Flipping Laboratory Sessions in a Computer Science Course: An Experience Report

Javier Troya, José A. Parejo, Sergio Segura, Antonio Gámez-Díaz,
Alfonso E. Márquez-Chamorro, Adela del-Río-Ortega

The final published version of this paper is available from IEEE's website:
https://ieeexplore.ieee.org/document/9186349

***Abstract*—*Contribution:* This paper presents an experience report on the application of flipped classroom to the laboratory sessions (henceforth lab sessions) of an undergraduate computer science course.**

***Background:* Hands-on work in computer science lab sessions is typically preceded by technical instructions on how to install, configure and use the software and hardware tools needed during the lab. In the course under study, this initial explanation took between 14% and 50% of the lab time, reducing drastically the time available for actual practice. It was also observed that students missing any of the labs had trouble catching up.**

***Intended Outcomes:* The application of flipped classroom is expected to increase the time for hands-on activities, and improve students' performance and motivation. This improvement is expected to be more noticeable in those students who are unable to attend all lab sessions.**

***Application Design:* The study compares the application of flipped classroom and a traditional methodology. It encompasses two academic courses and involves 434 students and 6 lecturers.**

***Findings:* The flipped classroom is suitable for lab sessions in computer science. Among other results, flipping the labs resulted in 24 more minutes of practical and collaborative work on average at each lab session. It was observed a significant improvement in the motivation of students, with 9 out of every 10 students preferring it over traditional methodologies. Also, the flipped classroom made it much easier for students to catch up after missing a lab, making the final grades less dependent on lab attendance.**

***Index Terms*—Computer science, experience report, flipped classroom, inverted classroom, laboratory, undergraduate.**

I. INTRODUCTION

The rationale behind flipped classroom (FC), also known as flipped learning and inverted classroom, is for students to learn the theoretical concepts prior to attending class, so that face-to-face class time is mostly devoted to hands-on activities where students can have immediate feedback and assistance from instructors [1], [2], [3]. This approach has become increasingly popular in recent years [4], [5], [6]. In the majority of both classroom reports and research studies on flipping [5], [7], [8], students' preparatory work takes the form of viewing web-based video lectures [9].

Laboratory sessions play a key role in computer science courses. At the beginning of each lab, instructors typically devote time to provide instructions on how to install, configure and use the software and hardware tools needed during the lab. This initial explanation is essential, but it can significantly reduce the time for practical activities in the classroom (up to 50% in the context of the course reported in this article). Motivated by this problem, this paper reports the application of FC to the *laboratory sessions* of a computer science course.

Empirical studies [5], [8] and surveys [3], [7], [10] on FC generally conclude that the application of this approach is at least as effective as its traditional-based format in terms of student achievement [6]. However, whether these results hold in the context of laboratory sessions is to be investigated. In this line, Karabulut-Ilgu *et al.* [8] recently concluded that the scarcity of archived journal publications indicates that research on FC in engineering education is still in its infancy. Giannakos and Krogstie reached the same conclusion in the domain of computer science, as they believe there is the need to collect and triangulate different types of data from different sources [5]; for instance, issues referring to any potential effect on students' attitudes on their learning performance or attendance.

Most studies on FC in the field of computer science apply the methodology to the theoretical classes, and therefore neglect its potential benefits and limitations in the context of labs. Only a few papers report results on flipping the labs of computer sciences courses, although they are not the focus of their work and so they miss the potential gains and drawbacks. For example, Day and Foley [11] completely flipped a human-computer interaction course and compared it to a traditional methodology, although the comparison also included theory lectures. Cupak and Riabov [12], [13] applied FC in some courses on computer science, lab sessions included. However, they did not perform any in-depth analysis of the methodology, and the conclusions of the study were mainly drawn from the students' final grades exclusively. The process followed for the creation of the videos and the lessons learned from it were not described either. Regarding online videos, they are used as learning material in remote laboratories or Lab at Distance (LAD) in engineering [14], which focuses on distance learning applied to the laboratory part. However, this is different to FC, so conclusions obtained from the study [14] cannot be extrapolated to FC. Maher et al. flipped four different computer science classes across multiple semesters over two years [15]. The flipping included labs, although they were not the focus. Regarding videos, some were produced by the instructors and some others were provided after curating online videos. Courses were flipped along different semesters, and a comparison of the FC with respect to traditional methodologies was not performed. Finally, Latulipe et al. [16] applied flipped classroom to a course in the first year of computer science,

Manuscript received July XX, 2019; revised Month DD, 20YY.
The authors are with the Department of Computer Languages and Systems, Universidad de Sevilla, Spain. (*Corresponding author: Javier Troya, jtroya@us.es.*)

which also included lab sessions. They did perform a thorough analysis; however, FC was not the focus.

This paper reports on the results of flipping a computer science course on software architecture. Its main novelty lies in the application of FC to the laboratory sessions exclusively, studying the potential gains and limitations of the methodology in the context of computer science labs. Specifically, the reported study was conducted in two subsequent editions of the course, taking place in 2017 and 2018. They involved a total amount of 6 instructors and 434 students, placing this among the largest studies on FC to date (together with studies such as [6], [15], [17]). Furthermore, it is the first work that presents a thorough evaluation of the application of FC strictly centered on the lab sessions of a computer science course by comparing its application to a traditional methodology. The traditional methodology—where lab sessions were arranged as a face-to-face explanation by instructors followed by practical exercises—was applied the first year (2017), whereas the FC was followed the second year (2018). More specifically, the introduction to each lab was moved to videos made by the instructors that students were asked to watch before the lab. Instructors followed a lightweight approach for video production, trying to overcome the limitations found in previous studies, such as too-long videos [18], [19]. To this end, Guo et al.'s guidelines on video production were followed [20]. The research questions to be answered in this study are the following:

- *RQ1:* How do students perform with either methodology and how attendance to lab sessions influences their performance?
- *RQ2:* How is students' attitude towards the course with either methodology?
- *RQ3:* How do students assess the FC methodology?

