ORIGINAL ARTICLE

Journal Section

Gender, Prior Knowledge and the Impact of a Flipped Linear Algebra Course for Engineers over Multiple Years

Anonymous Authors

Anonymous Affiliations

Correspondence
Anonymous correspoundence
Email: anon@example.com

Present address
Anonymous present address

Funding information
Anonymous funders

Abstract

Background: Research shows that active pedagogies could play an important role in achieving more equitable outcomes for underrepresented students in Science, Technology, Engineering, and Mathematics (STEM). Although Flipped Classes are a popular active methodology, there is a lack of high-quality studies assessing their impact in ecologically-valid settings, in particular for 'non-traditional' engineering students.

Purpose: This paper presents two *modified* replications of an experimental study investigating the impact of the flipped classroom approach on student learning in large-size classes of the first year of an Engineering Bachelor.

Methodology: A new strand, progressively flipped over three years, has been added to eight parallel occurrences of a high-stakes mandatory Linear Algebra course for engineers (1,700 students). The study followed a replicated-between-subjects design, with students in the flipped strand learning the same material as in the other strands and taking the same final exam.

Findings: Consistent with prior findings, our results indicate that the flipped format did not have any evident impact on students' learning overall. However, both replications in the flipped condition show a reduced attainment gap for female students and students with less prior knowledge in mathematics.

Conclusion: While Flipped Classrooms seem to have weaker effects on learning than other active methodologies, the evidence in this study indicates that they may have an impact on reducing the attainment gap between traditional and non-traditional students in engineering. They may therefore be particularly interesting to consider in efforts to achieve more equitable outcomes for underrepresented students.

KEYWORDS

Flipped Classroom, Engineering education, Underrepresented students, Gender, Linear Algebra, Experimental Research, Active Learning

* Equally contributing authors.

1 | INTRODUCTION

Science, Technology, Engineering, and Mathematics (STEM) fields have had a lingering problem with the equal representation of women and under-represented minorities. Studies –particularly in the western, developed world– have shown that despite policy-makers' efforts in the past 50 years to increase the representation of women and minorities, the progress has been slow (Lichtenstein, Chen, Smith, & Maldonado, 2015). This trend has held true for more or less all STEM fields, while in engineering education "*minimal progress has been made in recruiting and retaining students, and especially women and minorities*" (Lichtenstein et al., 2015), and some research suggests that "*engineering education is not simply numerically male dominated, it is also culturally associated with masculinity*" (Aeby, Fong, Vukmirovic, Isaac, & Tormey, 2019). This implicit bias in associating STEM fields with men has been found correlated with the disparity in the academic performance of male and female students in mathematics (Nosek & Smyth, 2011; Nosek et al., 2009). In addition, female students' perceived inclusion within their respective engineering programs has been shown to undergo a gradual decline over time (Marra, Rodgers, Shen, & Bogue, 2009). This phenomenon has been found to influence the female students' long-term persistence in the engineering domain (Beasley & Fischer, 2012).

Seymour and Hewitt (1997) demonstrated that the aspects of low perceived attractivity of engineering education and lower retention rates for the minorities, are influenced by their perceptions and attitudes towards the culture of educational practices, including teaching styles, within engineering education. They argue that attracting and (more importantly) retaining minorities would entail profound changes in existing classroom instruction methodologies. Active and interactive teaching approaches have been found to promote *inclusivity* within STEM education, to produce more equitable educational outcomes for students and in particular to reduce the achievement gap for underrepresented students (Theobald et al., 2020). This remains, however, an under-researched topic and a recent meta-analysis by Theobald et al. (2020) found that there were not yet sufficient studies to include gender in their analysis of interactive teaching and under-represented groups in STEM education. They also identified that further research was needed to distinguish which approaches to interactive teaching were most likely to have an impact.

Besides the aforementioned benefits of active and interactive learning strategies, the peer instruction (Crouch & Mazur, 2001; Fagen, Crouch, & Mazur, 2002) model has also been demonstrated to positively impact the learners' scores and their conceptual understanding of the subject matter. Moreover, the study conducted by Lorenzo, Crouch, and Mazur (2006) demonstrated that interactive and peer-instruction strategies that *a*) promote in-class interaction between teachers and students, *b*) inspire collaboration amongst students, *c*) discourage competition, and *d*) facilitate conceptual understanding, minimize the gender gap amongst students with regards to their academic performance. In addition, the authors observed that regardless of the initial disparity in academic performance of the students, both male and female students benefited (in terms of conceptual understanding and performance) from the interactive teaching approaches towards the end of the semester, however, the gains were significantly higher for the female students. Backed by these evidences, we hypothesized that interactive teaching and peer instruction model may lead to more inclusive pedagogy, and enable students to apply and transfer concepts to a wider range of contexts, and which has a relevance for engineering education.

Also, our engineering university develops and hosts a vast repository of online learning resources in the form of MOOCs (Massive Open Online Courses). Moreover, we observed a growing interest amongst our students, in the past years, towards more interactive teaching methods as compared to traditional lectures (an evidence of this enthusiasm was observed when 25% of students volunteered to participate in our study as illustrated in Section 3). These factors encouraged us to implement flipped classrooms, which combine the peer instruction model in an interactive context together with the online MOOC resources in our university, and also to examine their impact in well-designed experiments.

In this article, we examine the impact of a flipped classroom format on students' learning and academic achievement in an technical university. In particular, we examine the impact on a heterogeneous population of students of a "flipped" approach to a large, core and challenging *Linear Algebra* course. Our analysis considers *two* modified[1] replications of flipping the same course across subsequent academic years, where the students' exposure to the "flipped teaching" was changed incrementally from one year to another. The experiment presented in this article looked at the influence of a flipped class in an ecologically-valid setting, and addressed the following research questions:

1. What is the impact of flipped classroom learning on students' academic achievement and learning gain as compared to traditional (non-flipped) classroom learning?
2. Is there a differential effect of the flipped format on different student groups (specifically, gender, high-school background, and prior level in mathematics)?

In the following sections, *firstly*, we will conduct a literature review of past research on the flipped format, and illustrate the research gaps which we bridge through this article. *Secondly*, we will present our study context, design, participants, and results. *Finally*, the paper will conclude with a discussion of our results, including implications of our findings for the engineering education context.

2 | RELATED WORK

A shortage of well-skilled engineers and the lack of diversity in the profession are important problems confronting the STEM fields with far-reaching socio-economic ramifications, such as income inequality and decreased workplace diversity (Mckenna, Froyd, & Litzinger, 2014; Theobald et al., 2020). Lichtenstein et al. (2015) have provided a comprehensive account of US policy makers' efforts in the last 50 years to increase diversity in STEM fields. Similar measures have also been taken and documented within the UK and other European countries to attract and sustain more under-represented communities in STEM fields (Barnard, Hassan, Bagilhole, & Dainty, 2012; Powell, Dainty, & Bagilhole, 2012). However, regardless of these efforts and numerous programs for making engineering fields more inclusive, progress has been slow and disheartening (Aeby et al., 2019; Lichtenstein et al., 2015; Seymour & Hewitt, 1997). Moreover, the low retention rates of women and other minorities in engineering –particularly in undergraduate education– have been attributed to students' negative perception and attitude towards the prevalent culture of educational practices (Lorenzo et al., 2006; Nosek et al., 2009; Seymour & Hewitt, 1997). Secules (2017) argues that this slow progress in diversifying engineering education has been due to a misplaced focus, which has been "*more on the overlooked assets of minority groups than on the acts of overlooking, more on the experiences of marginalized groups than on the mechanisms of marginalization by dominant groups, more on supporting and increasing minority student retention than on critiquing and remediating the systems which lead minority students to leave engineering*". In order to address these fundamental and lingering problems in the engineering education context, researchers have argued in favour of an increased focus on the engagement and belonging of under-represented students by proactively engaging them through collaborative or active learning approaches (for example, Atadero, Paguyo, Rambo-Hernandez, and Henderson (2018); Lorenzo et al. (2006); Minin et al. (2016); Theobald et al. (2020)).

