

Article

# An Analytical Perspective of the Reliability of Multi-Evaluation in Higher Education

Antonio Rodríguez Fuentes \* and José Luis Gallego Ortega

Faculty of Education Sciences, Campus of Cartuja s/n, University of Granada, 18071 Granada, Spain; jlgalleg@ugr.es

\* Correspondence: arfuentes@ugr.es; Tel.: +34-655993832

**Abstract:** It is urgent to evaluate the rest of the renewed elements within the university didactic action, overcoming the hegemony of traditional methods in which the professor constitutes as the sole evaluator. If autonomous and cooperative group-based learning is encouraged, self-assessment and co-assessment must also be promoted, apart from the traditional lecturing and evaluation by others. The assessing competence of Teacher Training degree students ( $n = 175$ ) was researched, started with stratified sampling (in the second and fourth years), following a participant selection process in each group. The compiled data were subject to descriptive, inferential, and correlation analysis by means of statistical software. The results pointed to low execution levels as for the self-evaluation (individual and group), although a certain progress was identified in the four year students compared to those in their second year of study. A better execution in evaluation was observed in all students regarding co-assessment (among different work groups in the classroom) and assessment by others (towards the professor). The use of all types of assessment is proposed, having a certain awareness and training regarding self-evaluation, and counting with a full supervision and control over it. All in all, the advantages of multiple and democratic assessment surpass the drawbacks derived from them.

**Keywords:** university assessment; meta-assessment; multiple-assessment; assessors; assessment by others; self-assessment; co-assessment

**Citation:** Rodríguez Fuentes, A.; Gallego Ortega, J.L. An Analytical Perspective of Reliability of Multi-Evaluation in Higher Education. *Mathematics* **2021**, *9*, 1223. <https://doi.org/10.3390/math9111223>

Academic Editor: David Pugalee

Received: 8 May 2021  
Accepted: 26 May 2021  
Published: 27 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Assessment is an essential, particular, and delicate element within the learning-teaching process, with manifesting implications on other elements (aim, content, and competence identification, as well as for method, grouping, activity, and resource selection). It constitutes the first, the last, and the recurring element [1]. Traditionally, assessment has been more associated to teaching and, therefore, to the teacher, following the accountability model. The instructor subjects his students to diverse assessment tests (assessment by others), oriented to student marking and intended to provide merits and penalties, objective passing, or continuation, such as, student promotion to the next school year or grade repetition. This assessment process must be inevitably complemented with the assessment on the teacher's own practice (self-assessment), as well as with a group assessment of the student's educational teacher team (teacher co-assessment), focused on the teaching process feedback [2]. In fact, both successful elements (proposed to be consolidated) and failures (understood as improvement opportunities) must be linked to the teaching process. They should also be assumed as a whole, despite still counting with individual and group implications. Furthermore, this assessment ought to go beyond the limits of the educational centers to merge with the socio-political, community, and family context, in a systemic perspective of the educational system (holistic assessment).

However, the educational efforts do not end with the abovementioned elements, disregarding students. On the one hand, it would be useful to combine the assessment by others, designed as an assessment of different agents (teachers assessing students or vice versa), with the assessment by the students about their own teachers. This would propitiate a two-way communication that balances the power forces within the classroom and would in fact empower the student body. The benefits are not only for the students but also for all the involved agents, as they are included in a personal and professional improvement process, both individually and as a group. Thus, this benefit is achieved not only for the educational center but for society as a whole. We must also add the students' assessment about themselves (student self-assessment) and about their peers (student co-assessment), in order to complete the transition to a learning-oriented assessment model [3]. All of that should be achieved in a constructive and continuous way, so that it will entail a new opportunity to learn, encounter new aspects (such as critical and self-critical abilities), as well as improve the teaching action after a reorientation, if needed, of the didactic process. In fact, there should not be a separation of spaces, moments, and aims to learn and assess.

