

An Analysis of Shared Grading in Co-Assessment Practices by Teachers and Students

Análisis de las calificaciones compartidas en la modalidad participativa de la evaluación colaborativa entre docente y estudiantes

Gómez-Ruiz, M.A.⁽¹⁾ , & Quesada-Serra, V.⁽²⁾ 

(1) University of Cadiz (Spain). (2) University of the Balearic Islands (Spain).

Abstract

Current trends in educational assessment in different branches of higher education share the common goal of uniting learning with assessment. Most approaches and theoretical and practical developments in this field revolve around four main factors: feedback, democratization, alignment and relevance. This paper proposes the use of co-assessment as a means of ensuring dialogue-based, democratic and fairer evaluations. With co-assessment, the responsibility is shared by the teacher and the students, who negotiate and agree on the appraisal of student tasks and, in this paper, also on the awarded mark. The aim of this study is to analyse the relationship between a series of jointly agreed marks, following the co-assessment of four tasks, and the marks that the teachers and students would each have individually awarded. Two teachers and 100 students participated in the study, which follows a correlational design and analyses significant statistical differences. The results show a strong correlation between the jointly agreed marks and those assigned individually by the teacher, even though statistically significant differences were found between them. Conversely, no statistically significant differences were identified between the joint marks and the marks assigned individually by the students. These results call for reflection on the real possibility of adapting shared grading methods to students in university frameworks, where the repercussions of awarded marks go far beyond formative goals.

Keywords: Grading; Student Evaluation; Teacher Student Relationship; Alternative Assessment; Educational assessment; co-assessment.

Resumen

Dentro de la variedad existente en las tendencias actuales sobre la evaluación de estudiantes, se encuentra el propósito común de relacionar evaluación y aprendizaje. La retroalimentación, la democratización, la coherencia y la relevancia son cuatro tópicos aglutinadores sobre los que giran la mayoría de los planteamientos y avances teóricos y prácticos en este ámbito. Como forma concreta de cristalizar una evaluación dialógica, democrática y justa, se propone la modalidad participativa de la evaluación colaborativa en la que docentes y estudiantes se reparten la responsabilidad, negociando y consensuando de forma conjunta el valor de las tareas y en nuestro caso, también la calificación final. El propósito de este estudio, que ha involucrado a un total de 100 alumnos y 2 docentes, es precisamente la comprobación del grado de relación existente entre las calificaciones compartidas de 4 tareas universitarias con las que habrían aportado en solitario el docente y el grupo de estudiantes. Se ha seguido un diseño de investigación correlacional y se ha comprobado la existencia de diferencias significativas. Los resultados muestran la estrecha correlación entre las calificaciones compartidas y las calificaciones del docente, aunque se han hallado diferencias estadísticamente significativas entre estas. Por otro lado, no se han encontrado diferencias entre las calificaciones compartidas y las calificaciones de los estudiantes. Las repercusiones de estos resultados, hacen reflexionar, entre otras cuestiones, sobre la posibilidad real de ajustar dichas calificaciones abiertas a la participación de los estudiantes en contextos universitarios donde las repercusiones sobrepasan de largo el ámbito únicamente formativo.

Palabras clave: Calificación; Evaluación de estudiantes; Relación de docentes y estudiantes; Evaluación alternativa; Evaluación educativa; Coevaluación.

Received/Recibido 2020 January 27 | Approved /Aprobado 2020 June 03 | Published/Publicado 2020 June 24

In recent decades, theoretical and practical scientific developments in the assessment of learning processes, mainly in the field of higher education, have led to the emergence of a variety of new ideas and approaches. The emphasis has moved from what might perhaps be regarded as traditional formative assessment methods, focusing on diverse aspects and leading to multiple different trends and theories. These include learning-oriented assessment (Carless, 2007), sustainable assessment (Boud & Soler, 2015), assessment as learning and empowerment (Rodríguez & Ibarra, 2015), integrative assessment (Crisp, 2012), formative and shared assessment (López, 2009; 2012) and, even, from a more critical social standpoint, inclusive assessment (Santiuste & Arranz, 2009) or assessment for social justice (Hidalgo & Murillo, 2016), to cite just a few examples.

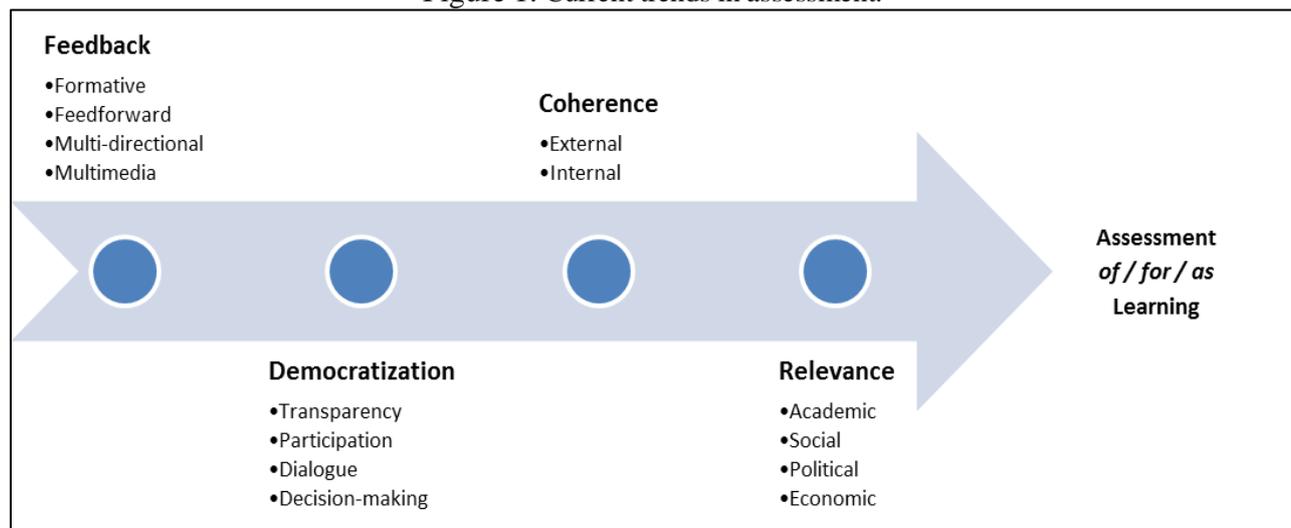
The common denominator to the majority of this broad spectrum of theories, with their wide-ranging names, common links and specific peculiarities, is the notion that assessment is closely tied in with the participants' learning process (i.e. both the students and teacher), whether it is to corroborate, modify or extend this learning (Hayward, 2015; Ibarra & Rodríguez, 2019). The strategies and concepts on which current notions of assessment tend to be based can be summed up in four factors: feedback, democratization, coherence and relevance (Figure 1).

