*Article*

# Using Peer Review for Student Performance Enhancement: Experiences in a Multidisciplinary Higher Education Setting

Juan Jose Serrano-Aguilera [1], Alicia Tocino [2], Sergio Fortes [3,†], Cristian Martín [4,†], Pere Mercadé-Melé [5,*,†], Rafael Moreno-Sáez [6,†], Antonio Muñoz [4,†], Sara Palomo-Hierro [5,†] and Antoni Torres [7,†]

1   Escuela de Ingenierías Industriales, University of Malaga, Campus de Teatinos s/n, 29071 Málaga, Spain; jj.serragui@uma.es
2   Algebra, Geometry and Topology Department, Faculty of Sciences, University of Malaga, Campus de Teatinos s/n, 29071 Málaga, Spain; alicia.tocino@uma.es
3   Departamento de Ingeniería de Comunicaciones, Universidad de Málaga, Andalucía Tech, Campus de Teatinos s/n, 29071 Málaga, Spain; sfr@ic.uma.es
4   Computer Science Department, University of Malaga, Campus de Teatinos s/n, 29071 Malaga, Spain; cmf@lcc.uma.es (C.M.); amunoz@lcc.uma.es (A.M.)
5   Department of Applied Economics (Statistics and Econometrics), University of Malaga, Campus el Ejido s/n, 29071 Malaga, Spain; sara.palomo@uma.es
6   Departamento de Tecnología Electrónica, Universidad de Málaga, Campus de Teatinos s/n, 29071 Malaga, Spain; rmorenosaez@ctima.uma.es
7   Department of Mathematical Analysis, Statistics and Operations Research, Applied Mathematics, Faculty of Sciences, University of Málaga, 29071 Malaga, Spain; a.torres@uma.es
*   Correspondence: pmercade@uma.es
†   These authors contributed equally to this work.

**Abstract:** Nowadays one of the main focuses of the Spanish University system is achieving the active learning paradigm in the context of its integration into the European Higher Education Area. This goal is being addressed by means of the application of novel teaching mechanisms. Among a wide variety of learning approaches, the present work focuses on peer review, understood as a collaborative learning technique where students assess other student's work and provide their own feedback. In this way, peer review has the overarching goal of improving the student learning during this process. Peer review has been successfully applied and analyzed in the literature. Indeed, many authors also recommend improving the design and implementation of self and peer review, which has been our main goal. This paper presents an empirical study based on the application of peer review assessment in different higher education BSc and MSc courses. In this way, six courses from different studies at the University of Malaga in Spain are subject to the application of peer review strategies to promote student learning and develop cross-wise skills such as critical thinking, autonomy and responsibility. Based on these experiences, a deep analysis of the results is performed, showing that a proper application of the peer review methodology provides reliable reviews (with close scores to the ones from the teacher) as well as an improvement in the students' performance.

**Keywords:** peer assessment; peer review; collaborative evaluation; higher education; rubric

## 1. Introduction

The integration of the Spanish university education system into the European Higher Education Area (EHEA) has entailed a paradigm shift in the teaching-learning process, leading towards a student-centred learning (SCL) approach where the learner and their needs are the primary focus. Under this perspective, active collaboration replaces passive knowledge transmission, with both teachers and students becoming mutual active contributors to the education process. Thus, successful mutual responsibility becomes critical in enabling the development of the learners' autonomy [1]. In Spain, this change in

the teaching-learning paradigm (or from the system of credits based on teaching hours to workload) has entailed a conceptual change in the higher education system [2,3].

The transition towards SCL has required the adoption of different or tailored forms of learning, resources and evaluation [4,5], so the transmission of technical knowledge and the methods of assessment and evaluation are aligned with competence development. Evaluation systems are one of the issues which have been most affected by the convergence towards SCL since they are one of the essential elements in the planning and execution of the teaching-learning process [6]. As indicated by Rodriguez-Esteban et al. [7], today evaluation is not understood only as a final act, being rather a process that is part of the same learning system. Not only must they serve to accredit learning, but they must also help students to learn and teachers to improve their teaching [8].

In this context and among a wide variety of learning approaches, *peer review* stands out as a key tool for SCL. Peer review, applied to the education field [9], is a collaborative learning technique where students assess other student's work and provide their own feedback with the overarching goals of improving the student learning process during this process, enhancing the understanding of the peer's work (maybe with a different approach) and improving the quality of the final product [10]. It is also referred to in the literature as peer evaluation, peer response, formative peer assessment or peer editing [11]. From now on we consider the term *peer review* in this article. This technique aims to involve students in the evaluation process, allowing a review of the material from a critical perspective, so that, when they analyze different approaches and points of view that may not have been considered when performing their task, students will have a more global vision of their learning in the course. During this collaborative process, the transfer of knowledge is mutual: in those who give the feedback and those who receive it, since this activity triggers critical reasoning and self-assessment in both sides [12]. Rubrics, as identified by Hafner et al. [13] are normally used. These provide a common scheme (usually prepared by the teacher in collaboration with the students) for assigning marks in each step of the assigned task, guiding students during this process.

### 1.1. Related Work

Peer review has proved to be a powerful evaluation strategy that brings a number of advantages over classical methodologies, such as facilitating the students' acquisition and development of skills and promoting their capacity of self-direction of their own learning [14]. Previous studies as Nicol et al. [15] have shown that, during the review process and the generated feedback, the evaluative judgment of the students is improved, both about their peers work and their own work. The analysis of other perspectives also greatly contributed to provide them with a deeper understanding of the course material [16]. A growing ability to give constructive feedback during peer review has also been shown in first-year students [17], who may be unaware of professional standards and expectations and reluctant for critiquing work and writing thoughtful feedback. Promoting independent learning, increasing student motivation, building problem-solving skills are also notably advantages to remark in this learning process [18].

Peer review has been widely applied and analyzed in the literature. In Saiz et al. [14], peer review at university is analyzed, highlighting the characteristics of this strategy and the conceptual, institutional and relational difficulties of its implementation. Challenges and benefits of the implementation of self and peer review and identifying potential inhibitors in practice are explored in Adachi et al. [19]. Moreover, authors also make recommendations about improving the design and implementation of self and peer review. Amendola and Miceli [12] propose a peer review methodology completely conducted through online technologies (in particular using the Moodle e-learning platform) showing the benefits such as lack of space and time in the standard lesson environment applying this methodology. In Indriasari et al. [20], a survey of peer review of source code in higher education showing how such activities have been implemented in practice, examining instructor motivations and the primary benefits and difficulties of this practice have been

reported. Authors also identify a wide variety of tools to facilitate the peer code review process. Peer review has been been widely adopted by major massive open online course (MOOC) platforms, but there is little evidence about if it is appropriate or under what conditions. In Meek et al. [21], student performance, participation and opinions of a peer review task of a science course in a MOOC are examined. According to Reddy et al. [22], the positive impact of training in peer review learning experiences on science students over three years in higher education is demonstrated. In Gaynor [23] the quality of peer feedback, the importance of assessments and student perceptions are investigated.

Apart from the aforementioned topics, there are other issues which have also been previously addressed by the peer review literature, such as the influence of cultural perspectives [24,25] or the analysis of students perceptions [26,27]. Moreover, Panadero et al. [28] investigated the impact of friendship on students scoring, finding a positive impact of using rubrics on scoring objectivity. However, the overscoring effect generally observed among students was amplified when the use of a rubric was accompanied by a high level of friendship between assessor and assesses. The influence of peer-related factors such as gender on peer-awarded marks is studied by Lagan et al. [29]. They observed the presence of a slight and positive gender effect between participants of the same gender over participants of different gender, with female evaluators being more consistent at awarding marks. Another branch of the peer review literature focuses on the assessment of the effectiveness and reliability of the use of this technique for educational purposes [30,31].

Within the Spanish context, the literature also provides some experiences in applying peer review to the higher education context. For example, one of the first documented experiences was conducted by Sánchez Rodríguez et al. [32] in an Education Science course at the University of Malaga. They found that although students' scores were slightly lower and more concentrated than those awarded by teachers, both were strongly correlated. Regarding students' perceptions about the vast majority of the respondents showed a positive attitude about the possibility of valuing and being valued by peers. Conde et al. [18] applied peer review to technological courses at the University of Leon to help students in technical studies to develop specific abilities such as critical thinking and get more involved. They evaluated student's opinions and performance and found that the application of peer review increased students participation and led to higher scores. Moreover, also by comparing students and teacher scores, they observed a significant and strongly correlated relationship. The quantitative data analysis about students' perception showed that, in general, the level of satisfaction regarding the methodology's appropriateness and its beneficial role in acquiring critical thinking is elevated. Only first-year students showed some concern regarding its use. Finally, it is also worth mentioning the study of Dopico [33] at the University of Oviedo, who found that emotional aspects related to personal criteria other than prearranged ones seemed to intervene in peer review processes. Besides, he also observed that the use of digital tools for the peer review exercise enhanced the motivation of the students.

