

Article

Effectiveness of Temporal Survey-Based Programs for Teaching Critical Thinking Skills in Engineering Courses: Analysis of Final Assessments

Víctor Revilla-Cuesta ¹, Manuel Hernando-Revenga ¹, Ismael Martín ², Marta Skaf ^{2,*} and Vanesa Ortega-López ¹

¹ Department of Civil Engineering, University of Burgos, 09001 Burgos, Spain; vrevilla@ubu.es (V.R.-C.); mhrevenga@ubu.es (M.H.-R.); vortega@ubu.es (V.O.-L.)

² Department of Construction, University of Burgos, 09001 Burgos, Spain; ismaelp@ubu.es

* Correspondence: mskaf@ubu.es; Tel.: +34-947-259399

Abstract

Peer- and self-critical skills are key to properly performing engineering work, as they allow engineering students to develop critical thinking regarding the quality standards required in this professional field. This research aimed to determine whether educational experiences based on temporal survey-based programs enabled the successful development of these skills in students enrolled in six courses in the final years of their engineering degrees. To this end, an educational experience of such a nature was implemented throughout a complete academic year, aimed at fostering peer- and self-critical skills through continuous formative assessment. The experience involved six student presentations evaluated by both teachers and peers using a Likert-scale survey encompassing four dimensions: explanatory ability, file quality, attitude, and overall assessment. Subsequently, these assessments were provided to students to encourage reflection on the scores assigned and their own work. The results revealed strong alignment between teacher and peer evaluations, with average deviations below 7%, demonstrating effective development of peer-critical competences. These results were also verified by means of analyses of variance. The greatest consistency was found in “explanatory ability” and “overall assessment,” while “file quality” and “attitude” showed wider variability, experience playing a key role in their precise evaluation. Peer evaluations tended to be more uniform than teachers’, reflecting students’ limited experience in discerning subtle performance differences. Additionally, 30% of students expressed willingness to repeat their final presentation to achieve a higher grade, evidencing substantial self-critical reflection. Qualitative analysis conducted through deductive cross-coding indicated that this motivation stemmed from both intrinsic self-improvement and peer-related responsibility. Overall, the results confirm that sustained peer- and self-assessment activities can effectively cultivate critical thinking skills among engineering students, although continuous practice is required to consolidate these competences. Future research could explore the more adequate course types, and students’ ages to perform such kind of educational experiences.

Keywords: university teaching; engineering education; critical thinking; self-criticism; peer-to-peer criticism; temporal survey; itemized assessment; Likert scale; statistical analysis; qualitative analysis



Academic Editors: Jordi Solbes and Teresa Lupión-Cobos

Received: 8 January 2026

Revised: 3 February 2026

Accepted: 11 February 2026

Published: 13 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Competency-Based University Teaching and Its Link to Engineering

University education must comprehensively prepare students across all dimensions required for the development of their future professional practice. Broadly, these dimensions can be classified into two main categories. The first is the knowledge dimension, which encompasses the acquisition of theoretical concepts, analytical procedures, technical skills, and other disciplinary tools that enable graduates to perform their professional roles effectively from an operational standpoint (Annala, 2023). The second is the competence dimension, referring to the development of transversal and social skills such as teamwork, critical thinking, effective communication, and public speaking, among others (Bezanilla et al., 2019). Contemporary educational paradigms increasingly emphasize competence-based learning and assessment, reflecting a shift from traditional content-oriented approaches toward holistic student development (Moreira et al., 2023). This trend is evident in the design of university curricula, where every course is explicitly linked to a set of professional and transversal competences (Cano et al., 2023). Furthermore, competence-based assessment has progressively gained prominence in pre-university education, in some cases exerting a greater influence on academic evaluation than knowledge acquisition alone (Ponomariovieniè et al., 2025). However, there are bibliographical gaps that show that it is still necessary to delve deeper into the effective development of some competencies in certain fields such as engineering education.

Effective engineering practice requires the acquisition and development of such competences, as engineering work is inherently collaborative (Cruz et al., 2020). The design and development of engineering projects are typically carried out within multidisciplinary teams composed of professionals with expertise in different technical areas (Kolmos et al., 2024). However, an engineering project should not be understood as a mere aggregation of independent parts; rather, the various components must be coherently integrated and interrelated to form a consistent whole (Kolmos et al., 2024). Moreover, project management constitutes a fundamental professional responsibility of engineers, who must be able to effectively coordinate and supervise multiple teams involved in the implementation phase (Cerezo-Narváez et al., 2019). Given the critical importance of these skills in professional engineering practice, a high number of studies and research efforts have focused on identifying and promoting educational strategies that foster the development of such competences among engineering students (González-Marcos et al., 2016).

To address this educational challenge, higher education institutions have progressively adopted active learning methodologies aimed at promoting both technical mastery and transversal competences among engineering students (Wang et al., 2023). Approaches such as project-based learning, cooperative learning, and problem-based learning have demonstrated significant effectiveness in fostering teamwork, communication, and critical thinking skills (O'Connor et al., 2025), while maintaining a strong focus on technical problem-solving (Revilla-Cuesta et al., 2020). These methodologies place students at the center of the learning process, encouraging autonomy, creativity, and the practical application of theoretical knowledge in real or simulated professional contexts (Ríos et al., 2010).

1.2. Critical Thinking Competency

One competence that is often overlooked but is of great importance in engineering practice is the development of critical thinking regarding the work performed. Although engineering tasks are generally supervised at different stages, the quality of the outcome frequently depends solely on the individual responsible for carrying it out (Litzinger et al., 2011). Therefore, developing the ability to conduct a critical assessment of both one's own work and that of colleagues ensures that professional activities are performed in the most

appropriate and effective manner (Turns et al., 2014). Engineers must possess the skills required to evaluate the work of others objectively and fairly, and to communicate this feedback appropriately, highlighting deficiencies or potential areas for improvement when necessary (Shuman et al., 2005). In addition, they should be able to apply the same level of scrutiny to their own performance, fostering a continuous process of self-assessment and professional growth (Litzinger et al., 2011).

Scientific research has shown that teachers are perceived as authority figures by students (van Gennip et al., 2010). As a result, corrections and assessments made by them are often interpreted as orders or actions to be carried out, rather than as learning opportunities for reflection on one's own work (Nicol & MacFarlane-Dick, 2006). Peer assessment, or co-evaluation, is a teaching methodology in which students evaluate and provide feedback on the work of their classmates (Sluijsmans et al., 2002). Although these evaluations do not contribute directly to the final grade of the course, they enable students to reflect critically on their own performance, since the feedback comes from individuals at the same level, without an authority position (van Gennip et al., 2010). This process can foster a self-critical attitude, encouraging students to seek ways to improve the quality of their work (Liu & Carless, 2006).

However, conducting a one-off peer assessment—that is, one which takes at a specific point in time without being clearly integrated into the development of the course—may lead students to approach it without the necessary seriousness and rigor (K. J. Topping, 2010). Nevertheless, it has been demonstrated that when peer assessment is carried out periodically and effectively integrated into the course structure, its usefulness increases significantly (van Zundert et al., 2010). During the first implementation, students may not perform the activity adequately, but over time, they tend to develop greater reflective skills with respect to the work of their peers, and the initial reluctance to critique peer's work and the tendency toward friendliness diminishes (Panadero et al., 2013). Thus, peer assessment applied over time helps to cultivate two essential competences in the engineering field. The first is peer evaluation skills, enabling students to assess their classmates' work accurately and objectively (K. J. Topping, 2010). The second is self-critical capacity, as repeated feedback from peers fosters a natural tendency to improve and to address the comments received (Nicol, 2021).

