OmniaScience

Journal of Technology and Science Education
JOTSE, 2024 – 14(2): 438-452 – Online ISSN: 2013-6374 – Print ISSN: 2014-5349
https://doi.org/10.3926/jotse.2341

IMPLEMENTATION OF EDUCATIONAL SEQUENCES BASED ON PEER ASSESSMENT FOR LEARNING KEY CONCEPTS OF STATISTICS

Ernest Pons*, Maria Elena Cano

Universitat de Barcelona (Spain)

*Corresponding author: epons@ub.edu
ecano@ub.edu

Received July 2023
Accepted November 2023

Abstract

This article presents the results of applying an educational sequence implemented with technological support on an LMS and focused on peer assessment that was designed specifically to address key concepts in statistics with first-year undergraduate students. Individualized information is available for a total of n=232 students to support the empirical conclusions that are drawn. Based on the comparison of the peer assessments and the academic performance obtained in the two previous academic years in which a different methodology was applied, differential effects are found in the quality of the assignments presented. This, together with the perception of the learning by the students, suggests the incorporation of peer assessment processes in future curricular design.

Keywords – Peer assessment, Self-regulated learning, Statistics.

To cite this article:

Pons, E., & Cano, M.E. (2024). Implementation of educational sequences based on peer assessment for learning key concepts of statistics. *Journal of Technology and Science Education,* 14(2), 438-452. https://doi.org/10.3926/jotse.2341

---

1. Theoretical Framework

1.1. Peer Assessment: Conditions and Benefits

Assessment is usually associated with a process of grading and accrediting that is uncomfortable for both students (Gijbels & Dochy, 2006; Van de Watering, Gijbels, Dochy & Van der Rijt, 2008) and faculty (García-García, Moctezuma-Ramírez, López-Francés & Pérez, 2021). Nonetheless, with sufficient assessment literacy (Carless & Winstone, 2023), we can understand that assessment may constitute a tool with great educational potential.

Formative assessment helps students to learn, making them aware of their mistakes and what they do well with regard to an assignment, and also of the process followed to achieve it. In order for assessment to be formative, it must be continuous (Näykki, Kontturi, Seppänen, Impiö & Järvelä, 2021) and it must be accompanied by feedback (Lui & Andrade, 2022).

However, the feedback is not yet perceived as information provided by the instructors to bridge the gap between real and ideal execution (Hattie & Timperley, 2007). This vision, which relegates students to the role of receivers, has been surpassed by practical feedback, understood as a process in which the students are actively involved, giving significance to the comments received (and which may come from different sources) and planning specific actions to apply them to the work and/or future learning processes (Carless & Boud, 2018).

This necessarily active role on the part of the students implies, among other possibilities, developing peer assessment processes, which have become an increasingly more common practice in higher education, since it provides a valuable alternative to traditional assessment methods. Through this focus, students have the opportunity to actively participate in the assessment of their classmates' work, promoting more interactive and cooperative learning (Ajjawi & Boud, 2017).

The benefits of peer assessment (Rodríguez-Gómez, Ibarra-Sáiz & García-Jiménez, 2013) are articulated around the active role of the student, the greater effect of the feedback processes and the development of critical thinking and evaluative judgment.

First of all, regarding the role of the student, peer assessment promotes active learning by involving the students in the assessment process and improving their understanding of key concepts. By assessing the work of their classmates, students acquire a broader perspective and develop a deeper understanding of the assessment criteria (Carless, 2019). This experience allows them to reflect on their own work and to make the necessary adjustments to improve their performance. Furthermore, the act of providing comments and suggestions to other students requires them to apply and consolidate their own knowledge, which reinforces their learning (Carless & Boud, 2018).

Secondly, receiving comments from their classmates has a specific value, because the comments that are offered tend to be more detailed and are provided in a more comprehensible language (Ajjawi & Boud, 2017). At the same time, the comments may be more readily accepted, since they come from a peer, which creates a supportive environment and the joint construction of knowledge (Anderson, El Habbal & Bridges, 2020). For these reasons, peer feedback can be more effective.

Finally, peer assessment allows critical thinking skills to be developed. By assessing and providing feedback on the work of their classmates, students must analyze, synthesize and assess the information in a critical manner. This strengthens their capacity to identify strengths and weaknesses in the work of others, and it also improves their capacity to justify and defend their proposals (Scott, 2017).

The last major benefit provided by peer assessment is the development of evaluative judgment, defined as:

> *"The capacity to make decisions about the quality of one's own work and that of others, including both understanding what quality consists of in a task or process (being very aware of the criteria and/or procedures to be carried out), and applying this understanding in the assessment of an assignment, either one's own or that of someone else."* (Tai, Ajjawi, Boud, Dawson & Panadero, 2018: page 472)

Peer assessment can also contribute to developing the SRL. According to Panadero, Alqassab, Fernández-Ruiz and Ocampo (2023), it is necessary to bear in mind that students can learn through modeling by the instructor, from examples that they can emulate, debating and/or application of the execution criteria in a structured manner and developing evaluative judgment, but they do not necessarily have self-regulation strategies. However, these are usually highly correlated concepts.