Feedback. All the approaches coincide in the idea that feedback is needed in formative assessment in the form of good-quality information on the student's work or performance, provided in due time and manner. Indeed, this feedback is considered to be the most important factor in fostering learning and self-regulation in students (Hattie & Timperley, 2007). In recent years, feedback has also come to be seen as an activity that can be extended to other tasks or spheres; that is, transferable feedback or feedforward, where assessments focus constructively on future performance (Canabal & Margalef, 2017).

The concept of feedback has also evolved in terms of the agents involved in it and the channels through which it is given. There has been a shift from the concept of unidirectional information passed from the teacher, as the expert, to the student, as the learner, to a multidirectional process where students take a more active role, either by providing useful information to their peers or to the actual teacher (López & Sicilia, 2017; Nicol, 2010). Student feedback to a teacher (Swaffield, 2011) can help the latter to reflect on the teaching method that is being used, the relevance of the assessment tasks, the achievement of academic goals, or the repercussions of the feedback given to students. Likewise, student feedback to their peers can benefit the recipients, since it is sometimes regarded as more easily understandable and more useful than the information provided by the teacher (Gallego et al., 2017; Topping, 2003), and it can also foster a better capacity for reasoning and assessment in the issuer of the feedback (Nicol, 2013). As for the channel through which it is given, thanks to technological developments, there has been a move from synchronous dialogue-based or asynchronous written feedback to a wide variety of formats and multimedia combinations, with audio and video feedback offering particularly useful potential (García et al., 2015).

Democratization. In formal education and as acts in themselves, assessing and grading are explicitly accepted to be acts of power (Leach et al., 2010). For some time now, there have been calls for institutional power to be redistributed in one way or another among students, at least partially (Quesada et al., 2019). From an informative perspective, assessment processes must be transparent (Ibarra & Rodríguez, 2019), based on fully accessible, understandable systems. The importance of student participation in assessment processes has also been upheld, whether it is in the design (the criteria and planning), execution (responsibility through participatory self, peer, or co-assessment methods) or in grading decisions, taken individually or jointly (Quesada et al., 2017).

Figure 1. Current trends in assessment.



Note: Source own

Whichever of the above participatory systems is used, through student involvement in assessment, the aim is to engage them in their own learning process, fostering independence and greater responsibility (Penuel & Shepard, 2016). Similarly, discussion, reaching a consensus and joint decision-making in the design and development of assessment methods are recurrent aspects of different current theories (López & Sicilia, 2017), in addition to the need to foster assessment literacy among students as a means of guaranteeing their well-grounded, effective, coherent participation (Smith et al., 2013).

Coherence. From the teaching staff's perspective, the assessment process has traditionally been the last factor to take into account in the design of teaching programmes, following the specification of the programme's objectives and teaching activities. Now, however, a student-centred approach to assessment tends to be taken as a springboard for success in tackling learning activities (Biggs & Tang, 2011). The idea is to make the whole process more coherent (external coherence), aligning the objectives, methodology, activities and envisaged assessment, while also paying special attention to the role that learning tasks and assessment play in the whole process. Since they determine the students' performance, they should be authentic and realistic (Swaffield, 2011); that is, useful and helpful in their education and training.

Emphasis is also placed on defining and aligning the different components of the assessment process (internal coherence); that is, the criteria, means, output, techniques and instruments, linking them in a coherent way and systematizing assessment procedures as far as possible (Ibarra & Rodríguez, 2019; Quesada et al., 2017).

Relevance. Increasing in-depth research is being conducted into different aspects of the assessment of learning practices. Given the obvious academic and cognitive repercussions, assessment has long been considered to be fundamental in teaching and learning at all levels of education. However, growing consideration is being given to the extent to which assessment affects students' lives (McArthur, 2019), in addition to the real social, emotional, political and economic repercussions. Its multi-dimensional relevance ties in directly with the axiological framework for teachers, raising the issue of the present and future consequences for students and for the society we are building (Hidalgo & Murillo, 2016).

Co-assessment as a participatory system

Current theories on assessment have not led to the development of one single practical model, given the wide variety of possible combinations, different emphasises, and the limited imperfect nature of assessment in the real world. Despite this, we believe that co-assessment—with its transparent design and

procedures and its guiding formative role—is a feasible means of achieving a democratic system of evaluation which fosters participation and decision-making. Discussion-based feedback plays a central role in co-assessment, coherently aligned with the methodology and assessment tasks, in an approach where inclusive, critical, optimistic fair guidance is used to transform the assessment process into a strategy aimed at social justice (Murillo & Hidalgo, 2015).

Co-assessment (*evaluación colaborativa*) in Spanish is synonymous with collaborative and cooperative assessment. It is also described in scientific literature in Spanish as *coevaluación*, *co-evaluación* and *evaluación compartida*. However, the terms *coevaluación* and *co-evaluación* (a literal translation of co-assessment) are also used in Spain to refer to peer assessment, while *evaluación compartida* (shared assessment) can also refer to broader frameworks (see López, 2012).

As a result, for the sake of greater clarity in scientific publications, in this study we use the term co-assessment to refer to participatory methods in which a consensus is reached through a joint appraisal and discussion of students' work by the teacher and students. The other two existing participatory methods are self-assessment and peer-assessment, although co-assessment is the least known, least commonly used type and the one that causes most insecurity among university teaching staff (Quesada et al., 2016).

Thus, co-assessment is a jointly negotiated process by a teacher and their students aimed at reaching a consensus, with shared responsibility and emphasis on dialogue. Three concepts are fundamental in this definition: responsibility, dialogue and consensus, reflecting a politically and socially committed approach to educational interaction. In this case, the teacher is not the only one in possession of the truth and the students are seen as being able to gauge the learning process and to take well-reasoned decisions. Students therefore play an empowered role in the learning process and they are responsible for

their academic situations (Quesada et al., 2019).

Hence co-assessment can be construed as entailing student self-assessment, peer assessment in the case of group work, and assessment by the teacher, all explicitly combined in a dialogue-based way (Quesada et al., 2016). Similarly, Kurt (2014) considers co-assessment to be a combination of self-assessment, peer assessment, assessment by the teacher, and negotiated assessment. Nevertheless, co-assessment should not be confused with practices in which work is self-assessed by the students and separately assessed by the teacher, with no discussion or negotiation. It is precisely through this process of comparison or triangulation that teachers manage to gauge how the learning process is going, and it boosts the students' potential for learning and their capacity for self-regulation by offering a better insight into their own performance and into the views of others.

