International Journal of Engineering Education Vol. 30, No. 3, pp. 711–721, 2014
Printed in Great Britain

0949-149X/91 $3.00+0.00

Peer Assessment and Self-assessment: Effective Learning Tools in Higher Education*

JUAN CARLOS GONZÁLEZ DE SANDE** and JUAN IGNACIO GODINO-LLORENTE
Circuits and Systems Department, ETSI de Sistemas de Telecomunicación, Universidad Politécnica de Madrid, Campus Sur UPM, Ctra. Valencia km 7, 28031 Madrid, Spain. E-mail: jcgsande@ics.upm.es

When used appropriately, self- and peer-assessment are very effective learning tools. In the present work, instructor formative assessment and feedback, self-assessment (SA), and peer-assessment (PA) have been compared. During the first part of a semester, the students followed a continuous formative assessment. Subsequently, they were divided into two subgroups based on similar performances. One subgroup performed SAs, and the other followed PA during the last part of the course. The performances of the two groups in solving problems were compared. Results suggest that PA is a more effective learning tool than SA, and both are more effective than instructor formative assessment. However, a survey that was conducted at the end of the experiment showed higher student confidence in instructor assessment than in PA. The students recognized the usefulness of acting as peer assessors, but believed that SA helped them more than PA.

Keywords: peer assessment; self-assessment, formative evaluation; evaluation methods

1. Introduction

During the last few decades, focusing on activities that promote student learning rather than on instructor teaching activities has emerged as a growing educational trend [1–4]. Examples of activities that promote student learning include peer learning, collaborative learning, group working, project working, and problem based learning. On the other hand, the manner in which students approach their learning is highly conditioned by the assessment method; thus, assessment is a fundamental part of the learning process. In this sense, assessment can be used as a learning tool. Among emerging forms of assessment in higher education, self-assessment (SA) and peer assessment (PA) are prominent in literature, with studies devoted to analyzing the results of many experiences and case studies [4–16].

Both SA and PA can provide the following benefits to students: improvement in critical thinking skills, greater sense of responsibility for their own learning, improvement in motivation, opportunity to observe and learn how peers address the same problem, opportunity to receive quick feedback regarding their performance and understanding of theory and key concepts, etc. [2, 10, 11, 17–19]. However, despite these positive aspects, SA and PA are not widely used in higher education, probably owing to fears regarding their reliability or validity, lack of ability of teachers to implement such assessment, and students' reluctance to criticize their classmates [11, 18, 20].

** Corresponding author.

Besides, several meta-analyses have shown that PA provides adequate reliability (different PAs of the same work are similar) and validity (PA resembles teacher assessment) in a wide range of applications [4, 8, 10, 11, 21, 22]. For example, with respect to PA, Falchikov et al. [11] reported a mean effect size of 0.02 (not statistically different from 0) for a collection of 24 studies, thereby 'indicating no consistent disagreement between faculty and peers on average'. On the other hand, with respect to SA, a mean effect size of 0.47 was found for a collection of 31 works, indicating 'that the average self-marker grades higher than approximately the 68% of faculty markers' [5].

From instructors' perspectives, SA and PA are advantageous in that the time-consuming task of assessment (especially when dealing with medium to large groups) is conducted by students. In the case of a formative assessment, students (and instructors) can doubt the validity and accurateness of the feedback provided. Thus, instructors must assume some control during the process: developing the correction criteria (with or without the participation of the students), guiding the application of such criteria, providing model solutions for each assignment, etc.

Besides, qualitative studies report an improvement in the writing skills of students who perform PA. PA improves learning both for the assessed and the assessor because the task of communicating feedback requires an explanation regarding how students can improve, and constructing such an explanation also enable the assessors' to improve their own writing skills [23–26]. In addition, in a recent study, McConlogue et al. [27] were surprised by their findings regarding giving such feedback:

* Accepted 22 December 2013.

'Some students improved their understandings of concepts as they looked for information to provide a "correct" feedback to their peers'.

Throughout this work, an attempt has been made to quantify the benefits of SA and PA in terms of the cognitive aspects of learning. The marks obtained by the same group of students in two different exercises were compared: the first comparison was made after the students received feedback only from the teacher; and the second was made after they had conducted SA or PA. The objective was to provide new evidence to support/reject the notion that PA is a more effective learning activity than SA, and that both processes are more/less effective than simply the formative assessment given by the instructor, at least for understanding new concepts and solving problems in engineering disciplines.

Section 2 describes the methodology and context of this experiment. Section 3 presents the results and Section 4 provides a critical discussion. Section 5 gives the concluding remarks.

2. Methodology

2.1 Research context

The present study has been developed in a second-year course (Signals and Systems) that is mandatory in the framework of the Electrical and Electronics Engineering (E&EE) degree given at the Escuela Técnica Superior de Ingeniería de Sistemas de Telecomunicación that belongs to the Universidad Politécnica de Madrid. The course was conducted during the fall term of the academic year 2010–11 and lasted 16 weeks during which regular classes were held. Both authors were involved in delivering and assessing this course. Students attended the lessons for 4 hours a week (divided in two sessions); approximately 55% of the course time was dedicated to practice, and the rest to theory and control exercises. Two additional weeks were allocated at the end of the course for final examinations. The assessment method applied in this course was a combination of formative and summative assessment; every week, the students were required to solve short exercises or to train themselves with automatically assessed and graded tests delivered using a b-learning system (Moodle) [28]. These activities represented 19% of the final mark. At the end of Weeks 3, 6, 11, 16, and 17, the students solved different exercises, which were assessed in a summative manner (the five classroom exercises represented 25% of the final mark). Additionally, using the b-learning platform available, at the end of Weeks 4, 7, 13, and 17, the students solved different tests regarding the same topics that they used for training (representing 16% of the final mark). The remaining 40% of the final mark was obtained through a final examination that was taken at the end of the semester. At the beginning of the course, students could choose between a single mark corresponding to the final exam or following the entire continuous assessment process. This course evaluation scheme was agreed upon by the teachers who were usually in charge of the course after an experimentation and reflection process [29, 30].