In their meta-analysis of over 220 STEM studies, Freeman et al. (2014) found that active learning positively influenced learners' academic achievement (6% higher as compared to traditional lecturing), and reduced learners' chances of failure as compared to traditional lecturing. Furthermore, cooperative learning and feedback are also known to have an important positive impact on learning, with effect sizes of respectively $d = .59$ (when compared to individual

[1] According to *APA Dictionary of Psychology* (`https://dictionary.apa.org/replication`, last visited on 08 September 2020), a **modified replication** is the one where "*a researcher incorporates alternative procedures and additional conditions*".

learning) and $d = .73$ in the meta-analysis by Hattie (2009). Since flipped classrooms represent an active learning methodology incorporating elements of cooperative learning and feedback to students (Cheng, Ritzhaupt, & Antonenko, 2019; DeLozier & Rhodes, 2017; Lo & Hew, 2019; Lo, Hew, & Chen, 2017; O'Flaherty & Phillips, 2015) it could well be expected that the flipped format would have an important positive impact on learning.

Recent years have seen a significant growth in the number of studies looking at the impact of flipped classes on learning and these in turn have been gathered in a number of recent meta-analyses in the field. While some of these have included a range of disciplines (Cheng et al., 2019), others have looked at the impact of the flipped format in engineering education (for example, Lo and Hew (2019)) and in math disciplines (for example, Lo et al. (2017)). These suggest a high degree of variability in both the results and courses which were flipped. Lo and Hew (2019) analysed 29 studies within the context of engineering education, published between 2008 and 2017, and show that while the flipped format had a positive –and significant– influence on students' achievements, the effect size was rather small ($g = .29$). Lo et al. (2017) conducted a meta-analysis which solely considered studies on the flipped format in the domain of mathematics education. Their analysis of 21 studies also revealed that the flipped format is moderately effective, with a significant and positive influence on learning as compared to the traditional format. However, the effect size was again modest ($g = .29$). This suggests that the flipped format does improve learning, but not as radically and profoundly as has been anticipated. In their meta-analysis, Cheng et al. (2019) observed that studies in engineering disciplines showed no statistically significant impact of flipping a class, and indeed, had a negative (if non-significant) effect. This result also led the authors to take a rather despairing position about the potential of the flipped format in engineering disciplines – "*engineering appears to not be a suitable candidate for the flipped classroom method when compared to other disciplines*". While a review by Kerr (2015), did find that students' grades improved in flipped courses, and that students reported a higher satisfaction with the flipped format, the majority of studies in Kerr's meta-analysis examine the impact of the flipped format in classrooms of small sizes. Indeed, it is common to look at flipped approaches in classes of 20-50 students (for example, Mason, Shuman, and Cook (2013); Schiltz, Feldman, and Vaterlaus (2019)). In the 2019 meta-analysis by Lo and Hew (2019), only 6 studies over 29 concerned classes were with more than 100 students.

The studies we have reviewed so far do not demonstrate a large effect size of flipped classes on student learning, particularly in the specific context of engineering education. A general tendency is that the reviewed studies show a high variability in the results, without clear and consistent moderating factors (except, perhaps related to the existence of transition activities at the start of the flipped class which has been identified as quite important in both Lo et al. (2017) and Lo and Hew (2019)). None of these meta-analyses have explored the effects of the background or gender of students on their academic achievement under the flipped format. A separate recent meta-analysis of interactive teaching in STEM education, has focused on studies that decompose the impact of interactive teaching in such a way that it is possible to look at the performance of under-represented minority students as well as low-income students (Theobald et al., 2020). They do find that active learning narrows achievement gaps with respect to these students. However they also find notable limitations in the existing data. *First* there are relatively few studies that report dis-aggregated data (as a result of which they were unable to include gender as a variable in their analysis) and *second*, poor quality of descriptions of classroom practices means they are unable to distinguish between different types of interactive teaching.

If there are few studies which have dis-aggregated data for interactive teaching in general, there are even fewer which have done so in the case of flipped classes. In one such study, (Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015) report on a repeated study, where students' academic achievement in a semester-long 'Physical Chemistry' course (designed for life science majors) was investigated. The last two iterations of this course were taught in the flipped format by the same instructor, and the differences in students' scores were examined across the different iterations of this course. Two separate but associated effects were identified. *First*, the results showed that, although male students performed significantly better than their female counterparts in the traditional format (first three iterations), the gender difference was no longer statistically significant when the course was taught in the flipped format. *Second*, there was also a positive impact on the attainment of students with lower prior performance. Overall, the authors concluded that "*the positive effects of the flipped class are more pronounced for students with lower grade point average and for the female students*" (Gross et al., 2015). Similar results were also found by Chiquito, Castedo, Santos, López, and Alarcón (2020). Another study by Dang and Gajski (2014) observed similar results with regards to students' prior attainment level in the context of engineering. In their study, students with lower prior attainment levels gained significantly, and as a consequence, the difference between previously high performing and previously lower performing students was no longer significant in the flipped condition.

A consistent message emerging from these studies is the need to address the quality of data reported in accounts of interactive teaching in general and flipped classes in particular:

- Many existing studies do not adequately specify how the flipped format was *implemented*, (including the type of learning activities used and their sequence, as well as the student workload and number of contact hours), and as a consequence do not allow the impact of different approaches to flipping a class to be evaluated;
- The *quality of the study designs* is highly variable, many studies do not describe the level of experimental control of the study conditions, including the comparability of teaching (teacher and content in particular), the type of evaluation (student feedback vs. evaluation of learning including the type of assessment), and the comparability of student groups (especially how students are assigned to groups and the control of their prior attainment);
- Numerous studies do not provide sufficient *data* to allow conclusions to be drawn (availability of detailed statistics, description of the type assessment of learning, etc.).

As a consequence, these meta-analyses often make design recommendations for studies on the flipped format, including: *1*) comparable student groups with random assignment of participants and control for previous achievement, *2*) the use of objective measures of learning with verified validity and reliability (see for example Freeman et al. (2014)), and *3*) experimentally controlling for teachers, content taught and study time (workload).

As a summary, previous research works on flipped classrooms, especially those relevant to engineering education, expose several research gaps, including *a*) the lack of rigorous empirical replication studies on the flipped format with well-controlled experimental conditions, *b*) the scarcity of flipped studies in classes of large size ($\geq$ 100 students), and *c*) the shortage of studies that examine whether the flipped format has a differential impact on students in terms of their gender and their prior academic attainment. In this article, we address these aforementioned gaps by presenting a controlled, replicated study to investigate the impact of flipping a large and challenging Linear Algebra course taught to engineering students.

