Heliyon 9 (2023) e20490

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Active methods in electricity and magnetism courses: Influence of degree, academic level and gender on student performance

María-Antonia Serrano [a], Ana Vidaurre [a], José M. Meseguer-Dueñas [a], Isabel Tort-Ausina [a], Susana Quiles [a], Roser Sabater i Serra [b], Tania García-Sanchez [b], Soledad Bernal-Pérez [b], M. Amparo Gámiz-González [c], José Molina-Mateo [a], José Antonio Gómez-Tejedor [a,*], Jaime Riera [a]

[a] *Departamento de Física Aplicada, Universitat Politècnica de València, Camino de Vera s/n, Valencia, 46021, Spain*
[b] *Electrical Engineering Department, Universitat Politècnica de València, Camino de Vera s/n, Valencia, 46021, Spain*
[c] *Departamento de Educación, Universidad Internacional de Valencia, C/ Pintor Sorolla, 21, 46002, Valencia, Spain*

ARTICLE INFO

Keywords:
Science
Technology
Engineering
And Mathematics (STEM) disciplines
Concept inventory
Gender differences
Physics
Retention
Student perception
First-year undergraduate students
Brief electricity and magnetism assessment (BEMA)

ABSTRACT

The performance of first-year students in electromagnetism (E&M) courses of different engineering degrees at a Spanish public university was measured using the Brief Electricity and Magnetism Assessment (BEMA), a standard research-based instrument to assess students' understanding after attending introductory courses in electricity and magnetism. In all cases, Flipped classroom (FC) built on information and communications technology was used. The objective of this paper is to analyse if the gain in the BEMA pre and post-test results is influenced by several factors such as the degree, the students' academic grade, and gender. Moreover, as some studies have shown that the students' retention of the concepts was significantly stronger in active learning than in traditional approaches, a third BEMA test was performed by the students to analyse the long-term retention gain dependence on the same factors. Students from different engineering degree programs were asked to complete two BEMA tests during the course and a third one after a few months. ANOVA tests were used to analyse the existence of significant differences in gain between student degree programs, student academic level and student gender. Results have shown no differences in the BEMA performance by degree program, but significant differences were found by academic level and gender. Retention did not depend on the degree course but on the academic level. Mean gain value by academic level, and gender was obtained and concluded that the best students presented the best gain results and that gain depends on the students' gender: males outperformed females in the BEMA tests, although there were no significant differences in the course grades. It is thus necessary to understand these differences and to implement measures in daily teaching work to improve women's performance.

* Corresponding author.
E-mail address: jogomez@fis.upv.es (J.A. Gómez-Tejedor).

https://doi.org/10.1016/j.heliyon.2023.e20490
Received 23 November 2022; Received in revised form 8 September 2023; Accepted 26 September 2023
Available online 28 September 2023

1. Introduction

1.1. Flipped classroom

Active teaching methods are now becoming more significant, and flipped classroom (FC) is among the most popular since it provides a series of benefits to the teaching-learning process. This method promotes autonomous learning by making the students control their own learning process [1,2]. Information and communications technology (ICT) offers a wide range of tools that play an essential role in FC implementation. The computer learning platforms commonly used at universities enable the organization of the course content into documentation, videos, tasks, and assessment. The FC creates a space for dynamic, responsive, just-in-time, and interactive team-based learning that builds on individual learning initiated outside the classroom so that the instructor's role is significantly different from traditional teaching [3–7]. In the FC built on the ICT context, the instructor guides students towards active learning and teamwork. However, for the method to be effective and obtain the expected results, it requires the commitment of the students and fluid teacher-student coordination, while the teachers should make an effort to design and prepare the appropriate material [8–10].

Numerous studies have reported that the FC positively impacts students' performance and enhances their soft skills such as social interaction, learning motivation, engagement and self-directed learning skills [11,12]. However, other studies found no significant effects on student learning compared to conventional instruction [13,14]. It has been suggested that FC encourages students' engagement and satisfaction [12], while others [15] showed that both students and educators may not be ready for flipped classes. The lack of pre-class preparation and little or no involvement, mainly in first-year university students, have been identified as the main difficulties found in this approach [16–18]. The design of both pre-class and in-class activities is crucial for creating active and experiential learning [19]. Tomas et al. [17] suggested that the FC should be progressively introduced in first-year university courses, as secondary school education is often based on the transmission model (traditional method). Gaming strategies can be merged in the FC in-class activities [20,21]. Besides providing diversification, this affects classroom dynamics and improves teacher-student interaction and student-student interaction [22].

Due to the COVID-19 pandemic, classes in many countries were forced to move to a virtual classroom. Teachers and students had to quickly adapt to the new situation and in the opinion of some authors the FC model built on ICT favoured this transition [23,24]. The strategy in pre-class activities was conserved while the in-class activities moved from face-to-face teaching to synchronous online with few modifications. In a previous study, it was found a positive student perception of the adaptation to the new situation [25].

1.2. BEMA conceptual inventory. Conceptual learning and retention study

Research-based conceptual inventories play an important role in assessing conceptual learning course development [26]. One of those most commonly used in Electricity and Magnetism (E&M) courses is Brief Electricity and Magnetism Assessment (BEMA), developed by Chabay & Sherwood [27,28]. BEMA test can be downloaded from the PhysPort website, with links from the National Science Digital Library (NSDL) and the American Association of Physics Teachers (AAPT), together with the methodology to administer it. In a previous study, in which BEMA was also used, it was found that this assessment tool has discriminatory capability and reliability when applied to first-year students of two engineering degrees at the Universitat Politècnica de València (UPV, Spain) [29]. In this study, BEMA results were analysed in terms of final course grades, and a retention analysis was included [30].

Retention is the ability to remember acquired knowledge over time in the same way as it was learned during instruction. Several standard tests have been used to measure this retention, like the Force and Motion Conceptual Evaluation (FMCE) test [31] and BEMA [32]. It has been established that performance decreases after training ceases and that interference occurs when two pieces of related information are learned [33]. Dori et al. [34] analysed long term retention one year and 18 months after finishing an introductory electromagnetism course. The results indicated that the students' retention of the concepts was significantly stronger in active learning than traditional approaches. In the same vein, a previous study suggested that active learning (in technical subjects) can help to increase retention for students with average or below-average scores [35].

Web-based assessment is now widespread mainly due to the COVID-19 pandemic. Ardid et al. [36] found that results of online exams were biased towards higher ratings in an unsupervised compared to proctored environment, while Bonham [37] found no difference in the scores on the force concept inventory for both groups, one administered in class and the second completed outside class on the web. The difference can be attributed to the use of the online exam as part of the evaluation process or not. Some authors [38,39] proposed strategies to improve test security and efficiency.

1.3. Gender perspective in the conceptual inventory test

Male students usually outperform female students on concept inventories in Physics, and this difference in scores is called the "gender gap" [40]. This gap has been extensively studied in various physics conceptual assessments [26,40–43]. Men's scores on two different electricity and magnetism concept inventories (Conceptual Survey of Electricity and Magnetism (CSEM) and BEMA) were on average 3.7% and 8.5% higher than women in the pre-test and post-test, respectively [40].

Madsen et al. [40] analysed 26 published articles comparing the impact of 30 factors that could potentially influence the gender gap and concluded that factors such as gender differences in background preparation and scores on different assessments could contribute to a difference between male and female responses. Radulovic et al. [44] used a standardised questionnaire to measure the gender differences in motivation towards Physics learning. They found that the weakest motivational components were the importance of

Physics as a science and self-efficacy. Verdugo-Castro et al. [45] performed a systematic literature review related to the choice of higher education in the STEM (Science, Technology, Engineering, and Mathematics) field in Europe. They found that gender stereotypes are strong drivers of the gender gap. Although it has been proved that male and female students react differently to testing conditions [41], some authors found no significant differences in the final qualifications of the course [46]. This could be explained by the positive results of females obtained from homework and participation in the classroom [43]. In addition, Lorenzo et al. [47] found that using interactive engagement methods could reduce the gender gap measured in the Force Concept Inventory.

1.4. Research purpose and questions

This paper analyses the gain in a conceptual inventory test, the BEMA test, obtained by engineering students that followed an ICT-based FC approach. The gain after a few months (often called retention) was also studied. A period of the study was performed during the COVID-19 pandemic, when, classes and tests were in a web-based scenario.

