



OPEN ACCESS

EDITED BY

Niroj Dahal,
Kathmandu University, Nepal

REVIEWED BY

Yi-Huang Shih,
Minghsin University of Science and
Technology, Taiwan
Laxman Luitel,
Kathmandu University, Nepal

*CORRESPONDENCE

Marta Skaf
✉ mskaf@ubu.es

RECEIVED 12 March 2024

ACCEPTED 23 April 2024

PUBLISHED 03 May 2024

CITATION

Revilla-Cuesta V, Hurtado-Alonso N,
Fontaneda I, Skaf M and
Ortega-López V (2024) Teaching self-
criticism and peer-critique skills to
engineering students through a temporal
survey-based program.
Front. Educ. 9:1399750.
doi: 10.3389/feduc.2024.1399750

COPYRIGHT

© 2024 Revilla-Cuesta, Hurtado-Alonso,
Fontaneda, Skaf and Ortega-López. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Teaching self-criticism and peer-critique skills to engineering students through a temporal survey-based program

Víctor Revilla-Cuesta¹, Nerea Hurtado-Alonso²,
Ignacio Fontaneda³, Marta Skaf^{2*} and Vanesa Ortega-López¹

¹Department of Civil Engineering, Escuela Politécnica Superior, University of Burgos, Burgos, Spain,

²Department of Construction, Escuela Politécnica Superior, University of Burgos, Burgos, Spain,

³Department of Organization Engineering, Escuela Politécnica Superior, University of Burgos, Burgos, Spain

Introduction: Engineering students should develop critical-thinking skills and insist on appropriate performance levels both from themselves and from their colleagues during their training. In doing so, they will adopt critical attitudes toward their own and others' work. This will help them to successfully perform their future professional work with the highest standards.

Methods: In this research, peer- and self-assessments of in-class presentations through a survey-based program were used to analyze the development of critical-thinking skills among engineering students. The program included two key features: firstly, formative assessments were repeatedly conducted over time; secondly, teachers' ratings were provided to students as comparative benchmarks. This approach encouraged students to reflect on their ratings over time using the reference of the teacher.

Results: From a general perspective, the analysis of survey responses showed that students assigned lower ratings in assessments conducted at a later stage, despite receiving higher ratings from their teachers over time. Therefore, students became more demanding throughout the experience in spite of the increased quality of their work according to the teachers' assessments. Moreover, students tended to closely attune their evaluations to their teachers' ratings. Comparing peer- and self-assessments, it was noted that students were more demanding toward the work of their peers in the long term, and especially their ability to explain concepts, than they were toward their own work. Nevertheless, high temporary increases were observed in students' own self-assessments of presentation-file quality. Finally, students did not conduct overall assessments with the same level of demand as their teachers at any time during the experience.

Discussion: The results demonstrate that this program and similar initiatives are useful to help engineering students to develop critical-thinking skills and to broaden their expectations with respect to their own and their colleagues' work. However, other relevant aspects could be evaluated in lengthier programs, such as whether the students' levels of critical thinking and expectations are maintained when the comparative reference of a teacher's assessment is unavailable.

KEYWORDS

engineering, higher education, critical thinking, expectations, Likert-type scale survey, peer-assessment, self-assessment

1 Introduction

Critical pedagogy is an educational stream aimed at fostering critical awareness in students, both in terms of self-, peer- and social criticism (May, 2024). This type of pedagogy intends students to develop critical-thinking skills regarding the surrounding social reality, as well as with the attitudes and work of others and themselves (Dessingué and Wagner, 2024). The scope of these skills will depend on the formative stage. Students in early stages of their education (infant and primary) should be able to assess whether a behavior is right or wrong, and determine whether any action could have been performed in a more correct way (Noula, 2018). In more advanced educational stages (secondary school), critical pedagogy includes the development of critical thinking not only about attitudes, but also about social realities and the own and others' work (Heckler et al., 2013; Lithoxidou et al., 2021). Finally, in the field of higher education, the implementation of this type of pedagogy is extremely useful, as it prepares students to excel in their future professional work through self- and peer-assessments (Echeandía et al., 2024), promoting a successful peer-to-peer work environment (Lithoxidou et al., 2021).

Engineering stands as a field in which the application of critical pedagogy can be particularly beneficial for students, given that it is a profession with a very clear civic vocation, in so far as its final aim is to respond to collective needs across a wide range of areas within society (Silveira et al., 2020; Kivimäki et al., 2023). Those areas range from proper management of a continuous supply of food, water, and electricity, to the design and the provision of goods, housing, and infrastructure (Revilla-Cuesta et al., 2023). Indeed, consumers and by extension society as a whole will often pronounce their own verdicts on the acceptability or otherwise of engineering works (Suomi et al., 2017). However, an absence of open evaluation may lead engineering professionals to adopt conformist attitudes, stifling creativity and resulting in the mere fulfillment of their duties. So, given the strong social focus of the engineering profession, it is essential that engineers give their utmost to their work and address the problems that may arise in the most effective way (Casper et al., 2021; Dias, 2023). In other words, engineers must adopt critical attitudes and hold high expectations toward their own work, striving at all times to excel.

Rather than the work of one individual, engineering projects invariably involve multidisciplinary teams today, in which a specific dimension is developed by each member (Sacks et al., 2022). An engineer must therefore employ communication and teamwork skills, working in coordination with colleagues, so that the final project may be perceived as a whole, rather than as a collection of interconnected parts (Al Hadithi, 2018; Say et al., 2022). In addition, based once again on the social focus of an engineer's professional work, engineers must be critical and demanding with regard to the work of their peers, so that teamwork can solve each problem in the most appropriate way (Mesutoglu et al., 2022). Although some engineers may be working to the best of their ability, their efforts might be in vain, if other team members are not doing so.

In view of the above, it is essential that engineers acquire self-criticism and peer-critique skills during their training (Mesutoglu et al., 2022; Santos et al., 2023). In doing so, future professional engineers will gain critical expectations toward their own work and the work of their peers, with the final aim of arriving at the most effective solutions (Siu, 1999), thereby guaranteeing the fulfillment of their professional commitments (Revilla-Cuesta et al., 2023).