Since long time ago, a democratic and participative assessment has been promoted; if commitment and involvement is requested from the students regarding their own learning process, they should be provided with responsibility in all their components, including their own assessment. The benefits derived from the self-assessment model are evident concerning the students' responsibility and planning. This is also true with respect to assessment by others, through co-assessment, as a shared task in the classroom, and as an opportunity to develop self-knowledge and awareness of their possibilities, talents, as well as being a critical and constructive [4] approach, one of active and autonomous work [5]. Blending these self and co-assessment modalities, regarding individual and even group efforts and work, perhaps becomes the only way to empower the students in their own learning process (individually and as a group), and to count with their absolute commitment and involvement. This is not inconsistent with an active, entertaining, motivational, participative methodology, based on the "learning to learn" and group work concepts; it is rather complementary with that, requesting active and responsible roles from the students likewise [6,7].

Another issue to be raised is assessment focus. Traditionally, it has been mainly cognoscitive and, occasionally, cognitive, prevailing a conceptual guideline. Nevertheless, it would be advisable to combine it with a procedural and attitudinal assessment, which would call for new processes, moments, spaces, and assessment tests, as well as for a new channel to conclude the assessing task, written and non-written, on-site and off-site. Among them, online digital platforms [1,8] can be pinpointed, which have already been employed for didactic tasks, but not for the assessing phase. This will bring about initial assessments, as well as continuous and procedural ones, in detriment of the traditional exclusive final or result assessment; that is to say, they should provide feedback but also feedforward on the process [9]. Nonetheless, there are still many practices and schools where the traditional assessment models are the ones mainly or exclusively employed [10]. Even though it is noted that every complementation effort with other models is praiseworthy [1,11,12], given the multiple aims of the assessment process, a multiple assessment is required, concerning results and efforts, moments and processes, techniques and instruments, spaces and contexts, etc. [8,13]. Perhaps the adjective "multiple" is a concept better describing the difference between the traditional unidirectional model and the alternate tripartite one, involving professor, student, and peers (Table 1).

**Table 1.** Polarized assessment models: traditional versus current.

Assessment	Traditional Model	Multiple Model
The purpose of the assessment	Marking and accountability	Process feedback and modification, when applicable
Assessing agent	Professor	Professor plus students and work groups
Assessed agent	Students	Individual and work teams
Nature	Assessment by others (“superior” towards “inferior”)	Shared (through self-assessment and group co-assessment)
Actions	Account for the expected results	Process-oriented in order to accomplish improvements
Nature	Summative and final	Procedural and continuous
Moment	Asynchronous	Asynchronous and synchronous
Place	On-site	Apart from on-site, off-site, and online
Content	Knowledge and results	Competences and possibilities
Custody	Only the assessing agent	Transparency, all agents participate in the process
Qualification	Quantitative	Qualitative and quantitative

This plural, tripartite and procedural assessment has been termed multiple, since it adds agents, moments, perspectives, places, aims, and new actions, combining them with the traditional ones. Other authors have called it socio-formative assessment [14]. This process includes, on the one hand, self-assessment, assessment by others and co-assessment; on the other hand, it involves a diagnostic, continuous and summative assessment. Other scholars [15] have named it comprehensive and collaborative due to the students’ involvement in it.

It is doubtless that this is the type of assessment requested by the new teachings within the European Higher Education Area [16] for universities in the European Union, focused on the development of the students’ competences [17]. That said, starting from that position, it is necessary to know if the educational system in general, and its agents in particular, are ready to perform such multiple assessment; this would require launching research processes in order to verify the agents and resources’ readiness. This would mainly apply to university students, given that the university professors’ expertise, as well as their optimal assessment competence are assumed in this study as axioms. In this occasion, research has been undertaken about the assessing competence requested from the students—in this case university undergraduates—regarding their own assessment and that of their classmates and professors. These experiences must but contextualized within particular contexts and levels, given the presumable differences among students. Hence, groups of students from intermediate years are compared with others about to graduate, in order to ponder the experimented progress. Research experiences on the students’ assessing competence are scarce, even among university students, who are presumed to possess a higher cognitive maturity and a more extensive development of the critical and constructive competence towards assessment. Thus, this line of research is considered original and necessary so as to keep on recommending alternative over traditional assessment. In other contexts, the threat of an absence or scarcity of training in such respect [18], as well as its subjectivity [14], has been observed.

## 2. Materials and Methods

This basic quantitative and non-experimental study has followed a cross-sectional descriptive and relational field design.

