whether students are really interested in doing peer assessments, even when this process is strongly beneficial and advantageous for their learning.

These questions are beyond the scope of this paper, because the answers depend on a variety of aspects: the group subjected to the experience, whether students are first-year university students or they are in final or intermediate courses, etc. We understand that, of course, answers could also vary, as in the case of students subjected to higher or lower overall academic pressure, or the period within the course in which the activity is proposed. Therefore, in this article, we focus on the experience in terms of the amount of peer assessment required, leaving the question of interest and benefits for further study.

## 2 RESEARCH QUESTIONS

This study attempted to answer the following research questions:

- Can differences be found among assessments made by different numbers of reviewers?
- Is there any way to process the information generated by the evaluators that is not excessively onerous in terms of time?
- Is there much difference between the results of peer assessment and those of professor assessment?
- What does the grade distribution look like? In other words, how are the marks distributed for each assignment?

We have restricted our investigation to reaching a final determination: Is there an optimal number of assessments per person in order to obtain reliable results in terms of grading?

## 3 EXPERIENCE

The peer assessment experience was carried out within the context of an Industrial Engineering course. More specifically, the course was entitled *Control and Industrial Automation*, and forms part of the framework of the European Higher Education Area (EHEA) or Bologna Process. The course is common to six different engineering degrees (Electronic, Electrical, Mechanical, Chemical, Biomedical and Energy Engineering) taught at the Barcelona College of Industrial Engineering (EUETIB) at the Technical University of Catalonia (UPC). The course was taught to six different groups of approximately 45 students each. Four of these groups received classroom instruction in the morning, and the other two in the afternoon. Students from the six different degrees were combined in the course groups, since it is a core course in the engineering program. With regard to assessing larger groups, the important thing to recognize is that they may require strategic solutions, which can only be implemented at the departmental or even institutional level, and which are beyond the control of individual tutors (Rust, 2001). In order to implement our experience, only two of these six groups were considered, one from the morning schedule and the other in the afternoon, because the overall performance of the students in the groups differed from morning to afternoon. Generally speaking, the students in the afternoon groups work and study at the same time. On the other hand, the students in the morning groups are dedicated to a single activity, i.e., studying. These two groups (the morning and the afternoon groups) have been used as a sample of each group, in representation of the remaining groups. This allowed us to prevent the transfer of opinion between students of both groups, and thus better isolate the two populations. One of the skills that a university graduate should possess is explicitly set out in Spanish law: "*Students can communicate information, ideas, problems and solutions to both specialist and non-specialist audiences*" (BOE, 2007). This competence is developed by all students and must be assessed by their classmates. Thus, it is critical for students to be

understood. This, in turn, is one of the characteristics of oral and written expression, which, without a doubt, is a fundamental competence for any university graduate.

Specifically, the assigned task required students to give a simple explanation of a technological issue of a certain degree of complexity, with the premise that anyone (a non-expert or layperson in the topic) could understand it. It should be considered that simple questions normally have simple explanations, while complex issues rarely have a simple explanation. Surely, for instance, it is not easy to explain in a nutshell the splitting of the atom to an audience that has absolutely no knowledge of what matter is made of. However, it is always possible to at least use comprehensible terminology, give examples and analogies, and make use of explanatory resources that can palliate the inherent difficulty of complex concepts.

In this work, the activity assigned to one of the groups consisted of describing how an ionic smoke detector works. It is based on the principle of the emission of ionizing radiation consisting of certain chemical elements, such as americium 140. This is a radioactive material that emits alpha particles and ionizes the air around it. This enables an electric current to flow between two electrodes. Thus, when the smoke particles fill the air around the material, the electrical current decreases and an electronic circuit detects the presence of these smoke particles. The topic assigned to the second group was the description of the term "phantom". In this case, this term applies to the power supply for capacitor-based microphones. In both cases, we need to understand both concepts very well in order to give a simple and concise explanation. As a matter of fact, it is only possible to explain something correctly, concisely, and completely if it is well known. Consequently, the aim of the proposed activity is for students to study these elements in detail. Only then can they give a competent and relevant explanation to a non-specialist audience.

Notwithstanding the difficulties already described, and others listed in Rust (Rust, 2001), including the problem of a small number of assessments, the nature of the activity is rooted in simplification, as the students in course had to make an effort in order to simplify the discourse and explanations. Thus, the activity was carried out using the social network Twitter. Accordingly, a limit of 140 characters was imposed to determine the student's comprehension (competence) to an even greater degree.