3 | RESEARCH DESIGN

In order to assess the impact of a flipped class in an ecologically-valid setting, we flipped one strand of the mandatory 'Linear Algebra' course, which was concurrently – on the same day and same hour – taught by different teachers in 9 different strands to the first semester bachelor students in a mainland European technical university. It is worth noting that the same teacher taught the course in the flipped manner (1 strand amongst 9), and the other 8 strands were taught in the traditional manner by 8 different teachers. Furthermore, the number of students from one strand to another may vary based on the number of student registrations in different engineering streams (for example, Mechanical, Electrical Engineering, etc.). The Linear Algebra course is a required course for all engineering and natural

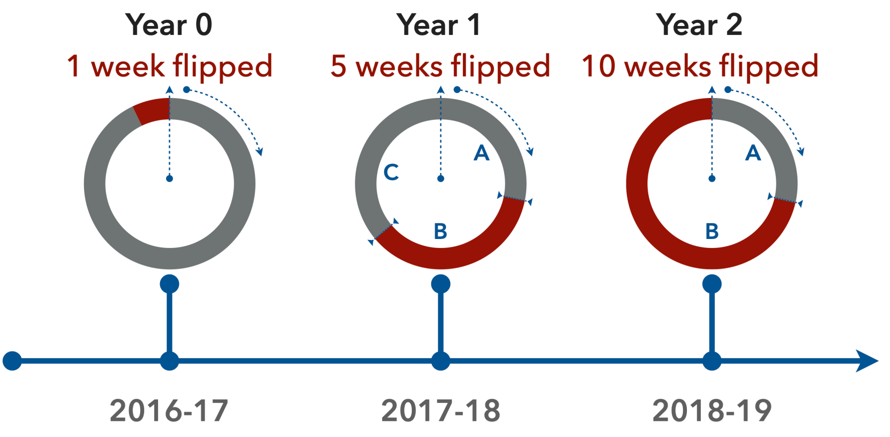


FIGURE 1 Incremental design of our flipped classroom study: The study officially started in the Autumn semester of 2017-18 (referred to as Year 1), but in the academic year 2016-17, one week towards the end of the semester was flipped (referred to as Year 0). Since Year 1, the number of weeks which have been taught in the flipped manner has changed incrementally. In Year 1, only 5 weeks were taught in the flipped format, whereas in Year 2, 10 weeks were flipped.

science programmes in the University. It is weighted at 6 ECTS credits (ECTS stands for European Credit Transfer System - the credit weighting system used throughout the European Universities region), which corresponds to a total of approximately 180 hours of work over the whole semester including both in-class and independent study time.

The comparison of the flipped format with students in multiple other strands taught by 8 other teachers reduced the possibility that what we were measuring was simply a 'teacher' effect. This was further verified by comparing the attainment of students in the experimental teacher's strand with students more generally in the years prior to the beginning of the flipped experiment.

Our experiment therefore followed a **replicated- between-subjects** design, where the students in the flipped strand learned the same topics and concepts weekly as their peers in other strands ('Control Condition'). Our experiment can be referred to as a 'modified replication' (see the definition from APA dictionary in Section 1) because although the duration of flipping the course varied from one academic year to next (Figure 1), the same teacher taught the course in the flipped format. Finally, all the students –in both the flipped and the traditional conditions– took a common, end-semester exam. This end-semester exam is composed of Multiple-Choice Questions (MCQs) which are 80% common among the different Linear Algebra strands (the remaining 20% of questions may differ from one strand to another).

3.1 | Procedure

Since Fall 2017, one strand of the Linear Algebra course has been taught in a flipped format. The course design of the flipped strand has been carefully planned by the professor working with pedagogical advisors, in order to ensure an evidence-informed approach for designing class activities. The implementation of the flipped classroom was carried out in an incremental manner as illustrated in Figure 1, and described below:

Year 0: In the Autumn semester of the 2016-17 academic year, the course was taught in the flipped manner for 1 week (the last or *fourteenth* week of the semester as shown in Figure 1). This was a pilot phase of the experiment, which was designed to *a*) inform the professor about the design of in-class activities and the ways of adapting their pace to match that of students', and *b*) elicit early feedback from students' about the teaching methodology and their perceptions regarding such experimentation in subsequent years.

Year 1: In the Autumn semester of the 2017-18 academic year, the course was taught in the flipped manner for 5 weeks (Part 'B' in Figure 1) – from *fifth* till the *ninth* week of the semester. The first 4 (Part 'A' in Figure 1) and the final 5 weeks (Part 'C' in Figure 1) were taught in traditional instructional format similar to other strands which were taught by different professors.

Year 2: In the Autumn semester of the 2018-19 academic year, the course was taught in the flipped manner for 10 weeks (Part 'B' in Figure 1) – from *fifth* till the end of semester. The first 4 weeks (Part 'A' in Figure 1) were taught in a traditional instructional format similar to other strands of the Linear Algebra course, and resembled the first four weeks of Year 1.

The rationale for incrementally increasing the duration of the flipped format (not including Year 0) was to *1*) pragmatically assess the impact of the flipped model in an ecologically-valid manner within what is an extremely high stakes course for the students taught in large classrooms, while mitigating the potential negative effects of such experimentation on students' learning, and *2*) enable the professor to scale the dissemination of pre-class materials (video lectures and suggested readings) and the organization of in-class activities, while making sure that the course remains aligned in terms of curriculum with the other parallel strands.

The flipped course was offered only to volunteering students. The professor informed the newly enrolled students about the possibility of participating in a flipped course by email 6 weeks before the start of the semester. The nature of the class and the experiment was explained as well as the expectations from volunteers. Table 1 illustrates the total number of students who volunteered to participate in the flipped course during its various recurrences across the two academic years. The number of students who were assigned to the experimental (flipped) condition were different from Year 1 to Year 2 (approximately 100 in Year 1 and 200 in Year 2). This was an intentional choice of the concerned professor, and was done so as to facilitate the management of the first instance of the flipped course.

Two weeks prior to the start of the semester, the professor and a pedagogical advisor collectively assigned the volunteers into either the **experimental** and the **control** group. Stratified Random Sampling was used (see Table 1 for the group sizes corresponding to different course years). The strata were defined based on **gender** and the **prior background** (secondary educational level) of students (as illustrated in Section 3.2). As a result, the proportionality of students' gender and their prior educational background was preserved within the experimental and the control groups. The students assigned to the experimental group (also referred to as the *flipped* condition) participated in the flipped Linear Algebra class during the Autumn semester. The students in the control group, attended the course taught by other professors in different strands in the traditional manner.

Course Year	Volunteers	Control Condition	Experimental (Flipped) Condition
Year 1	519	410	109
Year 2	373	171	202

TABLE 1 **Study participants:** The table illustrates the number of students who volunteered to participate in the flipped Linear Algebra course across its different replications (corresponding to the different academic year) as illustrated in Figure 1. In addition, the table also shows the assignment of students –after Stratified Random Sampling– in either Control (traditional course) or Experimental (flipped course) condition.

For the sake of clarity and presentation, in this article, we will present results from **Year 1** and **Year 2** only, **Year 0** being a pilot phase. It is worth noting that since each replication involved different populations of students with different assessments (exams with different sets of questions), consequently, **Year 1** and **Year 2** should be considered two separate experiments.

3.2 | Participants

The first-year bachelor population in our school is composed of both 'local' and 'international' students with varied high-school diplomas from a number of national education systems. As there is no selective entrance exam for 'local' students (i.e., those who are resident in the country in which the University is located), students' prior knowledge in different subjects –particularly physics and advanced (reinforced) mathematics– varies significantly.

3.2.1 | Identifying participants' background

Since we hypothesize that the flipped format may impact the students' cohorts differently, we classified the incoming bachelor students into *three* distinct categories based on their background as illustrated below.

International PAM (or INT-PAM): This category corresponds to the *international* students, who have completed a high-school diploma which included a strong component of Physics and Applied Mathematics (PAM). All the international students are subjected to a selection process, and so they all have a background in Physics and Applied Mathematics and were all high-performing within their respective high school system.

National PAM (or NAT-PAM): Students from the 'local' secondary education system are not subject to a selection process and so arrive with a diverse set of subject specialisms. This category (NAT-PAM), thus, corresponds to locally-resident students who studied Physics and Applied Mathematics (PAM) as their specialisation during high-school.