The research purpose of the present study was to determine the influence of several factors in the gain in the BEMA test. Therefore, the research questions that this work tried to answer were the following:

- Does the gain depend on the degree program?
- Does the gain depend on the students' academic level?
- Does the gain depend on the students' gender?

We also include the students' perceptions to obtain an insight into student satisfaction levels as supplementary material in Annexe 3.

The large number of students participating in this study (472), the use of the BEMA test, which is a valid and reliable instrument widely used in the literature, and a delayed post-test for the study of retention, justify the rationale for this study. To our knowledge, there is no this kind of systematic study in the STEM sector of the European educational field.

The paper is organised as follows: after the introduction in which the literature is reviewed, the methods, including participants, research context and BEMA test administration, are detailed in section 2. The results in terms of the degree program, academic level and gender are included in section 3. Discussion and conclusions, including study limitations and suggestions, complete the paper. In addition, an opinion survey was carried out to know students' motivation, confidence, and engagement related to the FC environment, which results are presented as complementary material in Annexe 3.

2. Methods

2.1. Participants and demographic information

This study took place at the public university *Universitat Politècnica de València* in Spain, a large higher education institution with almost 30,000 students that offers more than 40 BA engineering degrees. The participants were enrolled in Electricity and Magnetism courses in the following bachelor's degree courses: Electric Engineering (EE), Electronic Engineering and Industrial Automation (EEIA), and Aerospace Engineering (AE). The E&M courses were taught during the first and second semesters of the first year, with a duration of 35–45 h, according to the type of degree program. Data were collected between the academic years 2017/18 to 2020/2021. In these courses, 862 students (aged between 18 and 20) were enrolled, of which 472 participated in this study. Participation in this study is completely voluntary, and the students are knowledgeable about the research purpose of the data. The data were anonymized before processing. Although no demographic data were collected, it can be considered that the origin of the students at the engineering degrees presented in this study is mainly (>90%) from the Valencian Community, a Mediterranean region in eastern Spain, where the *Universitat Politècnica de València* is located. AE students have the highest university-access mark, with similar values between them. EE and EEIA students have lower university-access mark, with values within a wide range.[1] The study was approved by the *Comité de Ética en Investigación de la UPV* in session held on July 27, 2023.

2.2. Research context

Active learning teaching methods built on ICT, particularly FC, were used in the E&M courses in all the degree courses included. ICT-based FC involved instructional approaches based on working in groups to solve problems and lab sessions, activities in class after watching videos at home and computer-based gamification. The resources used in the classes were uploaded in the University's online teaching platform, based on the learning management system *Sakai* and known as *Poliformat* [48]. In this learning platform, students and instructors can share information on the different courses in the degree and use documents, videos, and tools such as assignments, tests, and quizzes, among others. The learning material consisted of various resources, including documents explaining the lesson objectives, purpose, content, videos focused on fundamental concepts and links to websites with relevant information. Different

[1] The Spanish students, after they finish high school, must pass a pre-university examination. After that, the students get a mark (university-access mark) that is calculated from this pre-university exam and their previous high school studies. Finally, the students enrol in the different degrees according to their preference and their university-access mark.

problems related to the theoretical concepts were proposed to the students, which they had to solve individually or to work on in groups. The web-app online quiz platform *Kahoot!* was used as a gamification tool [49]. It is also worth noting that the instructors, with extensive teaching experience, are part of an educational innovation group where the implementation of FC is discussed. The same methodology and educational material are used in all the E&M courses.

The data collected contained the results of the students' responses to BEMA tests, which were delivered to the EE, EEIA and AE degree students. More than 450 students (specifically 472, representing 54.5% of the total students enrolled) took some of the BEMA tests. In addition to BEMA data, the final grade of the E&M courses was also included. The students were divided into terciles for in-depth analysis according to their final grades (academic level). Three groups of approximately the same number of students were formed for each degree program, from highest to lowest grades (higher, intermediate, and lower tercile). Their gender was also noted.

2.3. BEMA students' performance

The BEMA test was administered following the recommended instructions (time limit of 45 min and the same grade for all students who completed the test, regardless of the score achieved) that were given to the students in advance. The BEMA pre-test (BEMA 1) was administered during the first week of the E&M courses, previous to instruction, while the post-test (BEMA 2) was administered at the end of the courses. A third test (BEMA 3) was administered to EE and EEIA students four months after completion of the courses to analyse their retention. AE students did not participate in BEMA 3 because when this test is administered these students are following another E&M course and the results could be biased. BEMA results were normalised in the range between 0 and 10 points. As stated previously, data were collected during several academic years: 2018/19 and 2019/20 for EE, 2017/18, 2018/19 and 2019/20 for EEIA and 2018/19 for EA (September to July). The final grade of the E&M courses was obtained from the academic years included in this study. The number of students who participated in each BEMA test, classified by degree, is included within the tables in the following section. Gender behaviour was analysed from the results of students that took BEMA 1 and BEMA 2 tests, in which of 341 participants, 64 were female students.

The comparison of BEMA 2 and BEMA 1 tests provided information on the effect of the courses on the students' knowledge of E&M, while the comparison of BEMA 2 and BEMA 3 provided information on the retention capacity of E&M concepts. The results were analysed in terms of degree program, academic level and gender.

2.4. Normalised gain

The students' normalised individual gain was calculated from the scores obtained in the BEMA tests according to the expression proposed by Bao [50,51]. The gain between the pre-test BEMA 1 and the post-test BEMA 2 scores was first calculated as:

$$g_{i(BEMA\ 2-BEMA\ 1)} = \begin{cases} \dfrac{BEMA\ 2 - BEMA\ 1}{10 - BEMA\ 1}\ if\ BEMA\ 2 \geq BEMA\ 1 \\ \dfrac{BEMA\ 2 - BEMA\ 1}{BEMA\ 1}\ if\ BEMA\ 2 < BEMA\ 1 \end{cases} \tag{1}$$

From the normalised individual gain (Equation (1)), $g_{i(BEMA\ 2-BEMA\ 1)}$, the average gain of the different groups was obtained and named G12. The results of the average gain were grouped by degree, terciles from the final grade of the E&M courses, and gender.

In order to analyse EE and EEIA students' retention, the performance between the BEMA 2 and the test delivered four months after completion of the course (BEMA 3) was studied by means of the normalised individual gain, defined as:

$$g_{i(BEMA\ 3-BEMA\ 2)} = \begin{cases} \dfrac{BEMA\ 3 - BEMA\ 2}{10 - BEMA\ 2}\ if\ BEMA\ 3 \geq BEMA\ 2 \\ \dfrac{BEMA\ 3 - BEMA\ 2}{BEMA\ 2}\ if\ BEMA\ 3 < BEMA\ 2 \end{cases} \tag{2}$$

From the normalised individual gain (Equation (2)), $g_{i(BEMA\ 3-BEMA\ 2)}$, the average gain of the different groups was obtained and named as G23. The results of the average gain were grouped by degree and terciles from the final grade of the E&M courses.

2.5. Statistical analysis

An initial screening was carried out after the BEMA tests to avoid data distortion, discarding any tests with no questions answered. The results were analysed on Statistical Package for Social Science Software-SPSS-IBM version 16 (raw data can be obtained upon request to the authors) using variance analysis (ANOVA). The data were anonymized and segmented into three groups by their academic level. ANOVA tests were used to analyse the existence of significant differences of gain between student degree program, student level and student gender The application of Leven's test justifies the correct application of the ANOVA test, although where the test showed non-homogeneous variances, the alternative nonparametric Krustal-Wallis test was applied, with results very similar to the parametric method, given the size of the sample. A 95% confidence level was considered statistically significant ($p < 0.05$).

3. Results

In order answer the research questions, set out in Section 1.4, the BEMA test results at the beginning and the end of the course were analysed for the three degrees. BEMA results were segmented according to the students' final grade in the subject. Genders were segmented to study its possible dependency of the BEMA results on this factor. The results of the BEMA tests conducted four months after completion of the course were studied to analyse retention by using student segmentation by academic level. The results of the student survey were then analysed.