II. Course Implementation

The study was performed in the compulsory course *Software Architecture and Integration*, taught in the second year of the software engineering degree at the University of Seville, Spain. The course is taught over 15 weeks and is divided into a theoretical and a practical part. The theoretical part is composed of 18 on-site lectures, whereas the practical one consists of 12 sessions in rooms equipped with personal computers—there are 8 lab sessions where new material is taught and 4 sessions dedicated to assessments. Both years, students were divided into 10 lab groups, so that they could fit in the lab classrooms and receive the appropriate individual attention from instructors when performing hands-on activities. All classes took 110 minutes, with one practical class per week.

The FC was applied to the practical part of the course (i.e., lab sessions), where students learnt to integrate software systems on the Web using Java and REST technologies.

A. Traditional Methodology

The first year, 2017, the course was composed of 225 students and 6 instructors. The laboratory part of the course was taught following a traditional methodology, where instructors used PowerPoint slides, the white board and the instructor's computer for explaining the lab content as well as how to install, configure and use the software needed during the lab. Students had to take notes of this. Depending on the lab and the instructor, the initial explanation took between 15 minutes (13.6% of the lab) and 55 minutes (50% of the lab), with an average of 33.2 minutes (30% of the lab). The remaining time was devoted to hands-on activities, namely practical exercises using classroom's computers or students' personal laptops. Students were allowed (and encouraged) to request instructors' help to solve any concern regarding the exercises.

B. Flipped Classroom

The second year, in which lab sessions where flipped, the course was composed of 209 students and the same 6 instructors. The initial face-to-face explanation was moved to videos. It was important to cope with the rapid obsolescence of the material to be included in the videos, since part of the material is often outdated in the following academic course, or even during the own course, due to changes in third-party tools and web services. As a result, videos showing outdated material must be updated accordingly. For this reason, instructors discarded contracting an external service for producing the videos. Instead, they applied a lightweight approach for video production following the guidelines proposed by Guo *et al.* in their empirical study on how video production affects student engagement [20]. In particular, videos were recorded taking into consideration the following recommendations: (i) videos should be short (indeed, some other studies have reported that long videos are not suited for FC [18], [19]), (ii) they should be recorded in informal settings, (iii) instructors should appear in the videos, (iv) videos should have good quality, (v) invest in pre-production. Each instructor prepared the videos of one or two labs in informal settings, namely their offices or homes. In total, 29 videos were produced, the shortest one with a duration of 2'34" and the longest one 9'45". The average length of the videos was 5'58", with 3.6 videos per lab on average. All videos included the instructor's talking head at the beginning of the video, and some of them also in middle parts. Some instructors' upper body was kept during the whole video in the right-down corner. All videos were posted on an unlisted playlist on YouTube (accessible on https://bit.ly/2JMFbVB) that no one outside the course could access, which would have jeopardized the evaluation of the FC.

Students were requested to watch the corresponding videos prior to lab sessions. Those who had not watched them were requested to do so first thing in the lab session (using their own headphones). This meant that these students would have less time than their classmates for doing practical exercises—this served as a powerful incentive to watch the videos before attending class. The sessions were opened with a 6-question quiz using the game-based platform *Kahoot!* in order to check if the videos had been properly understood. The questions and possible answers were displayed on the projector and students answered them using either their computers or cell phones with the *Kahoot!* app. The correct answer and the answers given by students for each question were displayed on the projector. Students got engaged by obtaining reward points based on the number of correct answers and their response times. At the end of each quiz, students were displayed in a ranking ordered by their points, making the quiz a fun and competitive task—yet another incentive to watch the videos before class. Any doubt or misunderstanding identified through the game-based quiz was briefly clarified by the instructor. The completion of the Kahoot quizzes plus the required clarifications took from 6 minutes (5.5% of the lab) to 20 minutes (18.2% of the lab), with an average of 9.1 minutes (8.3% of the lab). The remaining time was devoted to the very same hands-on activities as in the previous year, permitting to perform a fair analysis of the FC.

C. Students' Assessment

Students were identically assessed the two years of the study. The theoretical and practical parts had a weight of 40% and 60% of the final mark, respectively. This study is focused on the practical part, where students were arranged at the beginning of the course in teams of between two and four members in order to undertake the course's project, which consisted in implementing a web-based mashup using Java and REST technologies. The mashup had to meet several requirements, such as the integration of at least three real applications, and the mashup's topic could be chosen freely by each team. The requirements of the project were exactly the same in the two years of our study, so that the projects had the same level of difficulty. The project submission was arranged in three incremental deliverables, which were properly assessed and graded by instructors. Although the project was developed in teams, the final grade was individual since typically not all students contribute equally to the project and put the same effort. To this end, instructors had meetings with every group and asked each of its members to perform some modifications over the submitted project, in order to discover the degree up to which students had been working on the project. Additionally, students had the possibility of undertaking two optional on-site programming tests during the course, which could increase their final grade in the practical part up to 1 point, being 10 points the maximum achievable grade.