1.3. Research Approach and Objective

In a previous study (Revilla-Cuesta et al., 2024), the authors of this research conducted a temporal survey-based program aimed at helping engineering students develop critical thinking skills regarding both their own work and that of their peers. Over the course of a semester, students delivered a series of in-class presentations in which they evaluated their peers' work, which was also assessed by the teachers. The evaluated aspects included explanatory ability, the quality of the digital file (PowerPoint presentation), and attitude, in addition to providing an overall assessment. Student assessment of these dimensions is a pedagogical mechanism that activates fundamental critical thinking processes essential to engineering practice. Evaluating explanatory ability engages students in argumentation analysis, evidence appraisal, and causal reasoning (Barta et al., 2022). Assessing file quality requires the application of disciplinary standards and analytical judgment (Deo & Hölttä-Otto, 2024). The evaluation of the attitude relates to the dispositional dimension of critical thinking, including intellectual responsibility and openness to revision (Hong et al., 2021). Finally, providing an overall assessment demands integrative judgment and metacognitive regulation (Deo & Hölttä-Otto, 2024).

The results of this study (Revilla-Cuesta et al., 2024) showed that explanatory ability was the dimension in which students applied the highest level of rigor when evaluating their peers, due to the ease with which this aspect could be assessed without the need for

prior training. In contrast, overall assessment was the dimension where the greatest initial overrating was detected, likely due to its similarity to the concept of a “grade”, which made students more reluctant to initially apply the same level of rigor as they would in other dimensions. Nonetheless, overall assessments became more stringent over time. In general, this initial experience demonstrated the effectiveness of such programs, even leading in some cases to situations where students’ evaluations were stricter than those of the teachers.

Drawing on this established model, the present paper analyzes the final outcome of an extended, long-term experience, conducted over a full academic year applied to a new cohort of students. Specifically, the study analyzes the results of a final survey, assessing the development of students’ peer-review competence by comparing the evaluations provided by students and teachers, as well as their self-assessment competence through the consideration of their willingness to repeat and improve their own work. Therefore, this paper aims to serve as a validation tool for this type of procedure in engineering courses, which can be easily integrated into classroom activities to ensure adequate training of students in essential engineering professional skills (Litzinger et al., 2011; Shuman et al., 2005).

2. Materials and Methods

2.1. Program

The students who participated in this research were enrolled in the 4th year of the Bachelor’s Degree in Civil Engineering and in the 3rd and 4th years of the Bachelor’s Degree in Agri-Food Engineering and the Rural Environment at the University of Burgos, Spain, during the 2024/2025 academic year. Each participating student was registered in two courses belonging to the same year of the same Bachelor’s Degree, taking one course during the first semester and the second during the second semester. This structure facilitated a longitudinal study with the same group of students throughout an entire academic year, rather than a single semester, allowing the teaching experience to be developed over a longer period. According to the available literature, peer-assessment experiences are more effective when implemented over extended timeframes, which allow the minimization of the subjective factors that affect peer evaluations, thus justifying the approach adopted in this study (Tzeng et al., 2021). For this purpose, proper coordination among the teachers of the courses was essential (Törlind et al., 2023).

Each student gave three oral presentations per course—that is, three per semester—in front of their classmates and the teacher, with an approximate duration of 10 min each. The topic of each presentation was related to the subject matter of each course, defined individually by each teacher. However, their general approach was consistent across courses, requiring students to search for, collect, and synthesize information. Therefore, the degree of learning and understanding of course concepts was not directly assessed, as many students would not have been able to perform an adequate peer evaluation in that regard (Alqassab et al., 2023). These presentations were carried out individually and were scheduled approximately evenly throughout each course, depending on their specific organization, while always avoiding the concentration of presentations within specific time periods so that students could prepare all presentations under similar conditions (Muklason et al., 2017).

In each presentation, both the students and the teacher evaluated the presenter’s work across four dimensions: explanatory ability, quality of the prepared file (generally a Power-Point presentation), attitude, and overall assessment; that is, the general impression conveyed by the student during the presentation. Peers and teachers responded to the surveys simultaneously, thus avoiding bias arising from knowledge of each other’s scores. Furthermore, the peers responded to the surveys anonymously, indicating only the name of the colleague they were evaluating. This evaluation was conducted through a quantitative survey consisting of

17 questions, which were identical for both the teacher and the students. Each question was rated on a five-level Likert scale (from 1 to 5), where 1 meant “very bad” and 5 “very good”. This type of scale is commonly used in studies of this nature (Koo & Yang, 2025).

All surveys were collected by the teacher, who calculated the average peer evaluation score per dimension for each student and made these results available, along with their own evaluation, within two days of the presentation via the course’s online learning platform. The results remained accessible until the next presentation took place. This approach allowed students sufficient time to reflect on their performance after receiving feedback, identifying areas for improvement (Nicol & MacFarlane-Dick, 2006). Additionally, students could clearly compare their own evaluations with those of the teacher, which, through contrast, encouraged critical reflection on their peer assessment abilities and reduced the subjectivity when assigning their scores (Revilla-Cuesta et al., 2024).

The evaluation of the final presentation, conducted at the end of the second semester of the academic year, followed the same procedure. However, on this occasion, students were also asked whether they would be willing to repeat their presentation to improve it and obtain a higher grade after learning about their peers’ evaluations. Those who responded affirmatively were asked to provide a brief written justification (approximately five lines) explaining their reasons. The assessments and responses obtained in the last presentation constitute the results analyzed in this research, which are the least affected by subjective factors as efforts were made to minimize them during the whole experience.

The temporary organization of the program carried out for this research is detailed in Figure 1.

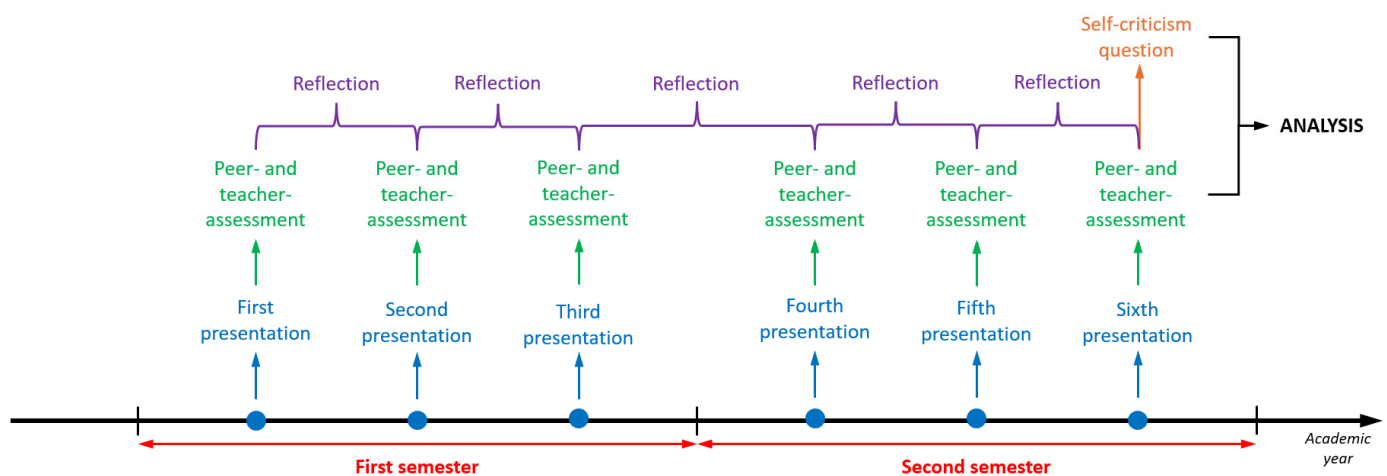


Figure 1. Program description.

2.2. Participants and Courses

All students participating in this educational research were enrolled during the 2024/2025 academic year in courses belonging to the Bachelor’s Degree in Civil Engineering and the Bachelor’s Degree in Agrifood Engineering and the Rural Environment at the University of Burgos, Spain. Specifically, the study involved six courses: two from the 4th year of the Civil Engineering program, one in the first semester and one in the second semester, and four from the Agrifood Engineering program, two from the 3rd year, one per semester, and two from the 4th year, distributed similarly by semester. The specific courses included in this study are detailed in Table 1.

These courses were selected not only to allow the temporal organization of the research according to the described planned program, but also for the following reasons:

- Continuous supervision: Conducting the study in courses at the University which the authors are affiliated allowed for ongoing and up-to-date monitoring of the research (Cetin, 2018).
- Manageable class size: All courses had a relatively small number of enrolled students, which made it feasible to conduct individual presentations without organizational issues and without any time constraints on the course (Mulryan-Kyne, 2010).
- Advanced academic level: The courses corresponded to advanced stages of university education. It was expected that students enrolled in these courses, having completed some prior university courses, would have a clear and appropriate understanding of what constitutes a high-quality presentation, both in terms of the prepared material and the presenter's attitude (Brew, 2001).