In any case, in order for these benefits to be obtained, the processes must be specifically designed, considering various conditions, among which we can highlight those indicated by Panadero, Broadbent, Boud and Lodge (2019): (a) clarifying and justifying peer assessment, as well as the expectations with regard to the students; (b) involving all the students in the decision, development and clarification of the assessment criteria; (c) pairing student participants in the process among peers, fostering a productive assessment; (d) specifically determining the peer assessment format (for example, with a number grade or

comments), as well as the method of interaction between the assessor and the assessee (for example, in person, on line, etc.); (e) providing an assessment instrument (rubric, checklists, etc.) for the assessment process; (f) specifying the continuous assessment activities and their scheduling, and (g) performing a thorough monitoring of the peer assessment process, supporting the students at all times.

These conditions have been systematized in the works of Hicks, Pandey, Fraser & Klemmer (2016) and Panadero et al. (2023), as described in the methodology section, in order to define the peer assessment process.

1.2. The Relevance of the Peer Assessment Processes in the Subject of Statistics

In the field of statistics education, it is not common to use peer assessment processes. Even so, some very interesting experiences have recently been reported, with good results (Udechukwu, 2020; Elinor, 2022).

Furthermore, peer assessment processes may prove to be a good strategy for developing complex learning, which requires connecting different concepts and skills. This is precisely one of the characteristics of the first year of the undergraduate degree in Statistics, where it is essential to learn key statistical concepts. The intent, above all, is to avoid the acquisition of misconceptions.

One of the basic concepts in statistics, but one that often generates misconceptions among students, is the concept of variability (Shiau & Zaleha, 2013). However, there are other basic concepts which, if not correctly developed, can condition the training of the future graduates in statistics to the point of implying the improper use of the inferential statistical instruments. This conception has led to the adoption of the term statistical thinking to refer to all the competences necessary for the correct application of the statistical models and techniques, beyond the strict knowledge of statistical inference on a formal level.

According to the American Statistical Association (ASA), statistical literacy must include, in a very synthetic manner and at the very least, the cognitive and attitudinal elements cited in the GAISE (2016).

All these elements make up what is commonly referred to as statistical thinking:

> *"Students must not make the mistake of perceiving of statistics as an unrelated collection of formulas and methods, (but rather as) a problem-solving and decision-making process that is fundamental for scientific research and essential for making the right decisions. Instruction must provide students with experiences that are linked to multivariate thinking, given that we live in a complex world in which the answer to a question often depends on many different factors. Students will encounter these situations within their fields of study and in daily life. Students must be prepared to answer challenging questions that require them to investigate and explore the relationships among many different variables. Doing so will help them to appreciate the value of statistical thinking and methods."* (GAISE, 2016: page 6)

In order to achieve this goal, it is important to work on these skills from the first undergraduate year on. This is the objective of the proposed activity, which is the topic of this research.

For the learning of the cognitive elements, a variety of methodological strategies can be used. In turn, the attitudinal elements that are key in this literacy process are much more demanding and difficult to work on at a methodological level. In this context, peer assessment can prove to be very useful. We will illustrate this with an example: the attitude, fundamental for this literacy process, of "adopting a critical stance towards quantitative messages that may be misleading, unilateral, biased or incomplete" (GAISE, 2016).

It is very complex to work on this attitude in the classroom, as it requires an individual process of reflection. For this reason, this research involves a strategy based on two elements:

a) On the one hand, there is the need for students to work intensively on a specific situation in which quantitative methods are applied. Furthermore, they must develop a set of conclusions.

b) Next, other students must analyze their arguments and conclusions, and assess the extent to which the narratives and messages are justified. In other words, they must develop this critical attitude that is so necessary for this literacy process.

In a similar manner, peer assessment is especially ideal for working on these attitudinal elements.

2. Methodological Framework

2.1. Purpose

The purpose of this study is to analyze whether a relationship exists between the development of peer assessment processes and the academic quality of the assessment tasks.

The study is framed within a more extensive project, the objective of which was to verify the differential effects of the monitoring technologies (chatbot and dashboard) on self-regulation. However, the students in the experimental group did not make sufficient use of the tools, and for the purposes of this article, the difference between groups has not been considered.

However, the study that is presented has made it possible to analyze the academic results and opinions of the students when an instructional sequence is applied that is designed for the purpose of generating peer assessment processes that boost self-regulation. These results and opinions were compared with those registered in previous years.

To reach this objective the following specific hypotheses are analyzed:

- The students in the experimental group would earn better grades on the assignment associated with this instructional sequence than the students in the control group.
- The students in the experimental group would earn better grades on the subsequent practice exercise, which did not form part of the instructional sequence, than the students in the control group.
- The students in the experimental group would earn better grades on the final exam than the students in the control group.

