ASSESSMENT OF WORK PERFORMED BY PAIRED STUDENTS: A PEDAGOGICAL EXPERIENCE

A. Souto-Iglesias, A. Cantón Pire, M. Pérez de la Portilla, M. Hermanns
Technical University of Madrid (UPM) (SPAIN)
antonio.souto@upm.es, alicia.canton@upm.es, marinap.portilla@gmail.com, miguel.hermanns@upm.es

.... to me a good school would be a different school, a school that had as principle that all its policies were aimed at keeping as many students as possible for as long as possible within the system. So, everything should be aimed at making students participate, to feel identified with the school, to have the sensation of doing things right. For me a good school is a school that keeps all the students working, committed and with the sensation that they will not fail.

Stephen Ball.

Abstract

"Teamwork" is one of the abilities most valued by employers. In [16] we describe the process of adapting to the ECTS methodologies (for ongoing assessment), a course in computer programming for students in a technical degree (Marine Engineering, UPM) not specifically dedicated to computing. As a further step in this process we have emphasized cooperative learning. For this, the students were paired and the work of each pair was evaluated via surprise tests taken and graded jointly, and constituting a substantial part of the final grade. Here we document this experience, discussing methodological aspects, describing indicators for measuring the impact of these methodologies on the educational experience, and reporting on the students' opinion of it.

Keywords: Competencies; Teamwork; Cooperation; Paired Students; Assessment.

1 INTRODUCTION

A precise definition of competence "Teamwork" is the willingness and ability to share knowledge and experiences that allow a person to work together with others towards a common goal, distributing and/or assigning responsibilities based on the strengths of each of the team members [2]. This is one of the generic skills most valued by employers, as can be seen for example in the White Papers (www.aneca.es) of degrees as representative as Computer Engineering and Industrial Engineering, which show statistical data to justify such statement.

In the training of our graduates, this competence is not usually treated with much rigor. In fact, there are not many references in the literature concerning the evaluation of continuous work in groups [1],[5]. In most cases the techniques come from the Cooperative Learning context [1] and its assessment is often based on a random selection of certain groups so that in a given time, one of its members, also selected at random, presents the work of this group. Among these approaches is that documented by Maroño Sanchez et al [13] in a course similar to ours. Other interesting approaches, even within this philosophy, are those who seek the added value of working as a team for carrying out certain activities, very difficult to do individually, and for which a continuous monitoring throughout the course is undertaken [6]. One possibility we believe that has not been sufficiently explored and paid attention to is that of examinations in groups that contribute to the individual mark of each component. This is one of the issues we will treat in the present paper. Furthermore, we have implemented an assessment of paired students work in which a significant part of the mark of the two members of the pair is the minimum of the marks obtained by those two members in exams taken individually. This way we aim at promoting a responsible behavior within each pair, at linking the transversal competence (team work) with the longitudinal ones (computing programming coding, validating), and at engaging with the students, trying to reduce abandon rates, which are especially high for 1st year students.

To do this, and starting from what is documented with a general perspective in [16] about the adaptation to ECTS methodology of a introduction to programming course in an Engineering degree

(Programming Languages, PL, Marine Engineering UPM), we focus on the present work in assessing the competence, "Teamwork. "

The teaching of computer programming has received extensive attention in recent years and a good overview of the most important work was done by Robins et al [11]. PL is a mandatory course, 3 face-to-face lectures a week, and is taught in the first semester (15 weeks) of the first academic year. The teaching is organized in three groups of varying sizes (between 40 and 70 students per class attending class regularly.) The language chosen for implementation of examples and exercises is the shell language MATLAB (or Octave free version). For the justification of that choice as well as for documenting other aspects of the evaluation we refer to reference [16].

In this paper we describe first the mechanisms that we used to enhance teamwork. Then we discuss the technique used for the evaluation of that work. Later we will analyze results, trying to characterize the impact that the methodology has had on different aspects of the educational intervention. Such analysis will be supplemented with a discussion of the problems we have encountered in implementing it. Finally some conclusions and some threads of future work will be summarized.