According to specialist literature, the benefits of co-assessment in student learning include (Boud & Falchikov, 2006; Cooper, 2017; Deeley, 2014; Dochy et al., 1999; Gómez & Quesada, 2017; Knight & Yorke, 2003; Quesada et al., 2016; Quesada et al., 2019) student reflection on their performance; more in-depth learning; better self-regulation, independence and decision-making; improved relations and communication between students and teachers; stronger efforts, motivation and engagement; and improved assessment literacy, self-esteem and self-confidence in appraisals.

However, there can be some risks or drawbacks to this system of assessment (*ibid.*), both for students and teachers. The complexity of student self-assessment has been highlighted, in addition to the tension and discomfort that some students feel when they have to discuss their work with teachers. In the case of teaching staff, it can increase their workload, even with small classes, and it is acknowledged to be a complicated practice with big groups.

Shared grading of assessed tasks

Co-assessment can include reaching a joint negotiated consensus on a mark. This is not an essential component though, because the most important factors are joint appraisals and decision-making based on dialogue, reasoned judgements and evidence (Van der Bergh et al., 2006). There is much debate on the possible culmination of the process in a joint mark, with some researchers suspecting that the mark might be under or over-estimated. Some authors also only recommend dialogue-based grading on completion of a subject (López, 2012).

The explicit consideration of the role that grading should play in new systems of assessment is a touchy subject for many teachers and researchers in this field. On occasions, formative theoretical approaches to assessment seem to be overlooked, whether it is intentional or not. It is sometimes also believed that grading continues to be the teacher's sole responsibility or even that involving students might be detrimental since it could distort the learning process. In university education at least, it is a legal and institutional requirement to grade the students' work for a subject with just a number. Teachers are still forced to use a mark to reflect the students' output and learning, and we must decide whether we include them in decision-making on their marks and those of their peers as part of an active participatory experience.

Involving students in the grading process can foster authentic, meaningful, deeper engagement in learning and assessment. In other words, it could be viewed as a clear sign of their involvement and decision-making in individual learning processes (Álvarez, 2001). A higher level of democratization might also be achieved, because students would take part in all the decisions that directly affect them. They might also perceive a certain logic and continuity to their involvement in assessment and grading as a more coherent rational alternative, moving away from current simplistic realities that foster competition, hierarchies, labelling and even the

marginalization of some students (Casanova, 2011).

However, some research studies have shown that this integration in the grading process encourages certain students to seek strategies that will allow them to boost their marks during negotiations (See Deeley, 2014). In studies that explore the opinions of students who have taken part in co-assessment initiatives (Gómez & Quesada, 2017; Quesada et al., 2019), the students also reveal a certain concern about the adequacy of the awarded mark due to their involvement in decisions. These are concerns as yet unsupported by scientific evidence.

In relevant literature on the subject, it is typical to find studies that analyse the reliability and validity of marks awarded in self and peer-assessment processes (Bretones, 2008; Falchikov & Goldfinch, 2000). However, given the relatively rare use of co-assessment and its nature, it was not possible to find equally solid research studies on the possible relationship between jointly awarded marks and individual marks by students and teachers.

Objectives

The main aim of this study is to analyse possible differences between joint marks awarded as part of a co-assessment process and the marks that the students who had done the assessed task or their teacher would each have individually awarded. More specifically, the objectives are outlined below:

- To analyse the correlation level between marks jointly assigned as part of a dialogue-based co-assessment process and the marks that would have been individually assigned by the teacher and group of students carrying out the task.
- To check whether there are statistically significant differences between the joint marks and the ones that the teacher and students would have separately awarded.

Method

The study was based on a non-experimental research design. A correlational approach was chosen where the researchers could not control or manipulate the variables under study, but where they attempted to establish the strength and direction of the relationship between the analysed marks (Hernández & Maquilón, 2010). This design was complemented by a difference of means test.

Procedure

Given the research study's practical focus, it was deemed particularly important to specify the whole appraisal process that was used. In it, the teachers proposed the assessment and grading of four group tasks (two per teacher). These tasks accounted for 30% of the total mark for the corresponding subject. The tasks were as follows:

- Task 1: To provide a creative response to questions through an audio, images or a video after analysing several texts and audio-visuals.
- Task 2: To write an essay on the importance of reflecting on innovations in teaching.
- Task 3: The design of part 1 of the research. Choosing a research subject, writing a brief theoretical framework, and the formulation of research questions.
- Task 4: The design of part 2 of the research. Methodology: sample, timeframe, techniques and data-gathering instruments.

The process for assessing and grading each task was adapted from other similar experiences (Gómez & Quesada, 2017; Quesada et al., 2019), albeit with some inevitable modifications in order to reflect the different marks and hence meet the study's objectives. Shown below are the different steps or stages:

i. Presentation of the task in class. To begin, each teacher gave a description of the activity to be performed, clearly specifying the criteria

that would be used for the joint assessment of the task.

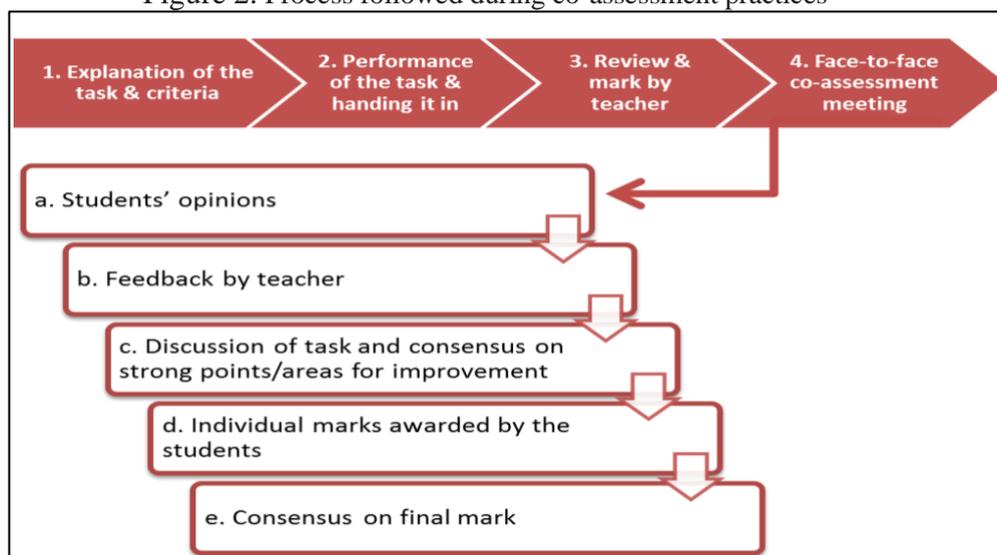
ii. Performance of the task. The students had two or three weeks to do the task and to hand it in online.