2.2. Intervention Design

The Descriptive Statistics course is required for first-year students; all students appear to be highly motivated by the course. In addition, they have a similar level of prior training. However, there are certain key skills for this instructional profile that are difficult to address. This is due to the fact that they require an intensive process of self-regulation on the part of the students. In the previous course, an attempt was made to achieve this self-regulation through the performance of practical activities by all students. Traditionally, they were addressed in groups. As a matter of fact, the course assessment model is based on diversified activities.

But they have proven to be insufficient to ensure that these mistaken pre-conceptions do not occur. For this reason, the decision was made to systematize the peer assessment processes, proposing a sequence of actions linked to Zimmerman's self-regulation process (2001), which may have positive effects on student performance and the development of their competences.

The designed sequence (Figure 1) consists of the following steps:

1. Display of the statement (description of the assignment and the performance criteria).
2. Forum debate on the meaning of the criteria, both those related to the assignment and to the proposed peer assessment process.
3. Planning questionnaire: "Now that you have read the assignment, indicate 3 actions that you are going to take to in order to respond."

4. First version of the work.
5. The work is assessed by a peer and the student is in turn the evaluator of another work.
6. Reception and reading of the feedback and response (using the "Retroaction" tool) to the question: "Now that you have received the feedback, say what you are going to change and how."
7. The same is done for the second cycle.
8. Final submission of the assignment with an online text assessing the process and its possible transfer to other future learning scenarios.

One of the main elements of the sequence is the peer review (step 5, which is repeated in the second loop). The specific design of the peer review is reported, following the example of the instrument offered by Panadero et al. (2023). Table 1 summarizes the specific characteristics of the peer review that has been designed.

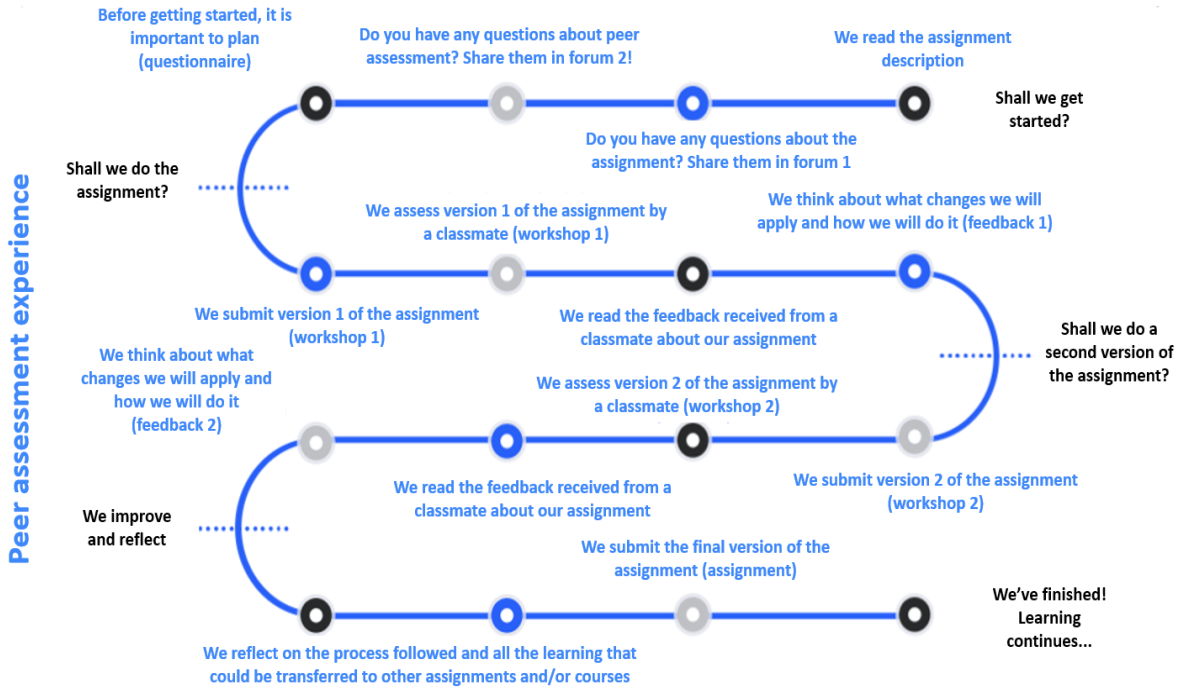

Figure 1. Instructional sequence designed as a section of a virtual classroom in Moodle

Context			
Category		**Description**	**Our study**
1	Subject domain	Subject domain in which the study was conducted (e.g., mathematics, instructional sciences, accounting, etc.)	**Statistics**
2	Place/Time	Where was the PA conducted? (In class or out of class?)	☐ In class/during class time ☐ **Out of class/during free time**
3	Setting	Formal or informal education setting?	☐ **Formal** ☐ Informal
4	Requirement	Was PA compulsory or voluntary for assessor/assessee?	☐ **Compulsory** ☐ Voluntary
5	Alignment	Was the PA activity aligned with the curriculum, learning goals or teaching?	☐ **Yes** ☐ No.