According to the PA typology conducted by Topping [8], the objectives of introducing SA and PA are reducing the time dedicated by instructors to assessment tasks and improving students' cognitive abilities. The focus was on formative assessment; therefore, students assessed written exercises, graded them (on a scale of 0 to 10), and gave feedback for wrong answers by comparing them with the model solution provided by the instructor. The grades of self- or peer assessed works replaced those given on the basis of instructor assessment and were considered to ascertain the final mark. One-way, one assessor-to-one assessed (changing for different exercises), and public peer assessment was implemented. After an assessment, assesses could ask assessors to provide reasons for the grades that were assigned to them, and the instructor acted as a moderator (the instructor decided in case of disagreement). All assessors and assesses were in the same course, and approximately 80% belonged to the same cohort of students; the remainder mainly belonged to the previous cohort of students, but neither this variable nor student capabilities were considered in this experience. Both SA and PA were conducted during formal classes and it was compulsory for everyone to follow the proposed assessment method (instead of only considering the final exam marks).

2.2 Research design

This work presents the results obtained from a quasi-experiment [14] and the students' opinions on the assessment method obtained through a questionnaire. The experiment compares the mean marks obtained by the students in an exercise that was solved in the classroom after receiving only the teachers' feedback, with the marks obtained by the same students in a second exercise after receiving the feedback from their SA or PA. The assigned homework as well as the classroom exercises consists of typical text book problems [31]. The instructor asked the students to fill out a questionnaire survey in the last session of the regular class, just after they had completed the last control exercise.

The experiment was divided into three phases (see Fig. 1). The first phase was developed during the first six weeks. Every week, the students solved a short problem (as homework) and gave it to the instructor, who assessed it and returned it to the

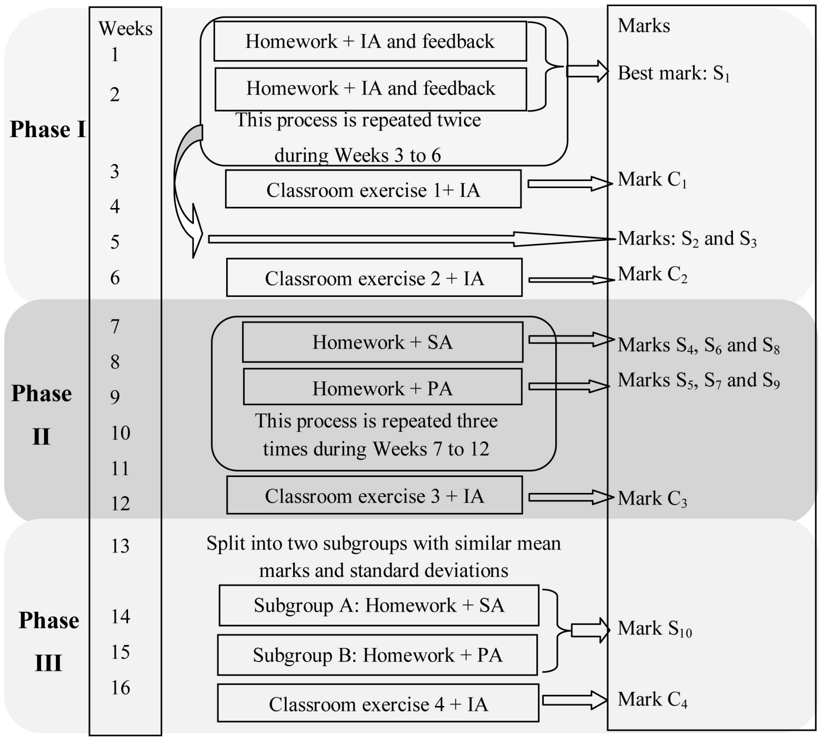

Fig. 1. Schematic of the quasi-experiment. IA: instructor assessment.

students during the next class session. The second, fourth, and sixth problems were similar to the first, third, and fifth, respectively. The higher mark of each pair of problems was the one that was counted for grading (S_1, S_2, and S_3). At the end of Weeks 3 and 6, two classroom exercises were developed (C_1 and C_2). Classroom exercises C_1 and C_2 were related to the homework that was done during Weeks 1 and 2 and Weeks 3 to 6, respectively.

The second phase of the experiment consisted of a training on SA and PA that lasted from Weeks 7 to 12; during odd-numbered weeks, the students self-assessed their own problems, and during even-numbered weeks, they assessed their peers' exercises (S_4 to S_9 marks). The students performed SA and PA by comparing the answers with a model solution given by the instructor. This model solution resembles the solved problems found as examples in many text books [31].

Finally, the third phase was developed during Weeks 13 to 16 and two subgroups of students were formed during this phase. During Week 14, one subgroup conducted SAs (SA subgroup in the following), and the other subgroup conducted PAs (PA subgroup) of their homework (mark S_{10}). Subsequently, during Week 16, both groups developed a classroom exercise that was related to the homework that was completed during Week 14 (mark C_4). These two subgroups were selected during Week 13 of the course in such a way that they approximately had the same number of students, with a similar proportion of male and females, who had shown similar performances in their work during the first 12 weeks (i.e., as of Week 12, they had obtained similar mean scores with similar standard deviation in the exercises and tests, including Moodle tests marks).

Table 1 resumes the collected marks corresponding to different assessment types and course weeks.

The students' opinions regarding instructor, self-,

Table 1. Assessment types, collected marks and weeks when these tasks were done

Assessment type		Mark	Week
Formative assessment	IA+feedback	S_1, S_2, S_3	1st to 6th
	SA	S_4, S_6, S_8	7th, 9th and 11th
	PA+feedback	S_5, S_7, S_9	8th,10th and 12th
	Subgroup A: SA Subgroup B: PA	S_{10}	14th
	Moodle tests	T_1, T_2,T_3, T_4	4th, 7th, 13th, 17th
Summative assessment	IA	C_1, C_2, C_3, C_4	3rd, 6th, 12th, 16th

Table 2. Survey items.