Therefore, the first part of the activity required the students to post their explanations on Twitter. Next, the second part of the activity involved the peer assessment of this explanation by some of their classmates. Over the course of a week, they scored each explanation, awarding between zero and ten points, based on whether it would be understood by a layperson.

In order to carry out the assessment activity, two patterns have been established, the first with a relatively low number of peer assessments, consisting of only three randomly chosen assessments. According to Race (Race, 2001), student peer-assessment can be anonymous, with assessors randomly chosen so that friendship factors are less likely to distort the results. However, in our case, we established a public list of assessments. However, since the students were coded, it would have been very difficult for anyone to find out the identity of the students evaluated or those that evaluated them. Thus, for all practical purposes, this implied randomness in selecting the reviewers.

On the other hand, the second assessment activity was massive, using the entire group of 37 students. Of course, it was assumed that assessment based on only three peers would yield different results than when the number was significantly larger, performed by 37 students in our case. Furthermore, in this second case, there was also a self-assessment component. When many assessments are made, it allows us to see whether the assessment carried out on oneself differs greatly from that performed by one's classmates. On the other hand, it should be noted that in the first case, we have preferred to limit the activity to peer-assessment (with no self-assessment) because, based on the authors' teaching experience, we believe that self-evaluation combined with only a few samples of peer-assessment (only three) could generate a bias effect on the final outcome. Data obtained from Twitter were processed using Microsoft Excel.

#### 4 RESULTS

As previously mentioned, we designed two different experiences with the same structure: the first part of the activity required the students to publish their explanations, and the second part consisted of the peer assessment of these explanation.

#### 4.1 Peer Assessment Carried Out By 3 Classmates (3-peer Assessment)

In one of the groups studied, it was established that the peer assessment of each student who had posted an explanation was to be carried out by three different classmates. The percentage of people posting an explanation was 89% (33 of 37). Thus, in principle, these were the students who could take part in the peer assessment. Finally, the population that took part in the peer assessment consisted of 29 out of the 33 students; i.e., 88%. Therefore, from the standpoint of participation, the minimum number of participants required to carry out the experience was exceeded. Thus, sufficient data were obtained for an accurate and reliable analysis.

The calculation of the results was based on the average of the marks given to each student by their peers, according to the method indicated by Brown, Bull and Pendlebury (1997): “An average for each student can be generated from the range of marks their peers give them”. Since not every member of the population took part in the peer assessment, some students were assessed only once or twice: 2 of the 33 were assessed only once, and 11 of the 33, only twice. The remaining 21 students received three peer assessments, and additional figures below are related to them. In addition, in this first case, the course professors analyzed all the given feedback, made any adjustments they considered necessary (Davies, 2006) and added an additional grade (the professor’s own score) for these students.

The average difference between the marks assigned by the course professors and the average grade given by students amounted to around  $\pm 1$  point, as is seen in Figure 1. It is curious to note that almost always, the marks assigned by the professor were more favorable (that is, the grades given by the professors are, in general, higher than those given by the course peers).

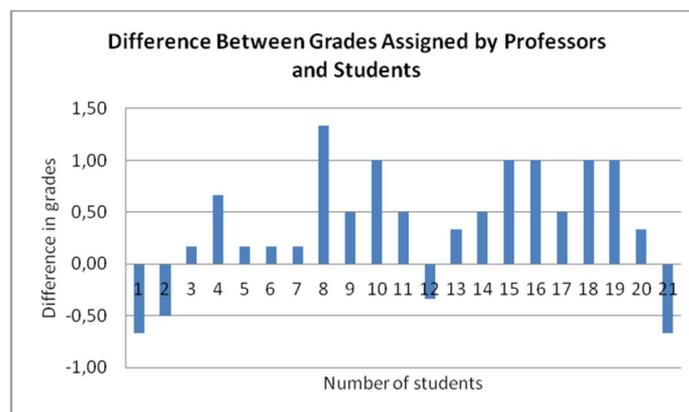


Figure 1. Difference between the grades given by the course professors and the average grade given by students for each of the explanations given (out of a total of 10 points)

Another element in our investigation was the analysis of the standard deviation between the grades given by the students. Figure 2 shows the concentration or dispersion of grades around the mean. In one case (student #21), the discrepancy was somewhat higher. Figure 2 can be understood as a measure of agreement or disagreement among the three students in terms of the respective average score.