iii. Review and mark by the teacher. After the task was handed in, the teacher corrected it, making comments in the document as feedback.

iv. Co-assessment meeting. At the next face-to-face session, the teacher met up with each work group and 15 to 20 minutes were spent jointly assessing the work in a dialogue-based, well-grounded way, as follows:

- a. Firstly, the students expressed their ideas and opinions about the task.
- b. Then the teacher's comments were shown to them and explained as feedback, based on the applied assessment criteria.
- c. The students could then express their agreement or disagreement with the teacher's opinions, queries were settled, and the students expressed their opinions and thoughts on the difficulties they had encountered in doing the task and on the adequacy of the work they handed in.
- d. After discussions between the students and teacher, the students were asked to consider what mark they would award the task, each writing down their own mark individually.
- e. Finally, there was a further opportunity for dialogue, exchanges of opinions and negotiation among the students and between the students and teacher to reach a joint consensus on the final mark, based on their personal opinions. When the meeting came to an end, the teacher recorded the individual marks given by each member of the group and the final assigned mark. Figure 2 summarizes the process:

Figure 2. Process followed during co-assessment practices



Participants

The data for the study was gathered within the framework of the subject "Educational Innovation and Research", a compulsory subject during the first semester of the second year of the University of Cadiz's degree in primary education (academic year 2016-2017). The participants were selected using intentional non-probabilistic sampling (Wood & Smith, 2018), based on the researchers' opportunities of access.

More specifically, a total of 100 students took part, divided into three sub-groups for the

subject's practical activities. 29 of them were men (29%) and 71 were women (71%). They were further divided into 22 work groups, normally made up of 4 to 6 people (86.4%). It is important to note that these students had already taken another subject with one of the teachers—teacher A—involving co-assessment (including shared grading), and so they already had some prior knowledge of this form of assessment. Tables 1 and 2 show the distribution of the students and work groups according to the number of participants.

Table 1. Distribution of the students taking part

| Group | Men | Women | Total |
|---------|-----|-------|-------|
| Group A | 11 | 22 | 33 |
| Group B | 5 | 32 | 37 |
| Group C | 13 | 17 | 30 |
| Total | 29 | 71 | 100 |

Table 2. Distribution of the work groups by the number of members

| Group | 2 members | 3 members | 4 members | 5 members | 6 members | Total |
|---------|-----------|-----------|-----------|-----------|-----------|-------|
| Group A | - | - | 3 | 3 | 1 | 7 |
| Group B | 2 | - | - | 3 | 3 | 8 |
| Group C | - | 1 | 4 | 1 | 1 | 7 |
| Total | 2 | 1 | 7 | 7 | 5 | 22 |

Two teachers took part, a man and a woman, both with about 10 years' university teaching experience. The teachers shared the subject and each of them was in charge of two of the four

tasks whose marks were analysed: teacher A (the man) supervised the assessment and grading process for tasks 1 and 2, and teacher 2 (the woman) for tasks 3 and 4.

Instrument

To gather the individual marks by the teacher and students and the joint marks from the shared negotiated assessment process, a record sheet was designed where the name of the group and task number were noted down, plus the mark given by the teacher in one column, the individual mark by each member of the

group in other columns, and the joint mark in the last one (see Figure 1).

As mentioned in the 'Procedure' section, the teacher's mark was recorded prior to the face-to-face co-assessment meeting when the individual student marks and final joint mark for the task were recorded. In all cases, the marks could range from a minimum of 0 to a maximum of 10.

Figure 3. Record sheet for the marks

| SHARED GRADING | | | | | | | | | |
|-----------------------|---------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| TASK: _____ | | | | | | | | | |
| GROUP | TEACHER | Student 1 | Student 2 | Student 3 | Student 4 | Student 5 | Student 6 | Student 7 | Joint mark |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

Data analysis

Once the data had been recorded, in order to meet the objectives of the study, a statistical analysis was performed, relating the three variables under study: the initial mark awarded by the teacher for each task, the mean mark of the individual ones given by each member of the group, and the definitive joint mark for the task, mutually agreed by the teacher and students.

With the aid of the JASP statistical software programme, an analysis of the descriptive statistics was first conducted, followed by the calculation of Pearson's correlation coefficient to check the degree of association among the different marks. As a complementary measure, the paired samples t test was also conducted to detect for statistically significant differences in the paired means. It was chosen so that the differences could be analysed group by group. Likewise, the effect size of the said differences was calculated using Cohen's d. Paired samples were used because the samples were in contact and they could influence one another

(remember that there was negotiation). Also, the members of one group could form part of another (for instance, the students in group CG helped to assign a joint mark in group CC, as did the teacher).

Results

Descriptive and correlational analysis of the marks

Given that four tasks were carried out by the 100 students, divided into 22 work groups, 88 tasks were graded in total. In other words, 264 marks were analysed: 88 awarded by the two teachers (44 by each), 88 final joint marks, and 88 marks that represent the mean values of the 400 individual marks given by the students.

When an analysis was made of the descriptive statistics for the three variables under consideration—the mark awarded by the teacher (CD), the mean value of the marks given by each member of the group (CG) and the final joint mark from the co-assessment process (CC)—, the results shown in Table 3 were obtained.