I.1.	It is feasible to understand and assimilate all the contents of the course.
I.2.	The weight of the homework in the final mark (19%) is adequate (in case you disagree, indicate if it should be increased or decreased).
I.3.	The weight of the Moodle tests in the final mark (16%) is adequate (in case you disagree, indicate if it should be increased or decreased).
I.4.	The weight of the classroom exercises in the final mark (25%) is adequate (in case you disagree, indicate if it should be increased or decreased).
I.5.	The weight of the final exam in the final mark (40%) is adequate (in case you disagree, indicate if it should be increased or decreased).
I.6.	Conducting peer assessment of my peers' exercises helps me to understand and assimilate the contents of the course.
I.7.	Assessment and feedback received from my peers helps me to understand and assimilate the contents of the course.
I.8.	Peer assessment of my peers' exercises is more helpful than self-assessment of my own exercises.
I.9.	I learned more from instructor assessment than from peer assessment.
I.10.	I always read the feedback received from my peers (not only the grade).
I.11.	I always read the feedback received from the instructor (not only the grade).

and peer assessment were analyzed using a five-point Likert scale questionnaire (5 = totally agree to 1 = totally disagree; see Table 2). Several questions regarding the weights that were assigned to the different activities were also included. The instructors asked the students to fill in the voluntary and anonymous questionnaires during the last session of the regular classes, just after they had completed the last control exercise. The questionnaires did not include any data that could identify the student and they just marked the selected answer.

The variables studied in the experiment were the type of assessments and feedback received by the students: instructor assessment and feedback, SA, or PA, and peer feedback. It must be noted that the students who were peer assessed also acted as assessors; therefore, they not only benefited from their peers' interpretation of the model solution but also from their own effort to understand their peers' exercises and complement them. The answers from the questionnaire allowed understanding the students' perceived usefulness of SA and PA (in comparison with the instructor's assessment).

A total number of 54 students who were registered in the course (14 females and 40 males) participated in at least one of the phases of the experiment; however, owing to its voluntary nature, only 49 questionnaires were collected (14 females and 35 males). Although the composition of the SA subgroup was 7 females and 18 males, the PA subgroup consisted of 7 females and 21 males. The ages ranged from 19 to 25 years with an average age of 20.5 and standard deviation of 1.5.

The following data were collected (see Fig. 1 and Table 1): (i) marks obtained by the students in their homework problems (S_1, to S_{10}), (ii) marks obtained by the students in the classroom exercises (C_1 to C_4), (iii) grades obtained in the first and second tests delivered using the b-learning platform (T_1 and T_2), and (iv) the survey data. With these grades, the following data were derived for each student: S_{1-3}: mean of grades S_1, S_2, and S_3 (instructor assessed exercises), C_{1-2}: mean of grades C_1 and C_2, and M: weighted marks of student performance until Week 12 (including short exercises, classroom exercises, and Moodle tests).

Histograms and 100% stacked column charts were used to represent and compare the different marks collected and answers to the survey. Mean values, standard deviations, and analysis of variance (ANOVA) were used to analyze, compare, and extract information from the data collected and derived. The effect size was taken as a measure of the 'benefits' obtained by the students as a result of instructor assessment and feedback, self-assessment, or peer assessment and feedback. This effect size is defined in the following manner [32]:

$$d = \frac{M_E - M_C}{\sqrt{\frac{\sigma_E^2 + \sigma_C^2}{2}}},$$

where M_E and M_C stand for experimental and control mean values respectively, and σ_E and σ_C are the experimental and control group standard deviations respectively. In this experience, the experimental values were the marks obtained by the students in the classroom exercises, which were graded by the instructor, and control values were the grades obtained by the same students in their respective homework, i. e. there is not a true control group. For this reason, this study must not be considered a true experiment, but a quasi-experiment [14].

3. Results

3.1 Performance after instructor assessment

Figure 2 shows the histograms of the marks obtained for the short homework (S_{1-3}) and classroom exercises (C_{1-2}) during the first phase of the experiment. The number of students who obtained high S_{1-3} marks (8–9 and 9–10 ranges) is larger than the number who obtained high C_{1-2} marks. The mean value of short homework exercises marks, S_{1-3}, is 6.85 with a standard deviation of 1.93, and the mean value of classroom marks, C_{1-2}, is 5.62 with a standard deviation of 1.27. The ANOVA of these two series of data gives a p value of .0002,

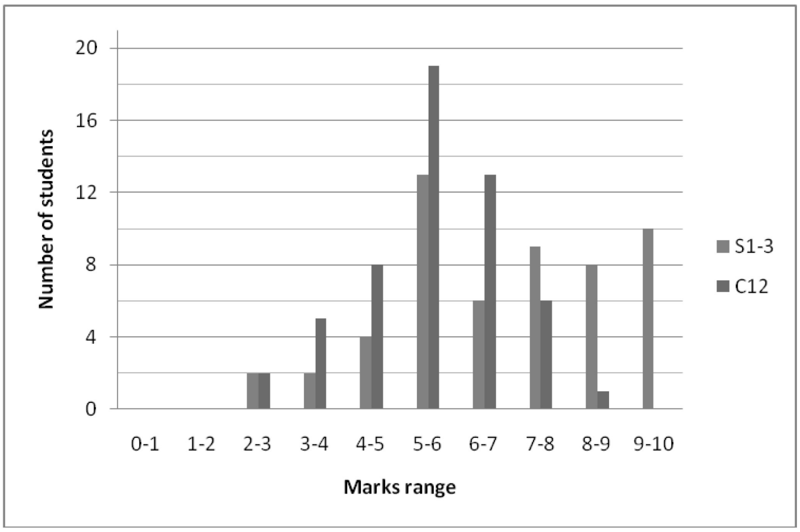

Fig. 2. Histograms of the mean marks obtained during initial weeks' homework S_{1-3} and first and second classroom exercise C_{1-2}. All these marks correspond to instructor assessment.