National Others (or NAT-OTH): Finally, this category corresponds to the local students whose high-school specialisation was in a subject other than Physics and Applied Mathematics (such as Philosophy, Economics, Biology, etc.). These students are also heterogeneous within this category, in that some of them may have followed advanced Mathematics courses while others only had basic Mathematics courses.

Furthermore, given the challenging nature of the first year of study, a reasonably large proportion of students are attempting the first year for a second time. We therefore also distinguished between **new** and **repeating** students.

3.2.2 | Cleaning participants' data

The initial data of volunteers from both Year 1 and Year 2 was cleaned and some participants were removed before we analyzed the data. The following steps elaborate how the data was cleaned:

1. The volunteering students who were minors (<18 years of age) at the time of data collection were excluded from our analyses in line with our ethics and data management approval. Consequently, 27 minors were removed from our dataset.
2. The volunteering students who were absent in the end of semester exam were also removed from the initial list of volunteers. 43 students were amongst the list of absentees.
3. Some of the volunteering students who were assigned to the *flipped* condition withdrew from the experiment

Course Year	Background	Total	Gender: Females	Gender: Males	Condition: Control	Condition: Flipped
Year 1	INT-PAM	207	63	144	168	39
Year 1	NAT-PAM	74	16	58	59	15
Year 1	NAT-OTH	70	36	34	54	16
	Sum (Year 1)	**351**	**115**	**236**	**281**	**70**
Year 2	INT-PAM	103	38	65	53	50
Year 2	NAT-PAM	48	12	36	24	24
Year 2	NAT-OTH	45	20	25	20	25
	Sum (Year 2)	**196**	**70**	**126**	**97**	**99**
Year 1 + 2	INT-PAM	310	101	209	221	89
Year 1 + 2	NAT-PAM	122	28	94	83	39
Year 1 + 2	NAT-OTH	115	56	59	74	41
	Total	**547**	**185**	**362**	**378**	**169**

TABLE 2 Study participants by background, gender and condition: This table summarizes the distribution of participating students over the different replications based on their background, gender, and condition. As compared to Table 1, this Table only shows the students' data which was subjected to filtering based on the steps elaborated in Section 3.2.2.

and consequently their data was filtered out, which represents 30 students.

4. Finally, the repeating students were filtered out (152 students). Owing to the fact that the repeating students have already finished their first semester once, their repeated exposure to the subject material may add a bias to our findings. As a result, only **new** students were part of the sample which was subjected to further analyses.

Table 2 summarizes the distribution of new volunteering students whose data was subjected to the analysis phase following the aforementioned filtering process.

3.3 | Materials and Instruments

The end-semester exam has a "common" part which includes 80% of the exam questions, in Multiple Choice Question (MCQ) format, and is designed collectively by several teachers who teach different strands of the Linear Algebra course. In particular, four teachers (amongst nine) collectively prepare the set of questions and their respective answers, which are then checked and validated by the other teachers. The remaining 20% questions are separately designed by each teacher for their respective strands and are based on the additional content taught by each individual teacher. In this paper we use the *common* exam score as a measure for students' academic performance, as this part of the exam was identical for control and flipped groups. Below, we illustrate the steps we took to compute this dependent variable.

3.3.1 | Computing normalized scores

Questions were negatively marked (+3 for a correct response, -1 for an incorrect response, and 0 for leaving the question unanswered). In order to ensure the validity of the measure, the following procedure was applied:

1. We removed questions which were designed to examine students on the themes/topics taught during the initial four weeks of the course (non-flipped part) of both Year 1 and Year 2 (part '**A**' in Section 3.1 and Figure 1). Although part '**C**' in 2017-18 was not flipped as such, it was felt that the problem solving methods addressed in part '**B**' would impact on students' learning in part '**C**'. Hence, questions from part '**C**' in 2017-18 were retained in the analysis.
2. We removed questions which did not effectively distinguish between students (for example, too easy, too hard or confusing in some way). We computed a *Discriminatory Index (DI)* value for each question based on the methodology described by Carneson, Delpierre, and Masters (2016, Section 4.6.2, pp. 16–17), and the questions with a low DI value were removed. Upon analyzing the DI for all the questions from both Year 1 and Year 2, we removed all the questions whose DI was less than 0.33. Our choice of this threshold was backed by *two* criteria: *1)* minimize the number of questions which are filtered out, and *2)* the boundary between questions which were retained and which were filtered out should be crisp, implying that the the DIs of two questions should differ by at least 0.1 (10%). Three questions were removed in Year 1 and two were removed in Year 2.

Finally, we normalized the scores of each (volunteering) student against all the first semester Bachelor students who took Linear Algebra.

3.3.2 | Identifying students' prior levels of attainment in Mathematics

The diverse origin of students and the lack of an entrance exam meant that there was no homogeneous metrics for quantifying students' prior knowledge in Mathematics. We therefore analysed their high school diploma and grades and designated them as having either a *high* prior performance or a *low* prior performance, as follows:

1. The students who only had a grade for basic mathematics, or no grade recorded for mathematics were excluded from further analysis. It is also worth noting that this step considerably reduced the size of our experimental population as many volunteering students had a basic or no certified background in reinforced mathematics.
2. The remaining students were further reviewed separately depending on whether they were *international* (INT) or *local* (NAT) students. The curriculum of the students' high school program was reviewed to determine the extent to which it taught mathematical content relevant to Linear Algebra. Based on this analysis, a series of score thresholds were identified for partitioning respectively the INT and the NAT students into "*Low Performing* (Low)" and "*High Performing* (High)" categories (separate thresholds were defined for the INT and NAT). On a normalized scale of [0, 1], with 0 being the lowest grade, and 1 being the highest, the thresholds for partitioning the students were:
   (a) **NAT**: *Low:* [0, 0.75] *High:* (0.75, 1.00]
   (b) **INT**: *Low:* [0, 0.85] *High:* (0.85, 1.00]

   For validation purposes, a median split was also performed. This yielded similar results to the qualitative categorisation, which suggested that the qualitative analysis and categorisation was valid.

The analysis of students' academic attainment across the Control and the Flipped conditions, based on their prior levels of attainment in reinforced mathematics is presented in Section 5.3.

4 | TEACHING DESIGN

As described above in Section 3, Linear Algebra is a 6 ECTS course. It is given over a 14 week semester with a weekly schedule of 4 periods of 45 minutes of lectures and 2 periods of recitation or exercise sessions with the teaching assistants (TAs), each split on two non-consecutive days (2 periods of lecture and 1 period of recitation/exercise session each). In addition, students are also expected to spend about 6 hours per week on individual study. This schedule and the overall workload for the flipped strand of the course was kept identical to that of the the other strands: the number of contact hours has neither been reduced nor increased.

In the following, we elaborate the type of learning activities which were used for preparatory work before class, in-class during the scheduled contact hours, and after class in the flipped strand of the course. To facilitate the categorization of these activities, we use the terms used by Lo and Hew (2019) whenever possible.

4.1 | Preparatory work (pre-class activities)

Students received instructions regarding the preparatory work for the whole week on the Friday of the week before. An indicative duration for the different tasks as well as the deadline for which to complete them (i.e. Day 1 or Day 2 of the scheduled in-class time) were provided. The typical preparatory work included:

- a list of sections from a Linear Algebra MOOC by Prof. NAME_ANON, which included video lectures and online quizzes; and
- an exercise worksheet.

The students were asked to take notes while watching the video lectures, like in a traditional lecture. The online quizzes enabled students to self-assess their learning and were not formally graded (although they were scored in the MOOC platform).