2 METHODOLOGICAL ASPECTS

2.1 Teaching method

In a Programming course in a non computing Engineering course, we think the students ought to have a very active role, experience, check, correct, perfectly described coded algorithms, consider third-parties codes, develop strategies to solve problems, etc. The difference with a Programming course in a computer degree is that it is not reasonable to allocate time to explain the fundamentals of Programming theory and Languages. In short, our goal is that students "do". In order to achieve this goal, classes take place entirely in the computer room, without separation of schedules between theory and practice. This way, the assimilation of fundamental concepts will take place as the students work directly in the computer.

As regards the organization of groups, for reasons of space, time of exposure to programming environment and the limited number of teachers who are available to supervise and assist in implementation of programs, distribution in the classroom is done in pairs, and Pair Programming technique [19] is promoted. To our knowledge, pairing students as a teaching methodology has not received much attention in literature. It has been discussed how pairing students at school and high school level can enhance their performance [9,12], but we have not found quantitative measurements nor references regarding university level. It is well known that as groups get larger usually one or two members of the group just do not meet the standards. Since we are dealing with first semester first year students, it is sensible to minimize eventual problems in these regards and therefore pairing them introduces teamwork competency without the risks involved with managing larger teams.

Students receive information about the organization of the course in the first face-to-face lesson. They can buy a hardcopy or download a PDF of the course guide [17], which contains notes, exercises, previous editions exams and all the information about the course objectives, resources and evaluation system. In particular, it is remarked that they have to organize themselves in pairs. To do this, we leave three weeks for students to meet and make friends (since they are newcomers to the university). During those three weeks, the importance of making a responsible choice for the couple, given the relevance that such choice has on course development and evaluation, is highlighted during the lessons. The issue on how groups are formed is relevant and other options have been explored. Hsiang [8] forms teams of 3 members by splitting the whole group in 3 subgroups depending on former semester grades. The groups have one member of each subgroup. In our specific case, this being a first semester first year subject, we think the most sensible option is to offer the students the possibility of organizing themselves in regards to choosing partnerships.

From third week couple arrangements are annotated by the lecturers. It is allowed, during the initial part of the course, to break up and swap partners with other couple or engage with a lone student.

As will be later discussed, the paired students joint work assessment is based on proposing quizzes throughout the course. To promote teamwork, at least one of these tests is done individually but the mark of the pair is the smallest of the grades of its members. Students do not know how many of these tests will be given. This way, we intend to motivate an evolution of the couple performance

throughout the course in which the stronger member will support the weak one fostered by its own interest.

Finally, as an additional aspect to motivate students to work together, it was arranged that an industry leader in the development of software applications for shipbuilding, SENER, sponsored a prize of 1200€, to be granted to the couple that presents a larger progress between the ECTS mark (in which the paired students work has a significant weight) in PL and the high school average mark, with which, the student applies competitively for accessing certain university degrees.

2.2 Assessment

Assessment is an essential part of the educational intervention, and probably characterizes it more than any other [14]. The assessment in PL is centered in the process [1], and evaluated items and their respective weights in the global mark have evolved since the 2004-2005 academic year, when this methodological change started. We will document the settings for the academic years 2008-2009 and 2009-2010, for which there are enough information for the analysis of the paired students work assessment procedure.

During 2008-2009, assessment included the following items, with their corresponding weight in the ECTS grade [17]:

1. Paired students work: 35%.
2. Final Exam: 35%.
3. Active participation in the lessons: 15%.
4. Participation in the B-learning platform discussion group, 5%.
5. Suggesting coding exercises, 5%.
6. Survey [18] (3%) y and debate in which survey outcomes are discussed (2%).

During 2009-2010, the item number 5 was eliminated and its weight was transferred to the paired students work. The reason for such change was that, as lectures, we could not use more time in assessment activities. Moreover, work in the competence of design problems them was not a core part of this introductory course. Therefore, in 2009-2010, the grade has had the following components [17]:

1. Paired students work: 40%.
2. Final Exam: 35%.
3. Active participation in the lessons: 15%.
4. Participation in the B-learning platform discussion group, 5%.
5. Survey [18] (3%) y and debate in which survey outcomes are discussed (2%).