Context			
Category		**Description**	**Our study**
Instructional design			
6	Purpose	What was the assessment purpose of the PA activity? (Formative, summative or both?)	☐ Formative ☐ Summative ☐ **Both**
7	Object	What was assessed? (e.g., written assignment, oral presentation, contribution to group work)	**Written assignment**
8	Product/ Output	What was the output of the PA? (e.g., score, written feedback, oral feedback, or a combination)	**Written feedback**
9	Relation to instructor assessment	Was PA done without instructor assessment (substitutional) or in addition to instructor assessment (supplementary)?	☐ **Substitutional** ☐ Supplementary
10	Official weight	Did participation in the PA activity or the grade given by peer(s) contribute to the learners' final grades?	☐ **No**. ☐ Yes - for participation in PA ☐ Yes - for PA grade ☐ **Yes - both (PA & participation)** ☐ Other: Click here to add text
11	Reward	Was there a reward for participation in PA?	☐ **No.** ☐ Yes - course credit ☐ Yes - incentives (e.g., free time, money, etc.) ☐ Other: Click here to add text
12	Directionality	Was the learner assessing another without being assessed (unidirectional) or acting as both assessor and assessee (bidirectional)?	☐ Unidirectional ☐ **Bidirectional**
13	Degree of interactivity	How did the assessee demonstrate engagement and response to PA?	☐ Reactive: assessee responds to assessor ☐ Reciprocal: same people assess each other on the same task ☐ Negotiated: PA was done more than once on the same task and both parties negotiated **it** ☐ **Lack of interactivity**
14	Frequency	How often was the PA of the same task done? (Once, twice, etc.)	**Twice before final delivery**
15	Group constellation	Did members of the same group assess each other (intragroup) or peers from another group (intergroup) or both?	☐ Intragroup ☐ **Intergroup** ☐ Both
16	Constellation assessor	The number of assessors assigned to each assessee	**One**
17	Constellation assessee	The number of assessees per assessor	**One**
18	Unit of assessment (assessor)	At what level did the assessor(s) perform PA? Individual, group or both?	☐ **Individual** ☐ Group ☐ Both
19	Unit of assessment (assessee)	At what level did the assessee(s) experience the PA? Individual, group or both?	☐ **Individual** ☐ Group ☐ Both

Table 1. Information about Peer Assessment (PA) design. At right, in bold, our decisions

It is important to stress that this intervention was carried out completely within an LMS. In this case, using the virtual instructional campus of the University of Barcelona (based on Moodle 3.11). This is relevant, since among other things, it ensures the scalability of the experience. Without this scalability, it would not be possible to apply the designed sequence to groups of 70 or 80 students.

Starting from the basis of the course assessment model described above, the instructional sequence described applies specifically to Practice 1. This sequence can be summarized by the following stages:

a) Organizing the students into groups of 3-4 students on a voluntary basis. Each group of students selects a case study (a problem to be solved) and develops a preliminary report. This report must include at least the following:
    - Establishment of objectives.
    - Specification of the hypothesis to be investigated.
    - Definition of relevant variables.
    - Collection of the necessary data.
    - Selection of tables.
    - Selection of figures.
    - Selection of statistics.
b) Each group evaluates the report of another group (random distribution). For this, they have a check list of criteria, which is what allow us to analyze the usefulness of the feedback.
c) Each group has the option of reviewing the decisions made and/or arguing why they are not reviewed.
d) Second feedback is made, using the same tool.
e) Finally, the instructor evaluates the final reports and provides feedback to each group through a report with comments.

It is important to bear in mind that this sequence is designed so that the feedback between the test and error allows the students to advance autonomously based on the necessary decision-making. In this context of feedback, the peer assessment allows the competence of "learning how to learn" to be specifically addressed.

2.3. Participants

This research involved a total of 232 students, spread out over three academic years. The students in the 22-23 academic year (n=75) make up the experimental group, which we are going to compare to the students from the 20-21 (n=86) and 21-22 (n=71) academic years, whom we use as the control group.

In spite of the fact that the peer assessment experience was initiated in the 21-22 academic year (it was initially optional for the students, as a trial phase), the full instructional sequence was not applied. Therefore, this group has been included within the control group. This decision was made because what we intend to compare is the impact of the full instructional sequence. This is further justified by the fact that in the 21-22 academic year, peer assessment was optional for the students. This thus provides us with more information. In fact, by including the 21-22 academic year in the control group, it is shown that there are significant differences between the experimental group and the control group, and thus the evidence is even clearer.

2.4. Instruments

In order to evaluate the impact of the pilot test, in addition to the reports from each of the groups of students, two other types of indicators were also collected:

- The grades (numerical, on a scale of 0 to 10) for each of the activities performed during the academic year.
- The results of a survey on the perception of learning by the participants.