### 1.2. Contributions of This Paper

Despite these clear advantages, there are very few courses at the University of Malaga (UMA) in Spain that apply strategies of self-assessment, coevaluation or peer review methodologies. The fact of identifying errors of peers or own mistakes during evaluation, provides a stable critical base to continue the construction of knowledge, while giving confidence to students on the skills acquired with a strengthening of them. In the literature [34,35] it is also mentioned some factors that limits the adoption of this technique such as: (1) not having the necessary maturity to evaluate, (2) not taking the evaluation seriously, (3) having negative attitudes towards an evaluation of these characteristics and (4) considering the evaluation as an additional load.

Although all these studies give a valuable insight into the use of peer review for evaluation and assessment in higher education settings, each of them performed the peer review for the evaluation of different types of assignments and in different scientific fields. Starting from these efforts, and by adding to the literature, the study described in this

article aims to gain more insight into the possibilities of formative peer review and its application in the classroom, setting the first steps towards continuing the much-needed research on the use and performance of peer-assessment methodology in the future. Thus, in order to assess the effectiveness and reliability of peer review, this study aimed to answer two key questions:

- Does the use of peer assessment improve students learning?
- Is peer review an effective and reliable method for evaluation?

By performing this peer review methodology we have also aimed at the development of certain transverse competences in students, such as responsibility and objectivity when making decisions. The study has been developed in six STEM (stands for science, technology, engineering and mathematics) courses at the University of Malaga, Spain. Students perform peer review of one or more tasks in these courses, such as the resolution of an exercise, the development of a report or the presentation of a project, which are carried out on the basis of similar statements and/or specifications with well-established rubrics. This methodology allows us to strengthen those concepts that teachers consider key in each of the courses using peer review learning processes based on error detection and observation of the different approaches that students expose when solving problems.

### 1.3. Structure of the Paper

The rest of the article is organized as follows. Methodology and developed experiments are detailed in Section 2. In Section 3 results are presented and discussed. Finally, our conclusions and future work are presented in Section 4.

## 2. Methodology

The methodology consists mainly of peer review evaluation of projects, class tests, works, activities, in addition to oral presentations. We have considered the case in which students know who they are evaluating [28]. One of the key objectives in the development of this activity is to ensure that students are able to make value judgments on the work done by their peers according to established criteria, in order to improve their degree perception of achievement of the subject. In order to make this possible, students need to have assimilated knowledge and be aware of what they are trying to evaluate.

The fact of knowing the students on whom evaluations are going to be made can lead to biased scores, depending on the affinity or not with the person being evaluated. It may even be that the rating is influenced by the possible consequences that may occur in personal relationships [28]. Therefore, in this case where the evaluated person is known, it becomes very necessary to have clearly defined and delimited the evaluation criteria, with specific headings for the qualification. In addition, it is convenient that there is an established weighting, where the qualification of the students influences the final grade of the task, but it is not determinant. In this line, we have not considered students weighting as we are on pilot experiences, but it is considered as ongoing work. However, the fact that student assessments have a direct impact on the score is a change in their usual role. It is important that students are aware that this activity prepares them, among other benefits, for the relevant critical thinking development in their upcoming professional life, so that without their involvement, the activity is meaningless.

### 2.1. Research Context

This study has been developed in the framework of a Teaching Innovation Project (PIE19-209) (https://www.uma.es/formacion/noticias/proyectos-de-innovacion-educativa-2019-2021/, accessed on 30 December 2020) grant funded by the University of Malaga. It has been carried out by a team of assistant professors from different STEM departments at the University of Malaga that have met with a common objective, as different weaknesses were recognized in particular learning points that we consider to be essential in our teaching activities. After researching and documenting it, we identified the peer review approach as an invaluable tool that could be developed in the context of the teaching

innovation call for projects that our University opens every two years. Thus, we decided to elaborate a proposal and this was submitted for evaluation. Then this project was granted one and a half years ago, although our activity on it goes back to January 2018. In order to assist the reader, details of our experiences are shown in Table 1.

**Table 1.** Summary of experiences.

| Nr. | Type | Course | Level | Degree | # | Type of Task | Modality | Grouping | Result Analysis |
|---|---|---|---|---|---|---|---|---|---|
| Experience 1 | I | Linear Algebra and Geometry | BSc | Mathematics and Computer Science | 77 | Worksheets | On-site | 3–5 members | Final test score improvement |
| Experience 2 | I | Algebraic Structure | BSc | Mathematics and Computer Science | 67 | Worksheets | Online | 3–5 members | Final test score improvement |
| Experience 3 | I | Statistic II | BSc | Marketing and Market Research | 214 | Worksheets | Online | Individual | Final test score improvement |
| Experience 4 | II | Wireless Networks | MSc | Telematics and Telecommunication Networks | 17 | Project Oral presentation | On-site | 2–4 members | Slight score deviation |
| Experience 5 | II | Renewable Energies | BSc | Energy Engineering | 23 | Report review | Online | 2–4 members | Slight score deviation |
| Experience 6 | II | Stochastic Models | BSc | Mathematics | 11 | Project Oral presentation | On-site | Up to 3 members | Slight score deviation |

In total there have been 409 students involved in this activity, distributed among the different courses offered by the University of Malaga. This provides a key multidisciplinary character to this work, starting from a relevant sample of students and applying peer review to a varied range of topics.

A total of six STEM courses have been chosen for this study. The peer review methodology had to be adapted to the nature of each of the subjects as well as to the preset teaching plans without disrupting the natural flow of the course. A different instructor is in charge of every experience since they belong to different STEM departments in the same institution (University of Malaga).

This work was motivated by the recurrent difficulty of a considerable proportion of students to acquire the necessary skills and thus solve problems that involve minimum requirements to pass the subjects. In this context, peer review is considered as a reinforcement activity that improves the way in which objectives are achieved. Additionally, given the wide range of possible solutions to science and engineering problems, students can propose different approaches that do not have to be strictly provided by the teacher. We believe that the peer review methodology can help the students to reach different solutions and understand multiple valid perspectives to address the presented challenges.

Finally, it has been observed that on many occasions it is easier for students to understand what their peers want to communicate to us than what the teacher is telling them. That is why a source to reinforce knowledge and contribute to their learning process has to be provided by their peers and we have tried to design a methodology to take advantage of this mechanism.

We propose students that voluntarily participate in this project to correct a series of activities of other classmates. In this way, they will review exercises using a detailed rubric provided at the time of carrying out their own exercises, so that they know the evaluation criteria. The evaluation of the students to their peers will be assessed according to the approximation of their scores to their peers with respect to that given by the teacher. Thus, we motivate the dedication on the part of the students. Moreover, it has to be noticed that this methodology provides a very interesting source of evaluation in the post COVID-19 environment as it provides data for a more enriched and constructive evaluation.

To sum up, we have conducted a series of experiments that contrast how the use of this methodology affects different scenarios with different conditions. In some cases, all students have taken part in the peer review voluntarily and in other cases there are groups in which it has been applied and others in which it has not.

*2.2. Research Method*

The method applied in this study has been divided into three well-differentiated steps in two types of experiences consisting of:

- Data collection: In this step, data from the students has been gathered. These data are of two different types. First, we used the numerical data of the student's score, both data of final score and data from each of the exercises related to the proposal of our project (peer review). In the Spanish educational system a 10-point scale is imposed, being 10 the maximum score. We have followed it, but it can be applicable to any other scale. Secondly, data has been extracted from the MSc and BSc degrees that the students who have voluntarily taken part in the project give to the peers they are reviewing. The students who take part in the project must provide scores to their classmates, in some cases using a rubric provided by the instructors. In this way, the students are always knowledgeable of the rubrics and their associated items and how they are going to be evaluated. Selection bias consisted in a self-assignment of students to treatment group or control group, since the final exam is identical for all students and is the only evaluation score to be analyzed. However, we notice that there were questions in the exam that were reinforced by peer review activities and students from the treatment group obtained better results. Peer review tasks were focused on reinforcing the learning process regarding critical content of the syllabus. Not only that, but also it allowed students to realize the most common mistakes in the most critical aspects of the academic subject, which provided them with an extraordinary validation tool of the knowledge acquired. Additionally, selection bias plays a major role, since the voluntary nature of the experience makes it so that the most motivated students are those willing to take advantage of this new tool. Nevertheless, some factors need to be considered since only scoring does not guarantee a successful learning process. Among them we are aware of no necessary assessment maturity, not taking assessment seriously or a negative attitude towards the evaluation of these characteristics that can restrain the success of this methodology. For this reason, activities are intentionally prepared by the instructor so that they cannot be solved without having a wide understanding of the resolution process. In other words, judging peers' solutions are not about comparing numerical results but following and reviewing the reasoning process. Instructors have also reported that this has always encouraged students to boost their learning process. Due to the classroom limitations, some courses are divided into different academic groups (shift). For this reason, in several courses the number of students exceeds the permitted limit. Therefore, two or more groups are created. Then all of these groups of the same course are within the paper academic group.
- Treatment of data: All data collected have been anonymized according to the privacy regulation. These data have been contrasted with each other according to their nature, so that in some cases all students have volunteered to take part in the activities related to the peer review and in other cases this has not been the case. For this reason, the analysis in each case has been dealt with differently and therefore data has been treated in a customized way for each one of the experiences.
- Processing results: Advanced software such as Matlab® with statistical tools have been used. Given the potential of this tool, we have obtained graphs that significantly show how the premises defined in our thesis have been satisfied in a tangible way.