which indicates a statistically significant difference between the S_{1-3} and C_{1-2} series. However, the correlation coefficient of these two groups of marks is relatively high (0.56), which means that the students who performed well in homework exercises also performed well in classroom exercises and vice versa. The marks of short homework exercises, S_{1-3}, could be considered as control grades of the experimental performance in a classroom exercise. In the comparison of the students' grades for classroom problems with their respective homework marks, a negative effect size of $d_{IA} = -0.75$ was found. This negative value cannot be interpreted as a negative effect of the formative instructor assessment. In fact, this negative value was expected since the students had no constraints for doing their homework (they had five days for doing the homework and were permitted to use any kind of resources to solve the problems; they were even allowed to work in groups), whereas the students solved the classroom exercises individually, within a time restriction of 40 minutes, and with limited resources. Consequently, a better performance was expected for solving problems with no constrains rather than solving similar problems in the classroom.

Figure 3 shows the histograms of the weighted marks (until Week 12) for the subgroup of students who performed SA and those who performed PA during Week 14. The mean value for the marks of

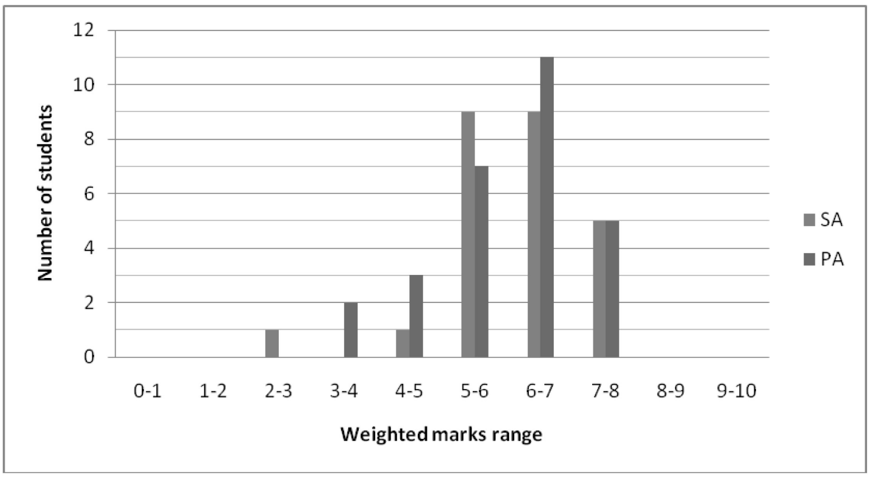

Fig. 3. Histograms of the mean marks obtained during Weeks 1 to 12 by the students who performed self-assessment (SA) or peer assessment (PA) during the last phase of the experiment.

the SA subgroup is 6.07 with a standard deviation of 1.05, whereas the mean value for those of the PA subgroup is 5.97 with a standard deviation of 1.16. Thus, since these two subgroups had performed in a similar manner during the first 12 weeks of the course, it can be expected that they also performed in a similar manner during the last weeks.

3.2 Performance after SA or PA

Figure 4 shows the homework marks obtained by both SA and PA subgroups for the third phase of the experiment (S_{10} marks). It must be noticed that the same homework was assigned to all students. In comparison with the histogram corresponding to the peer assessed homework, the histogram related to the SA marks is lightly biased to the high marks range. This displacement yielded a higher S_{10} mean for the SA subgroup (mean of 7.53 with 1.25 of standard deviation) than for the PA subgroup (mean of 6.97 with 1.8 of standard deviation).

Figure 5 shows the histograms of the classroom exercise related to homework, S_{10}, and the third phase of the experiment (C_4 marks) for both subgroups of students. This classroom exercise, which is the same for all students, was assessed by the instructor. The histogram of the PA subgroup is lightly biased to the high marks range (compared with the histogram of the C_4 marks for the SA subgroup). The mean of the C_4 marks for the PA subgroup was 7.04, with a standard deviation of 1.69, whereas the mean marks for the SA subgroup was 6.67, with a standard deviation of 1.73. ANOVA of these two groups yielded $p = 0.44$, which means that from a statistical perspective, there is no difference between the marks of these two groups. However, a statistically significant

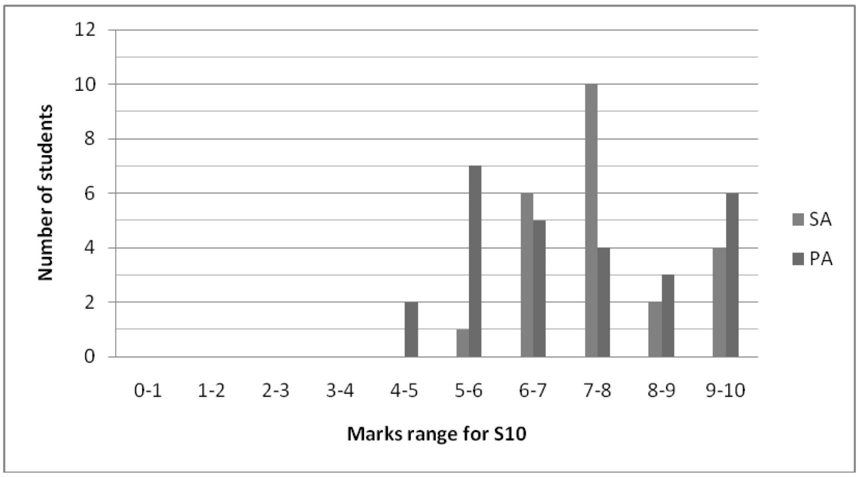

Fig. 4. Histograms of the students' homework S_{10} marks. These marks correspond to self-assessment (SA) or peer assessment (PA) during the third phase of the experiment.