Table 3. *Descriptive statistics according to the agents assessing the tasks*

| | Mark | CD - Teacher | CG - Members of group | CC - Joint mark |
|---------------|------------|--------------|-----------------------|-----------------|
| Task 1 | M | 7.93 | 8.15 | 8.18 |
| | SD | 0.84 | 0.59 | 0.64 |
| | Min | 6 | 7 | 7 |
| | Max | 9 | 7.25 | 9.25 |
| Task 2 | M | 7.5 | 7.76 | 7.8 |
| | SD | 0.76 | 0.66 | 0.67 |
| | Min | 6 | 6.5 | 6.5 |
| | Max | 9 | 9 | 9 |
| Task 3 | M | 6.93 | 7.33 | 7.17 |
| | SD | 1.04 | 0.91 | 0.98 |
| | Min | 5 | 6 | 5.5 |
| | Max | 9 | 9 | 9 |
| Task 4 | M | 7.26 | 7.68 | 7.35 |
| | SD | 1.75 | 1.44 | 1.76 |
| | Min | 3 | 3.67 | 3 |
| | Max | 9 | 9.1 | 9 |
| Global | M | 7.41 | 7.73 | 7.63 |
| | SD | 1.2 | 0.98 | 1.16 |
| | F | 2.61 | 1.91 | 2.63 |
| | p | .002** | .025* | .002** |

Note: *p< .05, ** p< .01; ***p< .001

Foreseeably, given other studies on student participation in the grading process (Acedo & Ruiz-Cabestre, 2011; Quesada et al., 2017), the lowest marks for the four tasks were awarded by the teachers, with a global mark of 7.41. This is 0.32 points below the individual marks given by the group members (M=7.73) and 0.22 points below the joint marks (M=7.63).

Although the teacher's mark was always lower, in tasks 1 and 2 (corresponding to teacher A), the final joint mark was higher than the mean value of the students' individual marks (+0.03 and +0.04), while for tasks 3 and 4 (corresponding to teacher A), exactly the opposite occurred (-0.16 and -0.33).

A review of the maximum and minimum marks for each sample show that they seem to be similar. In all cases, the minimum mark by the members of the group was higher than the teacher's mark. When each task is analysed, the minimum mark for task 4 stands out (3 out of 10). The standard deviation of the marks ranges from 0.59 to 1.76 points. There was greater

variability in the marks awarded for task 4 (Table 3) by the teacher, students, and both the latter jointly, with a standard deviation of between 1.44 and 1.76 points. This variability might be due to the greater complexity of the last task and to the period when it took place, which coincided with the end of the semester when the students had the heaviest workload. Both factors could lead to varying performance, with some groups getting high marks and others lower ones.

Table 3 shows that when Levene's test (F) was applied, there was no homogeneity of variance in the different marks awarded to each group (i.e. teacher's mark, the mean mark for the students and the joint mark).

Table 4. *Pearson's correlation coefficient*

| | Pearson's r | Sig. |
|---------------------|-------------|-------|
| Pair 1 Mean CD - CG | 0.893*** | < .01 |
| Pair 2 Mean CD - CC | 0.961*** | < .01 |
| Pair 3 Mean CG - CC | 0.955*** | < .01 |

Note: *p< .05, ** p< .01; ***p< .001

To examine the strength and direction of the association among the different marks, Pearson's correlation coefficient was calculated from the means of the marks for the four tasks. Table 4 shows the results, clearly demonstrating that the different marks are strongly and directly correlated, with a large effect size between the final mark and the mean mark for the members of the group ($r = .955$) and between the teacher's mark and the final mark ($r = .961$). From the evidence, there is therefore a close positive correlation among the different marks, with a simultaneous rise or fall depending on the tasks to be assessed and graded. In other words, it was demonstrated that all the agents assessing the tasks award a mark proportional to the standard of the submitted work.

If the marks are segmented by task and teacher, the following results are obtained when

Pearson's correlation coefficient is calculated (Table 5 & Table 6).

Table 5. Pearson's correlation coefficient, disaggregated by teacher

| | teacher A | teacher B |
|---------------------|-------------|-------------|
| | Pearson's r | Pearson's r |
| Pair 1 Mean CD - CG | 0.831*** | 0.909*** |
| Pair 2 Mean CD - CC | 0.892*** | 0.981*** |
| Pair 3 Mean CG - CC | 0.948*** | 0.956*** |

Note: * $p < .05$, ** $p < .01$; *** $p < .001$

In all the tasks, there is a high correlation among the three awarded marks. In task 4, a very strong correlation can be noted among the different marks, followed by task 3. When a review was made of the teachers' marks, a stronger correlation can be seen in the case of teacher B, who was in charge of correcting tasks 3 and 4.

Table 6. Pearson's correlation coefficient, disaggregated by task

| | Task 1 | Task 2 | Task 3 | Task 4 |
|----------------|-------------|-------------|-------------|-------------|
| | Pearson's r | Pearson's r | Pearson's r | Pearson's r |
| Pair 1 CD - CG | 0.837*** | 0.805*** | 0.850*** | 0.936*** |
| Pair 2 CD - CC | 0.924*** | 0.845*** | 0.951*** | 0.993*** |
| Pair 3 CG - CC | 0.919*** | 0.964*** | 0.933*** | 0.965*** |

Note: * $p < .05$, ** $p < .01$; *** $p < .001$

Analysis of the differences in the marks

To round off the study, the paired samples t test was used to compare the means of the marks for all four tasks. The mean marks are quite high (Table 3). As Table 7 shows, statistically significant differences were found between the mark awarded by the teacher and the individual marks for the students ($M(CD) = 7.41$; $M(CG) = 7.73$; $p < .05$; $d = 0.545$), with a medium

effect size, and between the mark awarded by the teacher and the jointly awarded mark ($M(CD) = 7.41$; $M(CC) = 7.63$; $p < .05$), with a medium-to-large effect size ($d = 0.670$). Nonetheless, no statistically significant differences were found between the joint mark and the individual marks for the students ($M(CC) = 7.63$; $M(CG) = 7.73$; $p > .05$), in this case with a very small effect size ($d = 0.148$).

Table 7. Paired samples t test for the mean marks for the four tasks

| | t | Sig. (bilateral) | Cohen's d | 95% CI Cohen's d | |
|--------------|-------|------------------|-----------|------------------|--------|
| | | | | Lower | Higher |
| Mean CD - CG | -5.14 | < .01*** | 0.545 | -0.775 | -0.324 |
| Mean CD - CC | -6.25 | < .01*** | 0.670 | -0.896 | -0.433 |
| Mean CG - CC | 1.55 | .12 | 0.148 | -0.045 | 0.378 |

Note: * $p < .05$, ** $p < .01$; *** $p < .001$

When a review is made of the disaggregated data and the differences in each of the tasks (Table 8), in three of the four tasks, statistically significant differences can be observed between the mark awarded by the teacher and the marks given by the members of the groups ($p < .05$). In task 4's case, although the teacher's mean mark ($M=7.26$) was still lower than the mean mark for the group of students ($M=7.68$), in contrast with the aggregate data, there are no statistically significant differences between both marks. However, there is a medium effect size ($p = .412$), as with tasks 1 and 2 ($d=0.458$ and 0.586 respectively), compared with a medium-to-large effect size for task 3 ($d=0.740$).