### 2.1. Aims and Hypotheses

The object stated in the previous section results in the following specific aims, derived from the general objective, which was to analyze the competence of Teacher

Training degree students to assess their own efforts and results, as well as those of their classmates and professors:

- Determine the assessing competence of the Teacher Training degree students (self-assessment, co-assessment, and assessment by others), evaluating their progress from their second to fourth year of study.
- Find out the reliability of self-assessment among the students in both levels, in order to evaluate its appropriateness of being employed.
- Research the suitability of self-assessment performed by the students on their classmates about the tasks carried out in practical seminars, for both levels.
- Analyze their ability to assess their professors, contrasting the assessment with that obtained in the subject, and establishing differences in the period from the second to the fourth year.

From these aims, the following null hypotheses are likewise derived:

- There is no correlation between the university students' expected marks in objective short answer tests and the marking given by their professors, which proves a good self-assessment competence for the students.
- The students within the study population do not demonstrate an optimal competence to assess other classmates, due to their independence regarding their belonging to other work groups or their own, and due to their coincidence in the marking assigned to the different groups.
- The students do not either exhibit an adequate competence to assess their professors. This can be inferred from the lack of coincidence between the marks assigned by the different students and the independence of such marking awarded to the professors.
- There is no relationship between the marks given by the students in their second year and those of the students in their fourth year of study, which is an indication of the students' lack of progress in their assessing competence.

## 2.2. Participants

A stratified sampling has been employed, in such a way that the second and fourth university years have been considered as strata. Within each stratum, a random participant selection process has been conducted. For the second year, the population rose to  $n = 155$ . The obtained sample was  $n = 124$ , which reveals an error rate of 3.9%. Regarding the fourth year, the population was reduced to  $n = 62$  and the study sample to  $n = 51$ , which correlates to a sampling error rate of 5.8%. For both cases, the error rate is close the commonly admitted 5% limit. Globally considered, the total population rose to 217 and the sample to 175, counting with a sampling error rate of 3.3%. The average age was  $M = 20.92$  ( $SD = 1.95$ ) and the predominant gender was the female one (75%).

Each year (stratum) has been considered as a case inferred to as a whole, even though intracase differences will be extracted according to the variables "gender" and "group". Nevertheless, the comparisons between "cases-years" (second and fourth years) will be the ones mainly performed, aimed at identifying discrepancies among them. Thus, this research can be considered as a comparative study.

## 2.3. Data Collection Procedure

Concerning the data collection process, participants from each university year were asked to assess their own effort in the subject (self-assessment) as well as that of the peers with whom they were working (intra-group assessment). They were also requested to assess the labor of the other work groups after watching their class presentations (inter-group assessment), and, finally, providing assessment on their professor too. This process has been a continuous assessment, as it has been carried out sequentially for every didactic unit included in the teaching program of the involved subjects. Such action was feasible thanks to the implication of the whole class group in the didactic units' tasks performed during the practical seminars of the corresponding subjects. In total, six di-

dactic units were involved in this study, resulting in six assessments for each student in every group. The only exception to this was the assessment on the professor, which was carried out globally before the students knew their final mark for the subject, so that they would not feel threatened by such an unusual variable.

Regarding the data collection instruments, the following were employed, and in the moments specified hereafter:

- When taking a written assessment test, based on both objective multiple choice and long answer questions, a section was created so that the students could indicate the mark they estimated for themselves. This estimation was compared with the actual mark given by the professor to measure its reliability. This measurement was only performed once, when taking the subjects' exam; this test was made up of the two tasks previously stated.
- Two assessment worksheets, duly validated by judges with a high unanimity degree (98%). They had to be filled individually, regarding the work carried out during the subjects' practical seminars [11] (p. 354), (a) intra-group assessment worksheet: personal self-assessment and assessment of the remaining group members (b) inter-group assessment worksheet: self-assessment of both the students' own work groups and all the other class groups.

#### 2.4. Data Analysis

The analysis was performed with the SPSS.v.22.0 software. It consisted in statistical calculations regarding central tendency of means ( $M$ ) and dispersion by standard deviation ( $SD$ ). It also carried out inferential analyses to identify the differences among means (Student's  $t$  and ANOVA's  $F$ ) and sizes, as well as calculations of the effect size (Cohen's  $d$ ) and relational size (Pearson's  $r$  correlation coefficient), in order to validate the relationships among the variables, All the above-mentioned analyses have been interpreted with a significance level of 0.05 ( $p < 0.05$ ), that is, accepting an error rate of 5%.