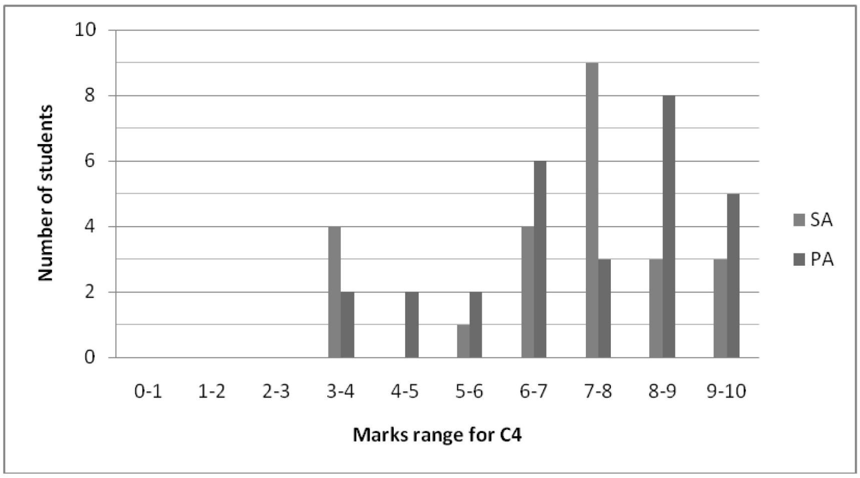

Fig. 5. Histograms of the marks obtained by the students in classroom exercise C_4 after following self-assessment (SA) and peer assessment (PA). These marks correspond to instructor assessment.

decrease from 7.53 to 6.67 ($p = 0.05$) was observed for the SA subgroup in the comparison of the marks obtained in the homework, S_{10}, and classroom exercise, C_4, whereas no meaningful differences were observed for the PA subgroup. In terms of the effect size, a negative $d_{SA} = -0.57$ was obtained for the SA subgroup (again, this negative value is owing to students' abilities to solve problems with no constraints, which is better than their ability to solve similar problems individually in the classroom).

A key point is the reduction of the effect size from 0.75 to 0.57 when introducing SA instead of instructor assessment. It must be highlighted that the topics that were studied at the end of the semester were more difficult to understand than the topics studied at the beginning of the semester; therefore, d_{IA} and d_{SA} were extracted under different conditions and they must be compared carefully. On the other hand, in a comparison of the performance of the students belonging to the PA subgroup in the homework, S_{10}, and in the classroom exercise, C_4, a very small [30] effect size $d_{PA} = 0.035$ is obtained. This means that the PA subgroup has a similar performance in solving classroom exercises and doing homework problems. Both d_{PA} and d_{SA} were extracted under the same conditions for the two similar subgroups of students; thus, it seems that students who performed and received PA and feedback improved their performance more than those who only performed SA.

3.3 Survey data

Figure 6 shows the responses of the students to items 1 to 11 of Table 2. The vertical axis represents the percentage of the responses that corresponds to each Likert scale category (from 5 = Totally agree to 1 = Totally disagree) for each item.

Items 1 to 5 are related to global issues of the course and to weights assigned to each kind of activity (homework, Moodle tests, classroom exercises, and final exam). It is worth noting that 40 students (82%) considered that the objectives of the course can be reasonably achieved (I.1). Most of the students were in agreement or total agreement with the weights assigned to homework (59% of students), Moodle tests (69%), and classroom exercises (84%). However, 18 students (37%) were in disagreement or total disagreement with the weight assigned to the final exam (15 students thought that it should be lower, and three thought that it should be higher than the actual weight). These results suggest that the students held a favourable opinion of the assessed tasks (especially regarding continuous assessment) and thought that the course objectives were achievable.

Items 6 to 11 of Table 2 are related to students' perceived usefulness of SA, PA, and instructor assessment. Most of the students reported that acting as peer assessors enabled them to understand and assimilate the course contents (60% of them were in agreement or total agreement with item I.6). However, there was a division of opinions regarding the usefulness of being assessed: 43% of students considered that being peer assessed was useful and 43% considered that this kind of assessment did not help them (item I.7). When comparing SA with PA, 55% of the students believed that SA was more helpful than PA, and only 16% were in agreement (seven students) or total agreement (one student) with the opposite (item I.8 of Table 2, PA is more helpful than SA). Instructor assessment was considered much more positive than PA; three out of four students were in agreement or total agreement with item I.9, wherein the instructor assessment was compared with PA. Finally, according to the survey

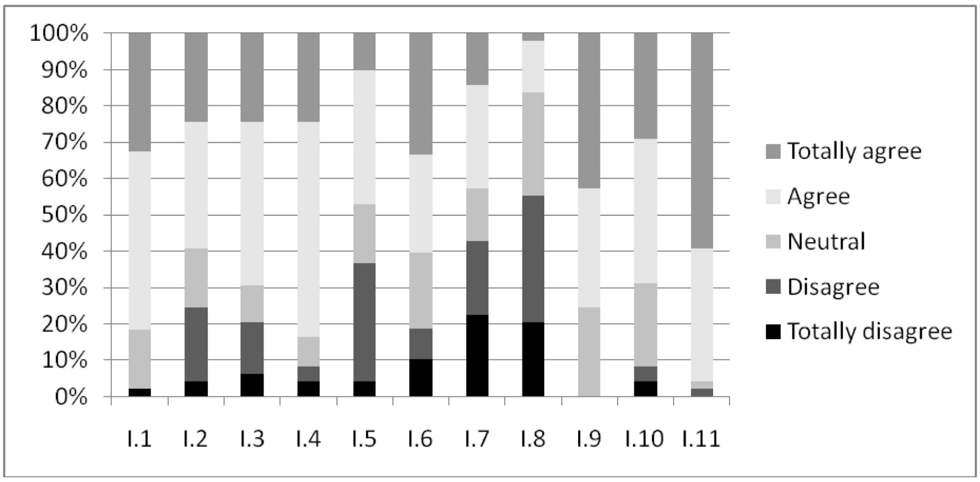

Fig. 6. Percentage of answers to items I.1 to I.11 in Table 2 corresponding to each Likert scale category.

answers, most of the students always read the instructor feedback (96%) and their peers' feedback (69%).