III. Data Collected

A. Lecture Duration

Instructors registered the time spent in the initial explanation (first year) and in the Kahoot quiz plus clarifications (second year). In total, 64 and 72 times were registered in 2017 and 2018, respectively. Not all instructors remembered to write down the times in every session, this is why there were not 80 times registered (there were 10 groups and 8 lab sessions). Despite this, a sufficiently large sample for statistical purposes was collected (the data excluded were below 20%).

B. Kahoot Quizzes

Students were asked to enter their university identifier when joining a Kahoot quiz, which allowed students' performance to be traced (cf. Section IV-B). When a quiz was finished, all answers were saved in csv format for the a-posteriori analysis. Answers registered with an invalid identifier, i.e., those not matching a student's registered identifier, were discarded. In total, 1027 answers were collected with a valid identifier, out of a total of 1163.

C. Students Questionnaires

At the end of each lab session during both years students were asked to fill in a questionnaire to express their perception of the lab. The questionnaire was almost identical during the two editions of the course, with slight changes the year the laboratories were flipped in order to include students' opinions concerning the videos. Questions related to students' perceptions were encoded using a Likert scale. The only exception was a question about the technical quality of the videos, which was encoded on a numerical scale from 0 to 10, so that students were able to evaluate the videos quality in an effective and accurate way. Students were asked to optionally write their names in the questionnaires, which would allow, again, to keep trace of their performance. A total amount of 2340 questionnaires were collected, 1166 (575 with students' names) in 2017 and 1174 (684 with students' names) in 2018.

D. Assessment Grades

The practical part of the course was assessed as explained in Section II-C. There are some students who gave up the course both years. In total, 185 final grades were collected in 2017, which constitutes an 82% of the 225 students taking the course that year. In 2018, 163 final grades were collected, which constitutes a 78% of the 209 students taking the course that year. These grades were calculated from the individual mark obtained from the project assessment plus the two optional programming tests.

E. Survey About Flipped Classroom

At the end of the FC experience, students were asked to complete a survey before they knew their final grade, so that this would not influence their answers. Some questions were answered using a Likert scale, while for others a free-text answer was required. Responses to most of these questions are detailed in Section IV-D. 144 surveys were collected.

The questionnaires and surveys for data collection described in this section are available on [21] (English versions are available, originals were in Spanish).

IV. Evaluation

This section first describes the statistical techniques that have been applied. Then, it presents the results of the evaluation according to the three RQs formulated in the introduction.

A. Statistical Techniques Applied

First, since data from different sources were collected, relationships between variables were studied using linear correlation. In particular, *Pearson's product-moment correlation coefficient* was computed [22], whose value can range from -1 to $+1$. The larger the absolute value of the coefficient is, the stronger the relationship between the variables will be. When the coefficient is of special interest or its value is not negligible, a *p-value* [23] on the existence of such correlation is computed. Such p-value determines whether the correlation coefficient is significantly different from 0 or not. If the p-value is less than or equal to 0.05, then the correlation is different from 0 and actually exists. Additionally,

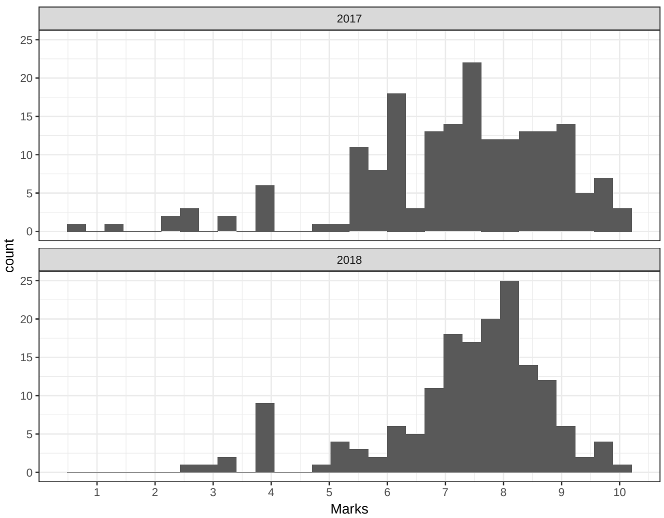

Fig. 1. Histograms of students' final grades per academic course

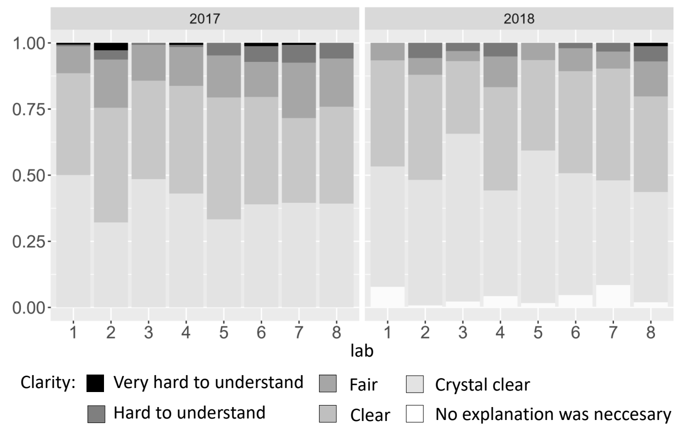

Fig. 2. Clarity of explanations in each lab session

the *Spearman's rho non-linear correlation coefficient* [24] was computed for all the relationships studied, and its values can be found at the on-line laboratory package of the paper [21]. The Spearman's correlation between two variables is equal to Pearson's correlation; however, whereas Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships. Spearman's correlation was applied in order to identify any correlation which was not identified by Pearson's correlation.