In all the tasks, there are statistically significant differences between the mark the teacher alone would have awarded (with mean marks from 6.93 to 7.93) and the agreed joint mark ($p < .05$) (with mean marks from 7.17 to 8.18). These differences have a medium-to-large effect size (Cohen's d of around 0.7) in the case the first three tasks and a medium effect size in the case of the last task ($d=-0.473$).

In contrast with the aggregate data, there is one task (task 3) where statistically significant differences were identified ($p < .05$), with a medium effect size ($d=0.453$) between the mark awarded by the members of the group ($M=7.33$) and the agreed joint mark ($M=7.17$).

Table 8. Paired samples t test for each task

| | t | df | Sig. 2-tailed | Cohen's d | CI 95% Cohen's d | |
|----------------|-------|----|---------------|-----------|------------------|--------|
| | | | | | Lower | Upper |
| TASK 1 | | | | | | |
| Pair 1 CD - CG | 2.15 | 21 | .04* | 0.458 | -0.893 | -0.013 |
| Pair 2 CD - CC | -3.40 | 21 | .003** | 0.725 | -1.189 | -0.247 |
| Pair 3 CG - CC | -0.64 | 21 | .53 | 0.137 | -0.555 | 0.285 |
| TASK 2 | | | | | | |
| Pair 1 CD - CG | -2.75 | 21 | .012** | 0.586 | -1.033 | -0.126 |
| Pair 2 CD - CC | -3.42 | 21 | .003** | 0.728 | -1.193 | -0.250 |
| Pair 3 CG - CC | -0.83 | 21 | .417 | 0.176 | -0.596 | 0.247 |
| TASK 3 | | | | | | |
| Pair 1 CD - CG | -3.39 | 21 | .003** | 0.722 | -1.187 | -0.245 |
| Pair 2 CD - CC | -3.47 | 21 | .002** | 0.740 | -1.206 | -0.260 |
| Pair 3 CG - CC | 2.13 | 21 | .046* | 0.453 | 0.009 | 0.888 |
| TASK 4 | | | | | | |
| Pair 1 CD - CG | -1.89 | 20 | .074 | 0.412 | -0.853 | 0.039 |
| Pair 2 CD - CC | -2.22 | 21 | .038* | 0.473 | -0.910 | -0.027 |
| Pair 3 CG - CC | 1.37 | 20 | .187 | 0.298 | -0.143 | 0.732 |

Note: * $p < .05$, ** $p < .01$

Lastly, when the statistical test was performed depending on the teacher in charge of each task, the differences between both could be seen, shown in Table 9. Statistically significant differences can be seen between each teacher's mean mark and the students' marks, with a medium effect size in each case (teacher A: $M(CD)=7.72$; $M(CG)=7.96$; $p = .001$; $d=0.526$; teacher B: $M(CD)=7.09$; $M(CG)=7.5$; $p < .001$; $d=0.572$). In both teachers' case, there are statistically significant differences between the mark awarded by the teacher and the joint mark, with a medium effect size in the case of teacher B and a medium-to-large effect size in the case of teacher A (teacher A: $M(CD)=7.72$;

$M(CC)=7.99$; $p < .001$; $d=0.731$; teacher B: $M(CD)=7.09$; $M(CG)=7.26$, $p < .001$; $d=0.607$). In both cases, the mean joint mark is higher than the mark that the teacher alone would have awarded.

As for a comparison of the individual student marks and the joint marks, in the case of teacher A, the mean marks for the students ($M=7.96$) are similar to the final mark ($M=7.99$;) and no statistically significant differences were found ($p=.317$; $d=0.153$). However, in teacher B's case, there were statistically significant differences between both means, with a lower joint mark and a small-to-medium effect size ($M(CG)=7.5$; $M(CC)=7.26$; $p = .02$; $d=0.37$

Table 9. Paired samples t test by teacher

| | t | df | Sig. 2-tailed | Cohen's d | CI 95% Cohen's d | |
|---------------------|-------|----|---------------|-----------|------------------|--------|
| | | | | | Lower | Higher |
| TEACHER 1 | | | | | | |
| Pair 1 Mean CD - CG | -3.49 | 43 | .001** | -0.526 | -0.839 | -0.208 |
| Pair 2 Mean CD - CC | -4.85 | 43 | < .001*** | -0.731 | -1.061 | -0.395 |
| Pair 3 Mean CG - CC | -1.01 | 43 | .317 | -0.153 | -0.449 | 0.146 |
| TEACHER 2 | | | | | | |
| Pair 1 Mean CD - CG | -3.75 | 42 | < .001*** | -0.572 | -0.892 | -0.247 |
| Pair 2 Mean CD - CC | -4.02 | 43 | < .001*** | -0.607 | -0.926 | -0.282 |
| Pair 3 Mean CG - CC | 2.47 | 42 | .02* | 0.376 | 0.064 | 0.683 |

Nota: *p< .05, ** p< .01; ***p< .001

Discussion and conclusions

This analysis confirms that the joint marks awarded during the co-assessment process and the teachers' marks are strongly and directly correlated. Even so, statistically significant differences were identified, with an effect size ranging from medium to large. Consequently, the marks are higher when they are jointly graded than when the teacher gives them.

The general means of the joint marks were +0.32 points higher than the ones that the teacher would have awarded, and they were -0.10 points lower than the mean individual student marks. This ties in with other studies of self-assessment marks—normally the highest kind of participatory assessment mark (Acedo & Ruiz-Cabestre, 2011; Quesada et al., 2016). However, in contrast with this finding, no statistically significant differences were identified between the joint marks and the mean marks for the individual group members, with a very small effect size.

When the data was disaggregated, only task 3 showed statistically significant differences between the mark that the individual group members would have awarded and the final joint one, with a medium effect size. This is why there are statistically significant differences between the mean mark for the group members and the final marks given by teacher B, with a small-to-medium effect size. It is important to reflect on this fact. This was the first task that teacher B assessed with the groups of students. As mentioned earlier, the students had already taken another subject with teacher A and they had some experience of

joint assessments and joint grading with him. Because the differences disappeared in the following task, the fourth—the only one where there were also no statistically significant differences between the mean mark for the group and the mark awarded by the teacher—, this could denote the existence of a learning curve during the co-assessment process, perhaps reflecting how the students gradually adapted to the teacher's way of assessing the tasks. These interpretations open up new fields of research into underlying learning processes, interactions, emotional and personal relations, and the teacher's group management skills in this kind of assessment process.