When implementing this assessment procedure, we find several problems. First, the students have some extra calls apart from the regular year one, in which they have the right to pass the subject by taking just one exam. Second, there were cases in which students could not attend a specific test due to a wide range of reasons, with or without adequate justification. For these two reasons and in order to avoid quarrelling with the students, we felt appropriate to maintain the final examination as sufficient evidence to pass the course. Therefore we decided that all students (regardless they would follow ECTS itinerary or not) always have the option to get the grade from the final exam. The mark would be hence the maximum of the ECTS grade and the final exam grade.

To evaluate the item "Paired students work", several exams (no more than 4 in general) throughout the course were set, during regular face-to-face sessions. One advantage of this type of exams is its simplicity, since they do not involve changes in classroom organization and can be performed during a regular lesson. These exams are not announced in advance in order to encourage couples to attend regularly and in order to promote that the students study the subject regularly and not just for specific tests. These exams consist in the implementation of several short exercises, similar to those proposed during regular lessons. An example is now presented and you can see the course Guide [17] for more examples.

"*Implement a function which has as input a number n (natural and larger than 1 by hypothesis) and returns as output a vector with the primes that are divisors of n (including n). It is not allowed to call auxiliary functions. For instance, if n=12, then the returned vector will be [2,3], which are the primes that divide 12. If n=7, then the function will return [7].*"

As can be seen, the exercises are not difficult but are not trivial either. They all come with an example in order to clarify what is requested.

The large weight (40%) that these tests have in the assessment procedure encourages students to take them seriously. We took advantage of this extra motivation to perform normally two "mock test", which are not corrected. Since the scenario mimics a real test, the motivation is high and the students do their best to solve the problems, which thrusts the learning pace.

The majority of the tests are performed in pairs. The bustle in the computer room is intense and the sensation of seeing couples working together is very uplifting. If a member of the couple misses the lesson when the exam is set, the grade of both members is zero. These situations can create conflicts between partners, whose educational value is indisputable [4]. To give a realistic character to the "mock tests" and encourage students to generate mechanisms to deal with these conflicts, "mock tests" are identified as so only the next day.

At least one of these "paired students work" examinations is set individually. In this case, the grade that corresponds to the pair is the smallest of the grades of its members. The possibility of taking individual test was once suggested by students in the debate, since it is a good training for the final exam. We took this opportunity to increase the percentage allocated to "students work" since we were sure that if the mark is good, then the two members of the pair have acquired enough skills. This way, we intend to link longitudinal with transverse competencies. Progressing in Programming learning implies progressing in "Paired students work" competence and vice versa. Computer Programming involves a limited number of concepts and it enables students to easily identify cooperative work actions needed to progress. In the next section we try to quantify, through a series of indicators, to what extent all this methodology has impacted on the intervention.

3 RESULTS

3.1 Analysis of results

The first indicator on the impact of this methodology relates to how it has influenced attendance to face-to-face lessons. We have aggregated information obtained from all subjects taught in the same semester by the same students that took our course during academic year 2008-2009. During that year an extensive analysis of attendance to lessons for all the subjects for first year students was conducted. The results are shown in figure 1. It can be seen that PL, the focus of this study, is the subject that has the highest percentage of students who always attend the sessions and has the highest ratio with respect to those that attend most of the time and even with respect to those that attend occasionally.