The exercise worksheet was taken from the course material of previous years. Students were strongly encouraged to work on the exercises by themselves before class but they didn't have to submit them and the exercises were not graded.

4.2 | In-class activities

In-class time was divided into time with the teacher (twice 2 periods per week, corresponding to the lecture time for traditional strands of the course), and time with the TAs (twice 1 period per week).

During class time with the teacher, there were *three* types of activities: *(i)* Quizzes to start the session – most of the time True/False questions. The score of the class helped the teacher to identify the common conceptual problems at the start of the session, and also enabled the students to review the pre-class learning (this functioned as an interactive explanation session, in line the findings of Lo et al. (2017) and Lo and Hew (2019)); *(ii)* Short, problem-solving exercises – The students were shown one exercise at a time and had to solve it in a given time-frame. Furthermore, the students were either asked to work in small groups for the development of a better understanding, or the teacher managed a class solution with students' interventions. This decision (between group or class-level work) was made based on the class' performance in the exercise in question; and *(iii)* Structured problems or proof-type problems – The students were asked to solve the problem individually in a given time-frame and, during this time, the teacher interacted with students and gathered partial responses (of different steps, for example) to enable interaction and discussion at the level of the whole class.

Classroom Response Systems, (or 'Clickers'), were used to facilitate the collection of students' feedback during the in-class activities. In addition to gathering students' understanding of the discussed topic, the Clickers also enabled the teacher to adjust the pace of the class. For example, when only a few students responded to a certain question, the teacher provided them with some additional time to finish their exercise. Furthermore, the aforementioned in-class activities, particularly of type *(i)* and *(ii)* were designed in a way so that students could finish them within 1 and 10 minute time-frame respectively.

During the recitation/exercise sessions with the TAs, students either worked individually or in small groups, and benefited from one-to-one help by the TAs (roughly one TA for every 28 students). The students had the opportunity to complete the exercise sheet, as well as the leftover exercises from the in-class activities. In these sessions, the TAs usually did not present the solution of the exercises, but only assisted the students with difficulties (both conceptual or procedural). A detailed written solution of the all the exercises was provided at the end of the week.

4.3 | Follow up work (post-class activities)

After the scheduled class time, students were asked to review the course material and finish any remaining exercises (unfinished by them or not done within the class due to time constraints) followed by verifying their work against the detailed solution.

5 | RESULTS AND ANALYSES

5.1 | Assessing the impact of the flipped class approach

A first question to address is whether the flipped class approach had any impact on overall student attainment. Results are presented in Table 3 and Table 4.

Although the data cleaning process (described above in Section 3.2.2) was intended to improve the quality of data used in the analysis, we also wanted to ensure it did not actually introduce unforeseen bias. The cleaning process described removed a large proportion of our effective data set. In Year 1, the size of our control group went from 410 to 281 and our experimental group was reduced from 109 to 70. In Year 2, the size of our control group was reduced from 171 to 97, while our experimental group was reduced from 202 to 99 participants (most of these reductions were

Non-Normalized Data				
Year	**Condition**	**N**	**Mean**	**SD**
Year 1	Control	281	31.00	15.60
Year 1	Flipped	70	31.70	15.00
t(109.67) = -0.37, p=.71				
Year 2	Control	97	33.20	17.50
Year 2	Flipped	99	31.90	16.60
t(192.77) = 0.55, p=.58				

TABLE 3 **Scores at the final exam:** The table illustrates the mean score and standard deviation attainment score for students in the Flipped and Control conditions. The non-normalized scores are used to elaborate the effects of removing questions with very low Discriminatory Index. In addition, the difference in students' mean score is also indicated as **Welch's 2 Sample *t*-Tests** for both the course years (Year 1 and Year 2).

Normalized Score (Parts A, B, & C)					Normalized Score (Parts B & C)				
Year	**Condition**	**N**	**Mean**	**SD**	**Year**	**Condition**	**N**	**Mean**	**SD**
Year 1	Control	281	-0.12	0.98	Year 1	Control	281	-0.09	0.97
Year 1	Flipped	70	-0.07	0.94	Year 1	Flipped	70	-0.03	0.96
t(109.67) = -0.37, p=.71					**t(106.80) = -0.42, p=.67**				
Year 2	Control	97	-0.19	1.01	Year 2	Control	97	-0.20	1.00
Year 2	Flipped	99	-0.27	0.95	Year 2	Flipped	99	-0.21	0.97
t(192.77) = 0.55, p=.58					**t(193.49) = 0.09, p=.92**				

TABLE 4 Normalized scores after question filtering, with or without questions from Part 'A': The table illustrates the mean score and standard deviation for students in the Flipped and Control conditions after the removal of questions with lower Discriminatory Index and the normalization against all the first semester Bachelor students who took Linear Algebra. The left side of the table shows the normalized scores including the questions from part '**A**' (first 4 weeks of the Linear Algebra course, which was taught as traditional lectures) while the right side of the table presents the normalized scores without Part '**A**'. In addition, the difference in students' mean score is also indicated as **Welch's 2 Sample *t*-Tests** for both the course years (Year 1 and Year 2).

due to the removal of repeating students). Nonetheless, our analysis indicates that removal of data did not impact on the overall pattern of findings. Table 3 summarizes the analysis of differences in students across the two experimental conditions on the non-normalized data.

In Year 1, we observed that the students' scores in the Flipped and the Control condition did not differ significantly (Welch's 2 Sample *t*-Test: t(109.67) = -0.37, p=.71). Similar results were observed for Year 2, where the differences in students' scores were again not statistically significant (Welch's 2 Sample *t*-Test: t(192.77) = 0.55, p=.58). On the basis of this data it appears as if the flipped class had no evident effect on the overall attainment of students.

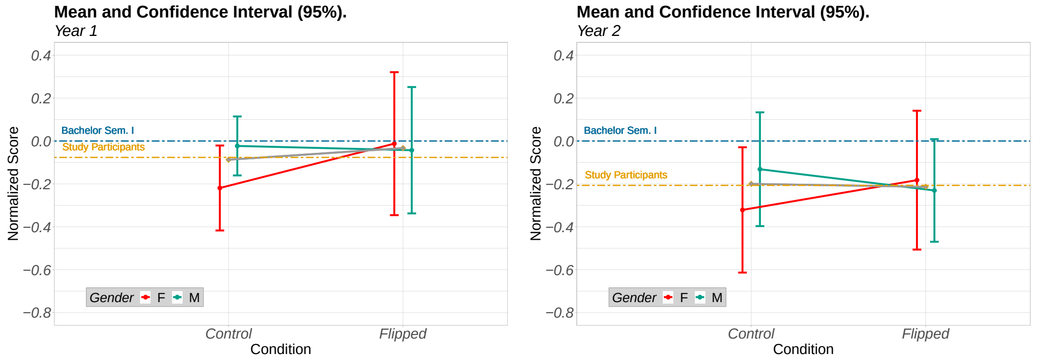
Since part of the exam in both Year 1 and Year 2 addressed material that was covered in the early 'non-flipped' weeks in the experimental setting, the data was also analysed with these questions removed. As before, the impact of removing these questions (part '**A**') from the analysis was also evaluated. The analysis is summarized in Table 4. This shows that the removal of these results do not change the overall findings: the flipped classroom format did not have any evident impact on the final attainment scores of students. Indeed, the removal of these scores do not affect the overall pattern of results to any notable extent.

5.2 | Inclusiveness of the flipped classroom model

Since our experimental population comprises different student cohorts with varying levels in prior mathematical knowledge (Reinforced Mathematics), we also wanted to explore if the flipped format had a different impact on these cohorts. In this section, we analyze the differential effects of the flipped format on different student groups.