In terms of educational methodologies, peer- and self-assessment in the context of formative assessment are effective tools to help students develop critical-evaluation skills (Hortigüela et al., 2015). The literature suggests that those assessments should never contribute to the final grade of a course (Lee et al., 2023), so that sort of mechanism helps students to reflect both on their own performance and on the performance of their peers, abandoning the common tendency to consider that minimum effort may at times be sufficient (Brazeal et al., 2021). However, time constraints and the absence of an objective reference often limit the effectiveness of those sorts of educational methodologies, in terms of developing critical thinking:

- Time is generally a fundamental variable in any learning process. On the one hand, time means that learners progressively comprehend and internalize concepts, so that they are capable of applying or explaining them later on (Chu and Chen, 2012; Bongers et al., 2020). On the other hand, in terms of skills and competencies, progressive changes among learners, whether in their thinking or their actions, also require time (Ryan et al., 2023). Peer- and self-assessments conducted at a specific point in time leave no room for learners to reflect on the adequacy and correctness of their previous assessments, and to apply the conclusions and insights that they have reached in future assessments (Soria et al., 2023). Repeated assessments over lengthier time scales can solve those sorts of issues (Pueyo and Alcalá, 2020).
- Any type of assessment is inherently subjective (Iskandar et al., 2023). If we solely focus on a numerical scale, the same rating in an assessment may correspond to two different levels of work or learning for two different people (Fagerlin et al., 2007). Furthermore, student ratings of their peers may often be influenced by friendship (Revilla-Cuesta et al., 2020) and many people cling on to the erroneous perception that their work is always of a high level of quality (Glenn and Pepper, 2023). Sharing an objective assessment, such as for example, the assessment of the teacher, can therefore serve as a reference element for the students to reflect upon both peer- and self-assessments.

Surveys are certainly the most effective element for that sort of assessment, because of their clear evaluation framework (Zheng and Xu, 2023). Two main groups can be distinguished. Quantitative-survey results show students their assessments on an objective scale, indicating whether improvement is needed, and leaving the student

to reflect upon the aspects to improve, as they are not specified (Edelhauser and Lupu-Dima, 2021; Leung et al., 2021). Qualitative survey results give students some knowledge of where improvements may be introduced (Hortigüela et al., 2015), although they offer no overall assessment and, if the explanations are not sufficiently clear, the results may disorient students as to where they need to improve (Revilla-Cuesta et al., 2023). A survey that requires quantitative assessments of the work within a wide range of dimensions is perhaps the most appropriate way to perform these assessments, as it offers the student a clear breakdown of the areas that are evaluated and the grades for each one (Lee et al., 2023; Soria et al., 2023).

In this paper, the utility of an innovative program is discussed. It is based on quantitative surveys to teach self-criticism and peer-critique to engineering students, addressing the utility of both over a relatively broad time scale, and using teacher assessments as a comparative reference. To do so, the students prepared several in-class presentations that were self- and peer-assessed, the ratings for which were subsequently published alongside the teacher's ratings, leaving the students with sufficient time for reflection before the next presentation, which was once again self- and peer-assessed. The idea is to underline that the implementation of this sort of activity, quite simple in itself, means that student expectations toward their own work and the work of others are broadened, which positively affect their performance and their awareness of the civic purpose of their future profession as engineers.

2 Materials and methods

2.1 Study program

The students participating in this program were enrolled on various Bachelor's and Master's engineering degrees at university. They delivered three in-class presentations on topics related to each course that their teachers had previously defined. Each presentation was the individual work of each student and involved gathering information on the proposed topics to prepare a presentation, which was subsequently presented in class within a timeframe of 8 to 10 min. The presentations were evenly distributed throughout the 12 weeks of each course. Thus, the first in-class presentation was conducted 4 weeks after the beginning of the course, and the two following ones were 4 weeks apart.

During the first in-class presentation, following the quantitative survey developed for this research and shown below, different aspects of the work and the attitudes of each student were rated by their peers and their course-teacher. A 5-point Likert-type scale (1, very bad; 5, very good), quite common in this kind of study (Oliveira et al., 2023), was used. In addition, when the students had finished their presentations, they self-assessed themselves using the same survey. The course teachers collected all completed surveys and posted the results on the teaching platform the day after the first presentation. They remained available until the second presentation. In that way, the students could reflect on the ratings they had received, on their self- and peer-assessments, and on their own work over a suitable period of time and when preparing the next in-class presentation (Hortigüela et al., 2015). The teacher's rating that served as a comparative reference when conducting this reflection was also posted on the student platform (Glenn and Pepper, 2023).

The same process was followed for the second and third presentations, although the third presentation was the last one for all the courses and any student reflections after that presentation were not monitored. The entire process of this study is depicted in Figure 1.

2.2 Study units and participants

The students were enrolled on study units that formed part of two Bachelor's degrees and a Master's degree at the University of Burgos, as detailed in Table 1. This is the university where the authors are affiliated, so conducting the study there enabled up-to-date follow-up. All the study units met two requirements. Firstly, they had a reduced number of students (maximum of 15 students enrolled), so that the students could complete the three in-class presentations individually in every one with no problems related to timing and course organization (Imran et al., 2022). Secondly, the study units corresponded to advanced years of the Bachelor's, and the Master's Degrees. At these educational stages, university students will have developed deeper insight into the quality of a presentation, whether it is their own work or the work of their peers (Furdui et al., 2021), which implies deeper and more demanding reflection (Revilla-Cuesta et al., 2020).

In all, 44 students enrolled on the study units shown in Table 1 participated in the study. Their average age was 22.58 ± 3.38 years old. Those students gave the three in-class presentations on some courses, so their results could yield valid conclusions. The students who gave only one or two presentations were not considered. The average age of the three teachers who participated in this research on a voluntary basis was 37.33 ± 8.08 years old.

2.3 Instrument: survey

All the assessments reported in this study, whether teacher-, peer-, or self-assessments, were conducted by completing the quantitative survey shown in Table 2. The researchers who conducted this study developed the survey on the basis of previous work (Revilla-Cuesta et al., 2021) and other studies available in the literature (Seifan et al., 2020; Feijóo et al., 2021). The Cronbach's alpha coefficient of the survey results was 0.795, which confirmed the internal consistency or reliability of the data at a level that, on the basis of previous experience, was considered adequate (Revilla-Cuesta et al., 2021).

The survey respondent was simply asked to evaluate aspects pertaining to three dimensions of the students' in-class presentations on a 5-point Likert-type scale (1, very bad; 5, very good), an easy-to-use quantitative scale (Oliveira et al., 2023). Those dimensions were as follows: explanatory ability, presentation-file quality, and attitude during the presentation. Around 5 different aspects were evaluated for each dimension. In addition, an overall assessment of the work was also requested at the end of the survey, which may be considered the fourth survey dimension. There was no evaluation of course subject matter, i.e., whether there had been any omission of key concepts, whether the discussion of all the aspects was correct, and whether there had been any errors. The reason was because the students were not supposed to have gained sufficient knowledge for an adequate assessment of those aspects, which might not therefore have been on a solid basis (O'Donovan, 2023). The above-mentioned dimensions

TABLE 1 Study units.

Bachelor's/Master's Degree	Study Unit	Year
Bachelor's Degree in Civil Engineering	Environmental Engineering	4th year
Bachelor's Degree in Civil Engineering	Project and Construction Management	3rd year
Bachelor's Degree in Agroalimentary Engineering & the Rural Environment	Agricultural Infrastructure	4th year
Bachelor's Degree in Agroalimentary Engineering & the Rural Environment	Engineering of Green Spaces	4th year
Master's Degree in Civil Engineering	Water and Waste Treatment	1st year

were therefore rated, which in no way depended on student knowledge gained during the course, but instead on the efforts of the students when preparing the presentation (Fagerlin et al., 2007).