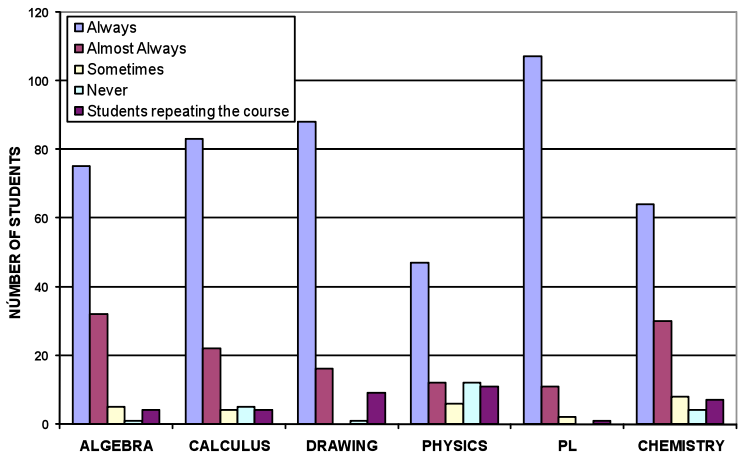

Figure 1. *Attendance for 1st semester courses.*

Interesting indicators can be developed by sampling the views of students from the anonymous survey [18] performed at the end of the course, after they know the final score, which was held by the majority of students. That survey contributed with 0.3/10 points to the mark and we kept that contribution for the extra calls that the students are granted after the regular one, so that everybody would feel compelled to fill the survey. That survey included the following questions:

- "*The aim of the paired student's exams is to encourage teamwork, by trying to make evident to the students the commitment that teamwork demands.*

    1. *To what extent (0 to 10) do you think the mark you have got matches the amount with which you have achieved those objectives (0 indicates strong disagreement, 10 indicates full agreement)?*
    2. *The weight in the global weight is 35% (2008-2009) or 40% (2009-2010). What do you think the weight should be?"*

In the year 2008-2009, the average values were 5.8 and 26% respectively. For the first question, the students that passed the course responded on average 6.7 while those who failed responded on average with 4.7.

In the year 2009-2010, the average values were 5.5 and 34% respectively. For the first question, the students that passed the course responded on average 5.7 while those who failed responded on average with 5.24, more balanced than in the previous year.

It can be deduced from these values that students are not very satisfied with this way of assessing the competence and believe it has too much weight in the grade. However, in the same survey they do not mention this way of assessment as a possible cause of failure in the subject.

3.2 Marks analysis, year 2008-2009

As regards the analysis of the grades, we have taken as a reference the average marks of the examinations performed jointly and compared them with the individual marks, both of the exam during the regular course and the final exam. We removed from the analysis those couples who lacked any of the marks, so that the comparison was consistent (in total, we have complete information of 54 pairs). We intend to see if there is a positive drag effect within the couples, by comparing the individual marks in the exam during the regular course (we will refer to this exam as the individual exam) and the final exam, after 16 weeks of joint work within each couple. For each student, the following variables are defined:

- I individual exam during the regular course mark.
- F final exam mark.
- P average marks of the joint exams.
- MI for each student, this is the maximum of the marks between its own mark in the individual exam (I) and the mark of his/her partner in the pair.
- MF for each student, this is the maximum of the marks between its own mark in the final exam (F) and the mark of his/her partner in the pair.
- LI for each student, this is the lowest of the marks between its own mark in the individual exam (I) and the mark of his/her partner in the pair.
- LF for each student, this is the lowest of the marks between its own mark in the final exam (F) and the mark of his/her partner in the pair.