A total of 50 students enrolled in the indicated courses voluntarily participated in this study after being informed about what it involved. Their mean age was 21.54 ± 2.97 years. Only students who delivered all the presentations over the whole academic year were included in the analysis of the results. In this way, it was ensured that the results and the conclusions yielded from them were based on complete data. However, the results obtained referred only to students who maintained a constant interest in the courses throughout the whole academic year, discarding those students who decided to abandon them due to lack of time or interest. The participating teachers had an average age of 43.48 ± 10.28 years.

Table 1. Participating courses.

Bachelor's Degree	Year	Semester	Course
Civil Engineering	4th	1st	Civil Engineering Projects
Civil Engineering	4th	2nd	Environmental Engineering
Agrifood Engineering	3rd	1st	Structural Technology
Agrifood Engineering	3rd	2nd	Agrifood Building
Agrifood Engineering	4th	1st	Agricultural Infrastructure
Agrifood Engineering	4th	2nd	Green Space Engineering

2.3. Assessment Survey

All evaluations performed in this research, whether conducted by teachers or peers, utilized the quantitative survey shown in Table 2. This survey was developed by the researchers based on previous work (Revilla-Cuesta et al., 2021) and other relevant articles in the scientific bibliography (Feijóo et al., 2021; Seifan et al., 2020). The Cronbach's alpha value for the survey was 0.795, indicating adequate internal coherence and reliability following the experience of previous research in relation to the development of critical thinking skills in engineering students (Revilla-Cuesta et al., 2024).

Respondents were asked to evaluate aspects of students' presentations across three primary dimensions using a five-level Likert scale (with 1 meaning "very bad" and 5 "very good"), a straightforward and widely used assessment tool (Koo & Yang, 2025). The three dimensions were: explanatory ability, file quality (quality of the PowerPoint presentation), and attitude during the presentation. About five specific aspects were assessed under each dimension. Moreover, at the end of the survey, an overall assessment of the presentation was asked for, which can be perceived as a fourth dimension.

The survey did not include any evaluation regarding the course contents, such as omissions of key concepts, correctness of discussions and explanations, or factual errors. This was due to the fact that students were not expected to have proper training at this stage to make informed judgments on these elements, which could have compromised the validity of such peer assessments (Alqassab et al., 2023). Instead, the dimensions evaluated focused solely on

the effort and quality of the students’ preparation and delivery, independent of their mastery of the subject matter of the course (Pascual-Gómez et al., 2015).

Table 2. Assessment survey, replicated from previous research (Revilla-Cuesta et al., 2024).

Explanatory ability
1. Preparation and deliver of the presentation was adequate, rather than simply read aloud.
2. The speaker demonstrated understanding of the material, so the speaker explained it and did not just recite it.
3. Concepts were explained properly.
4. Concepts were presented in a logical and acceptable order.
5. The speaker used their own words to explain the concepts.
6. I properly understood all the concepts that were explained.
File quality
7. The preparation of the file was careful.
8. The format of the file was attractive.
9. The organization of the file was well.
10. Images or graphs were used effectively to complement the text.
11. Text was concise and did not hinder clarity.
Attitude
12. The engagement of the presentation was well.
13. Some topics not included in the presentation file were addressed.
14. The interest of the audience was actively stimulated by the speaker.
15. The enthusiasm for the topic by the speaker was perceived.
Overall assessment
16. Speaker’s general evaluation.
17. Willingness to assist to another presentation by the same speaker.

After the students’ sixth and final presentation, in addition to completing the aforementioned assessment survey for both peers and teachers, the students were asked whether they would be willing to repeat the presentation to improve it in order to achieve a higher grade after receiving feedback on their peer and teacher evaluations. This question required a simple “yes” or “no” response. Students who answered affirmatively were asked to anonymously indicate, in no more than five lines, the reasons for their decision. In this way, the final development of students’ self-critical capacity was also assessed (Liu & Carless, 2006).

2.4. Data Processing and Analysis

The evaluations from both peers and teachers for the first through fifth presentations were analyzed to obtain the “average peer scores”. The purpose was solely to provide these average peer evaluations alongside the teacher’s assessment to the students, in order to promote student reflection and foster the development of critical thinking skills (van Zundert et al., 2010). Moreover, the effect of time on this type of educational experience had already been examined in previous research (Revilla-Cuesta et al., 2024).

Accordingly, this study aimed to evaluate the final outcome of an educational experience of this type in engineering students. Therefore, only the assessment survey scores from the sixth (final) presentation for both peers and teachers, as well as the open-ended responses regarding students’ willingness to repeat this presentation to improve their grade, were analyzed in detail. This analysis had two objectives. The first was to compare the evaluations made by peers with those of the teachers, in order to determine whether the students had developed critical evaluation skills regarding the work of their peers (K. J.

[Topping, 2010](#)). The second was to assess the development of self-critical ability in relation to their own work ([Nicol, 2021](#)).

The analysis of the survey scores was conducted separately for each of the four dimensions: explanatory ability, file quality, attitude, and overall assessment. For each student presentation and dimension, the mean score was calculated from all the questions belonging to that category (see Table 2), both for peer and teacher evaluations. Thus, it was possible to effectively compare peers and teachers. Furthermore, this dimension-based approach provided a broader and more comprehensive perspective of the results, avoiding an excessive focus on specific aspects that could deviate from the objective of achieving a general overview ([Carless & Winstone, 2023](#)). Once these mean scores were obtained for each dimension from both peers and teachers, they were analyzed as follows:

- First, an analysis of the deviation between the evaluations provided by the students' peers and those given by the teachers was conducted (Section 3.1.1). To this end, the mean score for each dimension was first obtained considering all student presentations. Subsequently, the variations were calculated by taking the teachers' assessments as the reference value. This enabled an effective comparison between peer and teacher assessments, as well as the identification of the dimensions in which the greatest discrepancies occurred ([Falchikov & Goldfinch, 2000](#)).
- Second, confidence intervals for both peer and teacher scores were calculated for each dimension at significance levels of 1%, 5%, and 10% (Section 3.1.2). These confidence intervals were obtained after verifying that the scores assigned by both teachers and peers followed a normal distribution. This approach allowed not only the analysis of mean values but also the dispersion of the evaluations, providing a broader understanding of the results. The objective was to determine not only whether the mean values were similar but also whether the ranges of peer and teacher evaluations overlapped, indicating potential agreement in their assessments ([Langan et al., 2005](#)).
- Third, the uniformity of the evaluations across the four dimensions—that is, the equality of the mean scores between dimensions—was analyzed for both teachers and peers (Section 3.1.3), again considering the average scores for each dimension calculated from all student presentations and the 95% confidence intervals. This analysis aimed to identify similarities in the evaluation trends between teachers and peers, assessing whether teachers and peers followed comparable patterns when rating the different dimensions ([Falchikov & Goldfinch, 2000](#)).
- Finally, an ANalysis Of VAriance (ANOVA) was conducted (Section 3.1.4) considering all experimental results to provide greater statistical robustness to the analysis ([Sridharan et al., 2019](#)). Previously, the normality and randomness of the residuals and the homoscedasticity of the experimental data were verified. The ANOVA was performed at a 95% confidence level and included two scenarios: (i) the comparison between the evaluations given by teachers and peers for each dimension, and (ii) the comparison of the evaluations across dimensions within each group (peers and teachers).

The responses given by students to the question regarding their willingness to repeat the presentation were analyzed in two steps. First, in terms of frequency, the percentage of students who answered affirmatively was determined (Section 3.2.1). Second, the reasons provided for such affirmative responses in the open-ended question were analyzed qualitatively. This included a cross-coding process of a deductive nature carried out using the Atlas.ti software version 8 (Section 3.2.2), allowing reasons to be grouped into categories suitable for subsequent systematic analysis ([Mastrobattista et al., 2024](#)), and the generation of a word cloud to identify the most frequently mentioned terms (Section 3.2.3), providing a quick and clear overview of their main underlying motivations ([McNaught & Lam, 2010](#)).