It is important to highlight that, for the purposes of calculating the deviation, Bessel’s correction, which considers  $N-1$  samples instead of population =  $N$ , was used to compensate for our small number of samples.

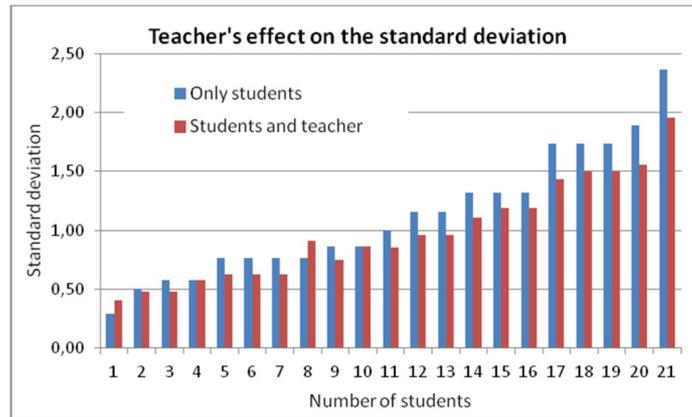


Figure 2. Standard deviation of the grades given by students and the effect when the instructor's grade is added (Only deviations >0 are shown)

An additional aspect of the study was the analysis of the grade distribution. Figure 3 shows a graph of all the grades and how their dispersion was distributed around the mean (standard deviation). A 6<sup>th</sup> order polynomial interpolation determined that the maximum grades for the different explanations fell between 7.5 and 8.5.

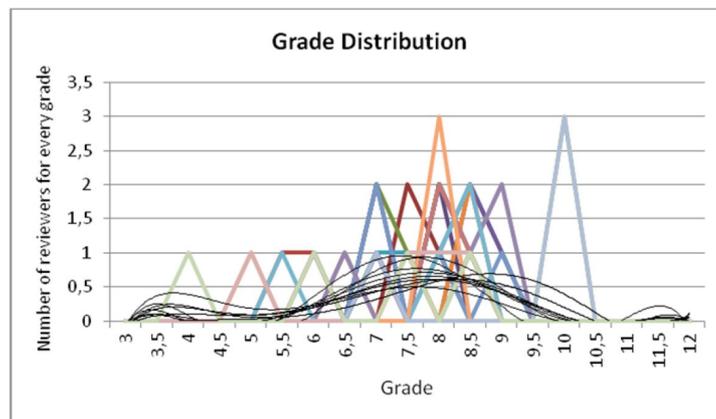


Figure 3. Distribution of the grades given to each explanation, with a polynomial interpolation curve of order 6 (the grades exceed 10 in the figure to allow proper interpretation of the interpolating curve)

#### 4.2 Whole-group Peer Assessment

In the other course group studied, the peer assessment was performed for each student who had posted an explanation. Thus, each student was required to assess *all* the submitted explanations, including his or her own activity (self-assessment). In this case, the authors did not take into account the difference between the grades given by the professor and the average grade given by the students for each of the given explanations. The reason is clear: the dispersion of results is so great that the professor's grade is quite insignificant in terms of the total.

86% of the students (37 of 43) posted an explanation, and therefore, they were considered to be the population that should be peer-assessed. In turn, 92% of the students (34 of 37) performed peer assessment. As before, from the standpoint of participation, the threshold was surpassed in order to consider this to be a reliable number of data for analysis.

In this case, as the course professor also assigned a self-grade for each student (i.e., self-assessment), it was found that these marks were higher in almost all cases than the average of the grades given by the rest of their peers, as shown in Figure 4. Except for one student, whose self-assessment was half a point below the average grade, the rest of students assigned themselves a higher mark, with more than four points of difference in some cases.

These errors in judgment lead us to suspect that the data should be considered, at best, doubtful, even when half of the students made an error of +1.5 points. This is somewhat reasonable, since one’s appreciation of oneself (and, thus, “self-assessment”) is usually more generous than that of one’s peers.