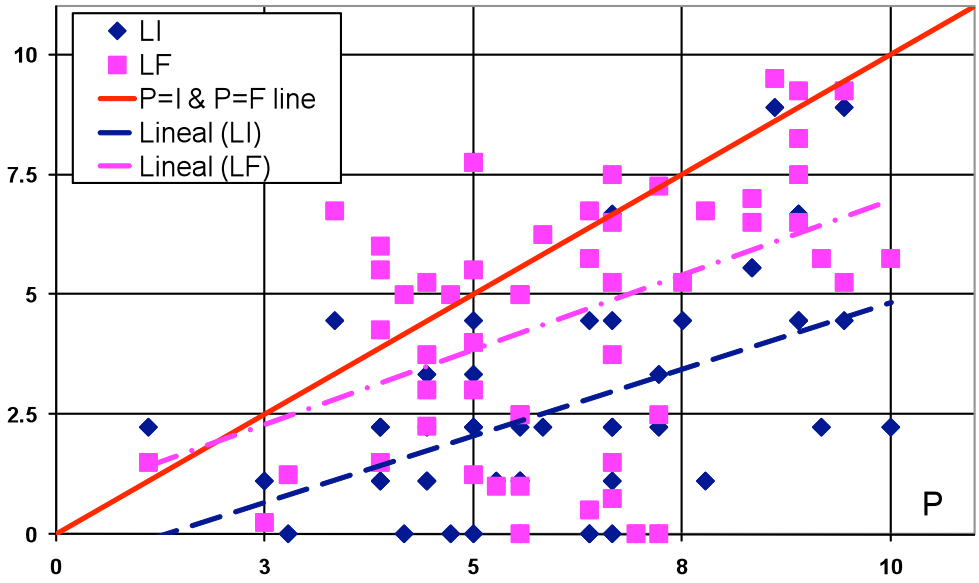

Figure 2. *Year 2008-2009. Average couples joint mark* P *(abscissa) vs. minimum mark in the individual and final exams.*

We present these data in figures 2 and 3. Analyzing the data illustrated by these figures we find that:

1. Among the 16 couples whose joint mark P is equal to or greater than 7 points, there is no couple in which the two members fail the final exam. This could be interpreted that joint work is a good motivator for individual achievement, which is one of the thesis or this paper.
2. Among those 16 couples, in 8 cases (50%), one of the members fails the individual exam but later passes the final exam. We think that this may be a sign of the progress during the course due to the joint work.
3. Among those 16 couples, in 6 cases (37.5%) the two members pass all the exams. Those cases could correspond to the best students, which do not contribute much to our analysis.
4. Among those cases, in 2 cases (12.5%) one of the members fails the individual exam and fails also the final exam. These two groups could be considered dysfunctional since the cooperation has not been effective in dragging the bad student to pass the final exam.
5. Among the 24 couples whose joint mark is between 5/10 and 7/10, in 13 cases (54%), one of the members fails the individual exam and fails also the final exam. The amount of dysfunctional teams is therefore larger for the segment of couples that pass the joint exam but with lower marks.

From these remarks, we think we can infer that the “drag” effect is larger when the joint mark is large but not as important when the joint mark is enough for passing but not large (5<=mark<7).

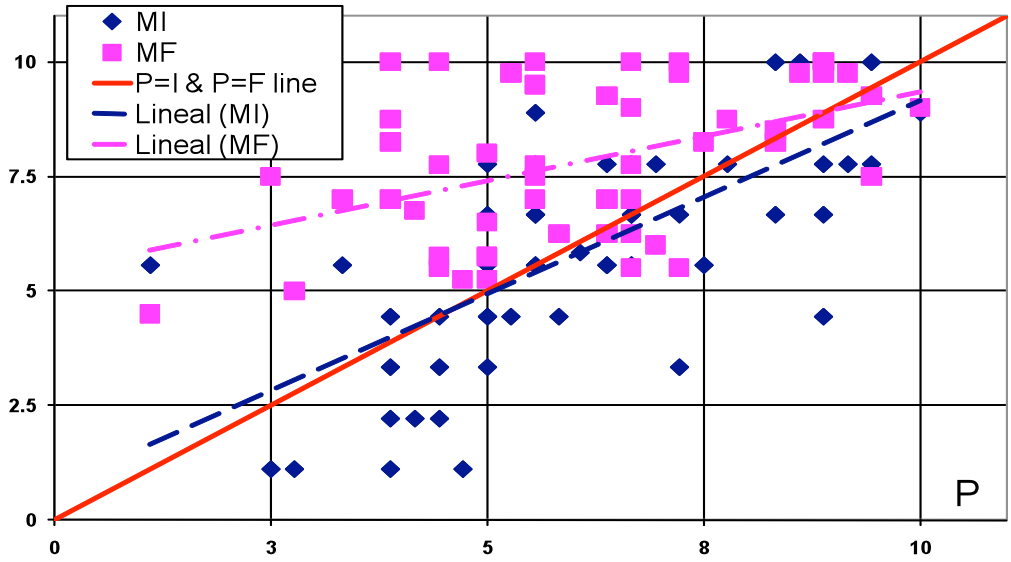

Figure 3. *Year 2008-2009. Average couples joint mark* P *(abscissa) vs. maximum mark in the individual and final exams.*

3.3 Marks analysis, year 2009-2010

In 2009-2010 the number of couples for which full information is available has reduced to 23. The marks are presented in figures 4 and 5. Analyzing this data we find that:

1. Among the 10 pairs whose joint mark P is equal to or greater than 7/10, there are two pairs (20%) in which the two members fail the final exam.