2.5. Data Collection and Analysis

The instructional sequence was implemented over the course of 6 weeks during the 22-23 academic year. It was applied to the completion of Practice 1. This duration coincides with the activity in the previous academic years (20-21 and 21-22).

Table 2 shows the descriptive statistics of the grades earned by the students in each of the two groups (control group and experimental group). In order to verify the hypotheses proposed in the research, it is necessary to carry out several equalities of means contrasts. For this purpose, the arithmetic means and standard deviation statistics have been calculated for both the control group and the experimental group.

	Control (group=0)		**Experimental (group=1)**	
	N = 157		**N = 75**	
Group	**Mean**	**S.d.**	**Mean**	**S.d.**
Practice 1	8.05	1.85	8.65	0.77
Practice 2	7.44	3.19	8.42	1.61
Final exam	6.22	2.04	6.35	1.91
Final	6.42	2.03	6.36	1.90

Table 2. Summary of the control group and experimental group grades

2.6. Ethical Aspects

Approval was obtained from the Universitat de Barcelona bioethics committee (IRB00003099).

All participants were properly informed and signed both informed consent and agreement with the data protection policy.

3. Results

The test evaluation made it possible to compare the grades of the students in each of the academic years, distinguishing between the control and experimental groups.

In the preliminary phase of analyzing the results, the grades earned by the control group and the experimental group were compared. Figure 2 allows us to see the difference between the two distributions.

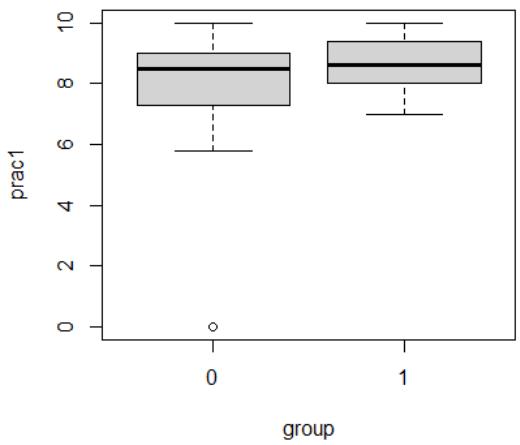

Figure 2. Comparison of the grades on Practice 1
(0 = control group / 1 = experimental group)

In this figure, we can see that the mean grade in the experimental group was higher than in the control group. But at first glance, we cannot conclude whether this difference is statistically significant or possibly the product of chance. Therefore, based on the data, we can carry out a formal statistical comparison. In this case, the comparison is made under the hypothesis of normalcy and equality of variance. This is assuming that the variance of the two distributions is unknown and does not necessarily coincide.

Group	Control (n=157)	Experimental (n=75)
Average grade	8.05	8.65
Standard deviation	1.85	0.77
t-statistic (comparison of means)	-2.6193	
p-value	0.0094	

Table 3. Comparison of means of the grades on the instructional sequence between the control and experimental groups

According to the results in Table 3, it is clear that the difference in means is statistically significant, operating with a level of significance of 1% and even lower. These results allow us to rule out the equality of the means between the experimental and control groups. We therefore conclude that the mean grade of the experimental group is higher than that of the control group. This is indicative of the fact that the designed instructional sequence has made it possible to improve the results. In any case, the effect of applying the instructional sequence does not have a great impact on the average grade.

In light of the finding that applying the peer assessment strategy has an effect on the grades, it is normal to conclude that this improvement in academic results will continue on to the final exam grades. But this hypothesis must be demonstrated. To do so, we next compared the grades of the experimental group with those of the control group, but based on the final exam grades. We therefore must assume the hypothesis that the difficulty of the final exam is the same in all three academic years involved.

Figure 3 reflects this comparison, and it can be seen that the means are practically the same.

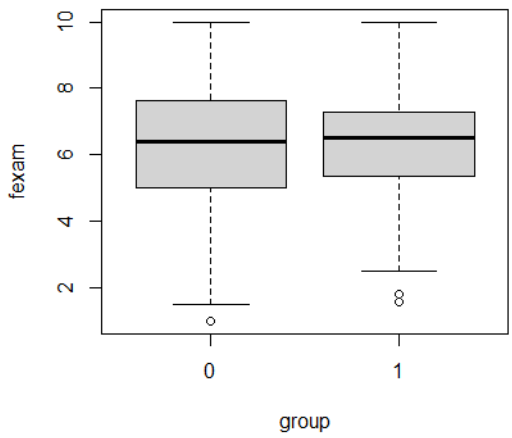

Figure 3. Comparison of the grades on the final exam (0 = control group / 1 = experimental group)

Once again, we apply a formal statistical comparison in order to verify whether there are any significant differences. The results shown below (Table 4) indicate that the differences in grades between the control group and the experimental group are not statistically significant. Therefore, in this case we cannot reject the hypothesis that the two distributions have the same means.