2.4 Results analysis and procedure

The results consisted of the responses that both the teachers and the students had given to the different points. Therefore, an analysis of quantitative variables could be conducted at all the stages of the study (Cirillo et al., 2016).

- When all the students had conducted their in-class presentations, the teacher-in-charge collected all the peer-assessments, and calculated the average ratings for each presentation. Those results, together with the self- and teacher-assessments, were posted on the teaching platform the day after. In that way, the students were able to reflect on the assessments that their peers had advanced before the next in-class presentation (Hortigüela et al., 2015), and were able to compare them with the self- and teacher-assessments (Glenn and Pepper, 2023).
- At the end of the study, the mean rating and the standard deviation of each survey dimension were separately obtained for each presentation number and for each rater (self-, peer-, and teacher-assessments). The objective was, on the one hand, to track the way that the self- and peer-assessments had changed over time and, on the other, to compare both the self- and the peer-assessments with the teacher- assessments. By doing so, it could be analyzed whether the students were developing their self- and peer-critique skills and whether their critiques were more demanding over time (Chu and Chen, 2012; Hortigüela et al., 2015). No distinction could be drawn between courses, because participant numbers were insufficient to do so, so the conclusions had to be drawn from a general perspective (Revilla-Cuesta et al., 2021). The results are presented in the following section.
- Finally, the significance of each factor in the study (dimension assessed, rater, and time point or presentation number) was analyzed with a three-way ANalysis Of VAriance (ANOVA). The

TABLE 2 Survey.

Explanatory ability
1. The presentation was well prepared and delivered (not merely read out aloud).
2. The speaker communicated an understanding of what was said and was not just "reciting a lesson."
3. The concepts were correctly explained.
4. The concepts were explained in an acceptable order.
5. The concepts were explained in the speaker's own words (i.e., not a 'copied' text).
6. I understood all the concepts that were explained.
Presentation-file quality
7. The presentation file was carefully crafted.
8. The presentation file had an attractive format.
9. The presentation file was well organized.
10. The presentation file had images or graphs to complement the text.
11. There was no excessive text that might hinder clear visualization of the presentation file.
Attitude
12. The presentation was interesting.
13. In the spoken presentation, the speaker touched on aspects not covered in the presentation file.
14. The speaker was proactive, seeking to stimulate the interest of the audience.
15. The speaker expressed enthusiasm for the subject matter that was addressed.
Overall assessment
16. The general assessment of the speaker was...
17. I would be willing to listen to the same speaker at another presentation.

ANOVA results added robustness to the conclusions of the study (Meier, 2022).

3 Results and discussion

3.1 Effect of time

One of the objectives of the study was to analyze whether reflection over time on the work and assessments could lead to a more critical perception among students of their own work and the work of their peers. Figure 2 shows, for each rater (peers, self-assessment, and teachers), the pattern of rating for each survey dimension throughout the three in-class presentations. Table 3 specifies the percentage variations of the ratings for each rater, in the second and third in-class presentations with respect to the first one.

Both peer- (Figure 2A) and self-assessments (Figure 2B) showed very similar trends. In general, the ratings decreased between the first and the second in-class presentations, although those variations were minimal, with a maximum decrease of -2.4% for the peer-assessments and -3.6% for the self-assessments. Those minimal decreases suggested that the ratings remained approximately constant. Much lower ratings were detected for the third in-class presentation, with maximum values of -8.8% and -7.9% for the peer- and self-assessments, respectively. People tend to improve the quality of a certain type of work when it was repeated (Jacob et al., 2022), as they already know how to perform it properly (Klein, 2012). Therefore, the lower ratings appeared to indicate that students became more

demanding over time toward their own work and the work of their peers, giving lower ratings for work that had in principle better quality, as all the in-class presentations required similar levels of effort.

Both peer- and self-assessments showed notable decreases over time in all the survey dimensions. In both cases, attitude was the dimension that experienced the lowest variations over time. Attitude is more properly assessed with experience and objectivity

(Choi and Kim, 2006) and any changes may not be properly perceived in a study within a relatively short time scale (Hofman and Kremer, 1983). It was therefore thought that the students never fully appreciated the temporal changes to attitude in this research. In the peer-assessments, the greatest decreases occurred in relation to explanatory ability, which was quite appropriate since the students who attended the presentations were the ones who really