2. Among those 10 pairs, in 1 case (10%), one of the members fails the individual exam but later passes the final exam.

3. Among those 10 pairs, in 3 cases (30%) the two members pass all the exams.

4. Among those 10 pairs, in 4 cases (40%) one of the members fails the individual exam and fails also the final exam. These pairs could be considered dysfunctional since the cooperation has not been effective in dragging the bad student to pass the final exam.

Compared to year 2008-2009 the tendencies are less evident and the need for better indicators appears as crucial for future work.

As a final remark regarding the analysis of results we must highlight indicate that our experience as teachers is positive. The impression we have is that students enjoy this type of assessment. In this sense the experience is very valuable, because as shown by Gardner [6], most often the assessment is associated with suffering, and in this case, these exams encourage a positive dynamic in face-to-face lessons and in the relationships between students.

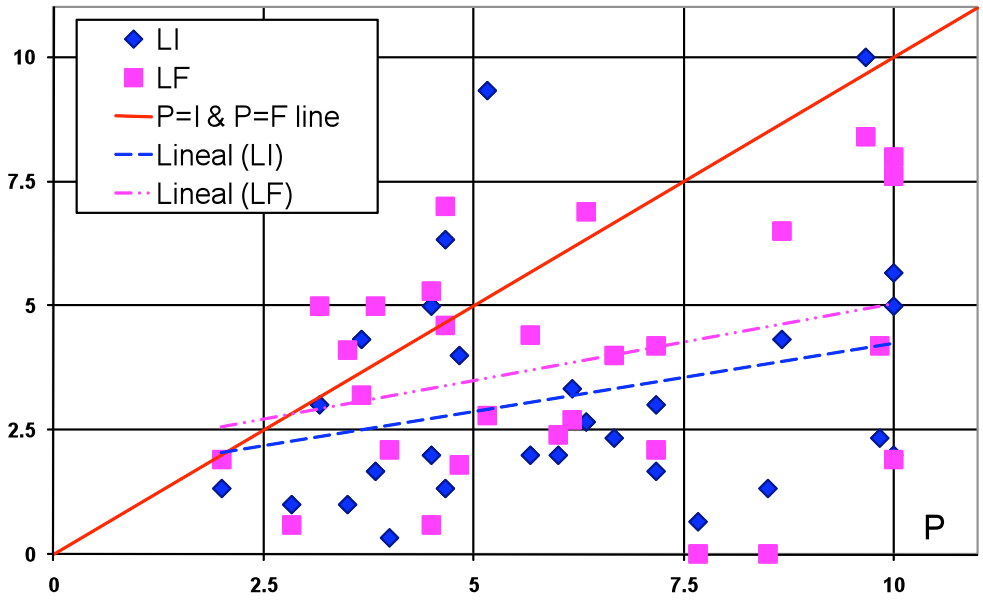

Figure 4. *Year 2009-2010. Average couples joint mark* P *(abscissa) vs. minimum mark in the individual and final exams.*