Group	Control (n=157)	Experimental (n=75)
Average grade	6.22	6.35
Standard deviation	2.04	1.91
t-statistic (comparison of means)	-0.4506	
p-value	0.6527	

Table 4. Comparison of means of the grades on the practice exercise following the instruction sequence between the control and experimental groups

What implication does this result have? It calls into question the sustained effect of using peer assessment. This is because while it may seem to have an effect on the immediate grade of the practice exercise, this effect does not seem to last over time. Furthermore, the doubt arises whether the criteria used to evaluate Practice 1 in the experimental year and the years that we considered to be the control group are the same. In fact, the customary method of evaluating this practice has not been based on a detail rubric. But it is also true that if we review the reports presented by the students as the final product of Practice 1, a greater quality is perceived in the comments for students in the experimental group.

In an attempt to improve the precision of the assessment, a more precise criterion has been applied for this assessment. And the comparison has been made based on the complete evaluation of the practices completed by the students over the three academic years. Naturally, in order to make this comparison, all these practices have been randomized.

	Item	Weighting
1	Work motivation	1/20
2	Identification and formulation of the problem	1/20
3	Search in the print and electronic literature for references	1/20
4	Definition of the objectives of the work	1/20
5	Formulation of specific hypotheses	1/20
6	Definition of statistical variables	1/20
7	Selection of the data source	1/20
8	Data search and collection	1/20
9	Data screening and organization	1/20
10	Quantitative analysis	1/20
11	Information synthesis with tables	1/20
12	Information synthesis with figures	1/20
13	Information synthesis with statistics	1/20
14	Interpretation of the results and discussion	1/20
15	Drawing of conclusions and recommendations	1/20
16	Coherence among objectives, methodology and results	1/20
17	Use of computerized resources	1/20
18	Written communication and presentation	1/20
19	Graphic communication and presentation	1/20
20	Critical analysis and integration of knowledge	1/20

Table 5. Criteria for the assessment of Practice 1

The following graphic (Figure 4) shows the distribution of the grades from Practice 1 in all three academic years, assessed according to these criteria. It can be seen that the mean grade in the experimental group was higher than in the control group. And most importantly, the differences are visually greater than in the initial assessment.

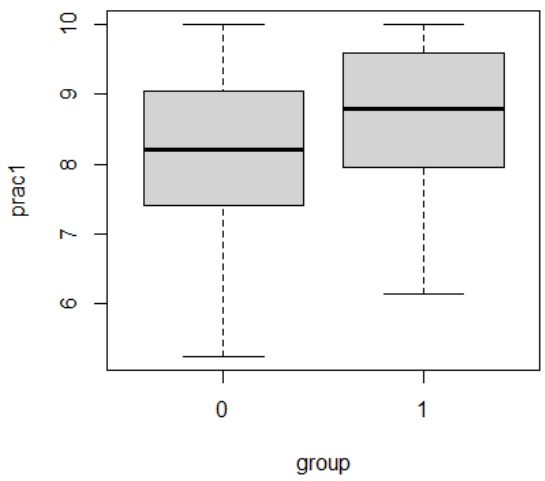

Figure 4. Comparison of the grades on Practice 1
(0 = control group / 1 = experimental group)

To check whether the difference is significant, we applied a new comparison.

Group	Control (n=157)	Experimental (n=75)
Average grade	8.17	8.69
Standard deviation	1.94	1.79
t-statistic (comparison of means)	-3.2818	
p-value	0.0012	

Table 6. Comparison of means of the grades on the final exam between the control and experimental groups

From the results in Table 6 it can be concluded that the difference in means is statistically significant. These results make it possible to confirm the positive impact of the peer assessment strategy.

In light of these results, we understand that a hypothesis can be proposed regarding the equality of the grades on the final exam, which can be analyzed with the data available in this study. One possible explanation in order to analyze the possible impact on the students' learning is that perhaps the final exam was not the proper instrument and a more precise method of evaluation should be used that would allow the measure of learning of each student to be determined much more accurately.

Finally, we remember that in addition to the grades, we also have a qualitative evaluation of the application of the instructional sequence in Practice 1 (2022-23 academic year), by means of a survey administered to the students in the experimental group.

Each student was asked to respond to a final satisfaction questionnaire. A total of 8 students (4 women and 4 men) responded to this survey. Based on the responses obtained, the means and the standard deviation of the scores were determined for all the students as a whole. Unfortunately, this questionnaire was only given to the students in the experimental group (2022-23 academic year), so it is not possible to make any sort of comparison.

This is not a representative sample, but it does allow us to better contextualize the results. The evaluation survey has allowed us to better know how the students evaluated the experience. In general, the overall satisfaction with the experience was high. However, some very important differences were detected in the assessment of different aspects. The results of this questionnaire must be considered with a great deal of caution, due to the small number of responses (n=8).

Curiously, the aspects that the students valued the most were the following (mean score between 0 and 5).