2.2.1. Experiences Type I

With these experiences we aim to observe the impact that peer review has on the final score of the course. Students were put into teams to perform different activities in the classroom in order to assimilate basic concepts. The tasks that are being evaluated by peer review consist of solving exercises in which students apply the concepts learned in class following a theoretical approach. For example, the instructor explains the concept of the derivative and its applications and students have to solve a concrete problem that involves

the derivative and that has not been previously solved in class. Therefore, the proposed tasks are very similar to the exercises to be solved in the exam. Afterwards, solutions were exchanged so that students could peer review each other. Finally, the teacher made a final correction of the results and the students' own corrections of their classmates. Accordingly, experiences 1, 2 and 3 have been conducted following a similar procedure so that the impact of peer review has been analyzed.

In these experiences, on-site activities have been carried out. Peer review was not implemented in all exercises as we considered that its application only on selected parts of the syllabus is a more effective way to preserve the appealing character of such a new methodology.

However, we have relied on the background and expertise of the instructor to choose those exercises that deal with fundamental concepts of the course and which, after previous years' experience, students find difficult to acquire. Between 2 and 3 activities have been carried out throughout the course, covering around 30–35% of the syllabus so as not to overload the students.

In addition, in these experiences, each student involved in the peer review has a period of time to carry out the exercises under the guidance of the instructor. These guidelines serve as a reference to accomplish the review of the exercises (between 2 and 3 peers) that takes place during the second stage. The number of exercises in each activity will be between 3 and 5. During this time, the student reinforces the knowledge that he/she has acquired after facing the proposed exercises they knew how to solve. In addition, although the student has the proposed solutions to those exercises, in some cases new possible solutions can be found. Finally, after the correction of the exercises, the instructor supervises all the scores in order to avoid deviations by students who mistakenly give wrong solutions.

### 2.2.2. Experiences Type II

Experiences belonging to this type are intended to test students' critical thinking and the capacity for objectivity with respect to the teacher. For them, different class tasks are presented and must be evaluated by their own peers following detailed rubrics shown in Tables 2–4. In order to assist the reader, the difference between the score issued by the teacher and the score issued by the peers (students) will be represented graphically so that final conclusions can be drawn. Experiences 4, 5 and 6 have been carried out for this purpose.

In these experiences, we selected activities in the form of projects to be presented by the students, at all times under the instructor guidance. These projects focused on the fundamental topic of the subject while reinforcing other transverse skills such as public speaking, synthesis, slide preparation, etc.

In addition, in these experiences, the students had a detailed rubric prepared by the instructor before defending each of their projects, so that they were aware of the items to be assessed, both by their classmates and by the instructor himself (see Tables 2–4). Before starting, it was proposed that the students review the rubric in depth in order to know how it should be applied and resolve any relevant doubt of the process. From this point on, students began to prepare the project under the supervision of the instructor. When the time came for the presentations, the students were in charge of assessing following both the rubrics and the exhibitions of their classmates' projects. It was the instructor's task to make their own evaluations of the presentations, as well as to review each student's.

**Table 2.** Experience 4-Rubric.

| CRITERIA | 4-EXCELLENT | 3-GOOD | 2-REGULAR | 1-INSUFFICIENT |
|---|---|---|---|---|
| | | TECHNICAL ASPECTS/ CONTENTS (C) | | |
| Difficulty/originality of the problem posed | The problem and objectives to be solved are highly complex from the point of view of communications and/or have not been addressed in previous cases. | The problem and objectives to be solved are challenging and it is difficult to find existing commercial deployments that cover them. | The problem and objectives to be solved are uncomplicated and it is easy to find existing commercial deployments covering the same scenario. | The problem and objectives are very common, there are many deployments that solve them and with documentation and information available in an open and detailed way in the Internet. |
| Use of radio technologies | The radio access technologies are perfectly adjusted to the proposed objectives, covering the needs of range, bandwidth, latency, etc. in a cost-efficient way. | The technologies adequately cover the use case although it may involve some restrictions to cover the requirements or is of a high cost (CAPEX or OPEX) with respect to other alternatives. | The radio technologies do not fully cover the objectives. There are other solutions clearly much more suitable to the problem. | The radio technology chosen does not allow the objectives to be covered even partially due to large bandwidth, latency or other limitations. |
| Architecture | The architecture is appropriate according to the technology, scalable and allows for the coverage of the required area. Its totality is detailed. | The architecture is adequate according to the technology and needs but some important details of it are left out. | The architecture presented does not include important elements of it and/or is clearly inefficient. | The architecture is not suitable for the technologies and/or objectives presented. |
| Cost Detail (not to consider the efficiency of the solution, only take into account the comm. components) | The CAPEX and OPEX costs of the solution have been calculated in a detailed and precise way for all the elements of the architecture, for an example case and taking into account costs of real suppliers and in comparison with other options. | Cost estimates have been made for most of the elements used and their area. | Unit cost values are indicated, but not system cost values. | No cost estimates are given or they are clearly wrong. |
| Environment | It takes into account in a detailed way social conditions, electromagnetic compatibility, security, installation and spectrum use license, etc. | It takes into account the main social, electromagnetic compatibility and safety conditions, etc. | It takes into account some social, electromagnetic compatibility, safety, etc. | It does not take into account environmental issues. |
| | | FORMAL ASPECTS (F) | | |
| Slides | Covers the topics in depth with details and examples. The visual and technical quality of the material is excellent. | Includes basic knowledge of the subject. The visual and/or technical quality of the material is acceptable. | The content is too basic and the visual and technical quality of the material is poor. | The content is too basic and the visual and technical quality of the material is poor. |
| Oral presentation | Great clarity, interesting and very well presented | Acceptable exposition, being able to maintain the interest of the audience. | Limited presentation, capturing little attention from the audience. | Poorly presented and did not get the attention of the audience. |
| Answer to questions | Knowledge of the subject is excellent and responds fluently to all questions asked. | Knowledge of the subject is acceptable and responds well to most questions. | Shows some lack of knowledge or ability to answer questions. | Answers are not answered correctly, has major gaps in basic concepts, and does not correctly define his or her work. |

**Table 3.** Experience 5-Rubric.

| CRITERIA | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| | TECHNICAL ASPECTS/ CONTENTS (C) | | | | |
| Installation proposed is suitable for its application | | | | | |
| System has been correctly sized | | | | | |
| System has been correctly placed (Orientation & Inclination) | | | | | |
| Converter is suitable for loads | | | | | |
| Loads are realistic for this kind of systems | | | | | |
| | FORMAL ASPECTS (F) | | | | |
| Recommended chapters | | | | | |
| Graphics have proper explanations | | | | | |
| Index of Content | | | | | |
| Index of Figures & Tables | | | | | |
| Appendix & Nomenclature | | | | | |
| Accurate References & Bibliography | | | | | |