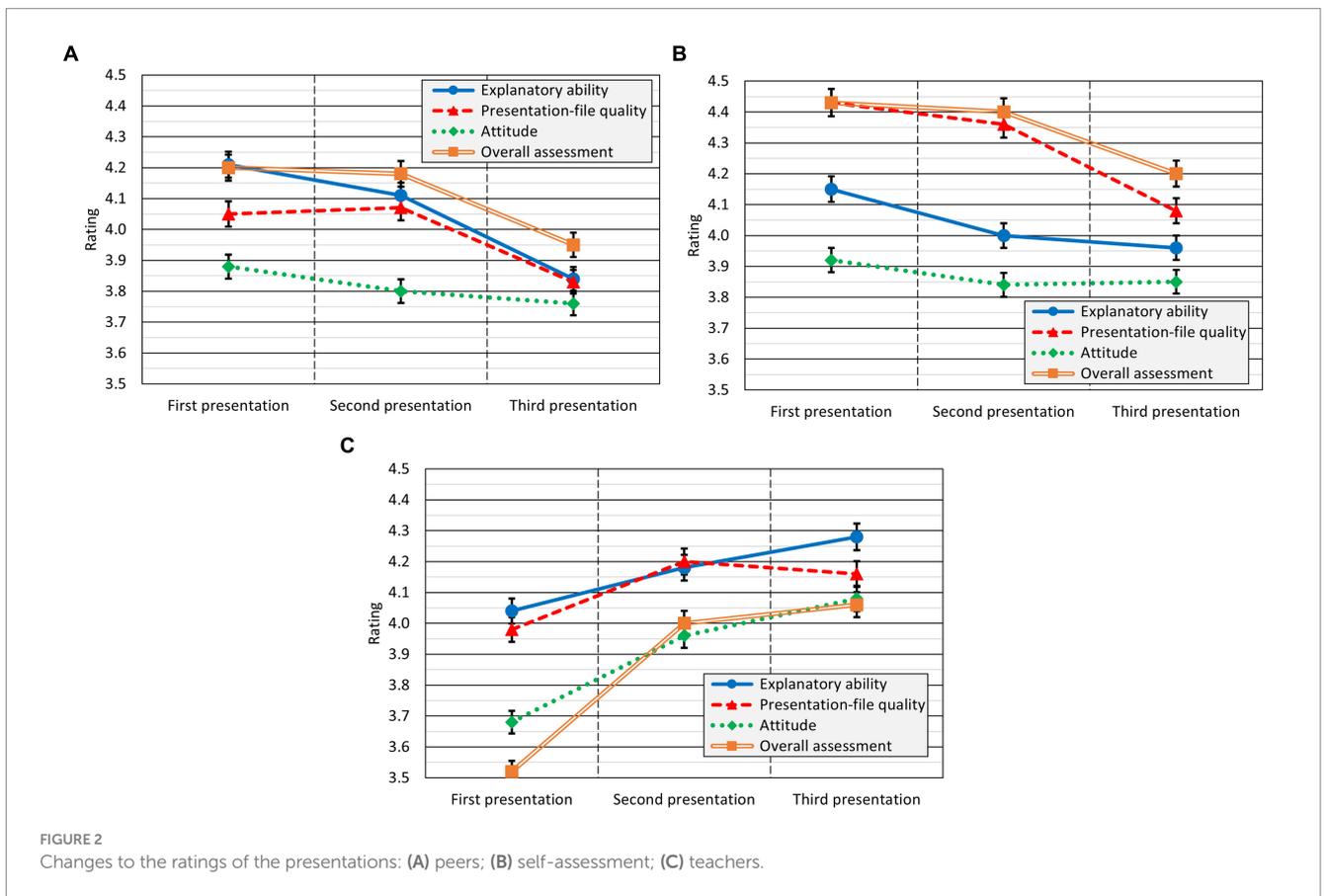
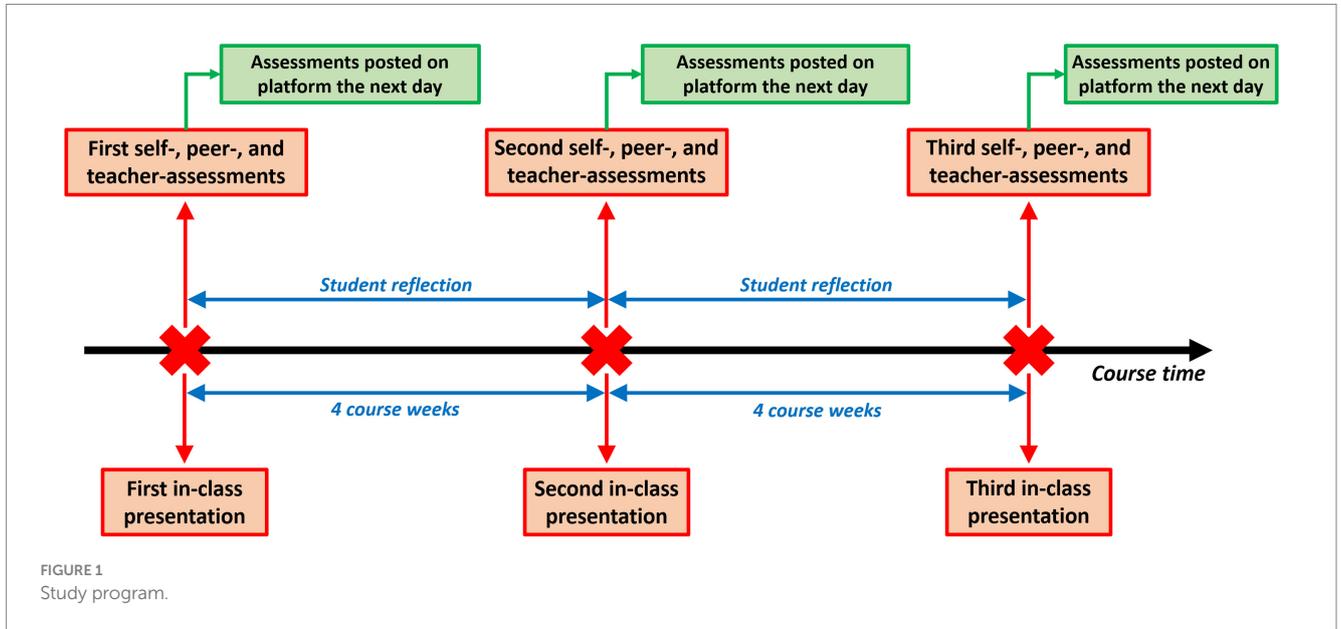


TABLE 3 Percentage variations of ratings with regard to the first presentation.

Survey dimension	Peers		Self-assessment		Teachers	
	Second presentation	Third presentation	Second presentation	Third presentation	Second presentation	Third presentation
Explanatory ability	-2.4	-8.8	-3.6	-4.6	+3.5	+5.9
Presentation-file quality	+0.5	-5.4	-1.6	-7.9	+5.5	+4.5
Attitude	-2.1	-3.1	-2.0	-1.8	+7.6	+10.9
Overall assessment	-0.5	-6.0	-0.7	-5.2	+13.6	+15.3

know whether it had been well or poorly delivered (Martínez and Ahumada, 2016). In the self-assessments, presentation-file quality was the most demanding dimension over time for the students, as they supposedly had the greatest previous knowledge of it and could develop the most accurate opinions (Contreras et al., 2022; Becerra et al., 2023).

If the percentage variations of the ratings of the peer- and self-assessments for the second and third presentations are compared (Table 3), then another relevant aspect can be noted. The self-assessment ratings experienced higher decreases than the peer-assessment ratings between the first and second in-class presentations (between -0.7% and -3.6%), while the decreases between the first and third in-class presentations were higher for the peer-assessment ratings (between -3.1% and -8.8%). It all indicated that students were slower at developing critical-thinking skills when assessing the work of their peers, due perhaps to feelings of friendship and, if the assessments were not positive, fear of negative attitudes toward them (Heckler et al., 2013; Soria et al., 2023). They therefore initially found it easier to be more demanding toward their own work. However, as time progressed, students gradually got used to the practice of co-assessment and those feelings faded, which led to greater demands that were focused on the work of others, which indeed responded to the usual tendency to consider one's own work as better (Glenn and Pepper, 2023).

In relation to the teachers' ratings (Figure 2C), they increased over time for all the survey dimensions. In the second in-class presentation, the ratings with respect to the initial ones increased between 3.5 and 13.6%, while the increases were between 4.5 and 15.3% for the third presentation. As discussed above, the process of delivering all in-class presentations was similar, and the students' work increased in quality, as they became familiar with the process and requirements (Klein, 2012; Jacob et al., 2022). That observation was consistent with the higher ratings of the teachers, which can be presumed to be more objective than the students' ratings. It may also primarily explain the improvement in the explanatory ability and the presentation-file quality of the students, as they could autonomously train and improve those dimensions (Feijóo et al., 2021). In addition, the improvement in the ratings for attitude may be attributed in part to the students' growing awareness of the peer-assessments of their work, which might have increased their motivation (Tenório et al., 2016; Ginsburg and Stroud, 2023). Finally, it was remarkable that the trend of the teachers' assessments was quite unlike the peer- and self-assessment ratings of the students, which decreased over time. It clearly shows that the students' capability to make critical judgments of both their peers' and their own work was increasing despite the fact that the objective quality of the presentations was also increasing, as was evident in the

teachers' assessments and as was also noted in the literature (Glenn and Pepper, 2023).