- It helps me to have all the information neat and organized (5.0)
- Understanding the assessment criteria of the assignment being assessed (4.5)
- I have become aware of the actions and processes that can allow me to improve my learning (4.4)

On the other hand, the aspects receiving the worst scores were:

- Becoming more involved in the learning process (3.0)
- Learning how to give feedback (3.3)
- Contributing to the development of the learning how to learn competence (3.5)

As a result, the aspect rated the highest is that related to statistics as a discipline, which is discussed in the following section.

4. Discussion

The first aspect of the study to highlight is the discovery of the positive effects of the peer assessment of the assignment. This was not true on the course final exam, however. This may advise a closer connection between the different assessment artifacts and/or reveal, on the other hand, the need for more continuous evaluation (Näykki et al., 2021). Reviewing the course assessment design in light of this evidence could lead to a closer alignment among the assessment criteria, assessment tasks and results of the intended learning (Biggs, 2005).

The results also seem to alert to the need to strengthen assessment literacy (Carless & Boud, 2018; Carless & Winstone, 2023), given that it is curious to note than the practical assessment of students regarding the usefulness of the instructional sequence has not been about what they can learn in relation to the feedback and self-regulation process. What they actually assessed is the dimension most closely related to the discipline: statistics. This is an aspect that should be investigated in greater depth. But it may have to do with the fact that these are students who have received little training or guidance in relation to study techniques and strategies and with learning and assessment processes (Panadero et al., 2019). Greater assessment literacy and better knowledge of the benefits of these practices would probably be positive and could affect their perception (Elinor, 2022; Van der Watering et al., 2008).

The evaluative judgment is the capacity to make decisions about the quality of one's own work and that of others (Tai et al., 2017). In the context of this research, it was proposed to improve the evaluative judgment through peer assessment. This means being capable of judging the quality of one's own work and that of others, not only in relation to an assignment or a course, but also throughout the learning process. This implies not only conducting peer assessment practices or self-assessment, it also means achieving a profound and authentic reflective practice (Suryadi & Kusairi, 2021; Valero, 2022). But it has been shown that peer assessment requires more prior training (of both students and instructors).

Furthermore, we identified a generally poor utilization of the instructional sequence. We understand that this is due to the fact that the strengthening of the evaluative judgment requires broader time frames and a continuity in the assessment proposals in which the student participate, in such a way that these are not isolated experiences, the effects of which can be compromised (Carless & Winstone, 2023). Therefore, we believe that the objective of developing statistical thought should be tested in future experiences in a more concrete and specific manner. The development of statistical thought forms part of the genesis of the design of the proposed assignment, but we suggest that it be tested in future research with a pre- and post-test designed specifically for this purpose.

We also observe from the students' scores that the "learning how to learn" category has not scored especially highly (3.5). We believe that this may be due to the way in which the students have understood this expression, which is far removed from the scientific field to which they belong. This

has led them to interpret the expression perhaps as something not very precise and/or valuable in their learning process (García-García et al., 2021). Furthermore, we have shown that the role of the assessor has been more appreciated than the role of the assessee, given that they have valued having a more critical vision (M=4.4). This confirms the findings of the previous literature on peer assessment (Scott, 2017).

Finally, the technology tools available (chatbot and dashboard) that could serve to identify the errors and overcome them in future learning processes have only been used by a minority of the students. Therefore, they could not make the best use of this advantage, and thus the strengthening of the digital competence beyond an instrumental sense would also be advisable. The high rating given by the students to the technology tools associated with the research in order to "have the information organized and available" (4.5) is also in agreement with the previous analyses regarding the little added value the technologies contribute to the learning processes of university students. Data management, the ease of grading or the immediacy of access are benefits that are reported by both students and instructors, but this leaves a wide margin for improvement in the use of digital technologies. All of this suggests that it is important to pursue this specific line of research.

5. Conclusions

Statistically significant differences have been reported between the results of the practice exercise completed by the students when the use of the instructional sequence (with the peer assessment process) is compared to the practice exercise in previous courses, without peer assessment. In spite of this positive finding, the other important conclusion is that the final course grades do not show any differences with the use of peer assessment. Nonetheless, with the use of a list of criteria that makes it possible to better assess the quality of the students' assignments, significant differences were detected when using the peer assessment process.

The limitations of the study include both the size and composition of the sample and the use of the same instrument during the double loop (list of criteria).

Finally, all these results in no way close this line of research, rather quite the opposite. Based on these results, at least two priority lines of research are proposed:

- Expanding the experimental design. Comparing, for the following academic year, two groups (pilot and control), and throughout the entire assessment process (not just one practice exercise). Collecting many more data during the semester (and over more time). This would make it possible to draw clearer conclusions about the impact of peer feedback, at least in this type of courses.
- Cooperating with other courses with similar characteristics, in order to be able to experiment with variations in the design of the instructional sequence/peer feedback.