### 3. Results

The collected information has been clustered around the research's three aims, corresponding to the following sections.

#### 3.1. Predictive Ability Regarding Exam Marks

The presented data correspond to the measurement of the self-assessing competence displayed by the participants in this research. This measurement consists in the comparison of the actual marks obtained by each student in the different tests intended to provide a more formal or traditional mark. This was based on assessment criteria previously stated and negotiated between professor and students, with the marks estimated by the students for each of their abovementioned tests, following their own judgement.

As for the students' competence towards the assessment of objective tests, their estimations have been slightly higher than their actual marks for both levels (Table 2). Nevertheless, the students in their fourth year came closer in their estimations than those in their intermediate level, their dispersion being lower too. The differences between estimated and actual marks are only significant for the second year students, and moderate in size, as the values of  $d$  and  $t$  denote. Conversely, the relationship between both variables is only direct (positive) and intense (high) for the fourth year students, according to Pearson's coefficient. Consequently, it must be deduced that the students in their fourth year possess a sufficient competence for this dimension of objective assessment. This is not the case for the second year students, who exhibit an over-estimation concerning their actual marks.

On the other hand, the students present a lower competence when estimating their written tests' marks. Even though the differences between estimated and actual marks are not significant for the students in their fourth year (as indicated by the  $t$  on Table 2),

the correlation between such marks is significant and direct, but only with a medium intensity (according to the correlation coefficient on the table). Conversely, there are significant differences for the second year students, moderate in size (according to the values of  $t$  and  $d$ , respectively) between these marks' estimation and the actual ones. Therefore, it can be confirmed that they do not display an optimal ability to assess non-objective tests, which could be explained by their higher subjectivity compared to the other group of students.

**Table 2.** Analysis of the students' predictive competence regarding test self-assessment.

		Multiple Choice Test		Long Answers Test	
		Estimated	Actual	Estimated	Actual
$M (SD)$	2nd year	6.7 (1.85)	5.63 (2.35)	6.76 (2.53)	4.95 (2.68)
	4th year	6.80 (1.94)	6.08 (2.01)	7.81 (1.59)	7.06 (2.07)
$t$ between estimated mark and actual mark	2nd year	1.58 *		1.86 **	
	4th year	0.33		0.66	
$d$ between estimated mark and actual mark	2nd year	0.36		0.41	
	4th year	0.24		0.26	
$r$ between estimated mark and actual mark	2nd year	0.2 *		0.19 *	
	4th year	0.69 **		0.49 *	

Note:  $t$  = Student's  $t$ ;  $d$  = Cohen's  $d$ ;  $r$  = Pearson's  $r$  correlation coefficient; \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

All in all, based on the data about actual and estimated marks, optimal assessing competences were expected, which were not very dissimilar between the students. Thus, a high correlative coefficient was anticipated, since among the performed tests, one was objective (multiple choice with four possible answers) and the other one requested for short answers (with exhaustive indicators about the nature of the answers). The competence did not prove to be high among the students in their fourth year, rather average, as it is reflected in the correlation coefficients ( $r = 0.69$  and  $0.49$ , respectively). However, that was the case for the second year students, for whom the coefficient was miniscule ( $r = 0.2$  and  $0.19$ , respectively). A certain progress along the academic year is observed, even though the final result is not very satisfactory, given the significance of this competence and the obtained values.

### 3.2. Co-Assessment Regarding Group Work during In-Class Practical Sessions

Two co-assessments tasks are considered for this occasion; one of them concerning the effort and results within their own groups, in which the members in each group agree on a mark about themselves (joint group self-assessment or intra-group). The other co-assessment task relates to the consensual group marking that they deliver for all the other groups in the classroom after they present their work (inter-group). Hence, group variations are identified (means for their groups against means of the different groups). In an optimal framework for the development of the assessing competence, the above-mentioned marks should be independent and unrelated, as it is deduced from the logical exercise.