5.2.1 | Differential effects of the flipped format across gender

Prior to undertaking this study, male students, on average, outperformed their female counterparts in the traditionally-taught Linear Algebra course. We hypothesized that this 'gender gap' would be reduced in the flipped condition. Therefore, as a first step, we analyzed the differential impact of the flipped format for male and female students.



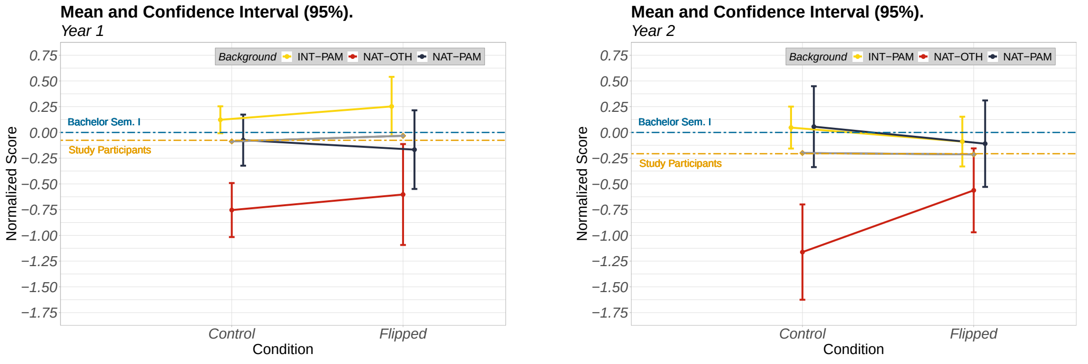
(a) Year 1 (b) Year 2

FIGURE 2 Gender differences in achievement across conditions: This figure shows the mean and confidence interval values for differences in normalized scores across Condition and Gender. The *blue* horizontal line at *y = 0.0* corresponds to the mean score in Linear Algebra for the whole Bachelor Semester I students. The *orange* horizontal line represents the mean score of all the **new** students who volunteered to participate in our experiment (repeaters were filtered out). Finally, the *gray* (solid) line represents the weighted mean across Condition.

In fact, in our data, the gender differences in students' scores was not significant (using a Kruskal-Wallis test – **Year 1**: χ^2(df=1)=2.13, p=.14, ε^2=0.006; **Year 2**: χ^2(df=1)=0.46, p=.49, ε^2=0.002). Nor were there significant differences between attainment of female students or male students in the Flipped and Control conditions. However, an examination of the data (see Figures 2a and 2b) within the study shows that female students on average, performed less well than male students in the Control condition, but this difference vanished in the flipped class (see Table 5 and Figure 2). This observation corresponds to a crucial pattern with regards to the performance of female students in the Flipped condition. Despite the lack of statistical significance, which can be attributed to the low size of the female population in our study, this pattern is worth considering. Furthermore the fact that a similar pattern emerges in both years is in itself notable. Provided that ours is a 'real-world' study, with a significantly smaller size of the female population as compared to male students, this repeating pattern for the Flipped condition emphasizes that the reduced gender gap must be taken into account despite the lack of statistical significance (discussed further in Section 6).

		Control			Flipped		
Year	Gender	N	Mean	SD	N	Mean	SD
Year 1	Females	92	-0.22	0.97	23	-0.01	0.82
Year 1	Males	189	-0.02	0.96	47	-0.04	1.03
Year 2	Females	35	-0.32	0.88	35	-0.18	0.98
Year 2	Males	62	-0.13	1.07	64	-0.23	0.98

TABLE 5 Gender differences in achievement across conditions: This table summarizes the mean score and standard deviation across Gender and Condition. The gender differences across Flipped and Control conditions are also illustrated graphically in Figure 2.



(a) Year 1 (b) Year 2

FIGURE 3 Background differences in achievement across conditions: This figure shows the mean and confidence interval values for differences in normalized scores across Condition and Background. The *blue* horizontal line at *y = 0.0* corresponds to the mean score in Linear Algebra for the whole Bachelor Semester I students. The *orange* horizontal line represents the mean score of all the **new** students who volunteered to participate in our experiment (repeaters were filtered out). Finally, the *Gray* (solid) line represents the weighted mean across Condition.

5.2.2 | Differential effects of the flipped format across prior mathematical background

In both Year 1 and Year 2, there were significant differences between student cohorts with different high-school backgrounds (Kruskal-Wallis Test – **Year 1**: χ^2(df=2)=39.28, p=2.95×10^{-9}, ε^2=0.112; **Year 2**: χ^2(df=2)=19.43, p=6.03×10^{-5}, ε^2=0.099) in the dataset as a whole. INT-PAM students and NAT-PAM students tended to outperform those with NAT-OTH backgrounds (see Table 6 and Figure 3).

Similar to our previous analysis of the gender gap, we examined these cohorts separately across the two conditions. We observed a decrease in the gap between scores of these *three* student groups in the Flipped condition as compared to the Control condition. We created subsets of our data based on the experimental conditions, and employed Kruskal-Wallis tests to assess the differences between the scores of INT-PAM, NAT-PAM, and NAT-OTH students. In Year 1, the difference in the scores of these student groups (see Figure 3a and Table 6) was found to be

		Control			Flipped		
Year	**Background**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Year 1	INT-PAM	168	0.12	0.87	39	0.25	0.92
Year 1	NAT-OTH	54	-0.75	0.98	16	-0.60	1.00
Year 1	NAT-PAM	59	-0.08	0.97	15	-0.17	0.76
Year 2	INT-PAM	53	0.05	0.76	50	-0.09	0.87
Year 2	NAT-OTH	20	-1.16	1.06	25	-0.56	1.04
Year 2	NAT-PAM	24	0.06	0.98	24	-0.11	1.05

TABLE 6 Background differences in achievement across conditions: This table summarizes the mean score and standard deviation across Background and Condition. The background differences across Flipped and Control conditions are also illustrated graphically in Figure 3.

statistically significant both in the Control (χ^2(df=2)=30.86, p=1.98×10^{-7}, ε^2=0.11) and the Flipped (χ^2(df=2)=9.97, p=.007, ε^2=0.14) conditions. Moreover in Year 2 (see Figure 3b and Table 6), we observed a statistically significant difference in scores of the three student groups in the Control condition (χ^2(df=2)=18.25, p=.0001, ε^2=0.19). However, this difference was not significant in the Flipped condition (χ^2(df=2)=4.07, p=.13, ε^2=0.04). It is worth noting that the gap in scores of the three student groups reduced in the Flipped condition for both Year 1 and Year 2. Furthermore, this reduction was more pronounced in Year 2. This finding may suggest that the flipped format is more conducive for classroom contexts with students from heterogeneous high-school backgrounds and varying levels of prior knowledge mathematics.

5.3 | Analysis of learning gains

In order to examine the influence of the flipped format on students' learning gain, we studied the effects of the experimental condition on students with varying prior knowledge in reinforced mathematics. Particularly, we examined the normalized scores of low- and high-performing students across conditions.

Table 7 and Figure 4 illustrate the mean attainment scores of students broken down by their prior level in reinforced mathematics. As expected, students with a strong-level in reinforced mathematics (High) attained significantly higher scores in the end-semester exams (Kruskal-Wallis Test: χ^2(df=1)=19.88, p=8.23×10^{-6}, ε^2=0.039).

Examining the differences in end-semester scores of High and Low prior knowledge across different conditions revealed that in the Control condition, there is a statistically significant difference in end-semester scores of students depending on their prior level in reinforced mathematics (Kruskal-Wallis Test: χ^2(df=1)=19.31, p=1.11×10^{-5}, ε^2=0.054). In the Flipped class however this difference is smaller and is not statistically significant (Kruskal-Wallis Test: χ^2(df=1)=1.90, p=.17, ε^2=0.013).