**Table 4.** Experience 6-Rubric. Mark: 1, 2 (Basic); 3, 4 (In Progress); 5, 6 (Achieved); 7, 8 (Highlighted).

| CRITERIA | Possible Aspects to Consider | Score |
|---|---|---|
| **TECHNICAL ASPECTS/ CONTENTS (C)** | | |
| **MOTIVATION/ CONTEXT** | - Aims of the analysis and motivation in the choice of the time series.<br>- Explanation of the process of obtaining data, its possible treatment and interest in the time series.<br>- Relationship and comparison with other processes with similar topics. | |
| **IDENTIFICATION** | - Models are identified that explain the behavior of the time series.<br>- Appropriate models are proposed from the analysis of the graphics, explaining their meaning.<br>- Suitable models are proposed according to a clear order, without unjustifiably ignoring possible candidate models. | |
| **ESTIMATION** | - Estimation of the parameters involved in the model according to the previous criteria.<br>- Application of significance tests for the model parameters.<br>- Use of unit root tests and their consequences.<br>- Formalization of the process in terms of theoretical equations. | |
| **VALIDATION** | - Model validation using estimators and residuals.<br>- Different tests application for the validation of the proposed model.<br>- Application of model selection criteria. | |
| **FORECASTING** | - Forecasting is applied by means of graphics and in a numerical form, contextualized with data of the process.<br>- The predictions are compared with the observed data. | |
| **SOFTWARE APPLICATION** | - Implementation of the basic tools for the modeling framework.<br>- Use of the adapted code according to the features of the process.<br>- Expansion of tools through those available in the software. | |
| **FORMAL ASPECTS (F)** | | |
| **GRAPHICS PRESENTED** | - Graphics are presented with basic outputs that allow the modeling.<br>- Presentation and explanation of the necessary graphics to support the methodology followed.<br>- Presentation and explanation of the reinforcement graphics with advanced tools to facilitate the monitoring of the presentation. | |
| **METHODOLOGY** | - Implementation of all phases of the Box-Jenkins methodology.<br>- Interest of the process for the methodological application.<br>- Clarity and order for the monitoring of the methodology. | |
| **EXHIBITION/ RESOURCES** | - Clarity and coherence with the slides and the contents of the work.<br>- It reflects the methodology used, facilitating the knowledge of the whole process until the results are obtained. | |
| **EXHIBITION/ PRESENTATION** | - Clarity and order in the development of the contents.<br>- It relates the parts of the presentation, the purpose and conclusions.<br>- Questions are answered showing knowledge of the contents.<br>- It reflects general knowledge on both the Time Series and the statistical methodology used in the modeling framework. | |

*2.3. Procedure Description and Applications*

Peer review to strengthen concepts in students through an active methodology from the teaching and learning process to evaluation. The aim of the project is to involve students in the evaluation process, allowing a review of the material from a critical perspective, so that, seeing the need to analyze different approaches and points of view that may not have considered when performing their task, students will have a more global vision of their

learning in the subject. This vision will allow us to identify possible errors, limitations and/or highlights, see different methods of resolution and also detect when there are successes or improvements to the techniques applied.

In short, the aim of the approach is that students have access to various forms of performance of the same task which will allow them to acquire a deeper knowledge of the subject. When students review projects or tasks of other classmates, they acquire a critical vision of the work, preparing them for their professional future, where any task is subject to public assessment. This is a motivating aspect for the students who see in the realization of this evaluation its application to the real world. This last concession does not intend at any time to discharge the responsibility of the teacher in terms of evaluation and qualification of students, since the teacher must monitor that both the qualification and the evaluation have been made based on objective criteria and, of course, the teacher has the ultimate responsibility.

The procedure is organized as follows. In a first face-to-face session, the knowledge test will be carried out. Students must identify themselves by name in the test that must be given to the teacher. However, in order to guarantee the anonymity of the answers, the name must be able to be replaced by an identification code that the teacher will establish.

The teacher must establish this identification code and associate it with the name of each student. This relationship will only be known by the teacher. The teacher will remove the student's name with the pertinent anonymous identification code. The teacher shall make the appropriate copies, so that each student can evaluate at least two tests from other students. In a second face-to-face session, and after the publication of the corresponding rubric, each student must evaluate the answers of at least two other peers. The possibility of having the rubric in the session is up to the teacher. In this second face-to-face session, the student will record the assessment he or she considers from the other classmates. The teacher will collect the evaluations on a corresponding card, which must contain the name of the student who has carried out the evaluation of the exercise and the code of the student evaluated. The teacher will have to make a definitive correction of the contents and determine how far they deviate from the score issued by the peers. In this sense, a smaller discrepancy between the teacher's and the student's score will guarantee that the student has assimilated the content and that the correction by pairs has been based on objective criteria.

During the academic year 2019–2020, the planned objective of the fieldwork by the teachers involved in this project and the students involved has been fulfilled for the realization of collaborative peer evaluation activities. A total of six experiences have been carried out in multidisciplinary subjects of different scores involving two of the three designed evaluation modalities: anonymous evaluation (the student does not know who he is evaluating) and public evaluation (the student knows who he is evaluating).

Under the mode of anonymous peer review, it is intended to limit the influence of bias due to any type of social relationship between students, deploying a series of protocols and procedures to ensure the anonymity of the contents of the peer evaluation. On the other hand, in the modality of public evaluation students are exposed to real evaluation situations that require the promotion of their critical sense under the application of evaluation validation mechanisms that ensure that the scores really reflect the acquisition of the objectives of the proposed task.

Next, we include a brief description of different experiences developed throughout the academic year 2019–2020. Derived from the multidisciplinary and enriching experiences developed during the last academic year, during the current academic year an exhaustive analysis of the results and information obtained will be carried out in order to obtain clear and concise information on the strengths and weaknesses of this peer evaluation system as well as to identify possible lines of improvement. This will contribute to a restructuring in the final phase of the project, where those improvements identified based on the results obtained in the first phase will be applied, if necessary, and the results obtained will be analyzed again to determine the effectiveness of the innovative proposal. Furthermore,

during the present course, new experiences will be developed that will contribute to improve the quality of the results obtained from the present PIE (Teaching Innovation Project) with respect to the collaborative peer evaluation.

*2.4. Experiences*

This section presents the six experiences grouped in two different types as we have previously described.

### 2.4.1. Experience 1

Participants: Students of the "Linear Algebra and Geometry" course in the "Mathematics" BSc and the "Mathematics and Computer Science" dual BSc.

Procedure: Contents included in this experience covers two of the total of four lessons of the course, where the other two followed a classic evaluation approach

The peer review took place in the classroom. The session took one hour and a half. Students were first grouped into 3–5 member teams to solve several activities in one hour. Once they solved the assigned problems, the solutions of each team were exchanged. At that time, the copies of the activities solved correctly were also handed out so that students had a correction guide. The instructor explained one standard way to solve each proposed activity and they appear in the correction guide. However, we are aware that in many cases there are different ways to solve a task, and this is highly valued.

Each group had to correct the activities solved by other groups in the last 30 min of the lecture. At the end of the session, the teacher collected all the activities, both the activities that were done and the corrections made by the students. It was possible to observe a great interest by the students to know if the corrections that they had proposed to their peer's exercises were correct or not. In addition, the attendance level for this sort of activities was very high. The fact of having their own peers at the time of being scored leveraged their motivation. From the 77 students in the group, 32 of them voluntarily performed the peer review task while another 45 students did not.

### 2.4.2. Experience 2

Participants: Students of the "Algebraic Structures" course in the BSc degree of "Mathematics" and the "Mathematics and Computer Science" dual BSc.

Procedure: As in the previous case, the experience covers two lessons of the total of four. The peer review took place during an hour and a half session. These lessons were held virtually due to the COVID-19 using the Zoom application. This tool allowed the creation of small rooms so that students could be divided into 3–5 member groups. Students spent an hour solving the activities. Afterwards, they exchanged solutions, also virtually, and spent the last half hour of class correcting. Finally, the teacher collected all the activities and corrected them. Since they are mathematical exercises, in many cases there are different ways to solve them. Nevertheless, the correction by the students of exercises solved in a different but correct way from the one provided by the teacher is highly valued. From the 67 students, 29 of them voluntarily participated in peer review activities while another 38 did not.

### 2.4.3. Experience 3

Participants: Student of "Statistics II" course in the "Marketing and Market Research" BSc degree.

Procedure: At first, the students had to group together and deliver the activities as a group. However, as a consequence of COVID-19, teaching was delivered online and the exercises were carried out individually. In this sense, the teacher created an appropriate task so that students would be able to manage all relevant issues under lesson 1. Specifically, a list of activities was given to the students before solving them in a virtual session. Students had to value both the resolution and the result of the activities of their classmates. From the 81 students in the group, 41 of them performed the peer review task voluntarily while

another 40 students did not. Since class attendance was not mandatory and these activities took place during on-site sessions, some students did not get involved in the experience. Others simply did not want to take part in them because of the additional workload. In addition, 133 students from two other academic groups of the same course did not do the peer review task and serves as a control group.

### 2.4.4. Experience 4

Participants: Students of the "Wireless Networks" course in the "Telematics and Telecommunication Networks" MSc degree.