First of all, interpreting the data on Table 3, it must be pinpointed that the self-assessment's mean for the students in their second year is higher than that of the other groups. This divergence is significant and large in size (according to the values of the  $t$  and  $d$  statistics). The overestimation about their own effort compared to that given to others is still happening. However, this is not true for the students in their fourth year, whose means are almost identical, being high in both cases. Besides that, the existence of any correlation between them should not be expected to prove a good assessing competence, as they are non-related efforts; the existing relationship is a median one, which constitutes an encouraging fact. Another positive aspect is the fact that there is not much dispersion among the inter-group assessments and, complementarily, the elements with

the highest marks are given such values by everybody and vice versa. Indeed, no differences were found among the inter-group assessments, according to the ANOVA calculations, neither for the students in their second year ( $F = 0.36, p = 0.162$ ) nor for those in their fourth year ( $F = 0.27, p = 0.09$ ). Finally, it also speaks favorably for the students' co-assessing competence the fact that the differences among the average marks awarded to other groups and those given by the professor were not significant ( $p > 0.05$ ), according to the Student's  $t$ -calculation, considering both groups together and separately. Furthermore, there was a relationship between the marks from the student groups and from the professor, measured through Pearson's correlation coefficient. Conversely, the differences between the marks coming from self-assessments by the groups and by the professor reached statistical significance for all the cases ( $p > 0.05$ ).

**Table 3.** Co-assessment analyses regarding their own groups (self-assessments) and all the other groups.

		Group Marks		Marks Given by the Professor to the Group Works
		Given to Their Own Groups (Self)	Given to the Other Groups	
$M (SD)$	2nd year	8.46 (0.52)	7.79 (0.51)	7.42 (0.97)
	4th year	8.53 (0.85)	8.51 (0.65)	8.30 (0.88)
$t$ between own_groups_marks and other_groups'_marks	2nd year	1.23 ***		0.76
	4th year	0.05		0.29
$d$ between own_groups_marks and other_groups'_marks	2nd year	0.92		0.15
	4th year	0.02		0.09
$r$ between own_groups_marks and other_groups'_marks	2nd year	0.47 **		0.71 **
	4th year	0.42 *		0.83 ***

Note:  $t$  = Student's  $t$ ;  $d$  = Cohen's  $d$ ;  $r$  = Pearson's  $r$  correlation coefficient; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

### 3.3. Marks Awarded to the Professor

Lastly, in order to judge the future teachers' competence development concerning the assessment towards their current professors, the two marks expected from the mastery of such competence independently were confronted: on the one hand, the final mark in the subject, resulting from several previous assessments; on the other hand, the quantitative assessment awarded by the students to the professor (Table 4).

**Table 4.** Analyses of the students' final marks and marks awarded to the professor.

		Mark – Professor Assessment	
		Student's Mark	Mark Awarded to the Professor
$M (SD)$	2nd year	6.92 (2.25)	8.93 (1.98)
	4th year	7.72 (2.28)	8.91 (1.90)
$t$ between mark_from professor to_students and vice versa	2nd year	-3.71 **	
	4th year	-1.27 **	
$d$ between mark_from professor to_students and vice versa	2nd year	-0.89	
	4th year	0.77	
$r$ between mark_from professor to_students and vice versa	2nd year	0.69	
	4th year	0.46	

Note:  $t$  = Student's  $t$ ;  $d$  = Cohen's  $d$ ;  $r$  = Pearson's  $r$  correlation coefficient; \*\*  $p < 0.01$ .

The variable "university year" did not seem to be determined in this case. In fact, no differences could be found between the assessments awarded to the professor from the second and fourth year students ( $t (123) = 0.15; p = 0.001$ ), which was interpreted favora-

bly, as they both refer to the same professor. Moreover, differences arose for both groups between their assessments awarded from and to the professor, as conveyed by the value of  $t$ . These divergences were significant, as it is added by the value of  $d$ . This was interpreted favorably for their assessing competence, since it indicates the independence between both marks. Finally, the correlation between marks ( $p > 0.05$ ) was not significant, as it is inferred from the calculation of  $r$ . Again, this implies positive data, in terms of independence and objectivity.