Finally, we examined if there were gender differences amongst students with High and Low prior-knowledge levels, and if these groups perform differently in the end-semester exam. We observed no gender differences in the Control condition, where the end-semester performances of female and male students in both Low and High prior-levels were equivalent. However, in the Flipped condition, the female students with Low prior-levels in high-school mathematics performed better than their male counterparts with Low prior-levels (Kruskal-Wallis Test: χ^2(df=1)=3.39, p=.06, ε^2=0.051).

		Control			Flipped		
Prior Maths Level	**Gender**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
High	Females	56	0.39	0.72	25	0.24	0.68
High	Males	123	0.36	0.85	59	0.22	0.84
High	**Sum**	**179**	**0.37**	**0.81**	**84**	**0.22**	**0.79**
Low	Females	65	-0.05	0.90	25	0.25	0.81
Low	Males	114	-0.07	1.02	42	-0.21	0.97
Low	**Sum**	**179**	**-0.06**	**0.97**	**67**	**-0.04**	**0.93**

TABLE 7 Differences in achievement taking into account prior maths levels: This table summarizes the mean score and standard deviation across Prior Maths Level, Gender, and Condition. The differences in students' prior level across their Gender and Condition are also illustrated graphically in Figure 4.

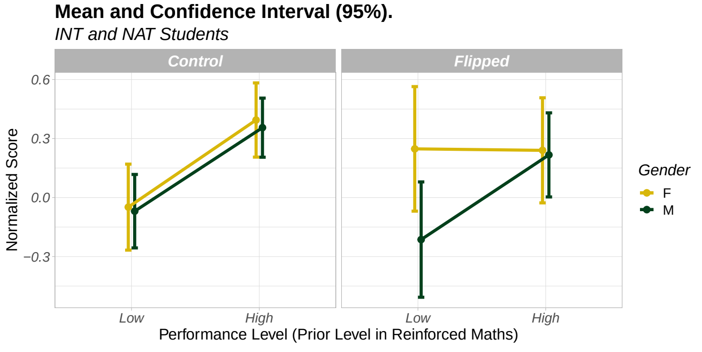


FIGURE 4 Differences in achievement taking into account prior maths levels: This plot illustrates the mean score and confidence interval for the scores of students across Gender, Prior Level in Maths, and Condition.

6 | DISCUSSION & CONCLUSION

There is very substantial interest in flipped class approaches in higher education and in scientific and engineering education, within the context of a broader enthusiasm about interactive approaches to teaching. While there is growing evidence that interactive teaching has a more positive effect on student learning and performance than traditional teaching in STEM disciplines (Freeman et al., 2014), weaker effects have been found in studies focusing specifically on flipped classroom approaches. Lo and Hew (2019) found an effect size of only *g = 0.29* in their work on flipped classes in engineering education while Cheng et al. (2019) found even weaker effects *(g = 0.19)*, with a weak positive effect for mathematics courses *(g = 0.21)* and a very weak negative effect for engineering courses *(g = -0.08)*. Hattie (2009) has identified that, for educational interventions, effect sizes of less than *0.40* should be regarded as indicating a low effect. Despite the fact that our replication has been modified from one year to another in terms of intervention time, our data is consistent with prior findings in that it shows no effect on average attainment from the flipped class – the grades of the control and experimental groups were effectively the same in both Year 1 and Year 2 of the study (see Table 3).

This finding is not surprising, given *a*) the apparently weak impact of flipped classes in general, and *b*) the short timeframe of the intervention (the intervention took place in one-third of one semester in Year 1 and two-thirds of one semester in Year 2). This finding may also be explained in part by virtue of the nature of assessment: weaker effects have been found for studies on interactive teaching in which studies are assessed using exams as compared to studies which measured impacts using concept tests (Freeman et al., 2014, p. 8412). It may be that interactive teaching is therefore more relevant when the focus is on application of concepts to physical scenarios and less relevant when the focus is on mathematical thinking and proofs. Finally class size may also be an issue. We had 70 participants in Year 1 and 99 in Year 2. Most existing studies on flipped classes are with classes in the 20-50 student size range.

There has been a growing interest in so-called null results (such as ours) in education studies in recent times, given that null results appear to be so common when moving from the 'efficacy' studies in highly controlled labs to more ecologically valid field-based randomised controlled 'effectiveness' trials (Kim, 2019). Kim (2019) notes that in one review of effectiveness trials designed to evaluate a previously validated educational intervention, only 11 of 90 trials yielded positive results (2019, p. 600). Jacob, Doolittle, Kemple, and Somers (2019, p. 580) argue that, rather than seeing these null results as an indication that something doesn't work, when designed and interpreted appropriately, null results have the potential to yield valuable information. In particular, Jacob et al. (2019) note that interventions which do not show a significant positive impact may still be worthwhile if the intervention is desirable for some other reason. In this case, the flipped classroom format is very popular with the students as evidenced by the fact that the flipped class experiment has been considerably oversubscribed each year in which it has been run. While there is considerable work involved in switching from a traditional to a flipped class approach, materials, once developed, can be reused. Discounting the initial costs over time in this way suggests that implementation of flipped class teaching may well be regarded as cost effective (Lo & Hew, 2019, p. 536). Indeed, taking into account that learning to give traditional lectures does, in itself, involve a steep learning curve, our experience suggests that the effort involved in becoming proficient in flipped class teaching is probably no greater than the effort involved in becoming proficient in traditional teaching. If so, and given that many students are looking for this kind of alternative to traditional teaching, if appropriate training can be offered to new faculty, and given that there is no real evidence of negative impact, it would seem strange not to offer this option to students.

Replication trials, even those that show little overall impact, can also shed light on previously unidentified interactions. One such interaction that is worthy of attention is the relationship between the impact of flipped classrooms and being a 'non-traditional' student. While we know that there is some evidence that interactive teaching can reduce the so-called 'gender gap' in science education (Haak, HilleRisLambers, Pitre, & Freeman, 2011; Lorenzo et al., 2006), this issue has not been well addressed in existing reviews on flipped classrooms in engineering settings (for example, Lo and Hew (2019)), in STEM education more generally (Lo et al., 2017), or in higher education (O'Flaherty & Phillips, 2015). The question of the differential impact of interactive teaching on under-represented minority students as well as low income students has been addressed (Theobald et al., 2020), however, this review found that the data was not sufficient to examine the problem of gender differences. The data presented here suggests that it is possible that flipped classrooms may have more positive impacts on the learning of some 'non-traditional' engineering students than on that of 'traditional' students. While female students performed a little worse than male students in each of the two years in our control group, they performed almost identically to male students in the flipped class in both years. While the differences between male and female attainment were not statistically significant, the fact that the same pattern emerged in both years was notable.

A similar –stronger– pattern emerged when differences in students' prior education are considered. While, in the control group, there are significant differences in attainment in the course between those students who have studied technical disciplines in high school (INT-PAM and NAT-PAM students in our sample) and those who have not taken a scientific strand in high school (NAT-OTH students in our sample), these differences were reduced and became non-significant in the flipped class group (see Figure 3 and Table 6). This pattern is even clearer when one looks at the experience of female students who enter with comparatively low high school mathematics grades. In the control group, both male and female students who come in with lower high school grades in mathematics tend to have a similarly weak performance in their end of semester exam. In the flipped class, this pattern does not hold true for female students; women who come into the flipped class with lower high school mathematics grades tend to do as well in the flipped class as both men and women who come in with stronger high school mathematics grades in both the flipped and control classes. While the difference between control and flipped settings is marginally non-significant *($p = 0.06$)*, the pattern is quite notable.