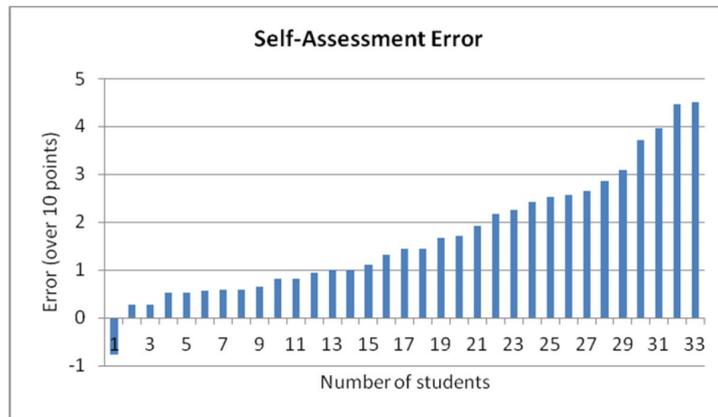


Figure 4. Self-assessment error

The calculation results are based on the average of the grades given to each student by his or her peers. In this case, not all students studied took part in the peer assessment. However, there were enough data (grades) on the explanations given by the students in the class and therefore this fact does not significantly influence the average results. One of the expected results was the variation range (i.e., the difference between the highest and lowest marks) found, which can be seen in Figure 5.

Range of the difference between maximum and minimum grades

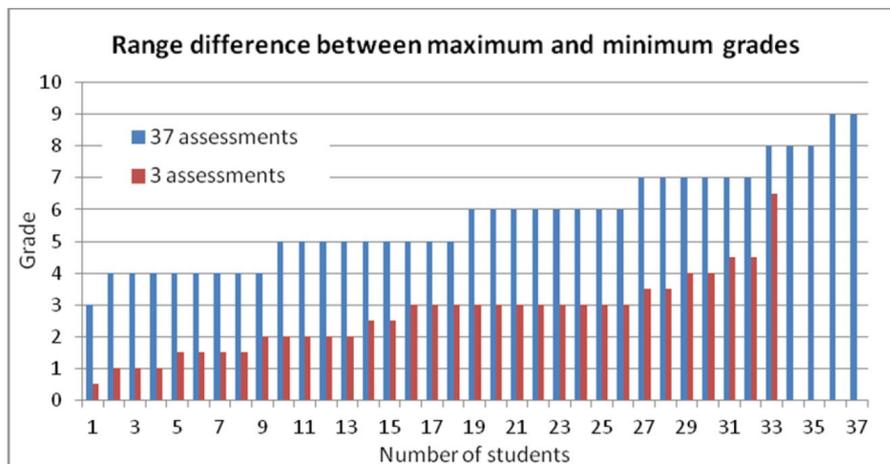


Figure 5. Comparison of peer assessments carried out for either 3 or 37 peers (whole-class group)

Regarding the analysis of the standard deviation of the grades given by the students, Figure 6 shows the concentration or dispersion of the grades around the average. Compared to the case of three-peer assessments presented in previous section, it can be seen that, in this second case, the standard deviation increases when the entire group is considered. As before, it is important to note that for the calculation of the standard deviation, Bessel’s correction was also considered.

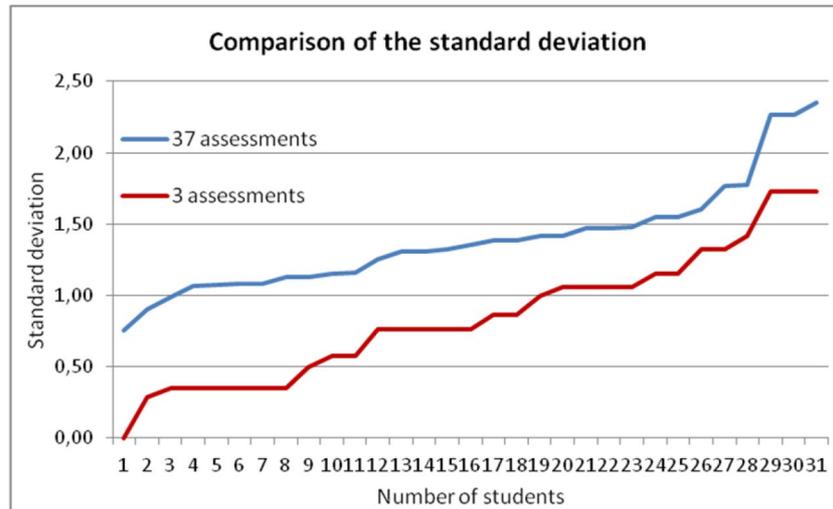


Figure 6. Standard deviation of the grades given by the students for the both cases considered in the study

When the dispersion of the distributions are compared in the case of 3 and the case of 37 peer assessments, the global deviation tends to increase as the number of peer assessments increases.