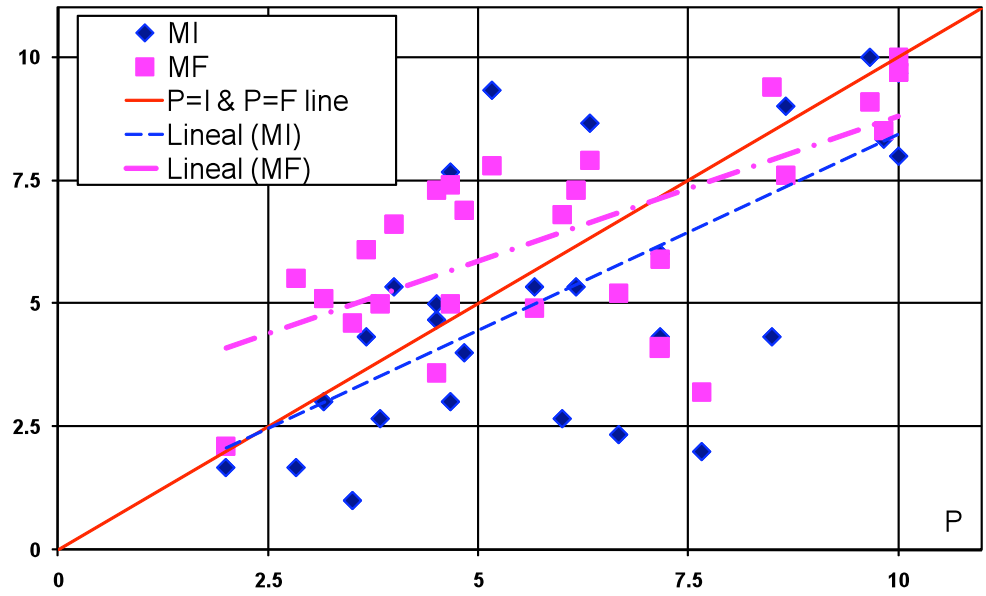

Figure 5. *Year 2009-2010. Average couples joint mark* P *(abscissa) vs. maximum mark in the individual and final exams.*

3.4 Problems description

Although overall positive, the experience is posing problems that should be listed and which is necessary to reflect upon.

1. During the formation of couples, it may be that if the number of regular attendees to classes is odd, someone is left without partner. Our response in this case has been to urge the lonely student to try to convince some of those who rarely attend class to attend. This problem is usually arranged by itself after other eventual abandonment and therefore it is not a big problem in the practical application of the methodology.
2. The students whose couple does not attend the day an exam is given feel discriminated since their mark will be 0/10. Even if there is a justified cause, it is not viable to repeat those exams due to the amount of extra work it would mean for the lecturers.

3. To lessen the impact of these situations, we leave always the final exam as a possibility to pass the subject. Notwithstanding this available exit, we insist that the smart attitude is to get to the final exam with already the pass in the pocket. This possibility, as explained in section 2 is necessary to make the subject compatible with the existence of mandatory extra calls, justified absences due to an illness or any other major problem, for which other test is mandatory by the chief of department, and finally due to the existence of students that take the subject for the second year and for which an exam is the simplest mean for passing the subject.
4. Correcting and marking all these tests means a large extra work. Usually it takes us much more time than what it would be convenient to have them corrected and marked which is bad news in order to provide prompt feed back to the students in regards to their achievements. We have to explore using multiple choice (MC) tests at least for some of the exams. This possibility has received some attention in the literature showing that the ability of MC tests in introductory programming courses is a delicate issue [*3*,*10*,*15*].

4 CONCLUSIONS

We have discussed in this article how we have assessed the competence "Paired students work" in an introductory course to computers programming in a non-computing oriented Engineering degree. We have attempted to document, in relation to this competence, how the course has been methodologically articulated, how the assessment has been performed and which outcomes have been obtained. One possibility we believe that has not been sufficiently explored and paid attention to is that of examinations in groups that contribute to the individual mark of each component. We have implemented an assessment of paired students work in which a significant part of the mark of the two members of the pair is averaged between joint exams marks and the minimum of the marks obtained by the members of a pair in exams taken individually. This way we aim at promoting a responsible behavior within each pair, at linking the transversal competence (team work) with the longitudinal ones (computing programming coding, validating), and at engaging with the students, trying to reduce abandon rates, which are especially high for 1st year students.

We have aimed at developing indicators that seek to assess the efficiency of the methodology in areas such as attendance and academic progress. We have not been able to describe a clear correlation between the variables put in place, which is reflected in the substantial differences between the outcomes in years 2008-2009 and 2009-2010. It is interesting as well to investigate the behavioral implications of this experience, which according to the comments from the students have been quite significant.

Notwithstanding this, we think the experience has been very positive, with significant indirect impact on attendance at face-to-face lessons and we believe that a significant impact also on academic progress. Added to this, observing couples working together during the joint exams, discussing and implementing the codes, has become a very rewarding pedagogical experience.