3.1. Results of the BEMA tests

The total number of students enrolled in the E&M courses in the three degree courses during the above-mentioned academic years was 862, and 55% of them (472 students) performed some of the different BEMA tests, as noted earlier. The number of students (N) performing each BEMA test in each degree is shown in Table 1, as well as the average BEMA grade (ABG) and its standard deviation (SD).

As shown in Table 1, the number of students performing the three tests in each degree program is not the same. Some students took BEMA 1 but did not take BEMA 2 and vice versa, and the same happened with BEMA 2 and BEMA 3. It was therefore decided to analyse those students who carried out the BEMA 1 and BEMA 2 tests (Table 2) to obtain G12 and those who did the BEMA 2 and BEMA 3 (Table 6) to obtain G23. The course grades (CG) data were considered in a deeper analysis to study the effect of the students' academic level on the BEMA results. This led us to eliminate from the sample the students who did not finish the courses (they did not obtain a final CG), even if they had performed any or all the BEMA tests.

3.2. Analysis of the gain between BEMA 1 and BEMA 2 and its dependence on the students' degree program and academic level

The students' understanding of E&M during the academic year was measured using the average gain. The gain between the pre-test BEMA 1 and the post-test BEMA 2 (G12, Equation (1)) was calculated considering the sample of students mentioned in Section 3.1. Table 2 shows the average value of G12 in each degree program, with its standard deviation (SD) and the number of students in the sample (N).

The students' data were segmented into three groups according to the final course grade to study whether the gain depended on the academic level. Table 3 shows the average course grades (ACG) and the standard deviation of the students that participated in the study (in each degree program) and the average and standard deviation of the grades in each one of the three terciles. Table 3 also shows the interval of the grades defining each tercile. It is worth noting that the ACG are similar for all degree programs. To a deeper understanding of the BEMA result, in Annexe 1 it is shown the difference between post- and pre-test by tercile.

The grade interval width of the intermediate tercile is the narrowest in all degree programs. This means that students in the intermediate tercile have a much closer average grade for the course. There are considerable differences between the grade interval widths of the terciles in each degree program. In EE, the lower tercile represents 61% of the full range of the grades, while in EEIA and AE the higher tercile represents around 50%. This means that there is high dispersion in the EE students' course grades belonging to the lower tercile and the course grades of the EEIA and AE students belonging to the higher tercile.

The average gain G12 has been calculated for each tercile for each degree program. Results are shown in Table 4. ANOVA test performed on the G12 results indicated that there are no significant differences between the three degrees (F (2, 340) = 2.38; $p = 0.094 > 0.05$), so that the degrees were grouped and divided into terciles according to the course grades. The last row of Table 4 shows the average normalised gains for each level with their standard deviation. The ANOVA analysis indicated that there was a significant difference between the normalised gain G12 of the terciles (F (2, 340) = 6.17 and $p < 0.01$).

Orthogonal contrasts were performed to delve into these differences. When applying these contrasts, two results were obtained. There were significant differences between the gains of the higher level students and the rest of the students belonging to the other two levels ($t = 3.26$; $p = 0.002 < 0.05$), while there was no significant difference between the gains of those belonging the lower level and those in the intermediate level ($t = 1.34$; $p = 0.18 > 0.05$). These results indicate significant differences in the normalised gains of the students belonging to the higher level with respect to each of the other two levels.

3.3. Gender perspective

The gender perspective was analysed in the average G12 gain for EE, EEIA and EA students, obtained from BEMA 1 and BEMA 2

Table 1
Number of students (N), average BEMA grades (ABG) and its standard deviation (SD) of the students performing the different BEMA tests in the different degrees (EE: Electric Engineering; EEIA: Electronic Engineering and Industrial Automation; AE: Aerospace Engineering).

	BEMA 1		BEMA 2		BEMA 3	
Degree	N	ABG (SD)	N	ABG (SD)	N	ABG (SD)
EE	92	1.79 (0.94)	96	2.47 (0.94)	78	2.6 (1.2)
EEIA	194	2.6 (1.3)	314	3.8 (1.9)	217	3.4 (1.7)
AE	56	3.4 (1.3)	56	4.3 (1.5)		

Table 2
Number of students (N) performing the BEMA 1 and BEMA 2 tests. Average of the gain and its standard deviation between the BEMA 1 and BEMA 2 tests (G12). The gain is between −1 and 1 point.

	N	G12 (SD)
EE	92	0.05 (0.19)
EEIA	194	0.13 (0.35)
AE	56	0.09 (0.30)

Table 3
Average and standard deviation of the final grades (ACG (SD)), grade interval widths, average, and standard deviation of the defined terciles (L, I, H). (Grades are between 0 and 10 points).

		L		I		H	
Degree	ACG (SD)	ACG Interval	ACG (SD)	ACG Interval	ACG (SD)	ACG Interval	ACG (SD)
EE	6.3 (1.3)	4.2	4.8 (1.2)	0.9	6.45 (0.27)	1.8	7.46 (0.40)
EEIA	7.0 (1.2)	2.5	5.79 (0.50)	1.1	6.92 (0.31)	2.5	8.23 (0.64)
AE	7.1 (1.1)	2.2	5.97 (0.40)	1.0	7.06 (0.28)	2.2	8.46 (0.63)

Table 4
Average of the gain G12 and its standard deviation corresponding to each tercile in each degree program.

	G12 (SD)		
Degree	L	I	H
EE	0.04 (0.20)	0.09 (0.14)	0.01 (0.23)
EEIA	0.04 (0.33)	0.10 (0.36)	0.24 (0.32)
AE	0.01 (0.31)	0.05 (0.27)	0.21 (0.28)
TOTAL	0.04 (0.29)	0.09(0.30)	0.16 (0.33)

tests. Sixty-four female students participated in the study, representing 19% of the total. This participation correlates with the percentage of female students in the different degrees (14%). The average G12 gain results for females and males from the three degrees are reported in Table 5.

The average G12 gain for female students varies between −0.05 from the EA degree to 0.03 for the EEIA degree. Male students obtain higher values, with average values ranging from 0.06 (EE degree) to 0.15 for the EEIA and EA degrees. A two-way ANOVA analysis was performed on G12 gain considering gender and degree program. As expected from the results collected in Table 5, also represented in Fig. 1, significant differences ($F(1, 340) = 9.329$, $p = 0.002 < 0.05$) were found in the average G12 gain between females and males. However, the results do not show any significant differences ($F(2, 340) = 0.818$, $p = 0.442 > 0.05$) for the degree program. It is worth noting that there is no interaction factor between degrees and gender ($F(2, 340) = 0.802$, $p = 0.449 > 0.05$), indicating that the G12 gain between female and male students could present the same behaviour in the three degree programs analysed.

To analyse whether the difference in performance between female and male students in the BEMA test was correlated with differences in the E&M course grades, the average course grades obtained by females and males are included in Table 5 and Fig. 1.

Unlike G12, these results show that female and male students obtained similar grades at the end of the E&M course (see Table 5 and Fig. 1). The ANOVA analysis of the course grades did not find any significant differences ($F = 0.257$ and $p = 0.797$) related to gender. These results point to the strong influence of gender on the BEMA test results, in contrast to what happens in the course grades.

Table 5
Number of students (N), average course grades (ACG (SD)), and average G12 gain (G12 (SD)) of female and male students that participated in this study. (Grades are between 0 and 10 points, whereas gains are between −1 and 1). The last column represents the percentage of female students.

	Female students			Male students			% Female
Degree	N	ACG (SD)	G12 (SD)	N	ACG (SD)	G12 (SD)	
EE	14	6.2 (1.1)	−0.02 (0.23)	79	6.2 (1.4)	0.06 (0.18)	15
EEIA	33	7.1 (1.2)	0.03 (0.38)	161	7.1 (1.2)	0.15 (0.34)	17
EA	17	6.9 (1.2)	−0.05 (0.27)	39	7.3 (1.1)	0.15 (0.29)	30
Total	64	6.9 (1.2)	−0.00 (0.32)	279	6.9 (1.3)	0.12 (0.30)	19

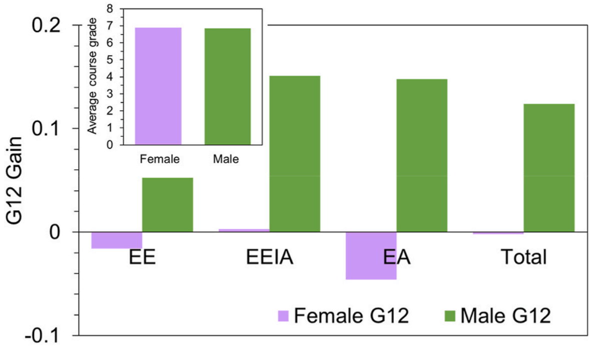

Fig. 1. G12 gain grouped by gender (green for male and purple for female). The insert represents the average course grade by gender.