Second, the *Cronbach's alpha (α) coefficient* [25] was computed for assessing the reliability or internal consistency of the responses to the survey about FC. Cronbach's α coefficient of reliability ranges from 0 to 1. Higher values indicate higher number of answers having a shared covariance on the question, meaning they measure better the underlying concept. Typically, α coefficients smaller than 0.5 are considered unacceptable. The value obtained from the surveys about FC was 0.88, meaning responses were not arbitrary and questions were not misinterpreted by students.

Third, the differences on surveys' and questionnaires' answers were encoded using and ordinal scale such as "bad", "fair", "good", where the exact distance between "bad" and "fair" and between "fair" and "good" is unknown, i.e., only a partial ordering function on the domain is defined. For the analysis of such answers, a *cumulative link model with a logit link function* [26], also known as *proportional-odds model*, was generated, and *Wald-based p-values* for tests of the parameters being zero [27] were used. This analysis permits to identify if the changes in the methodology have a statistically significant impact on the answers of students even when measured using ordinal scales.

Finally, for the differences among grades, the *Wilcoxon's rank sum test with continuity correction* [28] was applied, since the samples do not follow a normal distribution—*Shapiro-Wilk tests* were applied to evaluate normality.

B. RQ1 - Students' Performance with Both Methodologies

The results after applying Wilcoxon's test on students' final grades in both academic courses—since the data did not completely follow a normal distribution, although the variance was similar—indicate there is no evidence as to statistically assert that the methodology applied influenced students' grades, neither positively nor negatively (*p-value*=0.49). This goes in accordance with the conclusions drawn by related works [8], [29], [30], [31], [32]. However, some noticeable differences can be observed with a deeper scrutiny of grades' distributions. Figure 1 shows the histograms of students' final grades obtained with each methodology aligned vertically.

On the one hand, the distribution of final grades in the first year (2017) shows a wide and sparse distribution for the students who failed the course (marks between 0 and 4.9), and two prominent peaks around 5.8 and 7.5 for students who passed. There is also a long tail towards the maximum. On the other hand, the distribution of final grades in the second year (2018) shows a higher concentration of students around 4.0 for students who failed the course, and a single peak with a clear normal distribution around 8.0 for students who passed. Another noticeable change is that the number of students with a grade above 8.7 is smaller than in the previous year, i.e. the long tail of higher grades disappeared.

In order to analyze how students' attendance to lab sessions influenced their grades with either methodology, the authors analyzed students questionnaires and their final grades—only students who attended the sessions filled out the questionnaire. The first year there was a positive correlation (*r*=0.31, *p-value*=$1.2e^{-03}$), meaning those students who more often attended lab sessions generally obtained better grades. However, this correlation disappeared the second year. The first year there was also a correlation between attendance to lab sessions and the number of exercises finished per lab session (*r*=0.23, *p-value*=$1.89e^{-02}$), which disappeared again the second year.

An analysis on the time inside the classroom revealed that students had on average 24 more minutes for hands-on exercises in 2018—this increase was caused by the change of methodology as confirmed by applying a two-way ANOVA with factors lab (8 levels) and year (2 levels). For this reason, the percentage of students who finished all exercises in the lab sessions increased from 22.4% in 2017 to 35.9% in 2018. Similarly, the percentage of students who completed at least 75% of the exercises increased from 39.1% to 60.3%, and the