In order to promote self-regulated learning and to be able to see more far-reaching changes, the necessary time frames must be longer. Perhaps another type of curricular architecture that overcomes the fragmentation in disciplines would facilitate a more competence-based work, according to the type of learning that the students will need in order to continue life-long learning.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

This publication forms part of the R&D project PID2019-104285GB-I00, financed by MCIN/AEI/10.13039/501100011033.

References

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment and Evaluation in Higher Education*, 42(2), 252-265. https://doi.org/10.1080/02602938.2015.1102863

Anderson, O.S., El Habbal, N., & Bridges, X.D. (2020). A peer evaluation training results in high-quality feedback, as measured over time in nutritional sciences graduate students. *Advances in Physiology Education*, 44(2), 203-209. https://doi.org/10.1152/ADVAN.00114.2019

Biggs, J. (2005): *Calidad del aprendizaje universit*ario. Madrid: Narcea.

Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment and Evaluation in Higher Education*, 44(5), 705-714. https://doi.org/10.1080/02602938.2018.1531108

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150-163. https://doi.org/10.1080/13562517.2020.1782372

Elinor, T. (2022) A review of group-based methods for teaching statistics in higher education. *Teaching Mathematics and its Applications: An International Journal of the IMA,* 41(1), 69-86. https://doi.org/10.1093/teamat/hrab002

GAISE (2016). *Guidelines for Assessment and Instruction in Statistics Education: College report*. American Statistical Association.

García-García, F.J., Moctezuma-Ramírez, E.E., López-Francés, I., & Pérez, C.P. (2021). Learning to learn at university: Perceptions of teachers and students. *Estudios Sobre Educación*, 40, 103-126. https://doi.org/10.15581/004.40.103-126

Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, 32(4), 399-409. https://doi.org/10.1080/03055690600850354

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research,* 77(1), 81-112.

Hicks, C.M., Pandey, V., Fraser, C.A., & Klemmer, S. (2016). Framing feedback: Choosing review environment features that support high quality peer assessment. *Conference on Human Factors in Computing Systems - Proceedings* (458-469). https://doi.org/10.1145/2858036.2858195

Lui, A.M., & Andrade, H.L. (2022). The Next Black Box of Formative Assessment: A Model of the Internal Mechanisms of Feedback Processing. *Frontiers in Education*, 7, 751548. https://doi.org/10.3389/feduc.2022.751548

Näykki, P., Kontturi, H., Seppänen, V., Impiö, N., & Järvelä, S. (2021). Teachers as learners–a qualitative exploration of pre-service and in-service teachers' continuous learning community OpenDigi. *Journal of Education for Teaching*, 47(4), 495-512. https://doi.org/10.1080/02607476.2021.1904777

Panadero, E., Alqassab, M., Fernández-Ruiz, J., & Ocampo, J.C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*, 1-23. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., Broadbent, J., Boud, D., & Lodge, J.M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34(3), 535-557. https://doi.org/10.1007/s10212-018-0407-8

Rodríguez-Gómez, G., Ibarra-Sáiz, M.S., & García-Jiménez, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: Conceptualización y práctica en las universidades españolas. *Revista de Investigacion en Educación*, 11(2), 198-210.

Scott, G.W. (2017). Active engagement with assessment and feedback can improve Group-Work outcomes and boost student confidence. *Higher Education Pedagogies*, 2(1), 1-13. https://doi.org/10.1080/23752696.2017.1307692

Shiau W.C., & Zaleha, I. (2013). Assessing Misconceptions in Reasoning About Variability Among High School Students. *Procedia - Social and Behavioral Sciences*, 93, 1478-1483. https://doi.org/10.1016/j.sbspro.2013.10.067

Suryadi, A., & Kusairi, S. (2021). Developing computer-assisted formative feedback in the light of resource theory: a case on heat concept. *Journal of Technology and Science Education,* 11(2), 343-356.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018) Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education,* 76, 467-481. https://doi-org.sire.ub.edu/10.1007/s10734-017-0220-3

Udechukwu, J. (2020). Effect of self-assessment and peer-assessment techniques on the performance of undergraduate students in a basic statistics course. *International Journal of Innovative Mathematics, Statistics and Energy Policies,* 8(3), 69-75.

Valero, M. (2022). Challenges, difficulties and barriers for engineering higher education *Journal of Technology and Science Education,* 12(3), 551-566.

Van de Watering, G., Gijbels, D., Dochy, P., & Van der Rijt, J.A. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education*, 56, 645-658.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 1-37). Lawrence Erlbaum Associates Publishers.

Published by OmniaScience (www.omniascience.com)

Journal of Technology and Science Education, 2024 (www.jotse.org)

Article's contents are provided on an Attribution-Non Commercial 4.0 Creative commons International License. Readers are allowed to copy, distribute and communicate article's contents, provided the author's and JOTSE journal's names are included. It must not be used for commercial purposes. To see the complete licence contents, please visit https://creativecommons.org/licenses/by-nc/4.0/.