3.4. *Long term retention*

As stated previously, three BEMA tests were administered (see Fig. 2). BEMA 1, at the beginning of the E&M course, when students have basic E&M knowledge acquired in pre-university courses, BEMA 2 at the end of the course, and finally BEMA 3, four months after the end of the course in EE and EEIA degrees. The E&M course for the EE students was taught during the first semester (September to January), and the BEMA 3 test was administrated at the end of the second semester of the same academic year (early June). However, the course for the EEIA students was taught during the second semester (February to June), and the BEMA 3 test was administrated after the summer holidays, just at the beginning of the next academic year (early September).

Fig. 2 shows the sequence of different BEMAs through time. The gap between the BEMAs is approximately the same in both grades and there is no interference with any other subject. The study, therefore, focused first on the difference in degree programs. The gains between the BEMA 2 and BEMA 3 tests (G23 gains, Equation (2)), related to long term retention, are reported in Table 6, showing negative values in EE and EEIA degrees. An ANOVA test was performed to analyse the statistical significance between both degree programs. The results show no significant differences ($p = 0.973$), which led to not considering this criterion for further analysis.

The data were segmented again into three terciles based on the students' final grades obtained in the E&M course, L (lower tercile), I (intermedium tercile) and H (higher tercile), following the same procedure applied in the analysis of G12 gain. The results are collected in Table 7, classified by degree programs and the total students in the terciles. Applying the ANOVA analysis, the results show significant differences ($F(2,289) = 4.362$, $p = 0.014 < 0.05$) between the different terciles, indicating that this criterion can be considered relevant for a further understanding of the results. Orthogonal contrasts were performed to deepen these differences. Significant differences were found between the G23 gain of the student from the higher tercile (H) compared to the students of the two remaining levels ($t = 2.69$; $p = 0.008 < 0.05$). The Bland-Altman plot for analysing long term retention is also shown in Annexe 2.

These results can be seen more clearly in Fig. 3, where the total G23 gain is represented according to the tercile. All of them presented losses in the knowledge acquired at the end of the course, although the students in the lower tercile registered the greatest loss while those in the higher tercile registered the least.

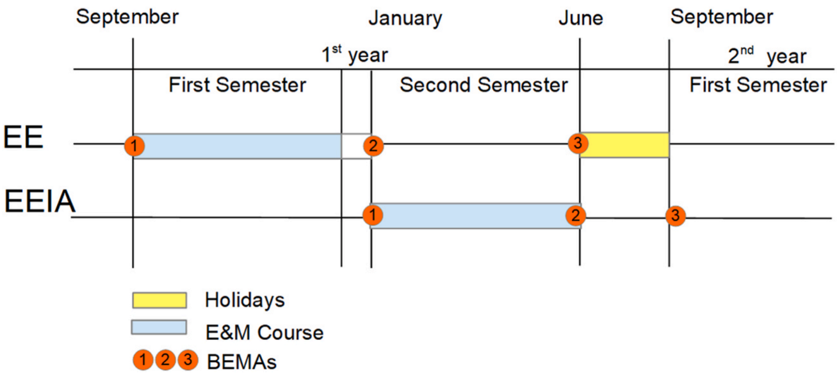

Fig. 2. Temporal distribution of BEMA 1 and BEMA 2, before and after instructions, and BEMA 3 after a few months of instruction to measure retention.

Table 6
Number of students (N) performing the BEMA 2 and BEMA 3 tests. Average of the gain and its standard deviation between the BEMA 2 and BEMA 3 tests (G23 (SD)).

Degree	N	G23 (SD)
EE	75	−0.11 (0.26)
EEIA	217	−0.11 (0.25)
Total	292	−0.11 (0.26)

Table 7
Average of the gain G23 and its standard deviation corresponding to each tercile in each degree program.

	G23 (SD)		
Degree	L	I	H
EE	−0.15 (0.27)	−0.11 (0.29)	−0.05 (0.20)
EEIA	−0.16 (0.29)	−0.10 (0.22)	−0.05 (0.26)
TOTAL	−0.16 (0.28)	−0.10 (0.24)	−0.05 (0.25)

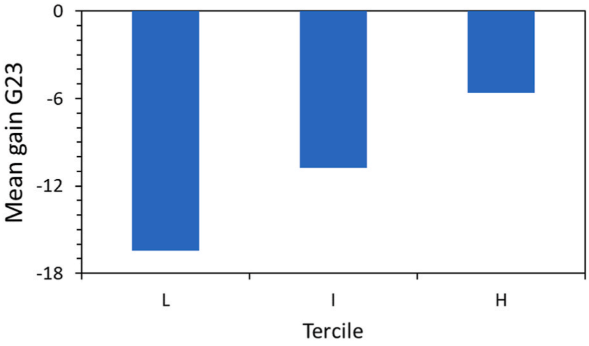

Fig. 3. Mean gain G23 by terciles.

4. Discussion

When applying the FC method, the number of students enrolled in the groups is an important factor. In this case the groups were numerous, given that first-year courses were analysed: between 50 and 70 students in the classroom. This can affect the academic results and the students' opinion of FC method, among other factors. One of the main benefits of this study is that the relatively large number of students involved in this work allows drawing conclusions that could be generalized in the university environment.

It can be first appreciate the high level of participation (Table 1). More than 450 students participated in this study, considering the different degree programs and courses. In some cases, the BEMA test was performed online, which facilitated taking the exams during the pandemic confinement, and as has been shown, its validity can be maintained [37–39]. It has been determined in previous work that BEMA has discriminatory capability and reliability when applied to first-year students of two engineering degrees at the *Universitat Politècnica de València* [29], all of which increases the validity of the results obtained.

Although the average normalised gain G12 obtained in the EE degree is lower than the other two degree programs (0.05 vs 0.13 and 0.09, Table 2), the statistical analysis shows no significant differences. Then, the degree program is not relevant to the results obtained in the normalised average gain G12. High standard deviations were associated with the mean normalised gain, indicating a high dispersion of the individual student gains around the mean. This may explain why there are no significant differences between the degree programs, while the students' current level, based on the course grade turned out to be a significant factor in the normalised gains obtained.

The average normalised gains obtained by students vary between 0.04 and 0.16 (Table 4) according to their academic level, although even lower values can be found in the literature. Kohlmyer et al. [28] found that post-BEMA test averages were significantly higher for the more specific M&I (Matter and Interactions) curriculum than the traditional curriculum. They argued that the revised learning progression offered by the M&I curriculum could be responsible for the better performance in the BEMA of the M&I students. In this study, the traditional class had normalised gains ranging from 0.15 to 0.40 and the M&I class from 0.30 to 0.52. The same trend can be seen in the results obtained by Hake [52] for the traditional course, since his article also showed lower normalised gains for the traditional class (0.08–0.24) than that obtained in interactive participatory courses (0.20–0.68). Kost-Smith et al. [43] obtained a normalised gain of between 0.33 and 0.48, with an average normalised gain of 0.40 for an E&M Physics course in the second semester, using the BEMA pre and post-test to calculate the gain.

The results of the G12 were also studied from the gender point of view. When separating the BEMA results for male and female students, it was found that the normalised gains did not depend on the degree program for either gender. However, significantly lower gains were obtained in women (0.00) than men (0.12), although the male and female course grades were not significantly different (Table 5 and Fig. 1). This trend coincides with that observed by Kost-Smith et al. [43], who obtained a significantly different normalised gain for men and women (0.42 versus 0.35) in the BEMA pre and post-test. They found that females scored about six percentage points lower than males on the post-test, although they performed equally well in the pre-test. These authors indicated that post-test gender differences could be attributed to differences in males' and females' prior physics and maths performance and their incoming attitudes and beliefs. They found that females outperformed males in homework and participation and males outperformed females in exams and competitions [53] and that the males and females' course grades were significantly different, as in the present study.