#### 4. Discussion and Conclusions

This research is intended to verify the university students' assessing competence by means of different evaluations within a multiple assessment model. Hence, individual self-assessments were carried out through objective written tests, by means of group self-assessments of both their own work groups and the other groups within the classroom and, finally, through assessment towards their professor. The study intended to lay solid foundations for an improvement in multiple assessment, combining the traditional heterogeneous assessment from the teacher with a continuous and multiple assessment. The latter would provide actual learning feedback and would allow for decisions to be made leading to teaching improvement. Firstly, it must be stated that the students always exhibited good disposition and acceptance when faced with this assessment model. This fact has been interpreted as a claim for a higher share of responsibility regarding their assessment, as well as empowering their own learning process, in line with the findings obtained by [8,19]. The students also admitted the importance of experience for their own training to becoming teaching professionals, which had also been previously demonstrated [11].

Nonetheless, the implication and their development of the multiple assessment do not justify per se the use of these assessment modalities; it was required to check their reliability first. For starters, it can be inferred that the participants in this research did not exhibit unanimously an optimal self-assessing competence. It was appreciated and exacerbated generosity regarding their individual self-assessment in tests, as was revealed by [13], as well as for their group practical exercises conducted in the classroom, as pinpointed by [18]. Their assessment lacked objectivity, as occurred in other studies [14]. However, it would be fair to admit that such self-assessments were closer to reality for the more mature students in their fourth year. It is also true that the students did not have the opportunity to carry out tasks to participate proactively and reactively in the self-assessment, a situation which is stated by [20]. Additionally, the authors of [8] provide a thread of hope as they conclude that the results improve in the future phases of this self-assessing process. This evidence emphasizes the value of experience too.

Nevertheless, they showed a better competence when assessing other groups, as revealed by the coincidence between the marks given mutually among groups and those awarded by the professor. This means that they successfully acquired their co-assessment competence. Some problems were identified related to this shared assessment, although they were overcome thanks to the students' potential, as it happened in the study by [21]. As an epilogue, an adequate competence concerning the assessments awarded to the professor's performance was identified. It was detected thanks to the coincidence between both classes and their independence from the marks received by him. These data concur with the findings by [21] about the concordance in the marks from professor and students, which highlights the students' optimal assessing skills. Studies such as those performed by [3,21] reveal divergences in the marks awarded by professors and those that students consider they deserve. This point, far from being negative, justifies the complementarity of this assessment model with more traditional ones.

Evidence has also been uncovered as for the differences between younger and more mature students, which pinpoints the suitability of this assessment model for the students in the degree's last year. It also emphasizes the need for a certain degree of training, testing, and monitoring concerning the self-assessments of students in the initial years.



The use of guiding instruments and the previous experimentation with tasks related to this self-assessment can guarantee the reliability of the obtained results. Likewise, it would be positive to insist more explicitly on the importance of self-assessment and its effects, as well as experiment in both the university (practical seminars) and lower level (school internships) classrooms and ponder on their adequacy and impact. Some studies [14] report an improvement on the students' academic achievement and their availability to assume this self-assessing role. Nevertheless, the process is not exempt from difficulties [14], as this research has revealed. These difficulties not only appear in the assessing process but also in the calculation of the multiple assessments; therefore, the use of specific platforms, designed by some authors [1], is advised. The bibliometric study conducted by [22] warns that it is also required to elucidate the typologies, models, and specific assessment practices alternative to the traditional ones by means of a formal exam. This is especially true for peer assessment and co-assessment [22]. Lastly, it would also be beneficial to make the students aware of the responsibility and commitment required to assess, as the last and first teaching competence. This would entail the development of both analytical and procedural abilities underlying in any kind of assessment. All in all, the benefits outweigh the difficulties and even the limitations, and this is perceived by the students as well [14,20]. Besides, this combination proves to be functional and satisfying, quoting [15].