3. Results and Discussion

3.1. Peer-Critique Skills

3.1.1. Peer–Teacher Comparison: Analysis of Deviations

The mean scores provided by both the teachers and the peers for the sixth presentation are detailed in Table 3. In addition, the standard deviation (SD) and the coefficient of variation (COV) are presented for both datasets as indicators of the variability of such evaluations. Table 4 also includes the deviations between the peers' and teachers' scores for each dimension. These deviations were calculated by taking the teachers' scores as the reference.

Table 3. Average scores.

Dimension	Peers			Teachers		
	Average	SD *	COV *	Average	SD *	COV *
Explanatory ability	4.36	0.34	0.08	4.29	0.72	0.17
File quality	4.37	0.33	0.08	4.15	0.67	0.16
Attitude	4.32	0.39	0.09	4.04	0.76	0.19
Overall assessment	4.38	0.34	0.08	4.37	0.60	0.14

* SD: standard deviation; COV: coefficient of variation.

Table 4. Deviations between peers' and teachers' scores.

Dimension	Deviation (%)
Explanatory ability	1.61
File quality	5.37
Attitude	6.82
Overall assessment	0.27

When comparing the average scores provided by the peers and the teachers, it can be observed that the peers' evaluations were consistently slightly higher. This trend is commonly reported in similar educational experiences, as students, despite specific training aimed at minimizing such biases, tend to be influenced by friendship ties and a certain reluctance to assign negative evaluations (Falchikov & Goldfinch, 2000; K. Topping, 1998). Nevertheless, it is noteworthy that in all cases, the peers' scores ranged between 4.0 and 4.5, which is below the “outstanding” threshold according to the Spanish grading system (grade from 0 to 10, with a grade equal to or above 9 being outstanding, i.e., equal to or above 4.5 on the Likert scale), and thus within the same range as the teachers' scores. This outcome may be attributed to the effective training in peer-critique skills, since the average scores remained fair when compared with those of the teachers, despite being slightly higher (Gielen et al., 2011). Additionally, the high scores may suggest the development of self-critical skills among students, leading to increased self-demand and an improvement in quality of their work with each presentation, as they performed the best when presenting the final one (Revilla-Cuesta et al., 2024).

When analyzing the scores by dimensions, the “overall assessment” dimension was rated almost identically by teachers and peers. This indicates that, through this experience, students effectively developed the ability to evaluate the work from a holistic perspective, thus providing a precise and critical overall judgment of the presentation, as also revealed in the literature (Cho & Schunn, 2007). The “explanatory ability” dimension was also accurately peer-assessed, showing only a 1.61% deviation from the teachers' evaluations. This result is consistent with the fact that any student can autonomously determine whether they understand what is being explained and assess this dimension accordingly (Lu & Law, 2012). The peer-assigned scores for “file quality” and “attitude” showed greater deviations compared to the teachers' scores. However, these differences were still modest, within the

5–7% range, which demonstrates that these peer assessments were also accurate and fair. The ability to critically evaluate the quality of a presentation file (or a document) generally develops through comparative experience, requiring the judgment of numerous examples. The same applies to the assessment of attitude, where experience plays a key role. It is evident that teachers, owing to their broader experience in such aspects, tended to be more demanding in these dimensions (Gielen et al., 2011; Lu & Law, 2012).

Finally, when examining the variability of the scores in terms of standard deviation and coefficient of variation, it was observed that these values remained approximately constant across dimensions for both peers and teachers. However, the variability values obtained for teachers were roughly twice as high as those of the peers. The teachers' experience enables them to more precisely discern the differences in the quality of work produced by each student, allowing them to provide more differentiated assessments (Lu & Law, 2012). This, in turn, explains the greater variability in their evaluations. In contrast, although students were able to assess their peers fairly on average, their evaluations tended to be overly uniform across classmates. This finding suggests that the ability to provide differentiated and nuanced assessments is a skill that requires prolonged practice and sustained development over time (Revilla-Cuesta et al., 2024; van Zundert et al., 2010).

3.1.2. Peer–Teacher Comparison: Analysis of Confidence Intervals

Figure 2 presents the confidence intervals of the evaluations provided by both peers and teachers for each of the assessed dimensions. Three confidence levels were considered: 90% (Figure 2a), 95% (Figure 2b), and 99% (Figure 2c).

Regarding the confidence intervals, the first relevant aspect is linked to the score variability discussed in the previous section. Thus, the confidence intervals for the teachers' assessments were wider than those obtained for the peer evaluations. As mentioned, teachers provide more accurate assessments due to their greater experience in this task, while students are not yet able to evaluate their peers' performance with the same level of critical judgment (Gielen et al., 2011). In this educational experience, critical thinking skills were developed through an academic year-long practice; however, the scores provided by the students still reflected these differences. Therefore, it would be essential to continue fostering such skills over time so that students can progressively develop them and eventually apply them effectively in their future professional practice within the engineering field (Litzinger et al., 2011; Shuman et al., 2005).

The confidence intervals for the “explanatory ability” and “overall assessment” dimensions were found to coincide between peers and teachers, regardless of the confidence level considered. In other words, the interval obtained for peer evaluations was always contained within that obtained for teachers. This indicates that, in addition to the mean scores being similar, their distributions were also comparable. Thus, statistically, the evaluations from both groups in these dimensions were equivalent (Cho et al., 2006; Sung et al., 2005).

Conversely, the confidence intervals for the “file quality” and “attitude” dimensions only partially overlapped, with the lower range of peer evaluations coinciding with the upper range of those given by teachers. This overlapping area increased with higher confidence levels, becoming nearly identical at the 99% confidence level. As discussed earlier, the assessment of these two dimensions largely depends on the evaluator's experience (Lu & Law, 2012), which explains why peers tended to assign higher ratings. However, the substantial overlap of the confidence intervals between peers and teachers is consistent with the fact that the year-long training aimed at developing critical thinking regarding the work of the classmates was effective, as peers' evaluations aligned closely with those of the teachers (Falchikov & Goldfinch, 2000). Nevertheless, it is evident that continued training

over time would be necessary to achieve fully consistent and robust evaluation skills with regard to these two dimensions (Panadero & Brown, 2017).