In the study performed by Henderson et al. [41], their collected data showed that male students outperformed female students in the pre-test (5%) and post-test (6%) Conceptual Survey of Electricity and Magnetism (CSEM). They found that male and female students performed equally well on quantitative test problems and concluded that the gender gaps were not a result of the general differences in physics ability. Henderson et al. [42] found that men had a 2.5%–3.5% advantage for pre-test scores on the CSEM test. A specific analysis examined differences in fairness between items and could not eliminate the possibility that the origin of the general advantage toward men was that most items were consistently unfair. Pollock [26] found that pre and post-test BEMA scores were significantly different by gender. Men improved 6% more than women in the pre-test and 10.7% more in the post-test, a result for which they do not have a mechanistic explanation. Their average final grades for males and females in this course and across the BEMA subpopulations were not statistically different. Sayre et al. [33] noticed a slight gender bias in the proposed tasks, with women slightly more likely to participate. This was statistically significant in the autumn and winter quarters, but the spring quarter showed a smaller statistically insignificant gap. Lorenzo et al. [47] reported that teaching with certain interactive strategies reduced the gender gap, and that in the most interactively taught courses the pre-instruction gender gap disappeared at the end of the course. In this study, although active methods were used, the gender gap is still present.

As observed in the G12 results, a positive effect of instruction was found, but this is reduced after completing the course (long term retention). As expected, a negative average G23 value was obtained, indicating that some knowledge was lost between BEMA 2 and BEMA 3 (Table 6). In the G23 gain it was found that there were no significant differences between the two degree programs. The students were divided into three levels according to the course grade. The students' G23 at the lower level was significantly lower than those at the higher level (−0.16 to −0.05 Table 7), indicating that the ability to retain knowledge at the end of the course depends on the students' level. Those with high or medium grades maintained their BEMA test results over time, while those with low grades (lower level students) could not. The intermediate level students show a slight drop in their BEMA results. The difference was smaller for the higher level students, indicating that practically all the knowledge acquired was maintained.

Present results coincide with those in the bibliography. Knowledge was lost in Sayre et al. [33], as in this study. A previous study (Sayre & Heckler, 2009) found that student understanding is dynamic and time-dependent and also that current instruction impacts previously learned knowledge, termed as "interference" [54]. They showed that this dynamism continues well beyond the immediate period surrounding instruction.

In general, it can be seen in this study that the neediest students do not receive all the help they need to improve their performance, and the courses are not specially designed to take this factor into account. In addition, from a gender perspective, it would be necessary to include more active and cooperative methodologies that make women feel more integrated than in the current competitive environment. So, instructors should apply teaching approaches that engage girls and encourage their learning to improve their self-efficacy and other motivational components [44]. In short, as indicated by other studies [44,55], it is about creating a learning environment that allows positive interactions with peers and that encourages students' interest in physics.

All these results lead us to consider that the design of the courses should take into account a gender perspective and also pay special attention to students with intermediate and low performance, who are those in whom there is a greater margin for improvement. Therefore, the inclusion of more active methodologies should be included; group work, class discussions and in general all those techniques that can help groups currently obtaining a worse performance.

5. Conclusions

In this paper the gain in the BEMA test was analysed obtained by first-year engineering students during the course and their retention after a few months in an active methodologies environment based on ICT. The influence of degree program, academic level and gender perspective were also studied.

A considerable dispersion was obtained of the data and a mean G12 gain value lower than those in the literature. The results showed that the G12 and G23 gains did not depend on the degree program. In contrast, it was found that these gains depended on the students' academic level. The highest level students presented the best gain results, while those at the lowest level showed the worst, with the gain results of the intermediate level being in between. Actions aimed at improving the teaching in STEM disciplines should therefore get special attention in this sector of the student body, since they not only start from the lowest level but their improvement is also lower in relation to their capacity for improvement.

As established in the literature, G12 and G23 showed a dependence on students' gender. Males outperformed females in the BEMA tests, although no significant differences could be found in the course grades. It is thus necessary to understand these differences and implement measures in daily teaching work to improve women's performance in questionnaires and tests.

6. Limitations and suggestions for future work

While the study included quite a large number of participants, it has some limitations. One of the most significant was that it only had data collected from faculty. Therefore, non-academic factors could be included in the study. Besides, student gender was determined using first name criteria before the anonymisation process. The number of instructors was relatively small, therefore, the impact of instructors on student performance was not analysed. Further research should include a greater number of factors to explain the gender differences to find appropriated pedagogical action to promote the participation of women in science and reduce the gender gap.

Furthermore, this study was conducted in an European state university; further studies would conduct similar research, including more state and private universities, to compare their results with those of the current work.

Summing up, instructors should promote an equitable and inclusive learning environment, favouring positive peer interactions and making all students feel they belong in Physics.

Author contribution statement

María-Antonia Serrano, Ana Vidaurre, José M. Meseguer-Dueñas, Isabel Tort-Ausina, Susana Quiles, Roser Sabater i Serra, Tania García-Sanchez, Soledad Bernal-Pérez, M. Amparo Gámiz-González, José Molina-Mateo, José Antonio Gómez-Tejedor, Jaime Riera: Conceived and designed the experiments; Performed the experiments; Analysed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data will be published in Data in Brief journal and author will made available on request.

Additional information

No additional information is available for this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Authors would like to thank the *Instituto de Ciencias de la Educación* (ICE) in the *Universitat Politècnica de València* for their help, through the Innovation and Educational Quality Program and for supporting the team of the *Innovación en Metodologías Activas para el Aprendizaje de la Física (e-MACAFI)*. This work has been supported by the UPV through the Project of Innovation and Educational Improvement Program (Projects PIME/18-19/76, PIME/18-19/88 and PIME/20-21/220 Convocatoria de Proyectos de Innovación y Convergencia de la UPV).

Annexe 1. Difference between post- and pre-test The difference between the post-test (BEMA 2) and pre-test (BEMA 1) grades has been depicted in Figure A1 for the defined terciles. A more appreciable dispersion can be seen in the results of the EEIA and AE degree students

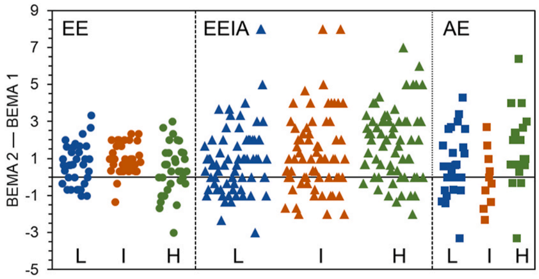

Fig. A1. Difference between the post-test BEMA 2 and the pre-test BEMA 1 scores for Electric Engineering (EE, circles), Electronic Engineering and Industrial Automation (EEIA, triangles), and Aerospace Engineering (AE, boxes) separated according to the segmentation done by academic level: lower (L, blue), intermediate (I, orange) and higher (H, green).

Annexe 2. Bland-Altman plot for analysing long term retention

As an additional tool for analysing long term retention, a Bland-Altman plot [56] was included with the data obtained from BEMA 2 and BEMA 3 (Figure A2). The X-axis is the semi-sum of BEMA 2 and BEMA 3, and the Y-axis shows the difference between both BEMAs (BEMA 3 − BEMA 2). For greater clarity, data are shown in three different graphs. The horizontal lines are the average difference value (solid blue lines). In all cases this average difference is negative, indicating loss of performance, as can be expected sometime after the end of the instruction. Turning to the distribution, it can be seen that the results in (L) and (I) are similar, although in (H) the students in this tercile reached higher scores than those of the lower and intermediate. It should also be noted that the students with the highest scores (semi-sum (BEMA 2 & BEMA 3) ≥ 6) show hardly any losses.