Despite the assumed suitability of the identified strategies as irrefutable improvements of the students' self-assessment ability, and the excellent impact on their professional training and academic achievement [14], it would also be convenient to extend this research topic, modifying participants, contexts, designs, and analyses. For instance, it could be employed longitudinal designs instead of the cross-sectional one used here. Thus, it would study the self-assessment abilities of the same group of students throughout all their university years, in order to remove certain involved variables. Another option would be to expand the study to students in their first and third year of the Teacher Training degree, with the purpose of following the analytical progress experimented from the beginning until the end of their university period. An alternative procedure could be increasing the participant groups and professors, aimed at identifying the incidence of the factor "professor" on the research results. Besides, the data analyses could be extended by adding different variables, colleges, universities, and contexts, using other programs and resources. Even the moments and procedures could be further studied, trying to collect data on marks awarded to the professor before and after that assessment has been carried out. This would intended to find out about the actual incidence of this aspect. Finally, employing other instruments to collect and analyze data, —even those being of a different sort, less nomothetic, and more idiographic—could be important for future studies. This would be achieved through content analysis of interviews, observation, and task or error analyses, among others. Hence, the limitations of this study, which go beyond the space assigned for its presentation, could be hurdled by conducting the abovementioned studies.

**Author Contributions:** Conceptualization, A.R.F. and J.L.G.O.; methodology, A.R.F. and J.L.G.O.; formal analysis, A.R.F. and J.L.G.O.; investigation, A.R.F. and J.L.G.O.; resources, A.R.F. and J.L.G.O.; data curation, A.R.F. and J.L.G.O.; writing—original draft preparation, A.R.F. and J.L.G.O.; writing—review and editing, A.R.F. and J.L.G.O.; visualization, A.R.F. and J.L.G.O.; supervision, A.R.F. and J.L.G.O.; project administration, A.R.F.; funding acquisition, A.R.F. and J.L.G.O. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Unidad de Calidad, Innovación y Prospectiva (Quality, Innovation and Foresight Unit) of the University of Granada as a Teaching Innovation Project.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Alain, L.; Gallego, J.L.; Navarro, A.; Rodríguez, A. Evaluation for Teachers and Students in Higher Education. *Sustainability* **2020**, *12*, 4078, doi:10.3390/su12104078.
2. Boud, D.; Falchikov, N. *Rethinking Assessment in Higher Education. Learning Longer Term*; Routledge: Oxfordshire, UK, 2017.
3. Saiz, M.C.; Bol, A. Aprendizaje basado en la evaluación mediante rúbricas en educación superior. *Suma Psicol.* **2014**, *21*, 28–35. Available online: <http://www.sciencedirect.com/science/article/pii/S0121438114700049> (accessed on 13 January 2019).
4. Gallego Noche, B.; Quesada, V.; Gómez, M.A.; Cubero, J. La evaluación y retroalimentación electrónica entre iguales para la autorregulación y el aprendizaje estratégico en la universidad: La percepción del alumnado. *Rev. Educ. Docencia Universtaria REDU* **2017**, *15*, 127–146, doi:10.4995/redu.2017.5991.
5. Barriopedro, M.; López, C.; Gómez, M.A.; Rivero, A. La coevaluación como estrategia para mejorar la dinámica del trabajo en grupo: Una experiencia en Ciencias del Deporte. *Rev. Complut. Educ.* **2016**, *27*, 571–584. Available online: <http://revistas.ucm.es/index.php/RCED/article/view/46811> (accessed on 1 May 2020).