Procedure: This experience has adopted a model based on the development of projects whose objective is the definition of the wireless elements, technologies and architectures to be applied in order to solve a real-world use case. The general lines of work are proposed (smart cities, smart buildings, security…) where the specific use is defined by the students (e.g., farm security sensors, healthcare monitoring in a hospital…). The project teams were formed by 2–4 students each.

The evaluation of the projects was based on oral presentations of 10 min by each team. These were performed in front of all the students and a jury formed by the teacher and three other invited members with professional experience in wireless networks.

Hence, from the presentation, the projects get two evaluations: one from the jury and one from the rest of the class. Both evaluations were based on the same rubric (see Table 2). The rubric was known in advance for the students, and the evaluation was gathered online after each exposition through a Google Forms based poll.

### 2.4.5. Experience 5

Participants: Students in the "Renewable Energies" course of the BSc degree "Energy Engineering".

Procedure: Students taking part in this experience were about to address the final year dissertation where they will be evaluated by means of a project report thesis and its defense.

On the first hand, students were grouped in teams of 2–3 members. Each team had to address a project of an installation based on renewable energies. This project had to be real and accurate, and it had to be properly developed and explained. For this purpose, students received previous instructions on how the report should be written. Technical aspects were explained during lessons in class.

Afterwards each team had to expose their project in a 10 min pitch, but because of the pandemic situation, on campus lessons were reduced to the minimum and virtual teaching was the recommendation. Because of that, an alternative activity was designed. This experience consisted in a peer review process where each student had to revise one project among those developed by the others teams. In that manner, every project would be evaluated by two or three students and the teacher. For this purpose a rubric (see Table 3) was designed and provided to each student so they would evaluate the project.

### 2.4.6. Experience 6

Participants: Students of "Stochastic Models" course in the "Mathematics" BSc degree.

Procedure: Following the aforementioned methodology an activity was developed consisting of the realization and presentation, by the students, of a project of analysis and forecast of a time series based on real data. For this purpose, groups of up to three participants were formed. Finally, each student evaluated both the work and the oral presentation of the other projects.

For the implementation of the proposed methodology, an evaluation rubric (see Table 4) was created with ten criteria and several aspects to be taken into account. In the evaluation form, open sections were also considered so that the students could include assessments and proposals for improvement, both of the rubric and of the development of the activity itself.

In order to have a comparison tool to assess whether the activity could be considered valid, the teacher also evaluated the projects and their exhibitions in the same way and under the same rubric used by the students.

## 3. Results

As previously detailed, a total of six experiences have been conducted, which are grouped in two categories. On the one hand, Type I experiences (1–3) aim to give new insights into the effectiveness of peer review, that is, how the application of peer review methodology can affect student performance. To that end we have analyzed the the scores of the final exam of the course as previously done by other authors, such as Amendola et al. [12], Conde et al. [18] and Li et al. [31]. In this way, groups with and without peer evaluation during the learning process have been compared.

On the other hand, Type II experiences (4–6) focus on evaluating the reliability of the scores provided by peers by analyzing them in terms of their statistical distribution and their relation to the marks provided by the instructor.
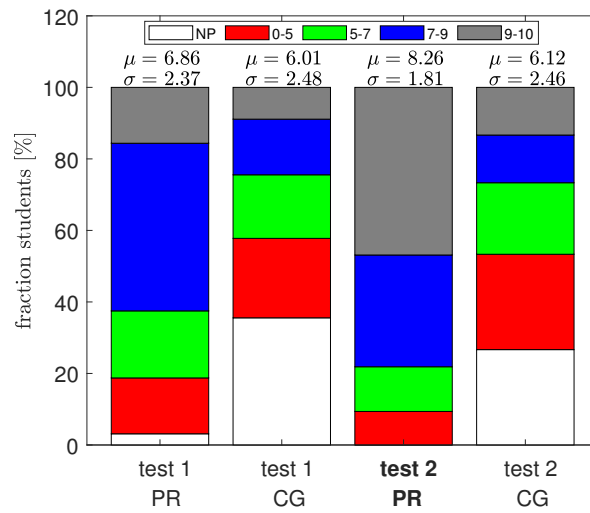
### 3.1. Experiences Type I

Results obtained from Type I experiences are displayed in Figures 1 and 2 according to the description given in experiences 1, 2, and 3 in Section 2.4. These diagrams display the number/fraction of students whose score is withing specific ranges. NP (standing for Non-Participant) refers to the fraction of students who did not attend to the final test. The rest of categories are defined following a 10-point scale. Since all the experiments were conducted in the context of the Spanish higher education system, it is considered that 5 is the minimum score to pass the exam. Accordingly, the following intervals [0–5), [5–7), [7–9) and [9–10] correspond to no pass, approved, outstanding and pass with distinction respectively.

Regarding experiences 1 and 2 (Figure 1), students were split into two subgroups: peer review (PR) and the control group (CG). As previously mentioned, the selection bias is based on a self assignment by the students themselves. Members of PR subgroup are those taking part in peer review activities during the learning process. On the other hand, members of CG are not taking any sort of peer review activity. In other words, they have followed a traditional learning process. In addition, the final exam of the course is formed by two tests: test 1 is related to the part of the syllabus where traditional learning process is used, whereas test 2 evaluates competences acquired in the second part of the subject (where only member of PR puts into practice peer review). This means that peer review activities have only been deployed by the members of subgroup PR in test 2 (bold in diagrams).

Based on this grouping, a comparison of the different subgroups based on mean scores and their standard deviation is of interest to check if the students' performance improves as a result of peer review activities. Left-hand side diagrams in Figure 1 describe the distribution of scores based on the number of students within each range, whereas right-hand side diagrams do the same but based on the percentage/fraction of students.
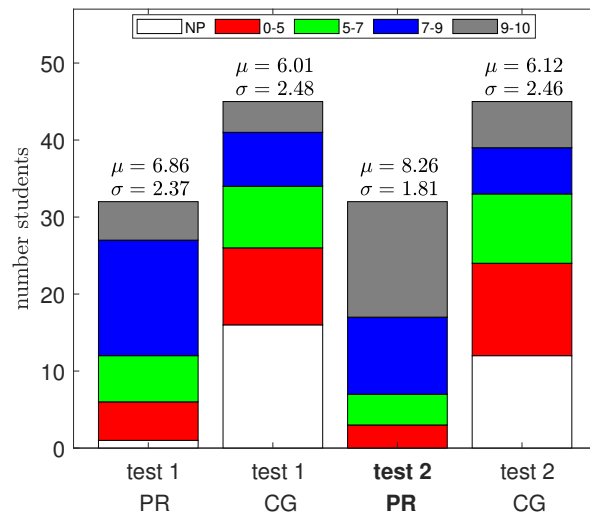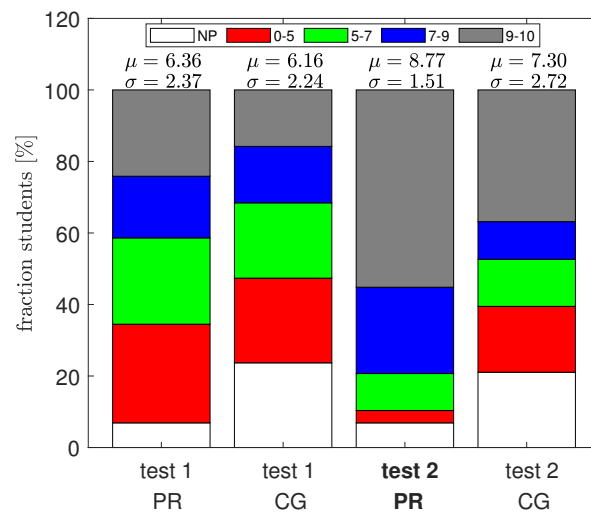
(**a**) Experience 1 (number of students).



(**b**) Experience 1 (fraction of students).
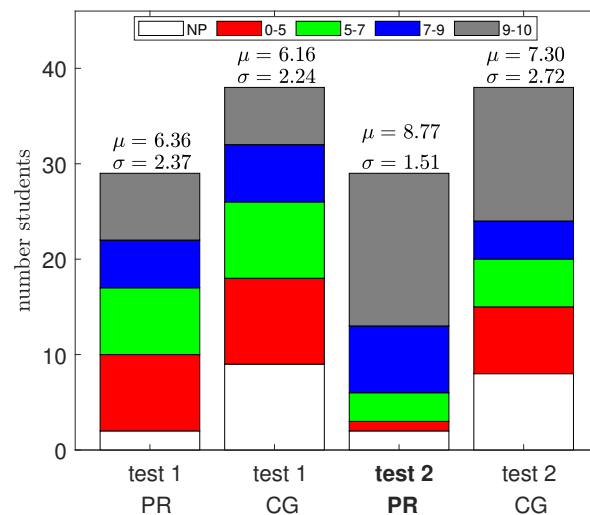


(**c**) Experience 2 (number of students).