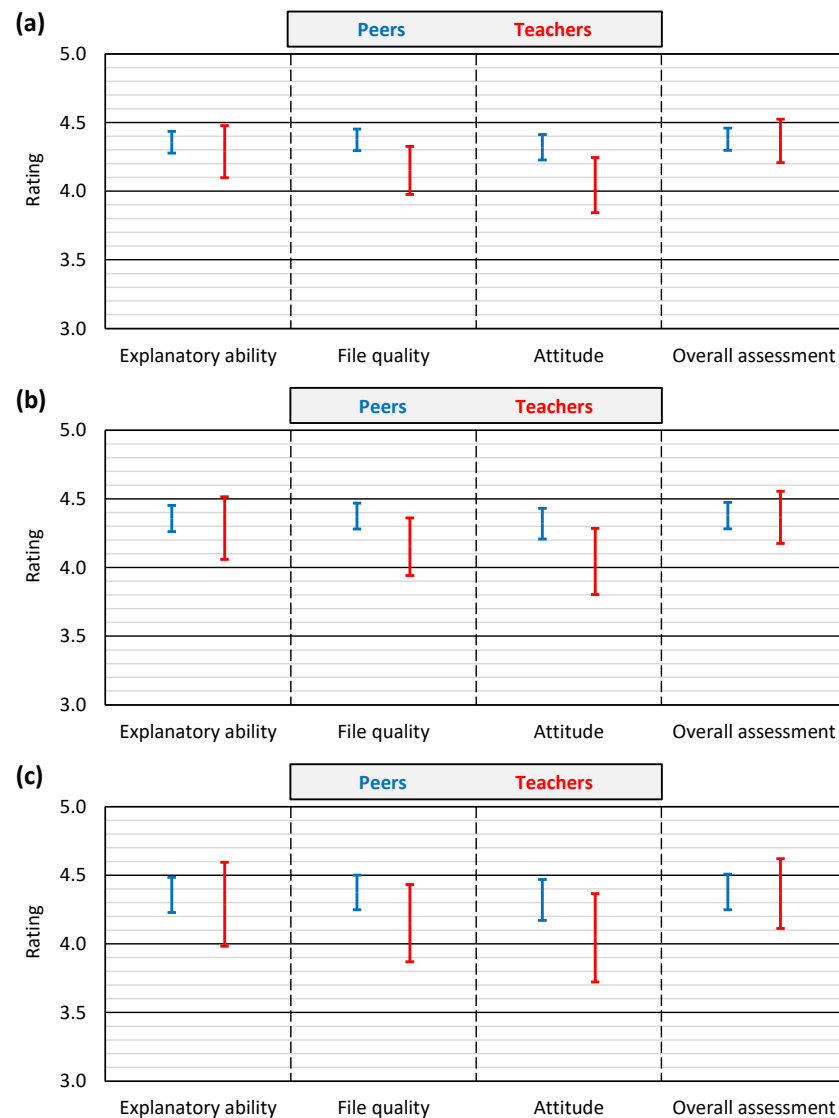


Figure 2. Confidence interval of peers' and teachers' scores: (a) 90% confidence (10% significance); (b) 95% confidence (5% significance); (c) 99% confidence (1% significance).

3.1.3. Inter-Dimensional Comparison

Previous sections showed that peers evaluated students more uniformly, presenting lower variability across dimensions compared to the teachers' assessments, a result attributed to the peers' limited experience in performing individualized assessments (Gielen et al., 2011). Building on these findings, this section shows an inter-dimensional comparison; that is, it analyzes the uniformity of evaluations across dimensions for both peers and teachers. To this end, data is presented comparatively in Figure 3, considering both average values (Figure 3a) and 95% confidence intervals (Figure 3b).

Again, the scores provided by peers were more uniform across the different dimensions, with all values ranging between 4.30 and 4.40 (a deviation of 1.4% between the highest and lowest scores). In contrast, the teachers' assessments showed greater variation across dimensions, with the lowest average score being 4.04 and the highest 4.37. This higher variability in teachers' inter-dimensional evaluations can be explained by their greater experience in assessing the "file quality" and "attitude" dimensions, as supported by scientific literature (Gielen et al., 2011; Lu & Law, 2012). Students, lacking similar

experience in evaluating these dimensions, even after training over an entire academic year, tended to assign scores similar to those given in dimensions where they exhibited greater understanding and critical thinking skills (“explanatory ability” and “overall assessment”) (Cho & Schunn, 2007). The 95% confidence intervals showed the same trend, being almost coincident across all dimensions in the case of the peers’ evaluations.

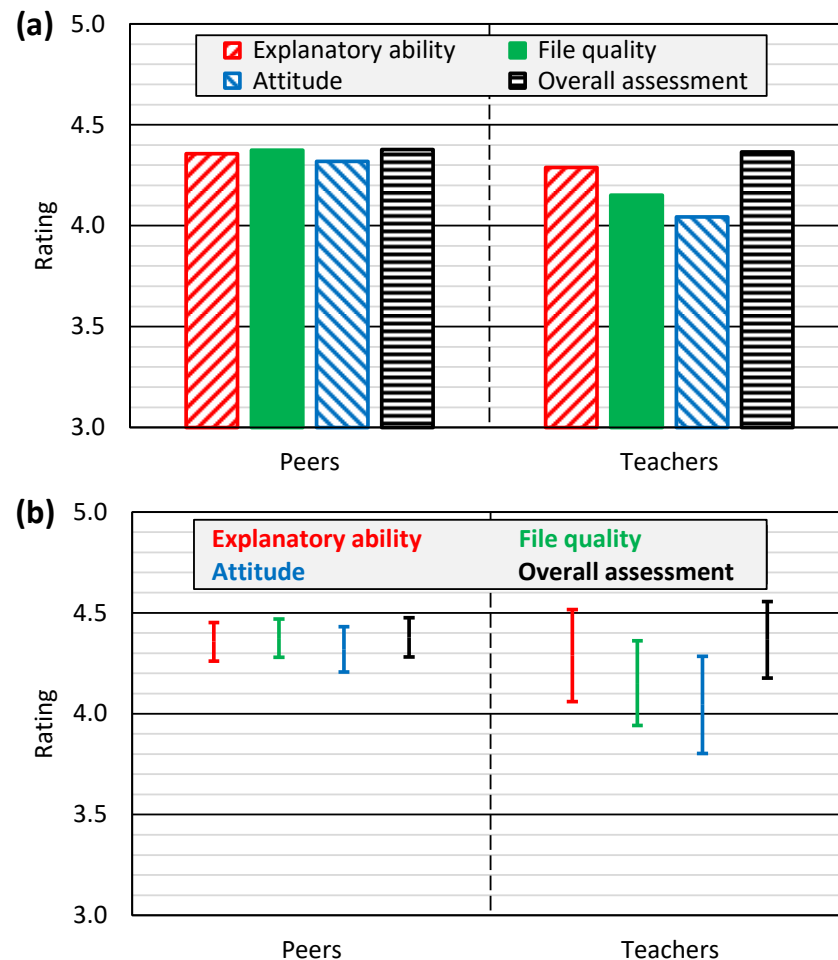


Figure 3. Inter-dimensional comparison for peers’ and teachers’ scores: (a) average values; (b) 95% confidence intervals.

3.1.4. Significance Analysis

Table 5 shows the p -values obtained from the one-way ANOVA regarding the significance of the evaluator when assigning scores to each of the four analyzed dimensions (peer–teacher comparison). Additionally, Table 6 presents the p -values from the ANOVA corresponding to the analysis of the significance of the evaluated dimension for each evaluator group (inter-dimensional comparison).

Table 5. One-way ANOVA for the comparison between peers’ and teachers’ scores.

Dimension	p -Value
Explanatory ability	0.5498
File quality	0.0411
Attitude	0.0288
Overall assessment	0.9058

In both cases, the ANOVA results confirmed the aspects discussed and highlighted in previous sections. On the one hand, the evaluations of the dimensions “explanatory ability”

and “overall assessment” by both peers and teachers were statistically equivalent, whereas they were significantly different for the dimensions “file quality” and “attitude.” On the other hand, the evaluation of each dimension by teachers differed statistically across dimensions, in contrast to peer assessments. These results can be explained by the lower experience of peers in assessing aspects such as document quality or attitude, which is considerably higher among teachers (Falchikov & Goldfinch, 2000; Panadero & Brown, 2017). Thus, it is evident that one academic year of training in evaluating these dimensions is insufficient for peers to achieve a level of critical assessment comparable to that of a teacher, indicating that such training should be extended over longer periods during university education (van Zundert et al., 2010). These are the dimensions that require more intense training and education to achieve proper peer assessments (Revilla-Cuesta et al., 2024).

Table 6. One-way ANOVA for the inter-dimensional comparison.

Evaluator	<i>p</i> -Value
Peers	0.8358
Teachers	0.0476

3.2. Self-Critical Skills

3.2.1. Frequency Analysis

The percentages of students who responded either affirmatively or negatively to the question of whether they would be willing to repeat their presentation to achieve a higher grade are shown in Figure 4. Specifically, 30% responded affirmatively, while 70% indicated that they would not. This figure also displays the average scores received from peers for each of the two student groups. The ratings were very similar for both groups, and even slightly higher for those who answered affirmatively; that is, those who were willing to repeat the presentation. In many cases, students who choose to redo an assignment do so because they have received a grade below the average. Thus, when comparing their performance with that of their peers, they may feel motivated to improve it (Liljeröd et al., 2025). In this case, however, such a situation did not occur. Therefore, assuming that the subjective factors that may affect the decision to repeat the work were the same for both groups of students, as all the students were in the same situation (Panadero et al., 2013), it seems that the students’ decision to repeat the work was not driven by a low grade but rather by a reflection on the quality of their work and on whether they felt capable of improving it.