6. Carrizosa, E.; Gallardo, J.I. Autoevaluación, Coevaluación y Evaluación de los aprendizajes. Ponencia presentada en III Jornadas sobre docencia del Derecho y TICs. In III Jornadas Sobre Docencia del Derecho TIC; 2015, pp. 1–14. Available online: [http://www.uoc.edu/symposia/dret\\_tic2012/pdf/4.6.carrizosa-esther-y-gallardo-jose.pdf](http://www.uoc.edu/symposia/dret_tic2012/pdf/4.6.carrizosa-esther-y-gallardo-jose.pdf) (accessed on 25 May 2020).
7. Maureira-Cabrera, O.; Vásquez-Astudillo, M.; Garrido-Valdenegro, F.; Olivares-Silva, M.J. Evaluación y coevaluación de aprendizajes en blended learning en educación superior. *Alteridad* **2020**, *15*, 174–189, doi:10.17163/alt.v15n2.2020.04.
8. Rodríguez Espinosa, H.; Retrepo, L.F.; Luna, G.C. Percepción del estudiantado sobre la evaluación del aprendizaje en la educación superior. *Educ. Electron. J.* **2016**, *20*, 1–17, doi:10.15359/ree.20-3.18.
9. García-Jiménez, E. La evaluación del aprendizaje: De la retroalimentación a la autorregulación. El papel de las tecnologías. *RELIEVE* **2015**, *21*, 1–24, doi:10.7203/relieve.21.2.7546.
10. Lisi, E.G. El Portafolios como herramienta para la evaluación de las competencias de los alumnos europeos. *Rev. Meta Avaliação* **2016**, *8*, 534–547. Available online: <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/1115> (accessed on 5 June 2019).
11. Gallego Ortega, J.L.; Rodríguez Fuentes, A. Alternancia de roles en la evaluación universitaria: Docentes y discentes evaluadores y evaluados. *Rev. Educ. Docencia Universtaria REDU* **2017**, *15*, 349–366, doi:10.4995/redu.2017.8780.
12. Pérez, R. ¿Quo vadis, evaluación? Reflexiones pedagógicas en torno a un tema tan manido como relevante. *Rev. Investig. Educ. RIE* **2017**, *34*, 13–30, doi:10.6018/rie.34.1.239381.
13. Ibarra, M.S.; Rodríguez Gómez, G. Modalidades participativas de evaluación. Un análisis de la percepción del profesorado y de los estudiantes universitarios. *Rev. Investig. Educ. RIE* **2014**, *32*, 339–361, doi:10.6018/rie.32.2.172941.
14. Cardona, S.; Vélez, J.; Tobón, S. Contribución de la evaluación socioformativa al rendimiento académico en pregrado. *Educar* **2016**, *52*, 423–447, doi:10.5565/rev/educar.763.
15. Fernández Sotelo, A.; Vanga Arévalo, M. Proceso de autoevaluación, coevaluación y heteroevaluación para caracterizar el comportamiento estudiantil y mejorar su desempeño. *Rev. San Gregorio* **2015**, *1*, 6–15, doi:10.36097/rsan.v1i9.52.
16. Rodríguez Fuentes, A.; Caurcel, M.J.; Ramos, A.M. *Didáctica en el EEES: Guías de Trabajo Autónomo*; Editorial EOS: Madrid, Spain, 2008.
17. Mateo, J.; Vlachopoulos, D. Evaluación en la universidad en el contexto de un nuevo paradigma para la educación superior. *Educación XXI* **2013**, *16*, 183–208, doi:10.5944/educxx1.16.2.2639.
18. Mok, M.M.C.; Lung, C.L.; Cheng, D.P.W.; Cheung, R.H.P.; Ng, M.L. Self-assessment in Higher Education: Experience in using a Metacognitive Approach in five case studies. *Assess. Eval. Higher Educ.* **2017**, *31*, 415–433, doi:10.1080/02602930600679100.
19. Gessa, A. La coevaluación como metodología complementaria de la evaluación del aprendizaje. Análisis y reflexión en las aulas universitarias. *Rev. Educ.* **2011**, *345*, 749–763. Available online: [http://www.revistaeducacion.educacion.es/re354/re354\\_30.pdf](http://www.revistaeducacion.educacion.es/re354/re354_30.pdf) (accessed on 26/05/2021).
20. Gómez Ruiz, M.A.; Quesada, V. Coevaluación o Evaluación Compartida en el Contexto Universitario: La Percepción del Alumnado de Primer Curso. *Rev. Iberoam. Evaluación Educ.* **2017**, *10*, 9–30, doi:10.15366/rie2017.10.2.001.
21. Lukas, J.F.; Santiago, K.; Lizasoain, L.; Etxebarria, J. Percepciones del alumnado universitario sobre su evaluación. *Bordón* **2017**, *69*, 103–122, doi:10.13042/Bordon.2016.43843.
22. Rodríguez, G.; Ibarra, S.M.; García, J.E. Autoevaluación, evaluación entre iguales y coevaluación: Conceptualización y práctica en las universidades españolas. *Rev. Investig. Educ.* **2013**, *11*, 198–210. Available online: <http://reined.webs4.uvigo.es/index.php/reined/article/view/183> (accessed on 26/05/2021).