The final element of this case study is the analysis of how the marks are distributed for each explanation. Accordingly, Figure 7 depicts four graphs (grouping students according to the dispersion of the grades given by their classmates and how their dispersion is distributed around each mean.

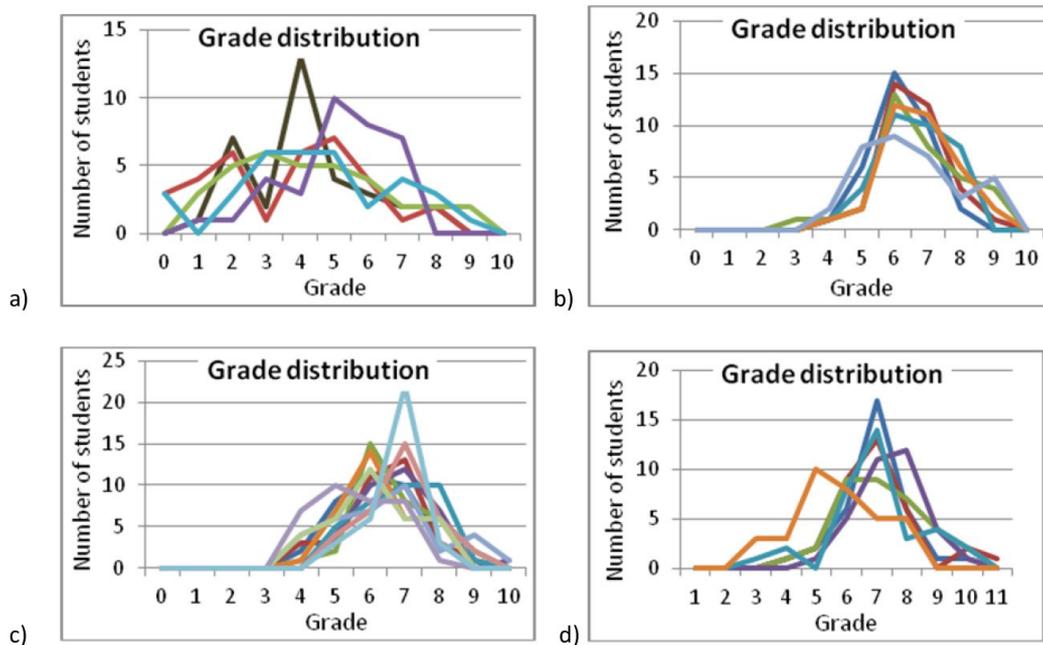


Figure 7. Distribution of the grades assigned to each explanation. a) Large spread of scores, b), c) and d) Higher or lower dispersion

The highest grade obtained through mass peer assessment was 6.9. However, in the case of 3-peer assessments, it reached 10 points (see Figure 8). It should be noted that students in the course have primarily learned the topic addressed in the first part of the activity, where their classmate were required to give a clear, concise explanation of a complex concept to a non-expert audience. However, it is interesting to note that the task of simply reading the different explanations (from the rest of the classmates) for the same concept produced greater learning, as compared to the understanding that the student initially had. Therefore, from the point of view of a student, we can conclude that the learning is both greater and richer: on one hand, thanks to

the act of performing the task itself, and on the other hand, thanks to the task of reading (and assessing) the explanations given by many classmates about the same topic.