Cheng et al. (2019) note that many of the existing studies of flipped classrooms are of questionable design and that quite a few do not provide adequate information about the study design to allow them to be effectively used to draw conclusions. Our aim in this study was to provide a clear account of both the research design and the instructional design to allow others to draw conclusions from our data. Our study explores what happens when flipped class approaches are used in a real teaching and learning setting in engineering education, with a high stakes course, addressing complex technical content. Such real life contexts can be messy, with students drawn from a variety of backgrounds and trajectories. These kind of ecologically-valid studies also have many potentially intervening variables including students' prior knowledge, self-efficacy beliefs, and motivations as well as teacher's skill and behaviour. This study was designed so that it meets the criteria for high quality studies already in use in the field, specifically, comparability between control and experimental groups in terms of assessment, students and instructors (see, Freeman et al. (2014, p. 8414)). Nonetheless, this study does have its limitations: classroom heterogeneity and the need to control for students' prior educational trajectory meant that even though we started with quite large numbers in both control and experimental groups, we were often left with rather fewer students in the final analysis. We would contend that this is a reasonable outcome given the desire to achieve both internal and ecological validity in our study design. Nonetheless, it should be recognised as a limitation and one which may well have impacted on our ability to identify statistically significant findings from the data. Furthermore, due to the incremental way in which the flipped class approaches were introduced over time, the potential impact of the intervention on student learning may well have been lessened. While we would contend that this approach was a realistic way of implementing a pedagogical change in the context of a high stakes and challenging course, it should be recognised that a more systematic and consistent use of flipped class approaches may well have a deeper impact on students' learning.

Our study also suggests some future trajectories of further research. It is notable that flipped class strategies seems to have, on average, a less positive impact on learning than the use of other types of interactive strategies. Lo and Hew (2019) also suggest that some 'explanation' component in the flipped class approach seems to have a positive impact on attainment (there was a 'flipped' explanation component in the design of the teaching in this study). Rather than simply comparing 'flipped' with 'traditional' courses, future research may wish to focus on whether different approaches to flipping classes may have different impacts. Given the significant challenges facing engineering education in attracting and retaining 'non-traditional' students (Aeby et al. (2019); Lichtenstein et al. (2015)) and given that this issue has been largely neglected by existing research on flipped classes (see Lo and Hew (2019)) we suggest this should be a priority for future research.

References

Aeby, P., Fong, R., Vukmirovic, M., Isaac, S., & Tormey, R. (2019). The impact of gender on engineering students' group work experiences. *International Journal Of Engineering Education*, *35*(ARTICLE), 756–765.

Atadero, R. A., Paguyo, C. H., Rambo-Hernandez, K. E., & Henderson, H. L. (2018). Building inclusive engineering identities: implications for changing engineering culture. *European Journal of Engineering Education*, *43*(3), 378–398.

Barnard, S., Hassan, T., Bagilhole, B., & Dainty, A. (2012). 'they're not girly girls': an exploration of quantitative and qualitative data on engineering and gender in higher education. *European Journal of Engineering Education*, *37*(2), 193–204.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448.

Carneson, J., Delpierre, G., & Masters, K. (2016). *Designing and managing multiple choice questions* (2nd. ed.). Cape Town, South Africa: University of Cape Town.

Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development*, *67*(4), 793–824.

Chiquito, M., Castedo, R., Santos, A. P., López, L. M., & Alarcón, C. (2020). Flipped classroom in engineering: The influence of gender. *Computer Applications in Engineering Education*, *28*(1), 80–89.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, *69*(9), 970–977.

Dang, Q.-V., & Gajski, D. D. (2014). Bringing in-class online-a hybrid solution. In *Fourth interdisciplinary engineering design education conference* (pp. 12–17).

DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational Psychology Review*, *29*(1), 141–151.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The physics teacher*, *40*(4), 206–209.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015, December). Increased Preclass Preparation Underlies Student Outcome Improvement in the Flipped Classroom. *CBE—Life Sciences Education*, *14*(4), ar36. Retrieved 2020-01-21, from https://www.lifescied.org/doi/full/10.1187/cbe.15-02-0040 doi: 10.1187/cbe.15-02-0040

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216.

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. *Tertiary assessment & higher education student outcomes: Policy, practice & research*, *259*, 275.

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M.-A. (2019). A framework for learning from null results. *Educational Researcher*, *48*(9), 580–589.

Kerr, B. (2015). The flipped classroom in engineering education: A survey of the research. In *2015 international conference on interactive collaborative learning (icl)* (pp. 815–818).

Kim, J. S. (2019). Making every study count: Learning from replication failure to improve intervention research. *Educational Researcher*, *48*(9), 599–607.

Lichtenstein, G., Chen, H. L., Smith, K. A., & Maldonado, T. A. (2015). Retention and persistence of women and minorities along the engineering pathway in the united states. In *Cambridge handbook of engineering education research* (pp. 311–334). Cambridge University Press.

Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education*, *108*(4), 523–546. Retrieved 2019-12-10, from https://onlinelibrary.wiley.com/doi/abs/10.1002/jee.20293 doi: 10.1002/jee.20293

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review*, *22*, 50–73.

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*(2), 118–122.

Marra, R. M., Rodgers, K. A., Shen, D., & Bogue, B. (2009). Women engineering students and self-efficacy: A multi-year, multi-institution study of women engineering student self-efficacy. *Journal of engineering education*, *98*(1), 27–38.

Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *IEEE transactions on education*, *56*(4), 430–435.

Mckenna, A. F., Froyd, J., & Litzinger, T. (2014). The complexities of transforming engineering higher education: Preparing for next steps. *Journal of Engineering Education*, *103*(2), 188.

Minin, S., Varodayan, D., Schmitz, C., Faulkner, B., San Choi, D., & Herman, G. L. (2016). Minority merit: Improving retention with cooperative learning in a first-year electronics course. In *2016 ieee frontiers in education conference (fie)* (pp. 1–5).

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125–1156.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … others (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593–10597.

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The internet and higher education*, *25*, 85–95.

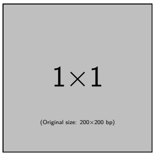
Powell, A., Dainty, A., & Bagilhole, B. (2012). Gender stereotypes among women engineering and technology students in the uk: lessons from career choice narratives. *European Journal of Engineering Education*, *37*(6), 541–556.

Schiltz, G., Feldman, G., & Vaterlaus, A. (2019). Active-learning settings and physics lectures: a performance analysis. In *Journal of physics: Conference series* (Vol. 1286, p. 012019).

Secules, S. (2017). *Beyond diversity as usual: Expanding critical cultural approaches to marginalization in engineering education* (Unpublished doctoral dissertation).

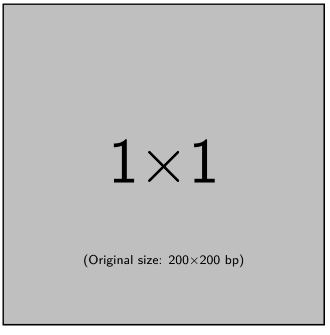
Seymour, E., & Hewitt, N. (1997). *Talking about leaving: Why undergraduates leave the sciences (pp. 115-116).* Boulder, CO: Westview Press.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., … others (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483.



A. One Please check with the journal's author guidelines whether author biographies are required. They are usually only included for review-type articles, and typically require photos and brief biographies (up to 75 words) for each author.

GRAPHICAL ABSTRACT



Please check the journal's author guildines for whether a graphical abstract, key points, new findings, or other items are required for display in the Table of Contents.