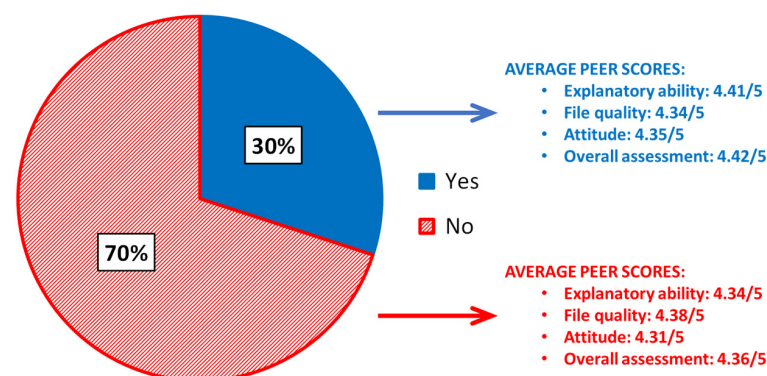


Figure 4. Ratio of students with affirmative and negative responses, and average peer scores.

Considering the conditions under which this presentation took place, the authors of this study regard the percentage of students who responded affirmatively as relatively high:

- First, this final presentation was conducted at the end of the academic year (Figure 1). A period when students face the final exams of all second-semester courses and, conse-

quently, experience a high workload. Moreover, fourth-year students were also in the final stages of completing their Bachelor's Thesis, which, at the University of Burgos (Spain), consists of a complete engineering project in the participating Bachelor's Degrees, further increasing the students' workload. Under such circumstances, students typically avoid taking on additional tasks (Jensen et al., 2023).

- Second, the grades obtained were high, exceeding 4/5 in all dimensions and regardless of the evaluator. Such high scores tend to reduce students' motivation to repeat the presentation or project in order to yield a higher mark (Liljeröd et al., 2025).

Thus, the fact that 30% of the students who participated in the study were willing to repeat the presentation indicates that the experience was successful in fostering students' self-critical abilities regarding their own work. This demonstrates their intention to improve the quality of their presentations, and consequently their grades, despite the high workload they faced at that stage of the academic year.

3.2.2. Qualitative Analysis

To gain deeper insight into the reasons why students who answered affirmatively were willing to repeat their presentations, they were asked to provide anonymous, open-ended responses explaining their motivations, in a maximum of five lines. The collected text fragments were qualitatively analyzed through cross-coding, a methodological approach in which the authors have prior experience (Revilla-Cuesta et al., 2022). This analysis allowed the identification of three main reasons underlying this attitude.

Firstly, some students indicated that they felt capable of improving identified deficiencies (code "capacity"); believing the enhancement was achievable, they found no reason to avoid repeating the presentation. From their perspective, making such improvements only required a small investment of time and effort, which they considered manageable despite their already heavy workload. These responses reflect a high level of self-demand that has arisen naturally, not forced by imposition, likely fostered by the continuous and periodic exposure to peer constructive critique throughout the course (Revilla-Cuesta et al., 2024). This is a positive reaction, as it encourages engineering students to consistently strive for excellence in their professional performance (Wu et al., 2021), as long as this attitude remains balanced, as excessive self-demand do not favor this goal (Bakker & Mostert, 2024).

"The aspects my peers criticized are things I believe I can improve with a bit of effort [...] " [...] it bothers me that the lowest score from my classmates was for the PowerPoint file [...] if I just spent a bit more time on it, it would be much better [...] "It's true that we now have all our final exams [...] but doing them well is still compatible with spending an afternoon trying to improve this project, especially since we are given the opportunity to do so."

Secondly, some students were strongly influenced by their peers' feedback, stating that the opinion that their classmates had of them was the main reason for wanting to repeat the presentation (code "peers"). For these students, it was important that their peers perceived that they had done a good job. This behavior can be interpreted as a sense of responsibility towards others, where students who exhibit this attitude wish their peers' efforts to be rewarded with a positive outcome (Simonsmeier et al., 2020)—in this case, by attending a well-delivered presentation. Therefore, this outcome can be understood as an indirect self-criticism, highly conditioned by others' opinions (Ardill, 2025; Keller et al., 2025), which can be particularly valuable in teamwork, an essential aspect of professional engineering practice (Cruz et al., 2020). However, this type of self-demand requires further development to ensure that students are able to be critical of their own work even when it is not going to be observed or assessed by others (Revilla-Cuesta et al., 2024).

"[...] It really bothered me that my classmates thought the presentation wasn't very good [...] "I didn't have enough time to do it as well as I could have [...] I even feel like I

made my classmates waste some of their time [...] " [...] When I didn't know how to continue and started reading from my notes during the presentation, I even felt a bit embarrassed [...] after spending the whole year doing this kind of presentations, my classmates saw that I did it poorly. [...]"

Finally, some students expressed feelings of frustration and unfairness regarding the assessments they received, indicating that, from their perspective, the work they had done deserved a higher score (code "feelings"). Consequently, they stated that they were willing to repeat the presentation to achieve a score more in line with their perceived performance. These responses also reflected a demanding attitude, though in this case directed not toward their own work, as in the previously discussed group, but toward the evaluations made by their peers. This type of reaction is quite common when students are asked about the peer assessments they receive and highlights that developing self-critical skills is not an easy task, as it is always more comfortable and straightforward to demand more from others than from oneself (Rasooli et al., 2025). In this case, working on these skills over a single academic year may not be sufficient for some students, even though peer-critical skills were successfully fostered. Therefore, it is necessary to work continuously on the development of these skills throughout the entire period of university education, as highlighted in the literature (K. J. Topping, 2010).

" [...] I believe that the score I received has been too low [...] I put in much more work than what was reflected." "I was surprised when I saw my score [...] I think I invested enough time to deserve a higher recognition." " [...] My classmates did not evaluate my work properly. They were not able to perceive all the effort and time I had dedicated to the PowerPoint presentation."

3.2.3. Word Cloud Analysis

Figure 5 shows the word cloud representing the most frequently repeated words in the students' responses. It can be observed that the two most common words were "classmates" and "time". This finding reinforces two key ideas. On the one hand, students who were willing to repeat the presentation were strongly influenced by their peers' opinions when considering whether they had to improve their work. On the other hand, time emerged as a crucial variable, both in terms of the time available to redo the presentation and the perceived correlation between the time invested in the project and the grade obtained. Other frequently repeated words, such as "improve", "effort", and "bit", are clearly associated with the idea that students believed a little more effort would allow them to enhance their performance, demonstrating a notable capacity for self-criticism regarding their work (Wu et al., 2021), thus validating the program's effectiveness on self-reflection.



Figure 5. Word cloud of the words most frequently used by students in their answers.

4. Conclusions

This article presented an educational experience conducted over a full academic year within university engineering education. This experience consisted of six in-class presentations delivered by students throughout the 2024/2025 academic year, which were evaluated using a Likert-scale survey considering the dimensions “explanatory ability”, “file quality”, “attitude”, and “overall assessment”. Each presentation was assessed both by the teacher and by the students’ peers. After each presentation, these evaluations were shared with the students, allowing them to reflect both on the appropriateness of the scores they had assigned to their classmates (development of peer-critical skills) and on their own performance (self-critical skills). After the final presentation, the same survey was administered again, and the resulting data are presented and analyzed in this article. In addition, once these evaluations had been received, students were asked whether they would be willing to repeat the presentation to achieve a higher grade. Those who answered affirmatively were asked to briefly explain their reasons in open-ended responses, which have also been qualitatively analyzed in this paper. From the whole analysis, the following conclusions can be drawn:

- On average, the evaluations provided by the teachers and peers were very similar, with those of the students being slightly higher. In fact, the maximum deviation between them was only 7%, which demonstrates the successful fostering of peer-critical skills among the students, as they consistently assigned ratings comparable to those given by the teachers.
- The assessments for the dimensions “explanatory ability” and “overall assessment” were almost identical between teachers and peers, with maximum deviations of 1.6%. However, the deviations for “file quality” and “attitude” were significantly higher. Thus, although peer-critical skills were successfully developed, this occurred more effectively in those dimensions where proper evaluation requires less prior experience and temporal training.
- The confidence intervals of teacher and peer evaluations completely overlapped for the dimensions “overall assessment” and “explanatory ability”. The overlap was only partial for “file quality” and “attitude”, where the lower bound of the students’ interval coincided with the upper bound of the teachers’ interval. Such range of common values increased as the confidence level rose, statistically indicating a high degree of agreement between both evaluators.
- The evaluations given by peers were more uniform than those of the teachers, both when analyzing variations between students and across dimensions. The reduced prior experience of the peers in perceiving differences between students’ presentations and assessing the dimensions “file quality” and “attitude” led them to rely more on comparative assessments, resulting in greater uniformity. The capacity to deliver differentiated and nuanced evaluations with regard to these aspects is a skill that develops gradually and requires sustained practice over time.
- Overall, 30% of the students were willing to repeat the final presentation to improve their grade. This percentage of students was high, considering that they had a heavy workload and taking into account that the scores they received from their peers were even higher than those of the students who chose not to repeat. All this indicates successful development of self-critical skills.
- The reasons for deciding to repeat the presentation were related to the development of self-critical skills, which were either naturally fostered through the training conducted or indirectly driven by a desire to make a good impression in front of their peers. However, some students also mentioned feelings of unfairness with the scores received, indicating

that the development of such abilities requires sustained work over time. Furthermore, this also shows that trust in assessment processes require continued development.

The teaching experience presented in this article demonstrates that initiatives of this kind can be effective for developing peer- and self-critical skills in engineering students. Training in these competences helps successfully develop critical thinking, thus preparing students for the responsible practice of their future professional roles. Peer- and self-assessment of the four dimensions considered prompt pedagogical mechanisms closely linked to critical thinking, such as argumentation analysis, analytical judgment, willingness to reconsider one's views, and integrative judgment. However, the study also shows that fostering such competences requires long-term effort, as their development is not possible without regular and systematic reflection by students on these aspects. Therefore, conducting educational experiences over extended periods of time is fundamental for the effective development of peer- and self-critical skills by students.

The results of this research can be further complemented by future studies that delve deeper into the usefulness of these types of educational experiences for the development of peer- and self-critical skills in engineering students. These include comparing the results of the study sample with those obtained in a control group, which would provide a more objective understanding of the usefulness of this educational initiative. Another possibility could be the use of 10-point Likert scales in the surveys, which would allow for a clearer elucidation of the differences in the scores given by students and teachers. From here, it would be appropriate to analyze which course types and age groups are most suitable for conducting these types of educational experiences.

Author Contributions: Conceptualization, V.R.-C., M.S. and V.O.-L.; methodology, V.R.-C. and V.O.-L.; software, V.R.-C., M.H.-R. and I.M.; validation, M.H.-R., I.M. and M.S.; formal analysis, V.R.-C.; investigation, V.R.-C., M.H.-R., I.M., M.S. and V.O.-L.; resources, M.S. and V.O.-L.; data curation, V.R.-C. and M.H.-R.; writing—original draft preparation, V.R.-C.; writing—review and editing, M.H.-R., I.M., M.S. and V.O.-L.; visualization, V.R.-C., M.H.-R. and I.M.; supervision, M.S. and V.O.-L.; project administration, M.S. and V.O.-L.; funding acquisition, V.R.-C., M.S. and V.O.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Burgos through the funding program “Convocatoria de ayudas a Grupos de Innovación Docente reconocidos para la elaboración de materiales docentes para los años 2025 y 2026”.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the fact that they are not mandatory, according to university regulations for conducting a teaching study such as the one described in this article, provided the anonymity of the participants is respected in the manuscript.

Informed Consent Statement: Informed consent was obtained from all subjects who responded to the survey whose anonymized results are described in this paper.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the need to maintain the participants' anonymity.

Acknowledgments: The authors would like to thank all the teachers and students of the University of Burgos, Spain, who participated in this research work for their collaboration and helpful attitudes.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Alqassab, M., Strijbos, J. W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review*, 35(1), 18. [\[CrossRef\]](#)
- Annala, J. (2023). What knowledge counts—Boundaries of knowledge in cross-institutional curricula in higher education. *Higher Education*, 85(6), 1299–1315. [\[CrossRef\]](#)
- Ardill, N. (2025). Peer feedback in higher education: Student perceptions of peer review and strategies for learning enhancement. *European Journal of Higher Education*, 15(4), 696–721. [\[CrossRef\]](#)
- Bakker, A. B., & Mostert, K. (2024). Study demands–resources theory: Understanding student well-being in higher education. *Educational Psychology Review*, 36(3), 92. [\[CrossRef\]](#)
- Barta, A., Fodor, L. A., Tamas, B., & Szamosközi, I. (2022). The development of students critical thinking abilities and dispositions through the concept mapping learning method—A meta-analysis. *Educational Research Review*, 37, 100481. [\[CrossRef\]](#)
- Bezanilla, M. J., García Olalla, A. M., Castro, J. P., & Ruiz, M. P. (2019). A model for the evaluation of competence-based learning implementation in higher education institutions: Criteria and indicators. *Tuning Journal for Higher Education*, 6(2), 127–174. [\[CrossRef\]](#)
- Brew, A. (2001). Conceptions of Research: A phenomenographic study. *Studies in Higher Education*, 26(3), 271–285. [\[CrossRef\]](#)
- Cano, E., Lluch, L., Grané, M., & Remesal, A. (2023). Competency-based assessment practices in higher education: Lessons from the pandemics. *Trends in Higher Education*, 2(1), 238–254. [\[CrossRef\]](#)
- Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150–163. [\[CrossRef\]](#)
- Cerezo-Narváez, A., de los Ríos Carmenado, I., Pastor-Fernández, A., Yagüe Blanco, J. L., & Otero-Mateo, M. (2019). Project management competences by teaching and research staff for the sustained success of engineering education. *Education Sciences*, 9(1), 44. [\[CrossRef\]](#)
- Cetin, S. K. (2018). Alternative observation tools for the scope of contemporary education supervision: An action research. *European Journal of Educational Research*, 7(2), 329–340. [\[CrossRef\]](#)
- Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, 48(3), 409–426. [\[CrossRef\]](#)
- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891–901. [\[CrossRef\]](#)
- Cruz, M. L., Saunders-Smiths, G. N., & Groen, P. (2020). Evaluation of competency methods in engineering education: A systematic review. *European Journal of Engineering Education*, 45(5), 729–757. [\[CrossRef\]](#)
- Deo, S., & Hölttä-Otto, K. (2024). Critical thinking assessment in engineering education: A scopus-based literature review. *Journal of Mechanical Design*, 146(7), 072301. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Feijóo, J. C. M., Suárez, F., Chiyón, I., & Alberti, M. G. (2021). Some web-based experiences from flipped classroom techniques in aec modules during the COVID-19 lockdown. *Education Sciences*, 11(5), 211. [\[CrossRef\]](#)
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment and Evaluation in Higher Education*, 36(2), 137–155. [\[CrossRef\]](#)
- González-Marcos, A., Alba-Elías, F., Ordieres-Mere, J., Alfonso-Cendón, J., & Castejón-Limas, M. (2016). Learning project management skills in engineering through a transversal coordination model. *International Journal of Engineering Education*, 32(2), 894–904.
- Hong, J. C., Hsiao, H. S., Chen, P. H., Lu, C. C., Tai, K. H., & Tsai, C. R. (2021). Critical attitude and ability associated with students' self-confidence and attitude toward "predict-observe-explain" online science inquiry learning. *Computers and Education*, 166, 104172. [\[CrossRef\]](#)
- Jensen, K. J., Mirabelli, J. F., Kunze, A. J., Romanchek, T. E., & Cross, K. J. (2023). Undergraduate student perceptions of stress and mental health in engineering culture. *International Journal of STEM Education*, 10(1), 30. [\[CrossRef\]](#)
- Keller, M. V., Daumiller, M., & Dresel, M. (2025). Relevance of student motivation for providing high-quality peer-feedback: Results of two field studies. *Learning and Instruction*, 99, 102152. [\[CrossRef\]](#)
- Kolmos, A., Holgaard, J. E., Routhe, H. W., Winther, M., & Bertel, L. (2024). Interdisciplinary project types in engineering education. *European Journal of Engineering Education*, 49(2), 257–282. [\[CrossRef\]](#)
- Koo, M., & Yang, S. W. (2025). Likert-type scale. *Encyclopedia*, 5(1), 18. [\[CrossRef\]](#)
- Langan, A. M., Wheeler, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., Lockey, L., & Preziosi, R. F. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment and Evaluation in Higher Education*, 30(1), 21–34. [\[CrossRef\]](#)