3.2 Effect of the comparative reference

The second objective of this study was to evaluate whether the simultaneous release of the teachers' ratings as a comparative reference with the results of the peer- and self-assessments led students to reflect on their own ratings and to adjust them over time. To illustrate this point, Figure 3 shows all the ratings over the three in-class presentations for each dimension under analysis. In addition, Table 4 also details the percentage differences between teachers' ratings and those of the peer- and self-assessments for each in-class presentation.

The first aspect that can be noted from Figure 3 is that in most of the dimensions (presentation-file quality, attitude, and overall assessment) that were analyzed, self-assessment led to higher ratings than peer-assessment. In this type of experience, people tend to rate their own work slightly higher than the work of others (Glenn and Pepper, 2023). The self-assessment ratings for the presentation-file quality and the overall assessment were 0.2–0.4 points higher than those of the peer-assessments, while the same difference for attitude was only 0.05–0.1 points, since it cannot be properly rated without levels of experience (Choi and Kim, 2006) that the students never had when this research was conducted. In spite of all this, it should also be kept in mind, as mentioned above, that the decreases of the self-assessment ratings were considerably higher in the first instance (between the first and second in-class presentations), due to an initial possible reticence on the part of students to criticize the work of their peers (Soria et al., 2023). The only dimension in which the peer-assessment ratings were slightly higher (around 0.05–0.1 points) than the self-assessment ratings was explanatory ability (Figure 3A). Assessing the quality of one's own explanations is more complicated than assessing the quality of other explanations, as an audience will usually be able to assess more accurately whether they have understood a presentation and, therefore, the quality of any explanations (Martínez and Ahumada, 2016). It may have led students to give lower ratings when self-assessing explanatory ability to ensure their acceptability.

If the peer- and self-assessments are compared with the teachers' ratings, a very clear trend can be observed. As mentioned in the previous section, the students gave lower ratings to both their peers and themselves for the in-class presentations over time, while the teachers gave higher ratings. Thus, the students' assessments were higher in the first in-class presentation than those of the teacher in all the dimensions, while they were lower in the third in-class

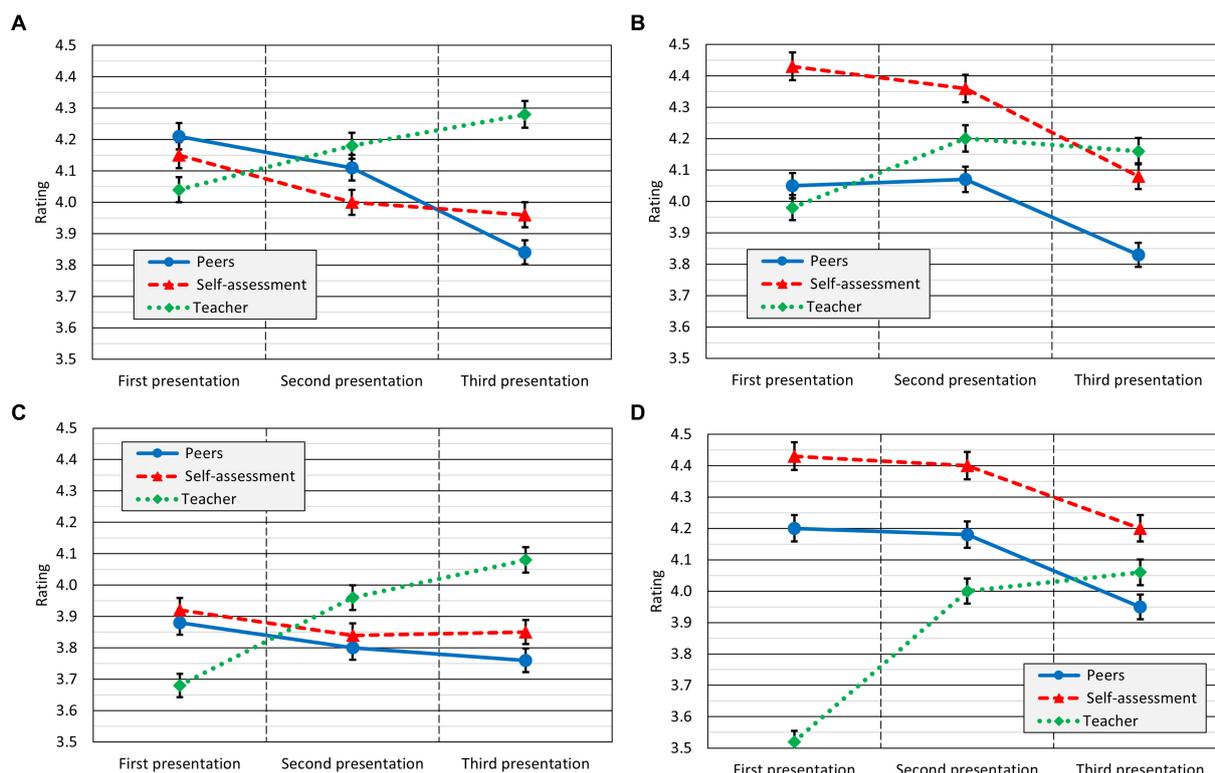


FIGURE 3 Rating patterns for all presentations by survey dimensions: (A) explanatory ability; (B) presentation-file quality; (C) attitude; (D) overall assessment.

TABLE 4 Percentage variations of ratings with regard to the ratings of the teachers.

Survey dimension	First presentation		Second presentation		Third presentation	
	Peers	Self-assessment	Peers	Self-assessment	Peers	Self-assessment
Explanatory ability	+4.2	+2.7	-1.7	-4.3	-10.3	-7.5
Presentation-file quality	+1.8	+11.3	-3.1	+3.8	-7.9	-1.9
Attitude	+5.4	+6.5	-4.0	-3.0	-7.8	-5.6
Overall assessment	+19.3	+25.9	+4.5	+10.0	-2.7	+3.4

presentation than those of the teacher in almost all cases. The availability of the teachers' ratings undoubtedly meant that the students could compare them with their own ratings and reflect upon the scale of their demands and expectations toward both their own work and the work of their peers (Siu, 1999; Becerra et al., 2023). As a result, the students' ratings were progressively adjusted, becoming more accurate over time, and successfully developing more adjusted critical-thinking valuations (Echeandía et al., 2024). Thus, in view of the discussion, it appears clear that the combination of a broad time scale with the use of a comparative reference meant that the students could develop a more critical view of the work performed and could adjust their ratings. However, the development of self-criticism skills was more complicated, since the self-assessment ratings were found to be less well adjusted to the teachers' ratings than the peer-assessment ones, according to the values shown in Table 4.