Despite the study's limitations in terms of the participants' lack of variety and the presence of statistically significant differences, the results are sufficiently interesting to reflect on their causes. The academic, social and economic consequences of the marks should probably not be overlooked but instead taken into account in order to gain an insight into the real influence of student participation in assessment processes, at least in terms of formal education at today's universities. To what extent should marks be adjusted when what is gained or lost is far more important than the cognitive consequences of the learning process? Despite efforts to systematize assessments and assessment literacy practices, can all the connotations of marks really be ignored? Even though student participation in the grading process is a source of problems, we do not believe that it is coherent or desirable to ignore this option, particularly when it comes to the possibility of transferring the logic of active

learning as an effective meaningful experience to the field of assessment. We agree with McArthur (2019) when she suggests that one possible solution would be to free assessments from the rigid confines of numerical marks, since this level of precision and differentiation does not fit in with the complex realities of most higher education tasks and, in turn, “it diverts attention from what really matters, which is the social application of this knowledge to foster greater individual and social wellbeing” (p.132).

At the same time, it is important to reiterate that, unlike former studies (Gómez & Quesada, 2017; Quesada et al., 2019), in this one the teachers started off by marking the tasks and then the students individually assessed their performance. Consequently, there was a prior stage of individual appraisals, and so the final mark was not just based on joint reflection and discussions. It is important to know how these modifications to a quantitative design in order to collect the necessary data might have positively or negatively influenced it or led to a series of limitations or biases which must be taken into account in future similar experiences.

Because the teacher had to assign and record a mark for each of the tasks prior to discussions with the students, it could feasibly have conditioned the process in one way or another or even have led to the same mark being awarded (although in our case at least, it was demonstrated not to have occurred). With this thought in mind, during joint assessment discussions, the teachers sometimes got the impression that the students were trying to guess what mark they had awarded instead of self-assessing the task.

Because each member of the group awarded an individual mark, a consensus on the joint mark was not always easily reached and, in some cases, it was suggested that the mean mark of their individual ones should be taken in order to settle the matter. This was reflected in the results, particularly in the case of teacher A, and this might explain the lack of any statistically significant differences between the

final marks and the marks given by the members of the groups who worked with this teacher, with ensuing repercussions on the general results.

In addition to the high correlation among the marks observed in this study, we believe that formative benefits like increased effort, motivation and engagement (Dochy et al., 1999), individual student reflections on their performance (Gómez & Quesada, 2017; Knight & Yorke, 2003; Quesada et al. 2019), and improved student self-confidence (Boud & Falchikov, 2006) are sufficient grounds to support this system of assessment and to call for ongoing scientific research. Given the results of our study, another key aspect to explore is the role of joint marks and participant attitudes in assessment processes. This could be analysed by recording the shared negotiated assessment sessions, whether they take place face to face or virtually by video-conference.

Lastly, in the design of co-assessment processes, in addition to technical and formal aspects, the issue of power relations must be analysed and considered, with the formulation of strategies to compensate for these imbalances, given the relevance of this factor when the process ends with a joint mark. The importance of acknowledging and tackling the complexities of assessments should not be under-estimated (Cooper, 2017), particularly when something as controversial and influential as grading in university education is involved.

References

- Acedo, M.A. & Ruiz-Cabestre, F.J. (2011). Una experiencia sobre la evaluación autónoma o participativa: autoevaluación y evaluación por los compañeros. *Arbor*, 187, 183-188. <https://doi.org/10.3989/arbor.2011.Extra-3n3142>
- Álvarez, J.M. (2001). *Evaluar para conocer, examinar para excluir*. Morata.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4^a Edición). McGraw-Hill-SRHE & Open University Press.
- Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning.

- Assessment & Evaluation in Higher Education*, 31(4), 399–413. <https://doi.org/10.1080/02602930600679050>
- Boud, D. & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-4013. <https://doi.org/10.1080/02602938.2015.1018133>
- Bretones, A. (2008). Participación del alumnado de Educación Superior en su evaluación. *Revista de Educación*, 347, 181-202. <http://www.educacionyfp.gob.es/dam/jcr:07d3f344-a3f9-4124-b4d1-c2b2d9f8bb27/re34709-pdf.pdf>
- Canabal, C. & Margalef, L. (2017). La retroalimentación: La clave para una evaluación orientada al aprendizaje. *Profesorado. Revista de currículum y formación del profesorado*, 21(2), 149-170. <http://hdl.handle.net/10481/47669>
- Carless, (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching international*, 57(66), 57-66. <https://doi.org/10.1080/14703290601081332>
- Casanova, M.A. (2011). Evaluación para la Inclusión Educativa. *Revista Iberoamericana de Evaluación Educativa*, 4(1), 79-89. <http://www.rinace.net/riee/numeros/vol4-num1/art4.pdf>
- Cooper, S. (2017). A collaborative assessment of students' placement learning, *Assessment & Evaluation in Higher Education*, 42(1), 61-76. <https://doi.org/10.1080/02602938.2015.1083093>
- Crisp, G.T. (2012) Integrative assessment: reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37(1), 33-43. <https://doi.org/10.1080/02602938.2010.494234>
- Deeley, S. (2014). Summative co-assessment: A deep learning approach to enhancing employability skills and attributes. *Active Learning in Higher Education*, 15(1) 39-51. <https://doi.org/10.1177/1469787413514649>
- Dochy, F., Segers M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>
- Falchikov, N. & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. <https://doi.org/10.2307/1170785>
- García, E., Gallego, B. & Gómez, M.A. (2015). Feedback and self-regulated learning: How feedback can contribute to increase students' autonomy as learners. En M. Peris-Ortiz & J.M. Merigó-Lindahl . *Sustainable Learning in Higher Education. Developing Competencies for the Global Marketplace* (113-130). Springer. https://doi.org/10.1007/978-3-319-10804-9_9
- Gallego, B., Quesada, V., Gómez, M.A. & Cubero, J. (2017). La evaluación y retroalimentación electrónica entre iguales para la autorregulación y el aprendizaje estratégico en la universidad: la percepción del alumnado. *REDU. Revista de Docencia Universitaria* 15 (1): 127-146. <https://doi.org/10.4995/REDU.2017.5991>
- Gómez, M.A. & Quesada, V. (2017). Coevaluación o Evaluación Compartida en el Contexto Universitario: La Percepción del Alumnado de Primer Curso. *Revista Iberoamericana de Evaluación Educativa*, 10 (2), 9-30. <https://doi.org/10.15366/riee2017.10.2.001>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hayward, L. (2015). Assessment is learning: the preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 27-43. <https://doi.org/10.1080/0969594X.2014.984656>
- Hernández, F. y Maquilón, J.J. (2010). Introducción a los diseños de investigación educativa. En S. Nieto (Ed.), *Principios, métodos y técnicas esenciales para la investigación educativa* (109-126). Dykinson.