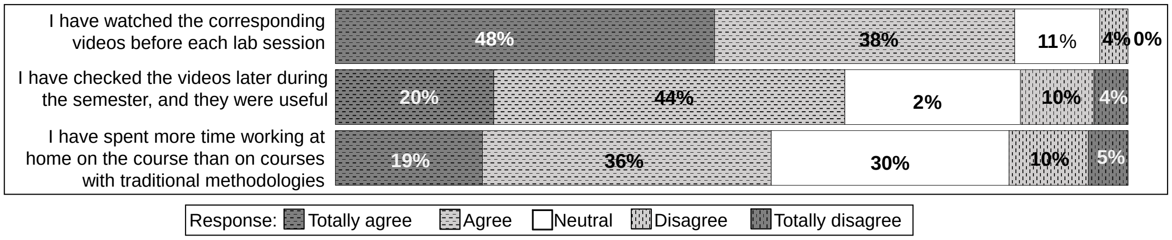

Fig. 3. Students' attitude towards flipped classroom

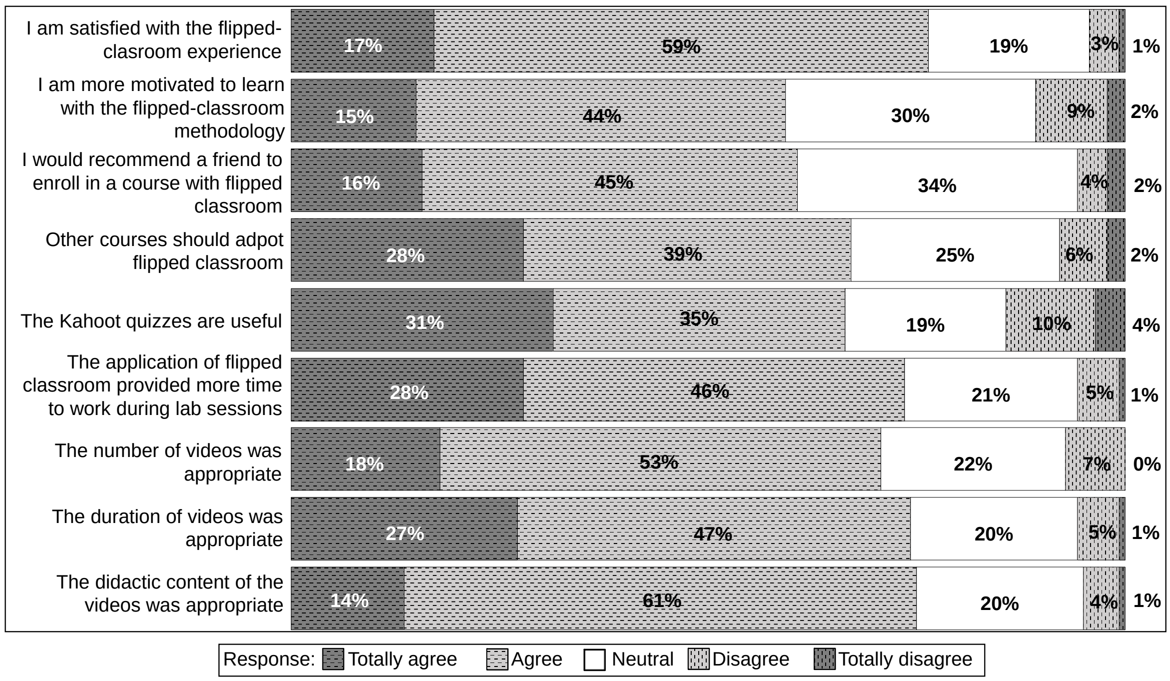

Fig. 4. Students' assessment of the flipped classroom

percentage of those who completed at least half the exercises increased from 61.2% to 85.2%. Finally, the second year there was a positive correlation (r=0.28, *p-values*<0.01) between students' final grades and the number of correct answers and reward points in the Kahoot quizzes.

C. RQ2 - Students' Attitude towards the Course

An analysis of the students questionnaires after each lab session (cf. Section III-C) revealed that 11% more students (from 31% to 42%) considered lab sessions interesting in 2018 with respect to 2017. Regarding instructors' explanations—in 2017 this refers to in-class explanations, whereas in 2018 to the videos together with in-class clarifications—, students' impressions are displayed in Fig. 2. It is noticeable that in 2018 the opinions that considered these explanations hard or very hard to understand decreased (from 5% to 3%), and those that considered them crystal clear increased (from 41% to 49%). After applying a cumulative link model with a logit link function (since the variable is ordinal) [26] on the answers from the surveys, *p-value* resulted as $5.47e^{-12}$, meaning there is a clear and highly significant difference between the clarity perceived on instructors' explanations in favor of the course where lab sessions were flipped.

In 2017, instructors used slides for their in-class explanations. The slides were available before class. In 2018, slides were replaced by videos, also available before class. Both slides and videos showed the same kind of content, although slides needed to be complemented with in-class instructors' explanations. In respect of the percentage of students who did some work prior to lab sessions—reading slides in 2017 and watching videos in 2018—, it increased from 32% to 83% according to students questionnaires. This result is reinforced with the answers students gave in some of the questions of the survey about FC (cf. Section III-E), displayed in Fig. 3. These reveal that 55% of students affirmed they spent more time working at home when the FC methodology was applied. Furthermore, 86% of students stated they watched (38%) or definitely watched (48%) the videos before attending class, and 64% of students indicated they watched (44%) or definitely watched (20%) the videos at any other point in time during the semester, which was useful to them.

D. RQ3 - Students' Assessment of Flipped Classroom

In the survey about FC, 87% of students (122 out of 144) indicated they prefer the FC methodology over traditional

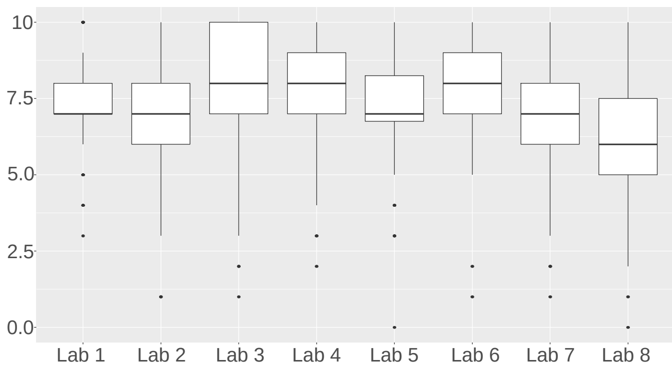

Fig. 5. Video technical quality assessment per lab session

methodologies. Regarding questions answered using a Likert scale—*Totally agree, Agree, Neither agree nor disagree, Disagree* and *Totally disagree*—, replies related to students' assessment of the FC are grouped in Fig. 4. They are in general very much in favor of the FC, as described in the following. First, 76% of students were satisfied (59%) or very satisfied (17%) with the FC methodology, whereas only a 4% were not. Besides, around 59% of students felt more (44%) or much more (15%) motivated to learn with this methodology and 61% would recommend (45%) or definitely recommend (16%) a friend to enroll in a course that applies it. This is reinforced by 67% of students believing that other courses should adopt (39%) or definitely adopt (28%) this methodology. As for the Kahoot quizzes performed at the beginning of every session, 66% of students found them useful (35%) or very useful (31%), whereas 14% of them found them not useful (10%) or not useful at all (4%). Finally, 74% of students considered they had more (46%) or much more (28%) time to do exercises during lab sessions when compared to previous courses they studied before enrolling in this course, since all of them followed a traditional methodology.