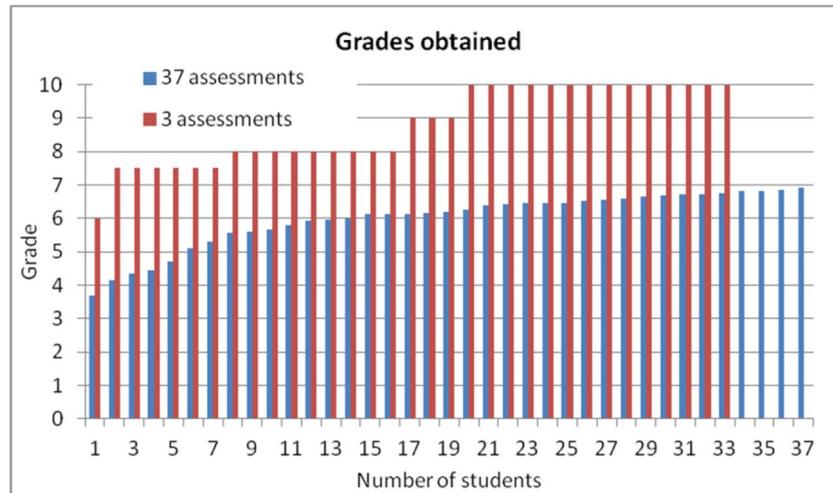


Figure 8. Marks obtained versus number of peer assessments

## 5 OBSERVATIONS AND DISCUSSION

It is interesting to note that, in the case of 3-peer assessments, it is publicly known whom is evaluating whom. Thus, it could be suspected that students who completed the activity later than others might know the grade that the other peer reviewers had already assigned to the classmate being assessed. They therefore might have had a reference in order to determine their own marks for their other classmates. In our case, however, the authors do not believe that this was the case, because the allocation of peer assessments follows no specific pattern, and logic would tell us that it would be more tiring to analyze the grades given by other students than to perform the assessment task assigned oneself.

In the case of global peer assessments, where the allocation of grades was an assignment, it may have been the case that some students chose one of the previous entries in Twitter and made slight modification to each grade. Nevertheless, if this did in fact occur, we have failed to observe the characteristic binomial distribution that was previously mentioned. Therefore, the authors believe that this effect has not taken place.

It is true, however, that if the professor passes around a sheet in class on which each student must write down a grade, a “memory” effect appears. Thus, the overall marks tend to resemble the first grade, since all the students know what the perceptions of the previous classmates are. In this way, classmate ‘2’ assesses in a way similar to classmate ‘1’; classmate ‘3’ assesses similar to classmates ‘2’ and ‘1’, and so on. This means that peer assessment results obtained by means of **public data** are not very reliable or desirable. Fortunately, the authors have found that, with the use of social networks, this effect dissipates somewhat.

If there are a large number of evaluations per person, most likely more than five, a high dispersion of results is observed, even if the task is simple and easy to complete. It should be noted that there are differences between the point of view of the student evaluator (four different evaluations carried out by the same person) and that of the evaluated student (evaluated by four different people).

From the point of view of the student evaluators, in the first experience, the authors noticed that students spent a certain amount of time when carrying out their first assessment. However, subsequent assessment times were faster, but less accurate. Therefore, a reasonable amount of peer assessments that students should be asked to do in order to obtain reliable results in terms of grading reliability is around three or four. We estimate that above this number, student evaluators will resort to random grading. In fact, in spite of its demonstrated virtues, peer assessment has a limitation to the number of persons engaged: the more people perform the evaluation, the less reliable the results are, which results in a greater dispersion of the ratings assigned by reviewers.

Conversely, from the point of view of the evaluated students, an optimal number of reviewers is not thought to exist. In the second experience, the four evaluations are more or less in agreement with one another in terms of the rating assigned by each evaluator, and the deviation is reasonable. Thus, the authors can conclude that

between three and five evaluations generate reliable assessments and, above this number, the quality of its virtues progressively deteriorates.

Finally, in Figure 2, it is curious that the results are almost always positive; i.e., more favorable grades are assigned by the professor. Thus, this confirms that the grades given by the professors are, in general, higher than those given by classmates.

## 6 CONCLUSIONS

Based on the study results, the authors conclude the following:

- a) If the number of peer assessments assigned to the students is reasonably low (around two or three), the students assign a grade for the task that is quite similar to that which would be assigned by the course professor.
- b) As seen in Figure 2, if the professor's grade is added to the calculation (thus, increasing the number of evaluators from 3 to 4), the overall distance from the mean decreases in most cases.
- c) From Figure 3, we can infer from the low deviation in the grades given by the students and the low error rate of the professors, the resulting grade could be truly representative of what each student has learned on a scale of 0 to 10. In addition, we can conclude that there is not a clear binomial distribution, a significant sign that students did not carry out a random evaluation for each assessment completed.
- d) From Figure 5, we conclude that assessments made by few students (the 3-peer assessment in our first case) for the same explanation result in less difference between the highest and lowest grades than those carried out by the whole group (the 37-peer assessment in our second case).
- e) From Figure 7, we can infer that the deviations of the grades given by the students are quite high, and a clear binomial distribution is evident. This is a clear sign that, for each assessment process, the data collected came from the students' own random (or at least pseudorandom) assessments.
- f) The remarkable dispersion of skills in the case of multiple peer assessments causes us to suspect that the students have not actually carried out the assessment activity, rather they have simply recorded numbers instead of giving reasons for their marks following a detailed reading of the explanations given by their peers. This is the reason for the huge differences (as much as nine points) in grades given for the same explanation, resulting in an evaluation range from zero to ten. In addition, more than half of the explanations exhibit differences of up to six points that, on average, fall three points above and three below the average value.
- g) With regard to the previous point, it is important to highlight that course students perceived the task of performing so many peer assessments to be excessive. Thus, they completed the task, but not in a serious manner, in terms of "scientific" peer assessment. They did not even use the evaluations previously conducted and published by other classmates as a reference.
- h) In fact, the number of peer assessments that students can reasonably be asked to perform, producing "reliable" grades that can be taken into account, is about three.
- i) By using statistical tools such as standard deviation, averages, variances, interpolations, etc., it is possible to determine the quality of the peer assessment carried out by students, especially in light of the impracticality of evaluating each on an individual basis and the fact that it has not been established as a general approach.
- j) With more than 3 peer assessments, instead of carrying out the desired learning process, students tend to assign a simple sequence of numbers with little sense and no actual qualifying value.
- k) In the case of mass peer assessments, the average grade assigned by student evaluators is 6.0. However, in the case of 3 peer assessments, this average mark is noticeably higher, i.e., 8.1.
- l) In the case of mass peer assessments, the trend was towards fairly similar grades in all cases, which is cause to suspect that the applicable assumptions of the law of large numbers cannot be valid, and thus the results are not reliable.
- m) The subject of peer assessment has been well documented, and the results reported in this article are predictable from a logical point of view.

## ACKNOWLEDGMENTS

The authors would like to express their appreciation for the useful comments received from RACEV group members (<http://blogs1.uoc.es/racev/>) who made suggestions and comments regarding this research and to the JOTSE reviewers for their useful suggestions.

## REFERENCES

- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-Assessment. *Assessment and Evaluation in Higher Education*, 32(2), 159-181. <http://dx.doi.org/10.1080/02602930600801928>
- BOE (2007). Nr. 260 of 30/10/2007 Real Decreto 1393/2007 of 29 October, on the organization of official university studies. Annex 1, paragraph 3.2.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D., & Falchikov, N. (Eds). (2007). *Rethinking assessment in higher education: Learning for the longer term*. London: Routledge.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing Student Learning in Higher Education*. London: Routledge.
- Cowan, J. (2006). *On Becoming an Innovative University Teacher: Reflection in Action*. McGraw-Hill Education.
- Davies, P. (2006). Peer assessment: Judging the quality of students' work by comments rather than marks. *Innovations in Education and Teaching International*, 43(1), 69-82. <http://dx.doi.org/10.1080/14703290500467566>
- Keig, L., & Waggoner, M.D. (1994). *Collaborative Peer Review: The Role of Faculty improving College Teaching*. Published by George Washington University. ASHE.ERIC Higher Education Report No. 2.
- Lin, S.S.J., Yuan, L., & Yuan, S.M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17, 420-432. <http://dx.doi.org/10.1046/j.0266-4909.2001.00198.x>
- Nilson, B. (2003). Improving student peer feedback. *College Teaching*, 51(1), 34-38. <http://dx.doi.org/10.1080/87567550309596408>
- Pigott, H. E., Fantuzzo, J.W., & Clement, P.W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *Journal of Applied Behavior Analysis*, 1(spring 1986), 19, 93-98. <http://dx.doi.org/10.1901/jaba.1986.19-93>
- Race, P. (2001). *The Lecturer's Toolkit* (2nd ed.). London: Kogan Page. Also published by Learning and teaching support Network (LTSN), Genesis 3, York Science Park, York, YO10 5DQ .
- Rust, C. (2001). *A Briefing on Assessment of Large Groups*. Published by Learning and teaching support Network (LTSN).
- Topping, K. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, 68(3)(Autumn, 1998), 249-276. <http://dx.doi.org/10.3102/00346543068003249>

**Citation:** Domingo, J., Martínez, H., Gomariz, S., & Gámiz, J. (2014). Some Limits in Peer Assessment. *Journal of Technology and Science Education (JOTSE)*, 4(1), 12-24. <http://dx.doi.org/10.3926/jotse.90>