Despite this general trend, the evolution of the level of demand and the critical skills of the students depended on the dimension that was rated. Thus, the following aspects should be highlighted:

- The ratings in the peer- and self-assessments of both explanatory ability (Figure 3A) and attitude (Figure 3C) were lower than those of the teachers in the second and third in-class presentations (Table 4). Those dimensions corresponded to aspects with which the engineering students were less familiar (Dias, 2023), so the level of their demands could be increased more easily. In addition, those dimensions, such as explanatory ability (Martínez and Ahumada, 2016) and attitude, also involved complicated self-assessments for students with little or no assessment experience (Choi and Kim, 2006).
- A different trend was noted in relation to self-assessment of presentation-file quality (Figure 3B). Peer-assessments showed that the students were as demanding toward their assessments of the presentation files as they were toward the explanatory ability and attitude of their peers. However, the self-assessment ratings for this dimension were 11.3% higher than the teachers' ratings in the first presentation. Furthermore, those self-assessment ratings were only lower than the teachers' ones in

the third in-class presentation. Presentation-file quality was the survey dimension over which the students had greater knowledge and control (Contreras et al., 2022; Becerra et al., 2023), which may have hindered them at the beginning from assuming their own mistakes and making fair assessments. However, it was the dimension with the lowest fall in self-assessment ratings between the first and the third in-class presentations, so the critical thinking of students may have developed on that point.

- Finally, the overall assessment (Figure 3D) was the item that clearly required the most effort from the students to adjust their ratings. Three aspects justify that last statement: (1) it was the item with the highest overvaluation in the first in-class presentation (ratings that were 19.3 and 25.9% higher than those of the teachers in the peer- and self-assessments, respectively); (2) it was the only item in which both the peer- and self-assessments ratings were higher than those of the teachers in the second in-class presentation; (3) only peer-assessments showed lower ratings than the teachers' assessments of the third in-class presentation. That behavior could be due to two aspects. First, the extremely general nature of the item, without focusing on any specific aspect, which may have detracted from any clear reference for students when assigning their ratings (Macken et al., 2020). In addition, it was the most closely linked element to the concept of "grade" (Revilla-Cuesta et al., 2021), which made students more reluctant to give a "low grade," especially for self-assessment.

3.3 Statistical validation of significance

A three-way ANalysis Of VAriance (ANOVA) was performed at a significance level of 5%, to ensure the accuracy of all the aspects discussed (O'Donovan, 2023), considering the three factors involved in this study: dimension (explanatory ability, presentation-file quality, attitude, and overall assessment), order of presentation (first, second, and third), and rater (peers, self-assessment, and teachers). The *p*-values shown in Table 5 demonstrated that the effect of each factor on the ratings was significant, as all of them were lower than the significance level (0.05) and led to the rejection of the null hypothesis of no significant effect. Furthermore, the effect of each factor was significantly different when varying the other factors, i.e., all

interactions were also significant. Therefore, all the trends discussed in the previous sections were valid, showing the actual behavior of the raters at assessing each dimension of analysis over the different in-class presentations (Meier, 2022).

4 Conclusion

The results of a teaching program with engineering students have been presented and discussed in this paper. In this program, the students performed three in-class presentations following different study units, four dimensions (explanatory ability, presentation-file quality, attitude, and overall assessment) of which were quantitatively assessed on a 5-point Likert-type scale by the students, by the teachers, and through self-assessment. All the ratings were shared with the students for them to reflect on their assessments during 4 weeks before going through the same evaluation process in the next in-class presentation. The objective was to determine whether a program that combined a broad time scale with a comparative reference (teachers' ratings) for students could be successful at helping them develop critical-thinking skills and increase their own demands toward their own work and the work of their peers. The following conclusions can be drawn from this experience:

- Peer- and self-assessment ratings decreased over the in-class presentations (time), while teachers' ratings increased. The quality of the students' work, on the basis of the teachers' ratings of the in-class presentations, was objectively increasing. Thus, the decrease in the students' assessments showed that they were progressively developing critical thinking and increasing their demand for the work done by themselves and others, the experience being successful in this regard.
- Initially, the sharpest decrease in student ratings was for the self-assessments. Students found it easier to criticize themselves initially, perhaps influenced by friendly relations with their peers. However, they lost that fear in the long term and became more critical toward their peers, considering their own work to be at least slightly superior.
- Students' ratings in both peer- and self-assessments were higher than the teachers' ratings for the first in-class presentation, although the increased demands that they placed both on their own work and on the work of their peers meant that their ratings were generally lower than those of the teachers for the third presentation. The peer-assessment ratings showed a tighter final adjustment to those of the teachers, because of the greater long-term demands developed with the work of the peers.
- Peer-assessments became more demanding than any other for explanatory ability. Naturally enough, people attending a presentation are better positioned than the speaker to assess the effectiveness of any explanations.
- The survey dimension presentation-file quality was initially the most overrated among the dimensions for self-assessment. However, the highest increase in self-demand was applied to that dimension, perhaps because it was an aspect about which the students knew most and were able to control by themselves.
- Finally, overall assessment was highly overrated throughout the whole program both in the peer-assessments and, especially, in

TABLE 5 *p*-values of the three-way ANOVA at a confidence level of 95%.

Element of analysis		<i>p</i> -value
Factors	Dimension	0.0000
	Presentation number	0.0006
	Rater	0.0000
Interactions	Dimension and presentation number	0.0013
	Dimension and rater	0.0000
	Presentation number and rater	0.0000

the self-assessments. The fact that it could be likened to the concept of a “*grade*” perhaps meant that students were more reluctant to be as demanding toward that survey dimension as toward the other dimensions. Therefore, further efforts should be made to provide students with skills that enable them to make general assessments in an adequate manner.

5 Limitations of the study and future research lines

From the point of view of the authors of this study and in accordance with the conclusions, the program has been successful, in so far as the students could progressively develop critical thinking with regard to both the work of their peers and themselves, and could adequately adjust their ratings over time. It requires repeated use of peer- and self-assessments within a broad time scale and the promotion of reflections among students on their ratings based on a comparative reference, such as teachers’ ratings. However, the program has also revealed some limitations, which are highlighted below:

- First, the time scale could be increased, repeating the procedure a greater number of times, in order to determine whether the students’ assessments stabilized over time and whether they properly matched the quality of the assessed work.
- Secondly, the behavior of the peer- and self-assessments of the students was not monitored when the teachers’ ratings were no longer provided once it was considered that the students had adequately developed critical-assessment skills. Thus, it remains uncertain whether they would continue adjusting to the teachers’ assessments or revert to the initial tendency to rate works more highly and in some cases much higher, such as for example the overall assessment.
- Thirdly, it was not verified whether the levels of critical judgment could have been maintained, had the type of work for student assessment been modified; in other words, if instead of evaluating an in-class presentation, a written text or a more complex project had been assessed.
- Finally, it was not verified whether students were able to maintain those critical-thinking skills over a long period of time after the end of the program. The program might have to be repeated, for example, 1 year after its completion.