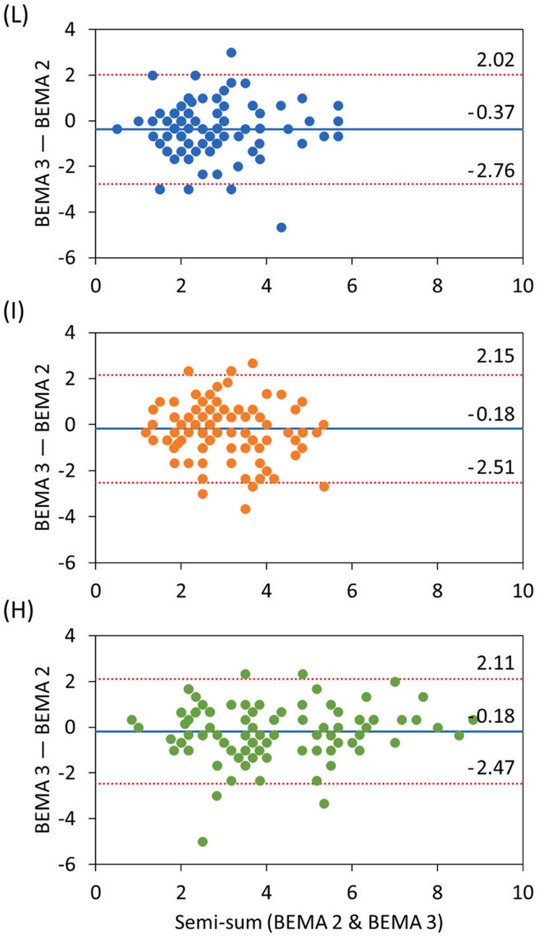

Fig. A2. Bland-Altman plot from BEMA 3 and BEMA 2 results: (L) Lower tercile, (I) Intermediate tercile, and (H) higher tercile. The values of the average difference and the limit of agreement are indicated over the corresponding lines.

Annexe 3. Students' perception

It is essential to understand students' motivation, confidence, and engagement related to the FC environment. Awidi [57] found that the FC provided a beneficial learning experience to university biology students. The students reported being supplied with the necessary resources and information and felt supported. However, motivation was not associated with participation, collaboration, assessment, or feedback, suggesting the need for improvements in learning design and communication. Burke and Fedorek [15] found that FC students felt less engaged than those in traditional and online classrooms, concluding the importance of delivering course content in the right way at the right time.

At the end of the course, a survey was carried out among a total of 120 EEIA and AE degree students ($N_{AE} = 61$, $N_{EEIA} = 59$) to evaluate their satisfaction with the E&M course and identify any areas for improvement. They were asked about: A) Their interest in the course, B) Quality of the online teaching component and assessment of the face-to-face activities such as teamwork and gamification, C) Preference for FC over traditional teaching, and D) Number of weekly hours dedicated to the study of the course.

The survey consisted of nine questions, in six of which a Likert scale with five levels of perception was used (strongly disagree, disagree, undecided, agree, strongly agree). Three open questions were included to collect suggestions for improvement: i) what aspects of the learning method do you like, ii) what aspects of the learning methodology you do not like, and iii) how can it be improved. To preserve their anonymity the students' names were eliminated and replaced by a numerical code. The complete set of questions are listed below.

1. Based on your learning experience during this course, please indicate whether you agree with the following statements:
   a) I am very interested in the content of this subject.
   b) After this semester, this subject will be very useful for me.
   c) I will need the contents of this subject in subsequent ones.
   d) I prefer flipped classroom format over a traditional class format.
   e) I would prefer to have more subjects with this type of methodology.
2. Assess the overall quality of the following aspects:
   a) Online component of teaching.
   b) Face-to-face component of teaching.
   c) The subject as a whole.
3. Evaluates different resources or activities carried out in the subject:
   a) Content synthesis task and resolution of doubts by the teacher.
   b) Documentation in video format.
   c) Documentation in PDF format.
   d) Exercises with PoliformaT tests.
   e) Exhibition of "monologues" on the contents of the subject.
   f) Peer evaluation.
   g) Presentations in PowerPoint.
   h) Team projects.
   i) Team problem solving and presentation.
   j) Gamification activities (Kahoot).
   k) Laboratory practices.
   l) Tutorials.
4. The weekly hours that I dedicate to the subject are:
5. Regarding the subject and its methodology, in your opinion:
   a) What are the strong points of the course:
   b) What are the points that should be improved in the subject:
   c) What proposals for improvement, which could be useful, do you propose:
6. Please indicate your degree of agreement with the following statements:
   a) I think that, despite the limitations of the situation we are going through, the subject has been able to function correctly.
   b) I think that, despite the limitations of the situation we are going through, I have managed to follow the subject reasonably well.
7. As for the characteristics of my home working place, from where I follow up on the classes:

a) I have sufficient technical means to follow the classes online.
b) I have a suitable working place (spacious and isolated) to be able to carry out my work.

8. We ask you to briefly assess the adaptation work of the teaching carried out by the professors of the subject.
9. We ask you to briefly assess your work and the conditions in which you do it in this situation.

The following data stand out among the results obtained in the survey.

A) In both degree programs, most students (90% in AE and 75% in EEIA) were interested in the Physics and Electricity courses. This positive predisposition is related to their opinion on the subjects' usefulness, as 80% agreed that they would need the contents in later courses.
B) The online component of the FC teaching method is very important. When students were asked about its quality, around 90% valued it positively in both grades (Figure A3).

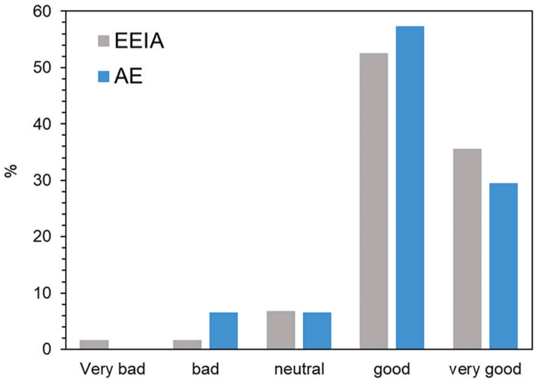

Fig. A3. Opinion of the AE and EEIA students on the online component of the teaching methodology.

Regarding other aspects of the method used, some results obtained in both degree programs coincide. For example, when it comes to teamwork (solving exercises and presenting them to the rest of the class), between 70% and 80% made a positive assessment. On the other hand, their opinion of gamification activities was quite different: 40% of the AE students gave it a positive rating (considering that almost 30% are neutral), compared to 70% of positive ratings by EEIA students.

C) When asked about their preference for FC or traditional teaching, in both degree programs the answers coincide: less than 50% of students prefer this teaching model to the traditional model. In addition, the AE distribution of responses presents two maximum values between students in favour of and against the FC (Figure A4).

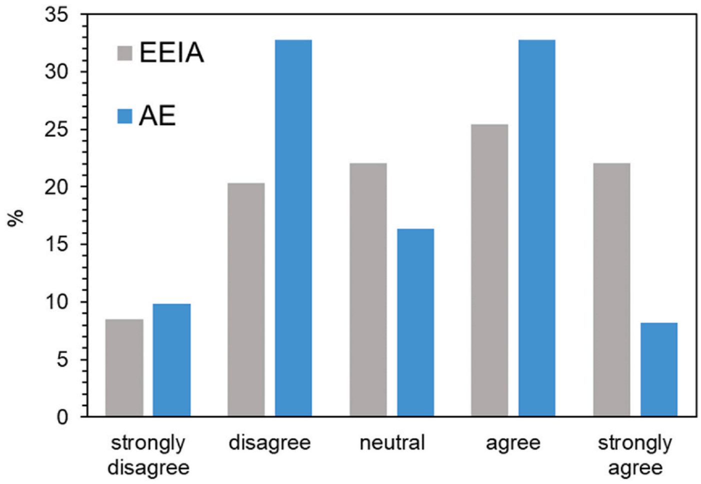

Fig. A4. level of agreement to the statement "I prefer this format of FC over a traditional class format."

To sum up, when asked about specific aspects of the method the assessment was positive, except for the FC methodology where their opinion was not so good.

D) A big difference was found in the number of weekly hours they spent studying the course. AE students spend an average of 2.6 h, less than 1 h per class hour, while EEIA students spend 6.3 h, almost 2 h per class hour. Although the success rate is above 90% in both grades, to achieve this the AE students declared that they needed less dedication.