**Figure 1.** *Cont.*

(**d**) Experience 2 (fraction of students).

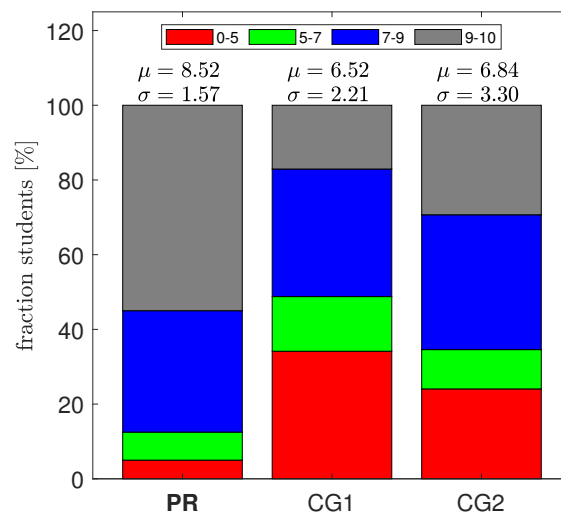**Figure 1.** Score distribution obtained in type I experiences 1 and 2.

Relevant conclusions can be drawn after from the results compiled after conducting experience 1 (see Figure 1a,b). Based on the mean scores, the control group (CG) obtained similar grades in both tests ($\mu = 6.01 \pm 2.48$ for test 1 and $\mu = 6.12 \pm 2.46$ for test 2). It can be explained since CG members have followed the same traditional learning process for both tests. In contrast, a significant difference has been reported when comparing results between test 1 and test 2 by PR members. Peer review activities allowed an increase in the mean score of test 2 ($\mu = 8.26 \pm 1.81$) vs. test 1 ($\mu = 6.86 \pm 2.37$). This shows that peer review has positively contributed to the learning process in the second part of the course being the process of learning from peer's flaws a key asset.

Experience 2 provides analogous results (see Figure 1c,d). One can observe that peer review activities deployed in the second part of the subject (test 2) contributed to an increase in the average score of PR members by 2.41 points (from test 1 $\mu = 6.36 \pm 2.37$ to test 2 $\mu = 8.77 \pm 1.51$). It is nevertheless noteworthy that results compiled by CG members have also shown an increase between test 1 and test 2 ($\mu = 6.16 \pm 2.24$ vs. $\mu = 7.30 \pm 2.72$) but less significant. This shows that, despite the fact that other factors may affect the average score (some parts of the syllabus may result more complex), the effectiveness of peer review has been proven, as it has shown a consistent improvement in the students performance expressed in terms of their final score.

Regarding experience 3 two figures have been shown (Figure 2a,b). Unlike experiences 1 and 2, only three subgroups have been formed, and a single test has been done covering the whole subject. The first subgroup (PR) is formed by those students who voluntarily decided to take part in the peer review experience (it is displayed in bold in figures). Alternatively, subgroup CG1 is associated to students who did not take part in such activities. It is important to mention that both subgroups (PR and CG1) belong to the same academic group/shift. In addition, subgroup CG2 represents students of the same course but from a different academic group/shift who did not take part in peer review activities (note that due to classroom limitations a course can be divided into different academic groups/shifts). A similar comparison can be set based on the academic results of the three different subgroups. Scores by subgroup PR was $8.52 \pm 1.57$, which is clearly higher than those obtained by members of CG1 ($\mu = 6.52 \pm 2.21$) and CG2 ($\mu = 6.80 \pm 3.30$). In line with experiences 1 and 2, experience 3 also indicates that students taking part in peer review activities benefit noticeably from the experience. All students have gone through the same final evaluation tests and course syllabus.
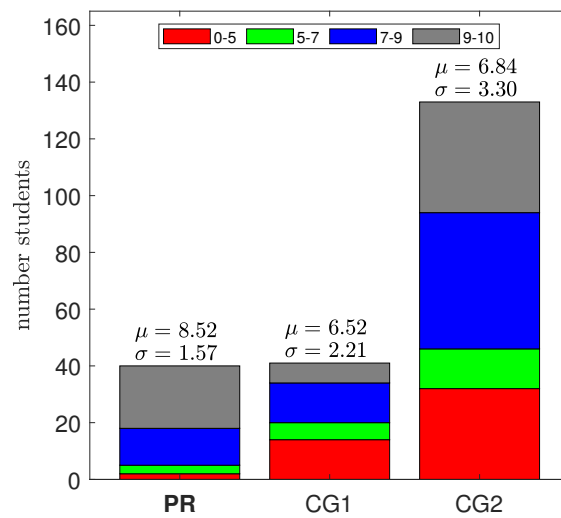
(**a**) Experience 3 (number of students).



(**b**) Experience 3 (fraction of students).

**Figure 2.** Score distribution obtained in type I experience 3.

### 3.2. Experiences Type II

Results concerning type II experiences were assessed by calculating the existing distribution of the score difference [12,30,31] obtained by subtracting the reference-score from the score provided by peers. Here, the reference-score is the one provided by the instructor (experiences 5 and 6) or the average between the score provided by a panel of experts and the instructor (experience 4). A positive difference indicates that the peer's score is higher than the reference-score. In order to unify the criteria, scores will be expressed on a 0–10 scale before calculating the difference.

The items detailed in the rubrics aim to cover most of the competencies playing a relevant role, but they have been grouped into two main categories. The first one refers to items related to the contents (C) of the task carried out by the student group. It basically refers to how students have applied the specific concepts related to the main scope of the course. In contrast, formal aspects (F) refer to whether students have properly applied the available resources and competences to communicate their conclusions and results.

Figure 3 displays the distribution of the score difference grouped in the two afore-mentioned categories. Left-hand side figures refer to scores related to the contents of the tasks (Figure 3a,c,e). In general terms, scores provided by peers are distributed around the reference-score (i.e., score difference is zero). However, it can be observed that mean

values are slightly positive (from $\mu = 0.71$ to $\mu = 0.08$ in experiences 4 and 5 respectively) meaning that scores provided by peers are slightly higher than the reference-score (set by the instructor). The adjusted underlying Gaussian probability distribution has also been included just for comparison, where the standard deviation shows that the dispersion of the score difference is around two points (0-10 scale) in the worst scenario.

Right-hand side Figure 3b,d,f show the distribution of the score difference concerning the formal aspects (F category). It shows how dispersed the distribution is of the score difference but considering only interdisciplinary competences (not exclusively related to the topic of the course). Analogous distribution is observed in this case, with standard deviation lower than 2 points in all cases and mean values slightly positive. Only experience 5 (Figure 3) has a relatively higher mean value ($\mu = 1.68$ points).
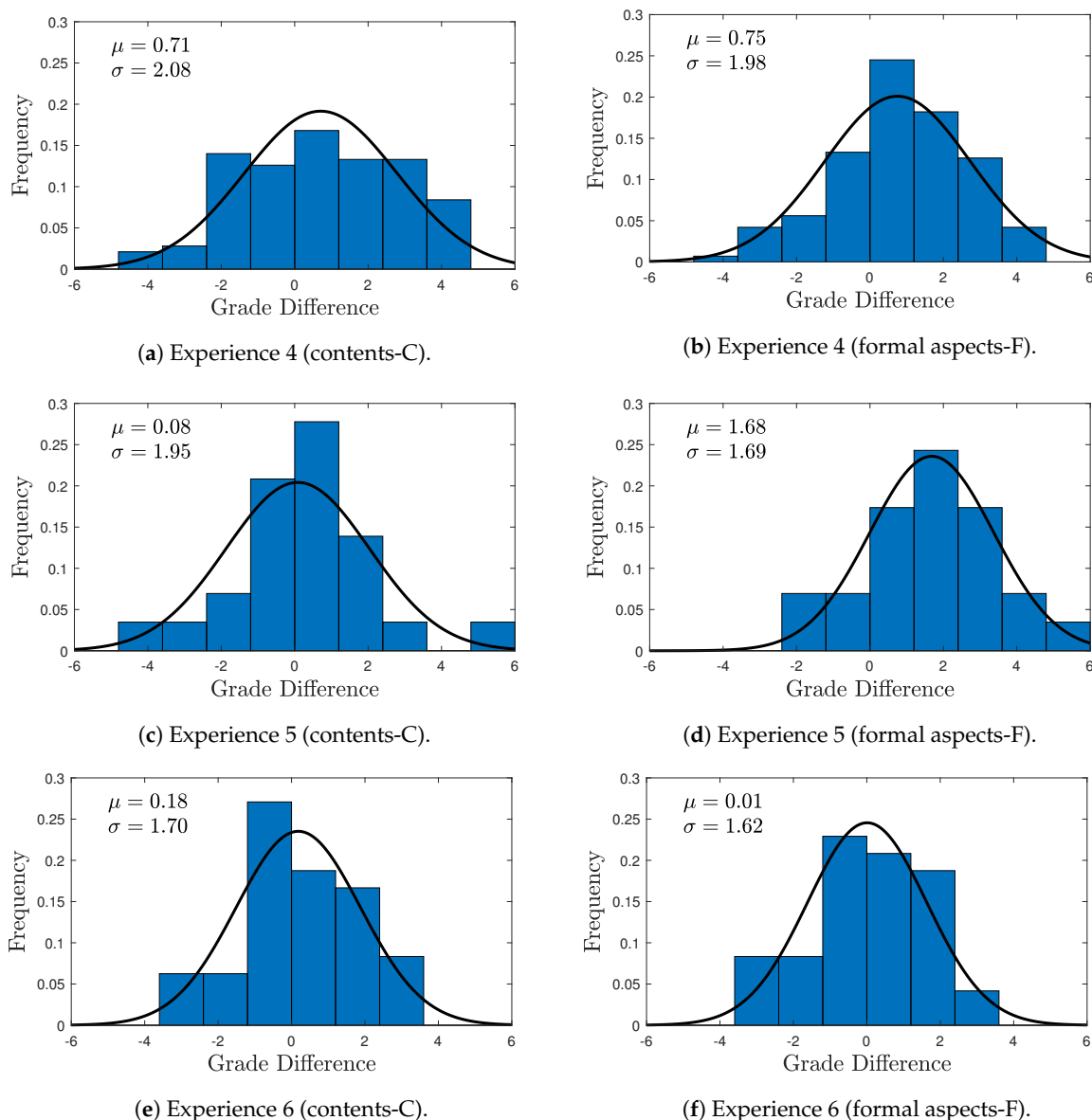