All the limitations mentioned above, which can also be seen as future research lines, have as a common factor the need for lengthier temporal survey-based programs. It is not a straightforward matter in view of the current length of time for university courses, usually of duration of one semester (de la Fuente Arias et al., 2010). Thus, programs that encompass several successive courses might be necessary, which could imply greater multi-disciplinarity in the development of critical thinking by students (Alves et al., 2017). The assessment of projects or presentations related to diverse topics could also motivate students when facing similar situations within the professional sphere of engineering (Revilla-Cuesta et al., 2020).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because all participants in this research, of legal age, gave their explicit written consent to voluntarily participate in the study as long as their anonymity was always guaranteed in all publication stages of this study. No approval of this educational research by an ethics committee was required. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

VR-C: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing – original draft. NH-A: Data curation, Formal analysis, Investigation, Visualization, Writing – review & editing. IF: Conceptualization, Investigation, Resources, Visualization, Writing – review & editing. MS: Conceptualization, Funding acquisition, Methodology, Writing – review & editing, Project administration. VO-L: Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors wish to express their gratitude to the Junta de Castilla y León (Regional Government) and ERDF [grant number BU066-22]; and the University of Burgos for the funding program “Convocatoria de Ayudas a Grupos de Innovación Docente reconocidos para la elaboración de materiales docentes para los años 2023 y 2024.”

Acknowledgments

The authors would like to thank all the students and professors who participated in this study for their availability.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Al Hadithi, B. I. (2018). An investigation into factors causing delays in highway construction projects in Iraq. *MATEC Web Conf.* 162:02035. doi: 10.1051/mateconf/201816202035
- Alves, J. L., Carvalho, B. R., Canavaro, V., and Monteiro, D. (2017). Transforming waste in industrial design products for social vulnerable groups: teaching industrial design based on real projects, a project based learning experience in faculty of engineering of university of Porto. In IEEE Global Engineering Education Conference, EDUCON. Athens, Greece: IEEE.
- Becerra, I. J., Reyes, R. C., Marín, A. A., and Vargas, L. D. A. (2023). ICT-mediated teaching models in university teaching: a systematic review. *Educ. Pesqui.* 49:e251276. doi: 10.1590/S1678-4634202349251276es
- Bongers, A., Flynn, A. B., and Northoff, G. (2020). Is learning scale-free? Chemistry learning increases EEG fractal power and changes the power law exponent. *Neurosci. Res.* 156, 165–177. doi: 10.1016/j.neures.2019.10.011
- Brazeal, K. R., Brown, T. L., and Couch, B. A. (2021). Connecting activity implementation characteristics to student buy-in toward and utilization of formative assessments within undergraduate biology courses. *J. STEM Educ. Res.* 4, 329–362. doi: 10.1007/s41979-021-00054-2
- Casper, A. M. A., Atadero, R. A., Hedayati-Mehdiabadi, A., and Baker, D. W. (2021). Linking engineering Students' professional identity development to diversity and working inclusively in technical courses. *J. Civ. Eng. Educ.* 147:04021012. doi: 10.1061/(ASCE)E12643-9115.0000052
- Choi, B., and Kim, C. (2006). A learning attitude evaluation system for learning concentration on distance education. *Lect. Notes Comput. Sci.* 3983, 808–817. doi: 10.1007/11751632_87
- Chu, H. C., and Chen, J. H. (2012). A time scale-based concept map approach to developing educational computer games for historical courses. Workshop proceedings of the 20th international conference on computers in education, ICCE 2012, Singapore: ICCE. 565–571.
- Cirillo, G., Nughes, E., Acanfora, A., Altavilla, G., and D'Isanto, T. (2016). Physical and sport education testing by quantitative and qualitative tools in assessment in senior school: a proposal. *Sport Sci.* 9, 97–101.
- Contreras, J. L. G., Torres, C. A. B., and Ojeda, Y. C. E. (2022). Using of ICT and LKT in higher education: a bibliometric analysis. *Rev. Complut. Educ.* 33, 601–613. doi: 10.5209/rced.73922
- de la Fuente Arias, J., Vicente, J. M. M., Sánchez, F. J. P., and Berbén, A. B. G. (2010). Perception of the teaching-learning process and academic achievement in diverse instructional contexts of higher education. *Psicothema* 22, 806–812
- Dessingué, A., and Wagner, D. A. (2024). Promoting dialogical critical thinking in education: examining teachers' practices and conceptualizations in the Norwegian school context. *J. Curric. Stud.* 22, 1–19. doi: 10.1080/00220272.2024.2334937
- Dias, D. (2023). Engineering learning outcomes: the possible balance between the passion and the profession. *Soc. Sci.* 12:37. doi: 10.3390/socsci12010037
- Echeandía, R., Murillo, J., and Palomino-Flores, P. (2024). Cultivating critical thinking among journalism students in the digital age. *RISTI* 2024, 370–382.
- Edelhauser, E., and Lupu-Dima, L. (2021). One year of online education in covid-19 age, a challenge for the romanian education system. *Int. J. Environ. Res. Public Health* 18:8129. doi: 10.3390/ijerph18158129
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., and Smith, D. M. (2007). Measuring numeracy without a math test: development of the subjective numeracy scale. *Med. Decis. Mak.* 27, 672–680. doi: 10.1177/0272989X07304449
- Feijóo, J. C. M., Suárez, F., Chiyón, I., and Alberti, M. G. (2021). Some web-based experiences from flipped classroom techniques in aec modules during the covid-19 lockdown. *Educ. Sci.* 11:211. doi: 10.3390/educsci11050211
- Furdui, A., Lupu-Dima, L., and Edelhauser, E. (2021). Implications of entrepreneurial intentions of romanian secondary education students, over the romanian business market development. *PRO* 9:665. doi: 10.3390/pr9040665
- Ginsburg, S., and Stroud, L. (2023). Necessary but insufficient and possibly counterproductive: the complex problem of teaching evaluations. *Acad. Med.* 98, 300–303. doi: 10.1097/ACM.0000000000005006
- Glenn, L. E., and Pepper, C. M. (2023). Reliability and validity of the self-rating scale as a measure of self-criticism. *Assessment* 30, 1557–1568. doi: 10.1177/10731911221106768
- Heckler, N. C., Forde, D. R., and Bryan, C. H. (2013). Using writing assignment designs to mitigate plagiarism. *Teach. Sociol.* 41, 94–105. doi: 10.1177/0092055X12461471
- Hofman, J. E., and Kremer, L. (1983). Course evaluation and attitudes toward college teaching. *High. Educ.* 12, 681–690. doi: 10.1007/BF00132424
- Hortigüela, D., Pueyo, Á. P., and Abella, V. (2015). Student perspective about traditional process and formative evaluation. Group contrasts in the same subjects. *REICE-Rev. Iberoam. CA* 13, 35–48.
- Imran, M., Baig, M., Murad, M. A., and Almurashi, S. H. (2022). Factors disturbing undergraduate students' interaction during lectures: a university-based survey. *Pak. J. Med. Sci.* 38, 1945–1951. doi: 10.12669/pjms.38.7.5101
- Iskandar, A., Indahingwati, A., and Ule, K. (2023). The development of automatic subject's test applications in universities during Covid-19 pandemic. *AIP Conf. Proc.* 2798:020009. doi: 10.1063/5.0155195
- Jacob, A., Faatz, A., Knüppe, L., and Teuteberg, F. (2022). Understanding the effectiveness of gamification in an industrial work process: an experimental approach. *Bus. Process. Manag. J.* 28, 784–806. doi: 10.1108/BPMJ-08-2021-0564
- Kivimäki, V., Ketonen, E. E., and Lindblom-Ylänne, S. (2023). Engineering students' justifications for their selections in structured learning diaries. *Front. Educ.* 8:1223732. doi: 10.3389/feduc.2023.1223732
- Klein, J. (2012). Repetition of lesson presentation as a tool for improving teaching efficiency. *Teach. Teach. Theory Pract.* 18, 733–746. doi: 10.1080/13540602.2012.746508
- Lee, S. W. Y., Tu, H. Y., Chen, G. L., and Lin, H. M. (2023). Exploring the multifaceted roles of mathematics learning in predicting students' computational thinking competency. *Int. J. STEM Educ.* 10:64. doi: 10.1186/s40594-023-00455-2
- Leung, A., Fine, P., Blizard, R., Tonni, I., and Louca, C. (2021). Teacher feedback and student learning: a quantitative study. *Eur. J. Dent. Educ.* 25, 600–606. doi: 10.1111/eje.12637
- Lithoxoidou, A., Seira, E., Vrantzi, A., and Dimitriadou, C. (2021). Promoting resiliency, peer mediation and citizenship in schools: the outcomes of a three-fold research intervention. *Particip. Educ. Res.* 8, 109–128. doi: 10.17275/PER.21.32.8.2
- Macken, S., MacPhail, A., and Calderon, A. (2020). Exploring primary pre-service teachers' use of 'assessment for learning' while teaching primary physical education during school placement. *Phys. Educ. Sport Pedagog.* 25, 539–554. doi: 10.1080/17408989.2020.1752647
- Martínez, D. J. C., and Ahumada, J. R. C. (2016). Subjective theories on professors in their professional training. *Rev. Bras. Educ.* 21, 299–324. doi: 10.1590/S1413-24782016216517
- May, E. (2024). Critical pedagogy and disability in participatory research: a review. *Inf. Learn. Sci.* 2:21. doi: 10.1108/ILS-02-2023-0021
- Meier, L. (2022). *ANOVA and mixed models: A short introduction using R*. New York: CRC Press.
- Mesutoglu, C., Bayram-Jacobs, D., Vennix, J., Limburg, A., and Pepin, B. (2022). Exploring multidisciplinary teamwork of applied physics and engineering students in a challenge-based learning course. *Res. Sci. Technol. Educ.* 43, 1–19. doi: 10.1080/02635143.2022.2154334
- Noula, I. (2018). Critical thinking and challenges for education for democratic citizenship: an ethnographic study in primary schools in Greece. *Educ. Real.* 43, 865–886. doi: 10.1590/2175-623674799
- O'Donovan, R. (2023). Missing the forest for the trees: investigating factors influencing student evaluations of teaching. *Assess. Eval. High. Educ.* 11, 1–18. doi: 10.1080/02602938.2023.2266862
- Oliveira, A., Brewer-Deluce, D., Akhtar-Danesh, N., and Wojkowski, S. (2023). Readiness for interprofessional learning among health science students: a cross-sectional Q-methodology and likert-based study. *BMC Med. Educ.* 23:583. doi: 10.1186/s12909-023-04566-w
- Pueyo, Á. P., and Alcalá, D. H. (2020). Is innovation always positive in physical education? Reflections and practical considerations. *Retos* 37, 579–587. doi: 10.47197/retos.v37i37.74176
- Revilla-Cuesta, V., Skaf, M., Manso, J. M., and Ortega-López, V. (2020). Student perceptions of formative assessment and cooperative work on a technical engineering course. *Sustain. For.* 12:4569. doi: 10.3390/su12114569
- Revilla-Cuesta, V., Skaf, M., Manso-Morato, J., San-José, J. T., and Ortega-López, V. (2023). Educating future agricultural engineers at the University of Burgos, Spain,