E) They were also asked for their opinions through three open questions about what they like and what they don't like about the method and how it could be improved. Opposing views have been found. Some appreciated the FC method's advantages such as autonomy, flexibility and continuous promotion of lifelong learning. They said they value positively "The freedom to adjust the study time to my needs". They also recognise the obligations involved in FC: "The fact that teachers send us tasks constantly, which forces me to study a little, is fantastic". They are aware of what it requires and value it positively: " … I understand things better, since I have already worked on them and I can clarify concepts in class". However, some thought the opposite; "It's hard to follow. It's easier if they explain how to do an exercise, and then with that model you try to do new exercises". The assessments of the online material used were all positive: " … the course documents and resources database are complete in all

> aspects and perfectly organised in topics to facilitate self-learning". However, the excess workload appears in comments such as: "It involves spending much more time than in other courses" or "Studying unknown content without help can be very difficult".

The opinion surveys on students in the EEIA and AE degrees provided information that led to more concrete conclusions about the results obtained. In total, more than 100 surveys were analysed and gave a good perspective on the students' opinions. The students generally enjoyed the course and appreciated the programmed activities. However, when asked about the FC, the opinions are more polarised, which shows that it is a more controversial method. This paper coincide with other authors in that the lack of pre-class preparation, mainly in the numerous first-year university course, represents the main difficulty [16–18] and affects their vision of the FC. The lower G12 values obtained in this study than the results obtained for the more active learning courses [28,43,52] could be related to students' lack of involvement.

Although the students had a favourable opinion of the subject in general on different aspects of the course, their opinion of the FC method varied, with some for and against it. Although it has already been shown in the literature that the FC improves student performance, considerable didactic work is also necessary for the students to see the improvements it entails and thus be able to assess it more positively.

References

[1] J. Bergmann, A. Sams, Flip Your Classroom Reach Every Student in Every Class Every Day, International Society for Technology in Education, 2012.

[2] Y. Chen, Y. Wang, Kinshuk, N.-S. Chen, Is FLIP enough? Or should we use the FLIPPED model instead? Comput. Educ. 79 (2014) 16–27, https://doi.org/10.1016/j.compedu.2014.07.004.

[3] W. He, A. Holton, G. Farkas, M. Warschauer, The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions, Learn, Instr 45 (2016) 61–71, https://doi.org/10.1016/j.learninstruc.2016.07.001.

[4] W. He, A.J. Holton, G. Farkas, Impact of partially flipped instruction on immediate and subsequent course performance in a large undergraduate chemistry course, Comput. Educ. 125 (2018) 120–131, https://doi.org/10.1016/j.compedu.2018.05.020.

[5] J. Baughman, L. Hassall, X. Xu, Comparison of student team dynamics between nonflipped and flipped versions of a large-enrollment sophomore design engineering course, J. Eng. Educ. 108 (2019) 103–118, https://doi.org/10.1002/jee.20251.

[6] T. Rahman, S.E. Lewis, Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis, J. Res. Sci. Teach. 57 (2020) 765–793, https://doi.org/10.1002/tea.21610.

[7] J.A. Gómez-Tejedor, A. Vidaurre, I. Tort-Ausina, J. Molina-Mateo, M.-A.A. Serrano, J.M. Meseguer-Dueñas, R.M. Martínez Sala, S. Quiles, J. Riera, Effectiveness of flip teaching on engineering students' performance in the physics lab, Comput. Educ. 144 (2020), 103708, https://doi.org/10.1016/j.compedu.2019.103708.

[8] N.T.T. Thai, B. De Wever, M. Valcke, The impact of a flipped classroom design on learning performance in higher education: looking for the best "blend" of lectures and guiding questions with feedback, Comput. Educ. 107 (2017) 113–126, https://doi.org/10.1016/j.compedu.2017.01.003.

[9] Y. Song, M. Kapur, How to flip the classroom - "productive failure or traditional flipped classroom" pedagogical design? Educ. Technol. Soc. 20 (2017) 292–305, https://doi.org/10.2307/jeductechsoci.20.1.292.

[10] I. Blau, T. Shamir-Inbal, Re-designed flipped learning model in an academic course: the role of co-creation and co-regulation, Comput, Educ. Next 115 (2017) 69–81, https://doi.org/10.1016/j.compedu.2017.07.014.

[11] Z. Zainuddin, H. Haruna, X. Li, Y. Zhang, S.K.W. Chu, A systematic review of flipped classroom empirical evidence from different fields: what are the gaps and future trends? Horiz 27 (2019) 72–86, https://doi.org/10.1108/OTH-09-2018-0027.

[12] S. Sergis, D.G. Sampson, L. Pelliccione, Investigating the impact of Flipped Classroom on students' learning experiences: a Self-Determination Theory approach, Comput. Hum. Behav. 78 (2018) 368–378, https://doi.org/10.1016/j.chb.2017.08.011.

[13] M.M. El-Banna, M. Whitlow, A.M. McNelis, Flipping around the classroom: accelerated Bachelor of Science in Nursing students' satisfaction and achievement, Nurse Educ. Today 56 (2017) 41–46, https://doi.org/10.1016/j.nedt.2017.06.003.

[14] A.J. Boevé, R.R. Meijer, R.J. Bosker, J. Vugteveen, R. Hoekstra, C.J. Albers, Implementing the flipped classroom: an exploration of study behaviour and student performance, High Educ. 74 (2017) 1015–1032, https://doi.org/10.1007/s10734-016-0104-y.

[15] A.S. Burke, B. Fedorek, Does "flipping" promote engagement?: a comparison of a traditional, online, and flipped class, Act. Learn. High. Educ. 18 (2017) 11–24, https://doi.org/10.1177/1469787417693487.

[16] T. Long, J. Cummins, M. Waugh, Use of the flipped classroom instructional model in higher education: instructors' perspectives, J. Comput. High Educ. 29 (2017) 179–200, https://doi.org/10.1007/s12528-016-9119-8.

[17] L. Tomas, N. Evans, T. Doyle, K. Skamp, Are first year students ready for a flipped classroom? A case for a flipped learning continuum, Int. J. Educ. Technol. High. Educ. 16 (2019) 5, https://doi.org/10.1186/s41239-019-0135-4.

[18] J. McCarthy, Reflections on a flipped classroom in first year higher education, Issues Educ. Res. 26 (2016) 332–350.

[19] C.K. Lo, K.F. Hew, The impact of flipped classrooms on student achievement in engineering education: a meta-analysis of 10 years of research, J. Eng. Educ. 108 (2019) 523–546, https://doi.org/10.1002/jee.20293.

[20] L. De-Marcos, E. Garcia-Lopez, A. Garcia-Cabot, On the effectiveness of game-like and social approaches in learning: comparing educational gaming, gamification & social networking, Comput. Educ. 95 (2016) 99–113, https://doi.org/10.1016/j.compedu.2015.12.008.

[21] Z. Zainuddin, Students' learning performance and perceived motivation in gamified flipped-class instruction, Comput. Educ. 126 (2018) 75–88, https://doi.org/10.1016/j.compedu.2018.07.003.

[22] A.I. Wang, R. Tahir, The effect of using Kahoot! for learning – a literature review, Comput. Educ. 149 (2020), 103818, https://doi.org/10.1016/j.compedu.2020.103818.

[23] M. Yasmin, Online chemical engineering education during COVID-19 pandemic: lessons learned from Pakistan, Educ. Chem. Eng. 39 (2022) 19–30, https://doi.org/10.1016/j.ece.2022.02.002.

[24] F. Delgado, Teaching physics for computer science students in higher education during the covid-19 pandemic: a fully internet-supported course, Future Internet 13 (2021) 1–24, https://doi.org/10.3390/fi13020035.

[25] I. Tort-Ausina, A. Vidaurre, J. Riera, M.A. Gamiz-Gonzalez, J.M. Meseguer-Dueñas, J.A. Gomez-Tejedor, S. Quiles-Casado, M.-A. Serrano, J. Molina-Mateo, Are we ready for a chronic crisis ? Reflections on the experience of teaching during confinement, in: Procedings EDULEARN 21 Conf, 2021, pp. 359–367.