(**a**) Experience 4 (contents-C).

(**b**) Experience 4 (formal aspects-F).

(**c**) Experience 5 (contents-C).

(**d**) Experience 5 (formal aspects-F).

(**e**) Experience 6 (contents-C).

(**f**) Experience 6 (formal aspects-F).

**Figure 3.** Results of Type II experiences according to rubric items category.

In order to answer the question about if peer review is an effective and reliable method for evaluation, the average of all the items of the rubric scores (for both categories) is represented in a single histogram for each experience. As displayed in Figure 4, a global perspective of the distribution is provided, where the mean value of the peer scores is significantly closer to the instructor's reference-grade. Note that in experiences 4 and 6,

mean values are relatively low ($\mu = 0.72$ and $\mu = 0.11$ respectively). Only experience 4 shows a mean value slightly higher ($\mu = 0.96$). However, regardless of the experience or classification assumed for the evaluated items, dispersion (standard deviation) is enclosed by 2 points out of 10. These results can be considered as reference values for the level of discrepancy and variability between the teacher and student qualifications in the peer review process.
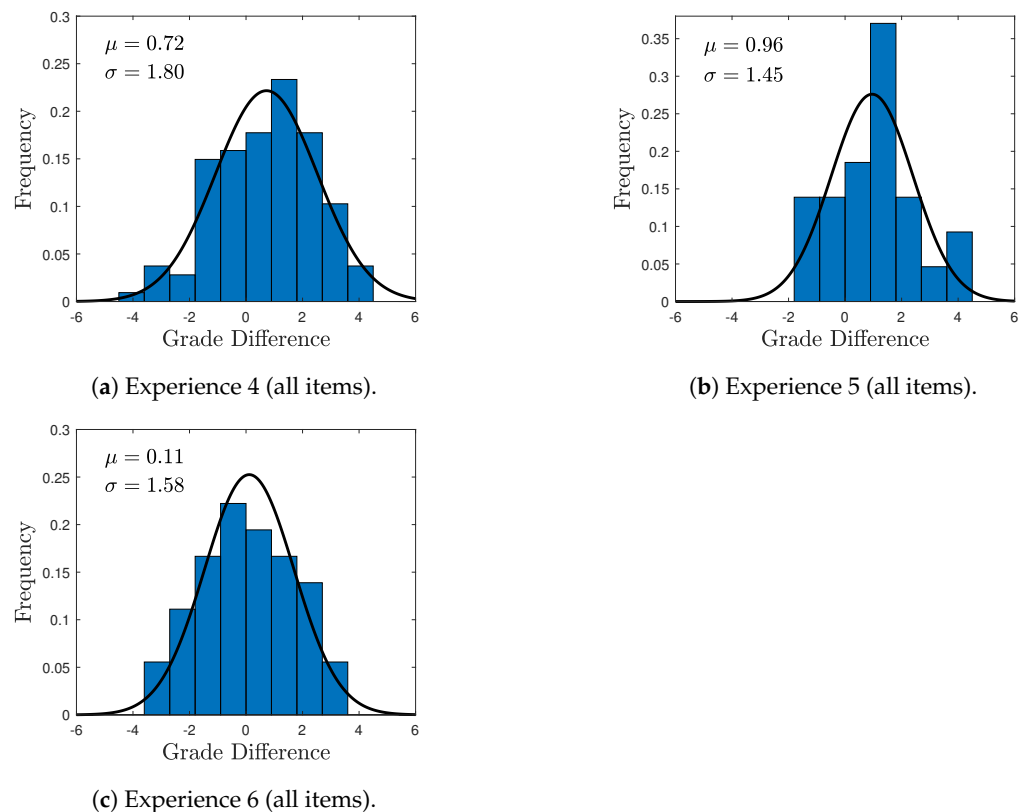


(**a**) Experience 4 (all items).                (**b**) Experience 5 (all items).



(**c**) Experience 6 (all items).

**Figure 4.** Results of Type II experiences including all the items in the rubric.

## 4. Conclusions and Future Work

When used for educational purposes, peer review methodology is considered a learning technique where students proactively evaluate the work of other students. In this article, an empirical evaluation of peer review has been carried out in a multidisciplinary set of six science and engineering courses in the University of Malaga in Spain, where more than 400 students were involved. In order to answer the two key questions of this study, experiences have been grouped in two different types: (i) those analyzing the impact that peer review methodology has on the students learning process (effectiveness), and (ii) those assessing students' critical sense and the capacity for objectivity in their provided reviews (reliability).

Regarding the effectiveness of the peer review methodology (first key question), it has been possible to validate how the experience has improved the results of those students who have participated. This has been confirmed for different degrees. Here, it can be highlighted that all experiences show a clear improvement in student performance. This is reflected in the general course scores of the participants, being the increase of more than two points (0–10 scale) in the best of cases and at least of one point but always achieving positive results with respect to the control group. Regarding the second key question, we found that the scores provided by the students are very similar to the score assigned by the teacher. Therefore, students are able to evaluate consistently enough according to our results, confirming the feasibility of peer review as a reliable evaluation methodology. Regardless of the type of experience, this methodology has shown to enhance

the learning process prior to evaluation. Not only do students assimilate contents for their first test, but they also develop a well based criteria to identify the mistakes made by other classmates, encouraging their critical sense, and identifying the more complex concepts. This way, students acquire the ability to identify the most common errors allowing them to complement their knowledge of the course more effectively.

After analyzing the application of the peer review methodology, it can be concluded that all the experiences have successfully performed from both academical and motivational perspectives. In this sense, the motivational aspect is key, especially considering remote education conditions. From our point of view, this fact is crucial to encourage students to participate in these initiatives as they clearly improve their academic performance. Hence, peer review can be a key approach for its implementation in the current pandemic scenario where the use of new technologies has become essential in the educational model. Moreover, peer review seems to be an advisable complement to the traditional evaluation process which is frequently considered one of the weak spots of online teaching. In addition, peer review provides the instructor with an invaluable source of information to support a more accurate evaluation.

As future work, a study on how anonymous evaluation affects the results is proposed. It is expected that friendship bias could be mitigated by deploying a series of protocols and procedures to ensure the anonymous nature of the evaluation process. For example, this can be guaranteed by the random allocation of the exercises to be scored following a double-blind review process. This reduces the likelihood that the answers will contain any sort of information about the author of the question. These can be done during the development of the course in question, either in written form or through any computer tool within a classroom setting. In this context, tests might be individual to help the anonymity of the review process.