On-line ISSN: 2013-6374 – Print ISSN: 2014-5349 – DL: B-2000-2012

## AUTHORS BIOGRAPHY

### Joan Domingo Peña

Received the B.Eng. degree in Electrical Engineering, specialization in Industrial Electronics, and the M.S. degree in Electronics Engineering from the Universitat Politècnica de Catalunya (UPC) in Barcelona, Spain, in 1983, and 1995, respectively. He received the Ph.D. degree in Electronics Engineering from the University of Barcelona (UB) in Barcelona, in 2001. Since 1983, Dr. Domingo develops his academic career in the College of Industrial Engineering of Barcelona (EUETIB-CEIB) of the UPC. He is currently an Associate Professor assigned to Systems Engineering, Automation and Industrial Informatics, ESAII Department of the UPC. He has taught different courses for undergraduate and graduate students in the areas of Industrial Electronics and Automation. He currently teaches in the area of industrial automation. His research interests include non-linear control, automatic systems and techniques for assessment and learning. He belongs to UPC RIMA-GIAC group for collaborative learning in classroom and RACEV group for collaborative learning in virtual environments.

### Herminio Martinez-Garcia

Received the B.Eng. degree (National Award) in Electrical Engineering, the M.S. degree (National Award) in Electronics Engineering and the Ph.D. degree in Electronics Engineering (all three with honors) from the Universitat Politècnica de Catalunya (UPC) (UPC) in Barcelona, Spain, in 1994, 1998 and 2003, respectively. During the period 1995-1998, Dr. Martinez-Garcia was a half-time Assistant Professor at the Department of Electronics of the College of Industrial Engineering of Barcelona (EUETIB-CEIB), where he became a full-time Assistant Professor at the same Department in September 1998. In September 2000 he joined the Department of Electronics Engineering of the Technical University of Catalonia (UPC), where he became an Associate Professor in 2006. Professor Martinez-Garcia currently teaches analog circuits design, communication systems, and data acquisition and control systems. His research focuses on the area of DC-DC power converters and their control, analog circuit design and techniques for assessment and learning. He belongs to RIMA-GIAC group for collaborative learning.

### Spartacus Gomàriz Castro

Received his Licenciatura, M.Sc. and Ph.D degrees in Telecommunication Engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain in 1990, 1995 and 2003 respectively. He is currently an Associate Professor with the Department of Electronic Engineering of the Universitat Politècnica de Catalunya and member of the research group "Remote acquisition systems and data processing (SARTI)". Since 1990, he has been a Teaching Assistant of Vilanova i la Geltrú School of Engineering (EPSEVG) of UPC. In 2010, he joined the staff of Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB) of UPC. His research interests include linear and nonlinear control theory, gain scheduled control, fuzzy control, design of navigation, guidance and control systems for underwater vehicles and different teaching and learning techniques. He is involved in national research projects and networks about underwater robotics and he is member of the following societies: IEEE Oceanic Engineering Society, Spanish Committee of Automation (CEA) of International Federation of Automatic Control (IFAC) and the Spanish Robotic Research Network on Marine Robotics and Automation (AUTOMAR).

### Juan Gámiz-Caro

Received the B.Eng. degree in Electrical Engineering, specialization in Industrial Electronics, from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1980, and the M.Eng. degree in Electronics Engineering from the University of Barcelona (UB) in Barcelona, in 1998. He also received the Ph.D. degree in Electronics Engineering from the UB in Barcelona, in 2004. He has been an Associate Professor at the areas of Electronics and Automatics of the College of Industrial Engineering of Barcelona (EUETIB-CEIB) since 1980, and he is attached to the Systems Engineering, Automation and Industrial Informatics, ESAll, Department of the UPC. Dr. Gámiz currently teaches microprocessor-based electronic systems, communication systems, data acquisition systems and automation and industrial informatics. His research focuses on the area of industrial applications with emphasis in industrial communications, and industrial process control and in active teaching techniques. Dr. Gámiz has authored or co-authored about 25 scientific papers in journals and conference proceedings and 15 books and book chapters.

Published by OmniaScience ([www.omniascience.com](http://www.omniascience.com))



Journal of Technology and Science Education, 2014 ([www.jotse.org](http://www.jotse.org))



Article's contents are provided on a Attribution-Non Commercial 3.0 Creative commons license. Readers are allowed to copy, distribute and communicate article's contents, provided the author's and Journal of Technology and Science Education journal's names are included. It must not be used for commercial purposes. To see the complete licence contents, please visit <http://creativecommons.org/licenses/by-nc/3.0/es/>