- Hidalgo, N. & Murillo, F.J. (2016). Evaluación de Estudiantes para la Justicia Social. Propuesta de un Modelo. *Revista Internacional de Educación para la Justicia Social (RIEJS)*, 5(2), 159-179. <https://doi.org/10.15366/riejs2016.5.2.008>
- Ibarra, M.S. & Rodríguez, G. (2019). Evaluación como aprendizaje. En J. Paricio, A. Fernández & I. Fernández (Eds.), *Cartografía de la buena docencia universitaria. Un marco para el desarrollo del profesorado basado en la investigación* (175-196). Narcea.
- Knight, P. T. & Yorke, M. (2003). *Assessment, learning and employability*. SRHE/Open University Press/McGraw-Hill Education.
- Kurt, M. (2014). Collaborative Assessment: Fostering ownership in assessment. *Education*, 134(3), 332-339.
- Leach, L., Neutze, G. & Zepke, N. (2001). Assessment and Empowerment: Some critical questions. *Assessment and Evaluation in Higher Education*, 26(4), 293-305. <https://doi.org/10.1080/02602930120063457>
- López, V.M. (2009). *Evaluación formativa y compartida en Educación Superior. Propuestas, técnicas, instrumentos y experiencias*. Narcea.
- López, V.M. (2012). Evaluación formativa y compartida en la universidad: clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society & Education*, 4(1), 117-130. <https://doi.org/10.25115/psye.v4i1.485>
- López, V. & Sicilia, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*, 42(1), 77-97. <https://doi.org/10.1080/02602938.2015.1083535>
- McArthur, J. (2019). *La evaluación: una cuestión de justicia social. Perspectiva crítica y prácticas adecuadas*. Narcea.
- Murillo, F.J. & Hidalgo, N. (2015). Enfoques Fundamentales de la Evaluación de Estudiantes para la Justicia Social. *Revista Iberoamericana de Evaluación Educativa*, 8(1), 43-61. <https://revistas.uam.es/index.php/rie/article/view/2975>
- Nicol, D. (2010). From monologue to dialogue: improving written feedback in mass higher education. *Assessment and Evaluation in Higher Education*, 35(5), 501-517. <https://doi.org/10.1080/02602931003786559>
- Nicol, D. (2013). Peer review: putting feedback processes in students' hands. *Perspectives on Pedagogy and Practice*, 4, 111-123. http://uir.ulster.ac.uk/26926/1/PERSPECTIVE_ON_PEDAGOGY_AND_PRACTICE.pdf
- Penuel, W.R., & Shepard, L.A. (2016). Social models of learning and assessment. En A.A. Rupp & J.P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (146-173). John Wiley & Sons.
- Quesada, V., García, E. & Gómez, M.A. (2017). Student participation in assessment processes: A way forward. En E. Cano & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (228-249). Hershey: IGI Global. <https://doi.org/10.4018/978-1-5225-0531-0.ch012>
- Quesada, V., Gómez, M.A., Gallego, B. & Cubero, J. (2019). Should I use co-assessment in higher education? Pros and cons from teachers and students' perspectives. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2018.1531970>.
- Quesada, V., Rodríguez, G. & Ibarra, M.S. (2016). What are we missing? Spanish lecturers' perceptions of their assessment practices. *Innovations in Education and Teaching International*, 53(1), 48-59. <https://doi.org/10.1080/14703297.2014.930353>
- Quesada, V., Rodríguez, G. & Ibarra, M.S. (2017). Planificación e innovación de la evaluación en educación superior: la perspectiva del profesorado. *Revista de investigación educativa*, 35(1), 53-70. <https://doi.org/10.6018/rie.35.1.239261>

- Rodríguez, G. & Ibarra, M.S. (2015). Assessment as Learning and Empowerment: Beyond Sustainable Learning in Higher Education. En M. Peris-Ortiz & J.M. Merigó-Lindahl (Eds.), *Sustainable learning in higher education, innovation, technology, and knowledge management* (1-20). Springer-Verlag. https://doi.org/10.1007/978-3-319-10804-9_1
- Santiuste, V. & Arranz, M.L. (2009). Nuevas perspectivas en el concepto de evaluación. *Revista de Educación*, 350, 463-476. http://www.revistaeducacion.educacion.es/re350/re350_20.pdf
- Smith C.D., Worsfold, K., Davies, L., Fisher, R. & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education*, 38(1), 44-60. <https://doi.org/10.1080/02602938.2011.598636>
- Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433-449. <https://doi.org/10.1080/0969594X.2011.582838>
- Topping, K.J. (2003). Self and peer assessment in school and university: reliability, validity and utility. En M. Segers, F. Dochy & E. Cascallar (Eds), *Optimizing new modes of assessment: in search of qualities and standards* (55-87). Springer. https://doi.org/10.1007/0-306-48125-1_4
- Van der Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D. & Vanthournout, G. (2006). New assessment modes within project-based education - The stakeholders. *Studies in Educational Evaluation*, 32, 345-368. <https://doi.org/10.1016/j.stueduc.2006.10.005>
- Wood, P. y Smith, J. (2018). *Investigar en educación. Conceptos básicos y metodología para desarrollar proyectos de investigación*. Narcea.

Authors / Autores

Gómez-Ruiz, M.A. (miguel.gomez@uca.es)  0000-0003-2418-9525

Lecturer at the Faculty of Education at the University of Cadiz. He is an active member of EVALfor Research Group (Assessment in formative Contexts, SEJ-509) and researcher in the UNESCO Chair of Assessment, Innovation and Excellence at the University of Cádiz. His main research interests on assessment are: Learning-oriented assessment, participative assessment processes, and dialogical and co-assessment.

Quesada-Serra, V. (victoria.quesada@uib.eu)  0000-0002-8881-2358

Lecturer at the Faculty of Education at the University of the Balearic Islands. Active member of GIFES research group (Educational and Social Research and Training). Her main research interests on assessment are related to Learning-Oriented Assessment, participative assessment practices and co-assessment.



Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation
[ISSN: 1134-4032]



Esta obra tiene [licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).
This work is under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).