4. Discussion

A qualitative comparison of Figs 4 and 5 show that there was a small number of students who obtained worse marks in the classroom exercise than in the SA homework; all short homework marks, S_{10}, for the SA subgroup were above 5 points, whereas several classroom exercises marks, C_4, were in the 3–4 range. This effect can directly be observed in Fig. 7, which represents S_{10} marks versus C_4 marks both for SA and PA subgroups (several squared dots are at the left of 5 on the horizontal axis, but significantly over 5 in the vertical direction). The following possible reasons can explain these results: the SA group of students could have over-marked their homework, used some help in doing their homework, or the classroom environment affected them more than it affected their peers [19, 22, 30]. Fig. 7 also shows a higher correlation between the short homework (S_{10}) and classroom exercise (C_4) marks for the PA (correlation coefficient equal to 0.52) than for the SA subgroup (correlation coefficient equal to 0.29), which suggests a deeper and more confident learning process for the PA than for the SA subgroup. These qualitative findings together with the facts that the mean of the C_4 marks for the PA subgroup was higher than that for the SA subgroup, and that a higher effect size was obtained for the PA than for the SA subgroup ($d_{PA} > d_{SA}$), seem to support the hypothesis that PA is a more effective learning tool than SA.

The mean of the classroom exercises marks that correspond to the first phase of the experiment (the mean of the C_{1-2} marks is 5.62) is lower than the mean of the marks that correspond to the third phase of the experiment (the mean of the C_4 marks is 6.87, including both SA and PA subgroups); this difference was found to be statistically significant with $p = 0.0002$. On the other hand, during the previous academic years, the topic Signals and Systems used to have a dropout rate ranging between 40% and 60% [33]. During the semester under study, 49, 50, and 50 students out of 54 participated in the last three activities that were planned before the final exam, respectively (and 51 students, i.e., 94%, performed the final exam). Subsequently, a drastic reduction of the dropout rate was observed. This reduction may be attributable to the differences in the methodology followed in the group under study with respect to the previous academic years: the introduction of the SA and PA assessment, faster feedback on their work (owing to the introduction of the SA and PA), and consideration of a higher number of exercises in order to obtain the final mark. Thus, it is reasonable to think that the introduction of SA and PA has been an important factor in this reduction. These facts, together with the mentioned negative effect size found for the instructor assessment ($d_{PA} > d_{SA} > d_{IA}$), indicate that both SA and PA are more effective learning tools than instructor assessment [18, 27].

Although analyzing the reliability of both SA and PA is not the main objective of this work, the 10th homework assignment was assessed by the students (S_{10} marks) and instructor for comparison purposes

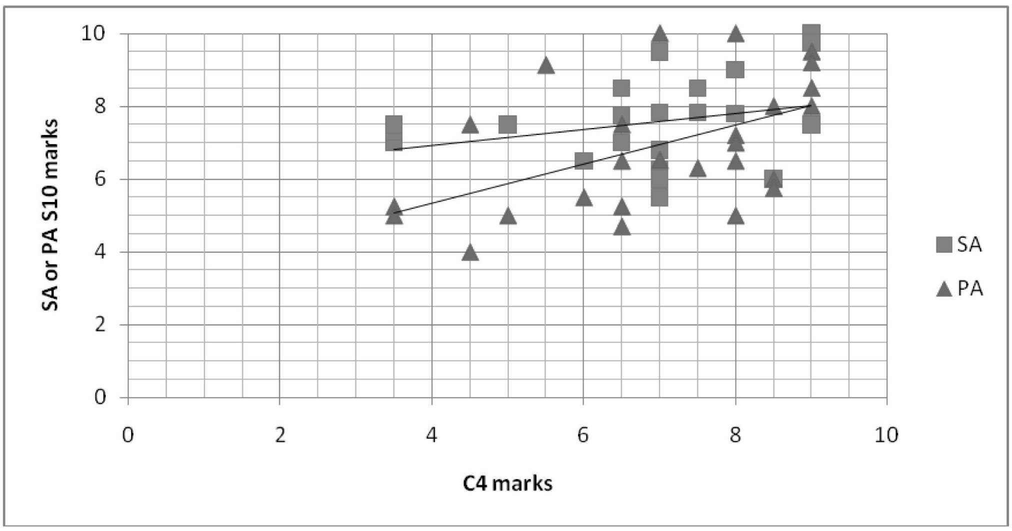

Fig. 7. Comparison of the marks given by the students to their own (self-assessment (SA)) or their peers' homework S_{10} (peer assessment (PA)) with those obtained in classroom exercise, which is assessed by the instructor C_4.

(see Fig. 8). The instructor assessment of this homework was not communicated to the students in order to avoid interferences with the development of the experiment. As it can be observed in Fig. 8, small differences were found between the instructor and the student's assessment, and considering the small weight that the S_{10} mark had in the final mark, these differences yield negligible differences in the final marks; therefore, omitting this information to the students was justified. Similar results are found in a different study of the reliability and validity of SA and PA under comparable conditions [22]. The instructor assessment of the students' homework in the SA group yielded a 7.43 mean value, with a standard deviation of 1.35, and a correlation coefficient with SA marks of 0.84. On the other hand, the instructor assessment for the students who performed PA yielded a 7.25 mean value, with a standard deviation of 1.70, and a correlation coefficient with PA marks of 0.92. These values gave us small effect sizes, that is, 0.077 for the SA subgroup and 0.16 for the PA subgroup, when the experimental marks given by the students to their own or to their peers' homework were compared with the 'control' marks given by the instructor. On a 0 to 10 scale, the maximum difference between the instructor assessment marks and SA or PA marks was 1.5 points. The following reasons could explain such high correlation coefficients, small effect sizes, and small differences: (i) Both groups of students had acquired some experience in SA and PA during the second phase of this experience; and (ii) the SA and PA were performed by comparing answers with a model solution. This data supports the reliability of SA and PA that has been reported in previous reviews [5, 8, 10, 11, 22].