[26] S.J. Pollock, Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA, AIP Conf. Proc. 1064 (2008) 171–174, https://doi.org/10.1063/1.3021246.

[27] L. Ding, R. Chabay, B. Sherwood, R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, Phys. Rev. Spec. Top. - Phys. Educ. Res. 2 (2006), 010105, https://doi.org/10.1103/PhysRevSTPER.2.010105.

[28] M.A. Kohlmyer, M.D. Caballero, R. Catrambone, R.W. Chabay, L. Ding, M.P. Haugan, M.J. Marr, B.A. Sherwood, M.F. Schatz, Tale of two curricula: the performance of 2000 students in introductory electromagnetism, Phys. Rev. Spec. Top. - Phys. Educ. Res. 5 (2009), 020105, https://doi.org/10.1103/PhysRevSTPER.5.020105.

[29] M.A. Gamiz-Gonzalez, A. Vidaurre, R. Sabater Serra, I. Tort-Ausina, M.-A. Serrano, J. Riera, J.M. Meseguer-Dueñas, J.A. Gomez-Tejedor, J. Molina-Mateo, T. Garcia-Sanchez, Evaluating reliability and discriminatory capability of BEMA in two Spanish engineering degrees, Educ. New Dev. (2019) 303–305.

[30] T. García-Sánchez, R. Sabater i Serra, A. Vidaurre, J.A. Gómez-Tejedor, M.-A. Serrano, J.M. Meseguer-Dueñas, S. Bernal-Perez, J. Riera, J. Molina-Mateo, V. Donderis Quiles, M.A. Gámiz González, Assessing Outcomes in Electricity and Magnetism Courses in Engineering Degrees. Students' Performance Analysed by BEMA, Educ. New Dev., 2019, pp. 144–148.

[31] B.R. Wilcox, S.J. Pollock, D.R. Bolton, Retention of conceptual learning after an interactive introductory mechanics course, Phys. Rev. Phys. Educ. Res. 16 (2020), 10140, https://doi.org/10.1103/PHYSREVPHYSEDUCRES.16.010140.

[32] S.J. Pollock, Longitudinal study of student conceptual understanding in electricity and magnetism, Phys. Rev. Spec. Top. - Phys. Educ. Res. 5 (2009) 1–8, https://doi.org/10.1103/physrevstper.5.020110.

[33] E.C. Sayre, S.V. Franklin, S. Dymek, J. Clark, Y. Sun, Learning, retention, and forgetting of Newton's third law throughout university physics, Phys. Rev. Spec. Top. - Phys. Educ. Res. 8 (2012) 1–10, https://doi.org/10.1103/PhysRevSTPER.8.010116.

[34] Y.J. Dori, E. Hult, L. Breslow, J.W. Belcher, How much have they retained? Making unseen concepts seen in a freshman electromagnetism course at MIT, J. Sci. Educ. Technol. 16 (2007) 299–323, https://doi.org/10.1007/s10956-007-9051-9.

[35] P.H. Kvam, The effect of active learning methods on student retention in engineering statistics, Am. Stat. 54 (2000) 136–140, https://doi.org/10.1080/00031305.2000.10474526.

[36] M. Ardid, J.A. Gómez-Tejedor, J.M. Meseguer-Dueñas, J. Riera, A. Vidaurre, Online exams for blended assessment. Study of different application methodologies, Comput. Educ. 81 (2015) 296–303, https://doi.org/10.1016/j.compedu.2014.10.010.

[37] S. Bonham, Reliability, compliance, and security in web-based course assessments, Phys. Rev. Spec. Top. - Phys. Educ. Res. 4 (2008) 1–8, https://doi.org/10.1103/PhysRevSTPER.4.010106.

[38] J.I. Yasuda, M.M. Hull, N. Mae, Improving test security and efficiency of computerized adaptive testing for the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. 18 (2022), 10112, https://doi.org/10.1103/PhysRevPhysEducRes.18.010112.

[39] T.M. Clark, D.A. Turner, D.C. Rostam, Evaluating and improving questions on an unproctored online general chemistry exam, J. Chem. Educ. 99 (2022) 3510–3521, https://doi.org/10.1021/acs.jchemed.2c00603.

[40] A. Madsen, S.B. McKagan, E.C. Sayre, Gender gap on concept inventories in physics: what is consistent, what is inconsistent, and what factors influence the gap? Phys. Rev. Spec. Top. - Phys. Educ. Res. 9 (2013) 1–15, https://doi.org/10.1103/PhysRevSTPER.9.020121.

[41] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, A. Traxler, Exploring the gender gap in the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. 13 (2017) 1–17, https://doi.org/10.1103/PhysRevPhysEducRes.13.020114.

[42] R. Henderson, P. Miller, J. Stewart, A. Traxler, R. Lindell, Item-level gender fairness in the force and motion conceptual evaluation and the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. 14 (2018), 20103, https://doi.org/10.1103/PhysRevPhysEducRes.14.020103.

[43] L.E. Kost-Smith, S.J. Pollock, N.D. Finkelstein, Gender disparities in second-semester college physics: the incremental effects of a " smog of bias, Phys. Rev. Spec. Top. - Phys. Educ. Res. 6 (2010) 1–17, https://doi.org/10.1103/PhysRevSTPER.6.020112.

[44] B. Radulović, V. Županec, M. Stojanović, S. Budić, Gender motivational gap and contribution of different teaching approaches to female students' motivation to learn physics, Sci. Rep. 12 (2022) 1–9, https://doi.org/10.1038/s41598-022-23151-7.

[45] S. Verdugo-Castro, A. García-Holgado, M.C. Sánchez-Gómez, The gender gap in higher STEM studies: a systematic literature review, Heliyon 8 (2022), https://doi.org/10.1016/j.heliyon.2022.e10300.

[46] M. Dew, J. Perry, L. Ford, W. Bassichis, T. Erukhimova, Gendered performance differences in introductory physics: a study from a large land-grant university, Phys. Rev. Phys. Educ. Res. 17 (2021), 10106, https://doi.org/10.1103/PhysRevPhysEducRes.17.010106.

[47] M. Lorenzo, C.H. Crouch, E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. 74 (2006) 118–122, https://doi.org/10.1119/1.2162549.

[48] PoliformaT. Universitat, Politècnica de València, 2003. https://poliformat.upv.es/. (Accessed 2 March 2020).

[49] Kahoot!, Kahoot! | Learning Games | Make Learning Awesome!, 2021. https://kahoot.com/. (Accessed 19 October 2021).

[50] L. Bao, Theoretical comparisons of average normalized gain calculations, Am. J. Phys. 74 (2006) 917–922, https://doi.org/10.1119/1.2213632.

[51] J.A. Gómez-Tejedor, A. Vidaurre, I. Tort-Ausina, J. Molina-Mateo, M.-A. Serrano, J.M. Meseguer-Dueñas, R.M. Martínez Sala, S. Quiles, J. Riera, Effectiveness of flip teaching on engineering students' performance in the physics lab, Comput. Educ. 144 (2020), 103708, https://doi.org/10.1016/j.compedu.2019.103708.

[52] R.R. Hake, Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. 66 (1998) 64–74, https://doi.org/10.1119/1.18809.

[53] A.M. Steegh, T.N. Höffler, M.M. Keller, I. Parchmann, Gender differences in mathematics and science competitions: a systematic review, J. Res. Sci. Teach. 56 (2019) 1431–1460, https://doi.org/10.1002/tea.21580.

[54] M.E. Bouton, Context, time, and memory retrieval in the interference paradigms of Pavlovian learning, Psychol. Bull. 114 (1993) 80–99, https://doi.org/10.1037//0033-2909.114.1.80.

[55] S. Cwik, C. Singh, How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses, Phys. Rev. Phys. Educ. Res. 17 (2021), 20143, https://doi.org/10.1103/PhysRevPhysEducRes.17.020143.

[56] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (1999) 135–160, https://doi.org/10.1177/096228029900800204.

[57] I.T. Awidi, M. Paynter, The impact of a flipped classroom approach on student learning experience, Comput. Educ. 128 (2019) 269–283, https://doi.org/10.1016/j.compedu.2018.09.013.