It remains as future work to study longer time series and to develop good indicators in regards to the success of the methodology. It is relevant to question whether this methodology, which relies on exams carried out jointly or individually but which impact uniformly the grades of the team members, is applicable in other contexts. It is interesting as well to reflect on how this methodology can be applied to larger teams. In addition, other aspects of the competence “Team work” deserve attention, as conflicts management, balance of task assignments within the group, leadership, etc

REFERENCES

[1] Adiego-Rodríguez, J., Vivaracho-Pascual, C. E., 2005, Una experiencia de AC en Fundamentos de Informática en Ingeniería Técnica Industrial. Quinta Jornada sobre Aprendizaje Cooperativo del GIAC., Universidad de Deusto. (in Spanish)

[2] Alonso, C. E. , Fernández-Combarro, E., Montaña, J. L. , 2008, Una propuesta genérica de trabajo en grupo y su aplicación en una asignatura de fundamentos teóricos de la Informática. XIV JENUI, Universidad de Granada. (in Spanish)

[3] Clark, D., 2004. Testing Programming Skills with Multiple Choice Questions. Informatics in Education 3 (2), 161-178.

[4] Del Canto, P., Gallego, I., López, J.M., Mora, J., Reyes, A., Rodríguez, E., Sanjeevan, K. , Santamaría, E., Valero, M., 2009, Conflictos en el trabajo en grupo: dos casos representativos. XV JENUI. UPC. (in Spanish)

[5] Erice, C., 2005, Estrategias de aprendizaje en la asignatura Matemáticas I de la Diplomatura de Ciencias Empresariales. ASEPUMA, XIII Jornadas. Universidad de Illes Balears. (in Spanish)

[6] Gardner, H., 1991, Assessment in context: The alternative to standardized testing, Changing assessments: Alternative views of aptitude achievement, and instruction, Ed. Kluwer Publishers.

[7] Hellriegel, D., Slocum, J.W., 1997, Managing, Cengage Learning, USA.

[8] Hsiung, C. M., January 2010. Identification of Dysfunctional Cooperative Learning Teams Based on Students' Academic Achievement. Journal of Engineering Education 89, 45-53.

[9] Knight, D., Pairing vs. Small Groups: A Model for Analytical Collaboration, The Teaching Professor, February 2007.

[10] Kuechler, W. L., Simkin, M. G., 2003. How Well Do Multiple Choice Tests Evaluate Student Understanding in Computer Programming Classes? Journal of Information Systems Education 14 (4), 389-399.

[11] Robins, A, Rountree, J., Rountree, N., 2003, Computer Science Education, Vol. 13, 137.

[12] Rowell, P. M., 2002. Peer Interactions in Shared Technological Activity: A Study of Participation. International Journal of Technology and Design Education 12, 1-22.

[13] Sánchez-Maroño, N., Fontenla-Romero, O., Bellas-Bouza, F., 2006, Aportaciones e ideas para el rediseño de la asignatura de Fundamentos de Informática al EEES. JENUI. Universidad de Deusto. (in Spanish)

[14] Santos-Guerra, M.A., Tendencias pedagógicas, no. 6, (2001), 89. (in Spanish)

[15] Simkin, M. G., Kuechler, W. L., 2005. Multiple-Choice Tests and Student Understanding: What Is the Connection? Decision Sciences The Journal of Innovative Education 3 (1), 7.

[16] Souto-Iglesias, A., Bravo-Trinidad, J.L, 2008, Rev. Educ., 346, 487, (in Spanish)

[17] Souto-Iglesias, A., Bravo-Trinidad, J.L., González, L., Cantón, A., 2009, LP. Guía ECTS, ETSIN, UPM. http://canal.etsin.upm.es/ftp/GUIALP0910.pdf. (in Spanish)

[18] Souto-Iglesias, A., Bravo-Trinidad, J.L., Pérez de la Portilla, M., Cantón, A., 2010, Encuesta de LP, ETSIN, UPM. http://canal.etsin.upm.es/ftp/encuesta0910.htm (in Spanish)

[19] Williams, L, Kessler, R., 2003, Pair Programming Illuminated, Addison Wesley.