- Liljeröd, H., Jönsson, A., Klapp, A., & Jonsson, A. C. (2025). Students' perceptions of how grades influence their motivation: Voices of upper secondary school students in Norway and Sweden. *Educational Assessment, Evaluation and Accountability*, 37(3), 385–409. [\[CrossRef\]](#)
- Litzinger, T. A., Lattuca, L. R., Hadgraft, R. G., Newstetter, W. C., Alley, M., Atman, C., DiBiasio, D., Finelli, C., Diefes-Dux, H., Kolmos, A., Riley, D., Sheppard, S., Weimer, M., & Yasuhara, K. (2011). Engineering education and the development of expertise. *Journal of Engineering Education*, 100(1), 123–150. [\[CrossRef\]](#)
- Liu, M. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–290. [\[CrossRef\]](#)
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275. [\[CrossRef\]](#)
- Mastrobattista, L., Muñoz-Rico, M., & Cordon-García, J. A. (2024). Optimising textual analysis in higher education studies through computer-assisted qualitative data analysis (CAQDAS) with Atlas.ti. *Journal of Technology and Science Education*, 14(2), 622–632. [\[CrossRef\]](#)
- McNaught, C., & Lam, P. (2010). Using wordle as a supplementary research tool. *Qualitative Report*, 15(3), 630–643. [\[CrossRef\]](#)
- Moreira, M. A., Arcas, B. R., Sánchez, T. G., García, R. B., Melero, M. J. R., Cunha, N. B., Viana, M. A., & Almeida, M. E. (2023). Teachers' pedagogical competences in higher education: A systematic literature review. *Journal of University Teaching and Learning Practice*, 20(1), 90–123. [\[CrossRef\]](#)
- Muklason, A., Parkes, A. J., Özcan, E., McCollum, B., & McMullan, P. (2017). Fairness in examination timetabling: Student preferences and extended formulations. *Applied Soft Computing Journal*, 55, 302–318. [\[CrossRef\]](#)
- Mulryan-Kyne, C. (2010). Teaching large classes at college and university level: Challenges and opportunities. *Teaching in Higher Education*, 15(2), 175–185. [\[CrossRef\]](#)
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment and Evaluation in Higher Education*, 46(5), 756–778. [\[CrossRef\]](#)
- Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. [\[CrossRef\]](#)
- O'Connor, S., Power, J., Blom, N., & Tanner, D. (2025). Engineering students' perceptions of problem and project-based learning (PBL): Comparing online and traditional face-to-face environments. *Australasian Journal of Engineering Education*, 29(2), 88–101. [\[CrossRef\]](#)
- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133–156. [\[CrossRef\]](#)
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195–203. [\[CrossRef\]](#)
- Pascual-Gómez, I., Lorenzo-Llamas, E. M., & Monge-López, C. (2015). Analysis of validity of peer assessment: A study in higher education. *RELIEVE—Revista Electronica de Investigacion y Evaluacion Educativa*, 21(1), 1–17. [\[CrossRef\]](#)
- Ponomariovienė, J., Jakavonytė-Staškuvienė, D., & Torterat, F. (2025). Implementing competency-based education through the personalized monitoring of primary students' progress and assessment. *Education Sciences*, 15(2), 252. [\[CrossRef\]](#)
- Rasooli, A., Turner, J., Varga-Atkins, T., Pitt, E., Asgari, S., & Moindrot, W. (2025). Students' perceptions of fairness in groupwork assessment: Validity evidence for peer assessment fairness instrument. *Assessment and Evaluation in Higher Education*, 50(1), 111–126. [\[CrossRef\]](#)
- Revilla-Cuesta, V., Hurtado-Alonso, N., Fontaneda, I., Skaf, M., & Ortega-López, V. (2024). Teaching self-criticism and peer-critique skills to engineering students through a temporal survey-based program. *Frontiers in Education*, 9, 1399750. [\[CrossRef\]](#)
- Revilla-Cuesta, V., Skaf, M., Espinosa, A. B., & Ortega-López, V. (2022). Teaching lessons learnt by civil-engineering teachers from the COVID-19 pandemic at the University of Burgos, Spain. *PLoS ONE*, 17(12), e0279313. [\[CrossRef\]](#) [\[PubMed\]](#)
- Revilla-Cuesta, V., Skaf, M., Manso, J. M., & Ortega-López, V. (2020). Student perceptions of formative assessment and cooperative work on a technical engineering course. *Sustainability*, 12(11), 4569. [\[CrossRef\]](#)
- Revilla-Cuesta, V., Skaf, M., Varona, J. M., & Ortega-López, V. (2021). The outbreak of the COVID-19 pandemic and its social impact on education: Were engineering teachers ready to teach online? *International Journal of Environmental Research and Public Health*, 18(4), 2127. [\[CrossRef\]](#)
- Ríos, I. D. L., Cazorla, A., Díaz-Puente, J. M., & Yagüe, J. L. (2010). Project-based learning in engineering higher education: Two decades of teaching competences in real environments. *Procedia—Social and Behavioral Sciences*, 2, 1368–1378. [\[CrossRef\]](#)
- Seifan, M., Dada, O. D., & Berenjian, A. (2020). The effect of real and virtual construction field trips on students' perception and career aspiration. *Sustainability*, 12(3), 1200. [\[CrossRef\]](#)
- Shuman, L. J., Besterfield-Sacre, M., & McGourty, J. (2005). The ABET "professional skills"—Can they be taught? Can they be assessed? *Journal of Engineering Education*, 94, 41–55. [\[CrossRef\]](#)

- Simonsmeier, B. A., Peiffer, H., Flaig, M., & Schneider, M. (2020). Peer feedback improves students' academic self-concept in higher education. *Research in Higher Education*, 61(6), 706–724. [\[CrossRef\]](#)
- Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education*, 27(5), 443–454. [\[CrossRef\]](#)
- Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education*, 77(5), 853–870. [\[CrossRef\]](#)
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education*, 45(2), 187–202. [\[CrossRef\]](#)
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. [\[CrossRef\]](#)
- Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, 20(4), 339–343. [\[CrossRef\]](#)
- Törlind, P., Larsson, L., & Eklöf, L. (2023, September 7–8). *Longitudinal evaluation of self-assessment and peer review in a capstone course*. 25th International Conference on Engineering and Product Design Education (pp. 415–420), Barcelona, Spain.
- Turns, J. A., Sattler, B., Yasuhara, K., Borgford-Parnell, J. L., & Atman, C. J. (2014, June 15–18). *Integrating reflection into engineering education*. ASEE Annual Conference and Exposition, Conference Proceedings, Indianapolis, IN, USA.
- Tzeng, A., Bruno, B., Cooperrider, J., Dinardo, P. B., Baird, R., Swetlik, C., Goldstein, B. N., Rastogi, R., Roth, A. J., Gilligan, T. D., & Rish, J. M. (2021). A structured peer assessment method with regular reinforcement promotes longitudinal self-perceived development of medical students' feedback skills. *Medical Science Educator*, 31(2), 655–663. [\[CrossRef\]](#) [\[PubMed\]](#)
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, 20(4), 280–290. [\[CrossRef\]](#)
- van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270–279. [\[CrossRef\]](#)
- Wang, Y., Liu, Y., & Wang, H. (2023). Competency model for international engineering project manager through MADM method: The Chinese context. *Expert Systems with Applications*, 212, 118675. [\[CrossRef\]](#)
- Wu, L. L., Fischer, C., Rodriguez, F., Washington, G. N., & Warschauer, M. (2021). Project-based engineering learning in college: Associations with self-efficacy, effort regulation, interest, skills, and performance. *SN Social Sciences*, 1(12), 287. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.