through a service-learning project on rural depopulation and its social consequences. *Educ. Sci.* 13:267. doi: 10.3390/educsci13030267

Revilla-Cuesta, V., Skaf, M., Varona, J. M., and Ortega-López, V. (2021). The outbreak of the covid-19 pandemic and its social impact on education: were engineering teachers ready to teach online? *Int. J. Environ. Res. Public Health* 18:2127. doi: 10.3390/ijerph18042127

Ryan, O., Fisher, M. J., Schibelius, L., Huerta, M. V., and Sajadi, S. (2023). Using a scenario-based learning approach with instructional technology to teach conflict management to engineering students. ASEE Annual Conference and Exposition. Baltimore, Maryland: ASEE Conferences

Sacks, R., Wang, Z., Ouyang, B., Utkucu, D., and Chen, S. (2022). Toward artificially intelligent cloud-based building information modelling for collaborative multidisciplinary design. *Adv. Eng. Inform.* 53:101711. doi: 10.1016/j.aei.2022.101711

Santos, R., Anderson, D., and Milner-Bolotin, M. (2023). Research trends in international science, technology, engineering, and mathematics education conference series: an analysis of a decade of proceedings. *Front. Educ.* 7:1099658. doi: 10.3389/feduc.2022.1099658

Say, B., Erden, Z., and Turhan, C. (2022). A team-oriented course development experience in distance education for multidisciplinary engineering design. *Comput. Appl. Eng. Educ.* 30, 1617–1640. doi: 10.1002/cae.22546

Seifan, M., Dada, O. D., and Berenjian, A. (2020). The effect of real and virtual construction field trips on students' perception and career aspiration. *Sustain. For.* 12:1200. doi: 10.3390/su12031200

Silveira, C., Reis, L., Santos, V., and Mamede, H. S. (2020). Creativity in prototypes design and sustainability - the case of social organizations. *Adv. Sci. Technol. Eng. Syst.* 5, 1237–1243. doi: 10.25046/AJ0506147

Siu, K. W. M. (1999). Criticism: a relatively neglected area in engineering teacher training programmes. *Eng. Sci. Educ. J.* 8, 206–208. doi: 10.1049/esej:19990504

Soria, M. M., Hortigüela-Alcalá, D., López-Pastor, V. M., Pascual-Arias, C., and Fernández-Garcimartín, C. (2023). Effects of the implementation of tutored learning projects and formative and shared assessment Systems in pre-Service Teacher Education. *J. High. Educ. Theory Pract.* 23, 240–257. doi: 10.33423/jhetp.v23i2.5827

Suomi, R., Nykänen, P., Vepsäläinen, T., and Hiltunen, R. (2017). Green turning brown-domain engineering for social and health services in Finland. *Stud. Health Technol. Inform.* 245, 803–807. doi: 10.3233/978-1-61499-830-3-803

Tenório, T., Bittencourt, I. I., Isotani, S., and Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Comput. Hum. Behav.* 64, 94–107. doi: 10.1016/j.chb.2016.06.020

Zheng, Y., and Xu, J. (2023). Unpacking the impact of teacher assessment approaches on student writing engagement: a survey of university learners across different languages. *Assess. Eval. High. Educ.* 48, 1240–1253. doi: 10.1080/02602938.2023.2219431