Regarding students' feedback related to the videos watched, more than 71% of students thought the number of videos per lab session and their duration was appropriate (53%) or very appropriate (18%), and only 7% thought contrarily. These percentages were even higher when related to the duration of the videos, where 74% of students thought it was appropriate (47%) or very appropriate (27%), and only 6% thought it was not. Finally, 75% of students reflected that the didactic content of the videos was appropriate (61%) or very appropriate (14%), and only 5% of them thought contrarily. Related to this last point, Fig. 5 displays how students assessed the videos technical quality. Although the answers were normally positive and, in some cases, very positive—such as for videos of lab 3—, it is also true that the technical quality of the videos of some labs was not well assessed by students, such as those in lab 8.

V. Discussion

Overall, the results obtained from studying the application of FC in the lab sessions of the course are encouraging. This section describes some interesting conclusions from the results obtained in the previous section following the same classification. Some limitations of the study are also discussed.

A. Analysis on Students' Performance

The results indicate that there is no statistical evidence to assert that the methodology applied influenced students' grades. However, as displayed in the histograms of Figure 1, it is worth noting that there were fewer students with lower grades when the FC methodology was applied. It is also worth mentioning that, in 2017, students who more frequently attended lab sessions generally obtained better grades and were able to finish more exercises in class, while this correlation disappeared in 2018. This conclusion suggests that students could achieve a deeper self-learning in 2018, since the videos of all lab sessions were available during the whole course, so they could watch them at any point. Therefore, if a student could not attend a lab session or was late, (s)he could catch up later more easily with the work her/his colleagues had done in the lab session. Consequently, the FC helps students who find themselves in this or similar situations to successfully cope with the course.

It is also worth mentioning the positive correlation between students' final grades and the number of correct answers and reward points in the Kahoot quizzes. This suggests that those students who consistently watched the videos before each lab—those with better results in the Kahoot quizzes—were also those with better grades. Besides, these students were normally more interested in the course and more eager to learn and achieve good grades, so the results obtained are reasonable. Also related to these quizzes, most students felt enthusiastic about them due to the game-based format and the competitive aspect—they tried to beat their classmates and friends—, so they were a good mechanism to foster video visualization prior to class.

B. Analysis on Students' Attitude

It is clear the improvement in the clarity with which students perceived explanations (instructors' explanations the first year and videos plus instructors' clarifications the second year), displayed in Fig. 2. A reason is that explanations given in videos tend to be clearer than those done on site. Videos require to prepare a script, examples and demos. They also allow performing a crafted post-production in order to edit and improve those parts that can be more confusing, until a satisfactory result is obtained. Besides, despite instructors spend more time and effort in preparing the lesson, this pays off because (parts of) the videos can be reused in future courses. Also, students can re-watch any part of the video on demand, and they are used to consume online video-based content nowadays, so watching a YouTube video explaining next day's lab session might not be felt as a burden. All these reasons make students have a good attitude towards the course when the lab content is offered in video format, as indicated by the percentage of students who watched the videos prior to lab sessions in 2018 (83%).

C. Analysis on Students' Assessment of Flipped Classroom

Fig. 5 analyzes students' responses related to the videos technical quality. It was concluded that they appreciated dynamic content in videos, e.g., they complained about the lack of demos in some videos. In particular, as extracted from the final survey in 2018, students appreciated very much watching exercises related to the lab content. The main reason is because the development environment they had to use for solving the exercises and developing their projects was shown in the videos. Therefore, it is recommended that videos used for flipping laboratory classes contain mostly dynamic content and demos.

Concerning students' general assessment of FC, the positive acceptance is noticeable: 87% of surveyed students preferred it over traditional methodologies. Videos were of prime importance in the course's project, from which students obtained their grades in the practical part. The statistics given by YouTube Analytics showed that students not only watched the videos before lab sessions, but also during the semester. Even more, also students who had to prepare the September project's retake watched them during their Summer holidays, when it is not possible to approach instructors for solving doubts. A chart obtained from YouTube Analytics displaying students' views of a video is available on [21].

D. Limitations of the Study

This study presents some limitations:

1) Even though all surveys were conducted in Spanish, the mother tongue of all students, the way of formulating the questions might have been a limiting factor. However, instructors took care in writing questions concisely, and students could answer most of them using a Likert scale, whose main advantage is that the questions involved use a similar method of collecting the data. This makes the questions easy to understand and answer, and students do not feel forced to express their opinion, allowing them to stay neutral.
2) Students' age and gender were not considered in the study. However, these factors must have likely not influenced the results because they were similar in the two editions of the course. Most students were between 20 and 23 years old and only about 10% were women.
3) Surveys were completed at the end of the lab sessions, so students might have felt fatigued when filling them in. Like before, since questions were written concisely and answered using a Likert scale, it is not very likely that students felt fatigued.
4) At the beginning of the course, students had to choose their group mates for undertaking the course's project. Students who could not find a group were grouped together. The way groups were formed could have an influence on their members' final grades (for instance, students who already know each other typically find it easier to work together).
5) Students were asked to optionally write their names in the questionnaires (in order to keep trace of their performance). Despite they were assured that their answers would not have any effect on their final grades, only about half of the students wrote their names on them. Having had all questionnaires with students' names could have thrown different results. In any case, a large enough number of questionnaires included students' names (575 out of 1166 in 2017 and 684 out of 1174 in 2018).
6) Part of students' assessment of FC was done through a survey at the end of the second year (cf. Section IV-D). Out of all students in the course, only those still attending filled out the survey. Specifically, 209 students enrolled in the course, 163 stayed until the end and were graded by instructors, and 144 (88% of those completing the course) filled out the survey. Despite having had opinions from all 209 students could have yielded slightly different conclusions, 144 students is still a large-enough number, especially considering that 163 students were graded—the remaining 46 gave up the course.

VI. Conclusion

The FC was applied to the laboratory sessions of a computer science course on software architecture and integration. Students' performance, attitude and perceptions were compared to the same course following a traditional methodology with a similar student cohort. The study took two academic courses and involved 6 instructors and more than 400 students. The results of the study are encouraging. They show that despite students' grades were generally similar both years, the way of learning with the FC showed to be more autonomous. The main reason is that videos are available during the whole semester, and they explain the contents of the lab sessions in a crafted manner (*RQ1*). Inside class, students had on average 24 more minutes for hands-on activities, so they were able to finish many more exercises. Results also show a very good students' attitude towards the FC, with a positive mindset towards video-based lectures: 86% of students watched the videos before class, and two thirds additionally watched them at any other point during the semester (*RQ2*). Finally, students' overall satisfaction with the FC was clear in that 9 out of every 10 students would choose it over traditional methodologies in the lab sessions of future courses (*RQ3*). For all these reasons, the FC can play a very important role not only in theoretical classes, but also in laboratory sessions, which are of prime importance in computer science degrees.

In this study, the FC methodology has been applied on a second-year undergraduate course. For future work, it would be interesting to evaluate the application of FC in courses taught in different years, and study if conclusions vary. For instance, it is interesting to see the difference between applying it in courses in the first year or in the last year, due to the different level of students' maturity. Other studies could evaluate whether students prefer their instructors to be the actors in the videos or they prefer someone else.

Acknowledgments

This work has been partially supported by the Program of Teaching Coordination and Innovation of the University of Seville 2017-2018 under Grant 21034.

References

[1] M. J. Lage, G. J. Platt, and M. Treglia, "Inverting the classroom: a gateway to creating an inclusive learning environment," *The Journal of Economic Education*, vol. 31, no. 1, pp. 30–43, 2000.

[2] J. Lowell, B. Utah, M. Verleger, and D. Beach, "The flipped classroom: a survey of the research," *Proc. of 120th ASEE national conference and exposition*, vol. 30, pp. 1–18, 2013.