In addition, based on our experience, we propose a continuous improvement in the rubrics in order to increase the quality and objectivity of the review process. This will be achieved by the redefinition of their format and instructions, the elimination of the points that can be considered weak and the inclusion of new items that can be interesting for the evaluation. These proposals for improvement of rubrics will be gathered from both the analysis of current rubrics as well as the obtained results accompanied with a survey fulfilled by those students that voluntarily participated in the peer review. Moreover, it has been identified that in some cases the correction of exercises requires an excessive amount of time considering the planned students workload. For these, an improvement is deemed necessary to adapt the amount of exercises to be performed during class in order to dedicate more time to the correction to be performed by the students. Another option being envisaged is to dedicate one session to the development of the activities and another to their correction.

Finally, it is planned to study the impact of the peer review methodology in the same six courses in the next years in order to assess its continued use. Moreover, its application to other courses and fields is also planned in order to widen the multidisciplinary nature of the analysis.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| EHEA | European Higher Education Area |
| SCL | Student-centred learning |
| BSc | Bachelor of Science |
| MSc | Master of Science |
| MOOC | massive open online course |
| PIE | Project Innovative Education |

## References

1. Calleja, J.; Avellino, R.; Farrugia, R.; Grech, V.; Grech, C.; Grech, O.; Mifsud, H.; Mizzi, B.; Vella, A. *Student-Centred Learning: A Bologna Process International Conference*; European Union: Brussels, Belgium, 2011.
2. Vidal, J. Quality Assurance, Legal Reforms and the European Higher Education Area in Spain. *Eur. J. Educ.* **2003**, *38*, 301–313. [CrossRef]
3. Agencia Nacional de Evaluación de la Calidad y Acreditación. *Programa de Convergencia Europea. El Crédito Europeo*; ANECA: Bossier City, LA, USA, 2003.
4. González, J.; Wagenaar, R. *Tuning Educational Structures in Europe*; University of Deusto: Bilbao, Spain, 2003.
5. Rodríguez-Izquierdo, R.M. Modelo formativo en el Espacio Europeo de Educación Superior: Valoraciones de los estudiantes. *Aula Abierta* **2014**, *42*, 106–113. [CrossRef]
6. De Miguel, M. *Modalidades de Enseñanza Centradas en el Desarrollo de Competencias, Orientaciones Para Promover el Cambio Metodológico en el Espacio Europeo de Educación Superior*; University of Oviedo: Bilbao, Spain, 2005.
7. Esteban, M.A.R.; Frechilla-Alonso, M.A.; Saez-Pérez, M.P. Implementación de la evaluación por pares como herramienta de aprendizaje en grupos numerosos. Experiencia docente entre universidades= Implementation of the evaluation by pairs as a learning tool in large groups. Teaching experience between universities. *Adv. Build. Educ.* **2018**, *2*, 66–82. [CrossRef]
8. Carless, D.; Joughin, G.; Liu, N.F. *How Assessment Supports Learning: Learning-Oriented Assessment in Action*; Hong Kong University Press: Hong Kong, China, 2006.
9. Pearce, J.; Mulder, R.; Baik, C. *Involving Students in Peer Review: Case Studies and Practical Strategies for University Teaching*; Centre for the Study of Higher Education, University of Melbourne: Melbourne, Australia, 2009.
10. Topping, K.J.; Smith, E.F.; Swanson, I.; Elliot, A. Formative peer assessment of academic writing between postgraduate students. *Assess. Eval. High. Educ.* **2000**, *25*, 149–169. [CrossRef]
11. Lundstrom, K.; Baker, W. To give is better than to receive: The benefits of peer review to the reviewer's own writing. *J. Second. Lang. Writ.* **2009**, *18*, 30–43. [CrossRef]
12. Amendola, D.; Miceli, C. Online Peer Assessment to improve students' learning outcomes and soft skills. *Ital. J. Educ. Technol.* **2018**, *26*, 71–84.
13. Hafner, J.; Hafner, P. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *Int. J. Sci. Educ.* **2003**, *25*, 1509–1528. [CrossRef]
14. Sáiz, M.S.I.; Gómez, G.R.; Ruiz, M.Á.G. La evaluación entre iguales: Beneficios y estrategias para su práctica en la universidad1 Benefits of Peer Assessment and Strategies for Its Practice at University. *Rev. Educ.* **2012**, *359*, 6–23.
15. Nicol, D.; Thomson, A.; Breslin, C. Rethinking feedback practices in higher education: A peer review perspective. *Assess. Eval. High. Educ.* **2014**, *39*, 102–122. [CrossRef]
16. Guilford, W.H. Teaching peer review and the process of scientific writing. *Adv. Physiol. Educ.* **2001**, *25*, 167–175. [CrossRef]
17. Dowse, R.; Melvold, J.; McGrath, K. Students guiding students: Integrating student peer review into a large first year science subject. A Practice Report. *Stud. Success* **2018**, *9*, 79–86. [CrossRef]
18. Conde, M.; Sanchez-Gonzalez, L.; Matellan-Olivera, V.; Rodriguez-Lera, F.J. Application of Peer Review Techniques in Engineering Education. *Int. J. Eng. Educ.* **2017**, *33*, 918–926.
19. Adachi, C.; Tai, J.H.M.; Dawson, P. Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assess. Eval. High. Educ.* **2018**, *43*, 294–306. [CrossRef]
20. Indriasari, T.D.; Luxton-Reilly, A.; Denny, P. A Review of Peer Code Review in Higher Education. *ACM Trans. Comput. Educ. (TOCE)* **2020**, *20*, 1–25. [CrossRef]
21. Meek, S.E.; Blakemore, L.; Marks, L. Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review. *Assess. Eval. High. Educ.* **2017**, *42*, 1000–1013. [CrossRef]

22. Reddy, K.; Harland, T.; Wass, R.; Wald, N. Student peer review as a process of knowledge creation through dialogue. *High. Educ. Res. Dev.* **2020**, 1–13. [CrossRef]

23. Gaynor, J.W. Peer review in the classroom: Student perceptions, peer feedback quality and the role of assessment. *Assess. Eval. High. Educ.* **2020**, *45*, 758–775. [CrossRef]

24. Arimoto, M.; Clark, I. *Cultural Perspectives on Classroom Assessment: A Path Toward the " Japanese Assessment for Learning Network "*; Annual Bulletin; Graduate School of Education, Tohoku University: Sendai, Japan, 2015. [CrossRef]

25. Bryant, D.; Carless, D. Peer assessment in a test-dominated setting: Empowering, boring or facilitating examination preparation? *Educ. Res. Policy Pract.* **2010**, *9*, 3–15. [CrossRef]

26. Lladó, A.P.; Soley, L.F.; Sansbelló, R.M.F.; Pujolras, G.A.; Planella, J.P.; Roura-Pascual, N.; Martínez, J.J.S.; Moreno, L.M. Student perceptions of peer assessment: An interdisciplinary study. *Assess. Eval. High. Educ.* **2014**, *39*, 592–610. [CrossRef]

27. Mulder, R.A.; Pearce, J.M.; Baik, C. Peer review in higher education: Student perceptions before and after participation. *Act. Learn. High. Educ.* **2014**, *15*, 157–171. [CrossRef]

28. Panadero, E.; Romero, M.; Strijbos, J.W. The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Stud. Educ. Eval.* **2013**, *39*, 195–203. [CrossRef]

29. Langan, A.M.; Wheater, C.P.; Shaw, E.M.; Haines, B.J.; Cullen, W.R.; Boyle, J.C.; Penney, D.; Oldekop, J.A.; Ashcroft, C.; Lockey, L.; et al. Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assess. Eval. High. Educ.* **2005**, *30*, 21–34. [CrossRef]

30. Falchikov, N.; Goldfinch, J. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Rev. Educ. Res.* **2000**, *70*, 287–322. [CrossRef]

31. Li, H.; Xiong, Y.; Hunter, C.V.; Guo, X.; Tywoniw, R. Does peer assessment promote student learning? A meta-analysis. *Assess. Eval. High. Educ.* **2020**, *45*, 193–211. [CrossRef]

32. Sánchez Rodríguez, J.; Ruiz-Palmero, J.; Sánchez-Rivas, E. *Análisis Comparativo de Evaluación entre Pares con la del Profesorado, Un Caso práctico*; Escuela Universitaria de Magisterio de Toledo: Toledo, Spain, 2011.

33. Dopico, E. Peer review as teaching method. *Educ. Learn. Open Access* **2016**, *1, 1-4.*

34. Van den Berg, I.; Admiraal, W.; Pilot, A. Design principles and outcomes of peer assessment in higher education. *Stud. High. Educ.* **2006**, *31*, 341–356. [CrossRef]

35. Brown, S. *Peer Assessment in Practice*; SEDA Birmingham: Kenilworth, UK, 1998.