Regarding the survey results, it is obvious that the answer to some of the questionnaire items could be biased because the students may attempt to portray themselves in a good light with the instructor and answer questions accordingly. This could occur with items I.1, I.9, and I.11, which consider some aspect of the instructor work. However, for the rest of items, especially for those comparing SA and PA, the authors do not find any reason to expect a biased response from students.

Recent studies on students' perception of SA and PA assessment claim that students have a positive opinion of the PA system and usefulness of peers' feedback [4, 15, 25–27, 34]. In the present work, the students appreciated the learning benefits of acting as assessor; however, they preferred to assess their own work instead of their peers. On the other hand, the number of students who thought that PA and feedback helped them in understanding and assimilating the course contents was equal to the number of students who preferred PA to assessing their own work.

O'Dwyer [26] reported a division of opinions when comparing learning benefits of PA with those of instructor assessment (mean values of approximately three on a five-point Likert scale). In the present work, 75% of the students thought that they had learned more from the instructor assessment than from PA, and the remaining 25% of students had no preference, but this observation must be considered carefully because the students' answers to item I.9 may be biased. Additionally, 96% of the students said that they read the instructor feedback (these answers are the most susceptible to being biased), whereas only 69% read the feedback from peers. This data could suggest a lack of confidence in their peers' assessment and feedback in comparison with the instructor's feedback [4].

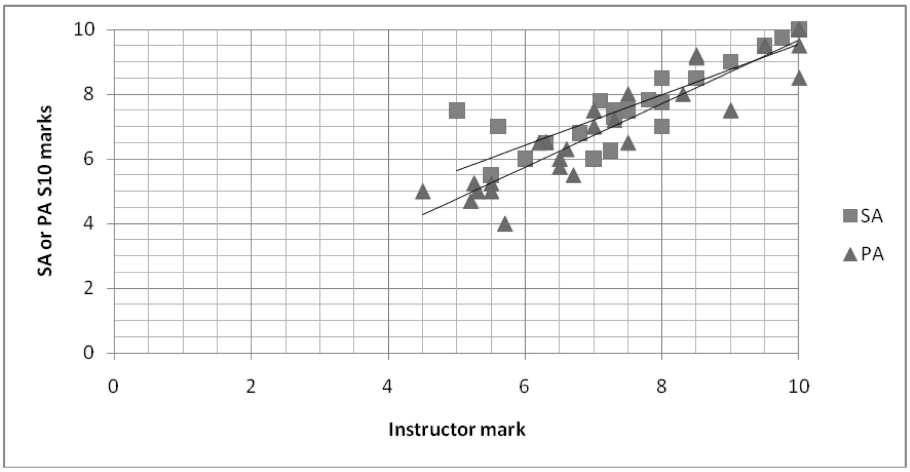

Fig. 8. Comparison of the marks given by the students to their own (self-assessment (SA)) or their peers' homework S_{10} (peer assessment (PA)) with those given by the instructor for the same homework.

These negative feelings toward PA in comparison with the instructor assessment must not be interpreted as a negative opinion of the students regarding the course or general assessment method because they showed positive opinions to these items (see Fig. 6, items I.1–I.6).

5. Conclusions

The comparison of problem solving abilities of two groups of students who had previously participated in PA or SA showed that the students who participated in PA performed better than those who participated in SA. The comparison of the problem solving abilities of the same students after receiving only the instructor assessment supported the idea that SA and PA were more effective than the instructor formative assessment. Results also show that SA and PA are highly reliable. Moreover, a larger correlation between peer-assessed assignment marks and instructor-assessed classroom exercise marks than the correlation between self-assessed assignment marks and instructor-assessed classroom exercise marks could also mean a better understanding of the course contents when students participated in PA than when they performed SA of their own assignments. On the other hand, the answers to the survey at the end of the course indicate higher confidence in instructor assessment and feedback than in PA and peer feedback. In addition, the survey reveals that although the students recognized the usefulness of acting as peer assessors, they believed that SA was more helpful than PA.

Acknowledgments—This work has been funded by the Universidad Politécnica de Madrid under Project IE105902139.