[3] J. O'Flaherty and C. Phillips, "The use of flipped classrooms in higher education: a scoping review," *Internet and Higher Education*, vol. 25, pp. 85–95, 2015.
[4] L. Abeysekera and P. Dawson, "Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research," *Higher Education Research & Development*, vol. 34, no. 1, pp. 1–14, 2015.
[5] M. N. Giannakos, J. Krogstie, and N. Chrisochoides, "Reviewing the flipped classroom research: reflections for computer science education," *Proc. of CSERC'14*, pp. 23–29, 2014.
[6] C. K. Lo, C. W. Lie, and K. F. Hew, "Applying "first principles of instruction" as a design theory of the flipped classroom: findings from a collective study of four secondary school subjects," *Computers and Education*, vol. 118, pp. 150–165, 2018.
[7] C. K. Lo, K. F. Hew, and G. Chen, "Toward a set of design principles for mathematics flipped classrooms: a synthesis of research in mathematics education," *Educational Research Review*, vol. 22, pp. 50–73, 2017.
[8] A. Karabulut-Ilgu, N. Jaramillo Cherrez, and C. T. Jahren, "A systematic review of research on the flipped learning method in engineering education," *British Journal of Educational Technology*, vol. 49, no. 3, pp. 398–411, 2018.
[9] S. L. Dazo, N. R. Stepanek, R. Fulkerson, and B. Dorn, "An empirical analysis of video viewing behaviors in flipped CS1 courses," *ACM Inroads*, vol. 7, no. 4, pp. 99–105, 2016.
[10] M. K. Seery, "Flipped learning in higher education chemistry: emerging trends and potential directions," *Chemistry Education Research and Practice*, vol. 16, pp. 758–768, 2015.
[11] J. A. Day and J. D. Foley, "Evaluating a web lecture intervention in a human-computer interaction course." *IEEE Transactions on Education*, vol. 49, no. 4, pp. 420–431, 2006.
[12] J. J. Cupak and V. V. Riabov, "Applying "Flipped Classroom" Methodology in Computer Science Courses," *InSight: Rivier Academic Journal*, vol. 13, no. 2, pp. 1–6, 2017.
[13] J. J. Cupak Jr, "Flipped Classroom in Teaching Computer Science," *Journal of Computing Sciences in Colleges*, vol. 33, no. 6, pp. 176–178, 2018.
[14] R. Mhiri, M. Saad, M. Dodo Amadou, V. Nerguizian, G. Brady, S. Ouertani, S. Sahli, and H. Saliah-Hassane, "The experience of a collaborative project on remote laboratory: From development to operation," *Proc. of ICELIE'12*, pp. 116–120, 2012.
[15] M. L. Maher, C. Latulipe, H. Lipford, and A. Rorrer, "Flipped classroom strategies for CS education," *Proc. of SIGCSE'15*, pp. 218–223, 2015.
[16] C. Latulipe, N. B. Long, and C. E. Seminario, "Structuring Flipped Classes with Lightweight Teams and Gamification," *Proc. of SIGCSE'15*, pp. 392–397, 2015.
[17] K. Missildine, R. Fountain, L. Summers, and K. Gosselin, "Flipping the classroom to improve student performance and satisfaction," *Journal of Nursing Education*, vol. 52, no. 10, pp. 597–599, 2013.
[18] G. C. Gannod, J. E. Janet, and M. T. Helmick, "Using the inverted classroom to teach software engineering," *Proc. of ICSE'08*, pp. 777–786, 2008.
[19] E. F. Gehringer and B. Peddycord, "The inverted-lecture model: A case study in computer architecture," *Proc. of SIGCSE'13*, pp. 489–494, 2013.
[20] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: an empirical study of MOOC videos," *Proc. of L@S'14*, pp. 41–50, 2014.
[21] J. Troya, J. A. Parejo, S. Segura, A. Gámez-Díaz, A. E. Márquez-Chamorro, and A. del-Río-Ortega, "Experiments and analysis performed," [Online] https://exemplar.us.es/demo/Troya2019FlippedClassroom, 2019.
[22] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Pearson correlation coefficient*. Springer, 2012, pp. 1–4.
[23] J. M. Bland and D. G. Altman, "Multiple significance tests: the bonferroni method," *BMJ*, vol. 310, no. 6973, p. 170, 1995.
[24] L. Myers and M. J. Sirois, "Spearman correlation coefficients, differences between," *Encyclopedia of statistical sciences*, vol. 12, 2004.
[25] J. R. A. Santos, "Cronbach's alpha: A tool for assessing the reliability of scales," *Journal of extension*, vol. 37, no. 2, pp. 1–5, 1999.
[26] A. Agresti and M. Kateri, *Categorical Data Analysis*. Berlin, Heidelberg: Springer, 2011, pp. 206–208.
[27] R. H. B. Christensen, "Cumulative link models for ordinal regression with the r package ordinal," *Submitted to Journal of Statistical Software*, 2018.
[28] J. Cuzick, "A wilcoxon-type test for trend," *Statistics in medicine*, vol. 4, no. 4, pp. 543–547, 1985.
[29] B. Mason, S. Gregory, T. Rutar, K. E. Ieee, B. Strayer, J. Learning, B. Herreid, C. Freeman, and N. A. Journal, "Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course," *IEEE Transactions on Education*, no. 4, pp. 430–435, 2013.
[30] R. S. Davies, D. L. Dean, and N. Ball, "Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course," *Educational Technology Research and Development*, vol. 61, no. 4, pp. 563–580, 2013.
[31] B. Love, A. Hodge, N. Grandgenett, and A. W. Swift, "Student learning and perceptions in a flipped linear algebra course," *International Journal of Mathematical Education in Science and Technology*, vol. 45, no. 3, pp. 317–324, 2014.
[32] S. Velegol, S. Zappe, and E. Mahoney, "The evolution of a flipped classroom: evidence-based recommendations," *Advances in Engineering Education*, vol. 4, pp. 1–37, 01 2015.

Javier Troya is an Associate Professor at the University of Seville, Spain, where he participates in several teaching innovation programs. Previously, we received his Ph.D. in Computer Science from the University of Malaga, Spain, and spent two and a half years as a postdoc at the TU Wien, Austria. His research focuses on the use of innovative methodologies for teaching, model-driven engineering and software testing.

José A. Parejo is a Professor at the University of Seville, Spain. He has 7 years of experience in the software development industry and more than 12 years of experience teaching undergraduate and master courses on various topics related to Software Engineering. His research is focused on Search-Based Software Engineering, and Empirical Software Engineering.

Sergio Segura is a Professor at the University of Seville, Spain. He has more than 13 years of experience teaching undergraduate and master courses on various topics related to Software Engineering. He has supervised two doctoral theses, both awarded. Currently, he seeks to improve student participation and motivation using innovative teaching techniques such as inverted class and the use of serious games.

Antonio Gámez-Díaz is a Predoctoral Researcher at the University of Sevilla, where he received a BS degree and an MSc degree in Software Engineering in 2016, with a competitive predoctoral fellowship (FPU) granted by the Spanish government. His research interests are focused on Service-Oriented Computing. He collaborates with leading organizations such as the OpenAPI Initiative, Docker or GitHub Education as an ambassador for the technological and industrial knowledge diffusion.

Alfonso E. Márquez-Chamorro is an Assistant Professor at the University of Seville, Spain, and a member of the ISA Research Group. His current research areas include machine learning, business process management and process mining. Previously, he worked in bioinformatics and evolutionary computing.

Adela del-Río-Ortega is an Associated Professor at Universidad de Sevilla, Spain, where she got her Ph.D. in Software Engineering in 2012. She has 8 years of teaching experience in Bachelor and Master degrees in Software Engineering and Information Systems. She has supervised two theses. She is also involved in several teaching innovation programs. Her main research topic is process performance management.