References

1. D. Boud, R. Cohen, and J. Sampson, Peer learning and assessment. *Assessment and Evaluation in Higher Education*, **24**, 1999, pp. 413–426.
2. K. Topping, Peer assessment between students in colleges and universities, *Review of Educational Research*, **68**, 1998, pp. 249–276
3. M. Keppell, E. Au, A. Ma and C. Chan, Peer learning and learning-oriented assessment in technology-enhanced environments, *Assessment and Evaluation in Higher Education*, **31**, 2006, pp. 453–464.
4. H. Ertl and S. Wright, Reviewing the literature on the student learning experience in higher education, *London Review of Education*, **6**, 2008, pp. 195–210.
5. N. Falchikov and D. Boud, Student self-assessment in higher education: A meta-analysis, *Review of Educational Research*, **59**, 1989, pp. 395–430
6. L. A. J. Stefani, Peer, self and tutor assessment: Relative reliabilities, *Studies in Higher Education*, **19**, 1994, pp. 69–75.
7. E. Mazur, *Peer instruction: A user's manual*, Prentice Hall, Upper Saddle River, NJ, 1997.
8. K. Topping, Trends in Peer Learning, *Educational Psychology*, **25**, 2005, pp. 631–645.
9. U. Zoller and D. Ben-Chaim, Student Self-Assessment In HOCS Science Examinations: Is There A Problem? *Journal of Science Education and Technology*, **7**, 1998, pp. 135–147.
10. F. Dochy, M. Segers, and D. Sluijsmans, The use of self-, peer and co-assessment in higher education: a review, *Studies in Higher Education*, **24**, 1999, pp. 331–350.
11. N. Falchikov and J. Goldfinch, Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks, *Review of Educational Research*, **70**, 2000, pp. 287–322.
12. D. Boud and N. Falchikov, *Rethinking Assessment in Higher Education Learning for the Longer Term*, Routledge, 2007.
13. A. D. Vidic, Assessment in Problem-based Learning Incorporated into Traditional Engineering Education: Difficulties and Evaluation, *International Journal of Engineering Education*, **26**, 2010, pp. 554–563.
14. J-W. Strijbos and D. Sluijsmans, Unravelling peer assessment: Methodological, functional, and conceptual developments, *Learning and Instruction*, **20**, 2010, pp. 265–269.
15. M. Van Zundert, D. Sluijsmans, and J. Merriënboer, Effective peer assessment processes: Research findings and future directions, *Learning and Instruction*, **20**, 2010, pp. 270–279.
16. Y. Cinar and A. Bilgin, Peer Assessment for Undergraduate Teamwork Projects in Petroleum Engineering, *International Journal of Engineering Education*, **27**, 2011, pp. 310–322.
17. P. Race, A Briefing on Self, Peer and Group Assessment, *LTSN Generic Centre*, 2001.
18. M. Dziedzic, P. R. Janissek and A. P. Bender, Assessment by peers—An effective learning technique, *38th ASEE/IEEE Frontiers in Education Conference*, Saratoga Springs, NY, October 22–25, 2008, pp. T2F-1-5.
19. J. Swingler, Peer assessment and developing students' professional ethics, *The Higher Education Academy Engineering Subject Centre, EE2008*, 2008, pp. 1–16.
20. M. L. Wen, C. C. Tsai and C. Y. Chang, Attitudes toward peer assessment: a comparison of the perspectives of pre-service and inservice teachers, *Innovations in Education and Teaching International*, **43**, 2006, pp. 83–92.
21. J. A. Marín-García, C. Miralles and M. P. Marín, Oral Presentation and Assessment Skills in Engineering Education, *International Journal of Engineering Education*, **24**, 2008, pp. 926935.
22. J. C. G. de Sande, J. I. Godino-Llorente, V. Osma-Ruiz and N. Sáenz-Lechón, Analysis of the Validity of e-Assessment and Self-Assessment in Formal Assessment in Electrical and Electronics Engineering Studies trough a Case Study, *IEEE 12th International Conference on Advanced Learning Technologies (ICALT 2012)*, Rome, Italy, July 2012, pp. 31–35.
23. N. Falchikov, Involving students in assessment. *Psychology Learning and Teaching*, **32**, 2003, pp. 102–108.
24. K. Cho, C. D. Schunn and K. Kwon, Learning writing by reviewing in science, *Proceedings of Computer-Supported Collaborative Learning*, New Jersey, 2007.
25. K. Ellery, Assessment for learning: A case study using feedback effectively in an essay-style test, *Assessment and Evaluation in Higher Education*, **33**: 2008, pp. 421429.
26. A. O'Dwyer, Learning and assessment of student communication skills on engineering programs: Some experiences, *IEEE Transforming Engineering Education Conference*, Dublin, Ireland, 2010
27. T. McConlogue, J. Mueller and J. Shelton, Challenges of developing engineering students' writing through peer assessment, *The Higher Education Academy Engineering Subject Centre, EE2010*, 2010.
28. Moodle.org: Open-source community-based tools for learning, http://www.moodle.org, [Accessed 20 July 2013].
29. J. C. G. de Sande, L. Arriero, C. Benavente, R. Fraile, J. I. Godino, J. M. Gutiérrez-Arriola, D. Oses and V. Osma, A case study: Final exam versus continuous assessment marks for electrical and electronics engineering students, *Proceedings of ICERI 2008 Conference*, Madrid, Spain, November, 2008.
30. J. C. G. de Sande, R. Fraile, L. Arriero, V. Osma, D. Oses and J. I. Godino, Cheating and learning through web based tests, *Proceedings of ICERI 2010 Conference*, Madrid, Spain, November 2010, pp. 5004–5009.

31. A. V. Oppenheim, A. S. Willsky and S. H. Nawab, Signals and Systems, Prentice-Hall, New Jersey, 1997.
32. J. Cohen, Statistical Power Analysis for the Behavioral Sciences (2nd ed.), *Lawrence Erlbaum Associates*, Hillsdale, NJ, 1988.
33. J. C. G. de Sande, L. Arriero, C. Benavente, R. Fraile, J. I. Godino, J. M. Gutiérrez-Arriola, D. Oses and V. Osma, Evolution of efficiency and success rates for electrical and electronic engineering students at EUITT from Universidad Politécnica de Madrid, *Proceedings of INTED 2009 Conference*, Valencia, Spain, March 2009, pp. 4725–4732.
34. B. Basnet, L. Brodie, and J. Worden, Peer assessment of assignment, *40th ASEE/IEEE Frontiers in Education Conference*, Washington, DC, 2010.

Juan Carlos González de Sande received the M.Sc. degree from Universidad Autónoma de Madrid (UAM), Madrid, Spain, in 1987, and the Ph.D. degree from Universidad Complutense de Madrid (UCM), Madrid, Spain, in 1994, both in applied physics. From 1989 to 1994, he was with the Department of Optics, UCM, as an Assistant Professor. From 1994 to 1996, he was with the Physics Department, Universidad Politécnica de Madrid, Madrid, Spain, and in 1996 he joined the Circuits and Systems Department as an Associate Professor. His research interests include optoelectronics, digital signal processing, and optical characterization of materials.

Juan I. Godino-Llorente received the M.Sc. degree in communications engineering and Ph.D. degree with honors in computer science from the Universidad Politécnica de Madrid, Madrid, Spain, in 1996 and 2002, respectively. Currently, he is Full Professor with the Universidad Politécnica de Madrid, where he was the Head of the Circuits and Systems Engineering Department from 2005 to 2009. His main research areas are in the fields of assistive technology and biomedical signal and image processing. He has led numerous projects and contracts in these fields.