CONTINUOUS ASSESSMENT WITH FLIPPED LEARNING AND AUTOMATED ASSESSMENT

Nicolás Serrano, Carmen Blanco, Kevin Calderón, Iñigo Gutierrez, María Serrano

University of Navarra, TECNUN School of Engineering, San Sebastián, Spain

ABSTRACT

Numerous studies claim that one of the practices that improves student performance is continuous assessment, since it encourages work on the subject and allows for better class performance. Often, this continuous assessment is implemented with 1 or 2 partial exams, which causes an improvement of the performance, but sometimes they become for the students "early final exams" with the consequent disturbance in the usual preparation of the semester. In this article, we describe and analyze a method of continuous assessment implemented in a second-year programming subject of the School of Engineering of the University of Navarra since the 2016-2017 academic year. Moreover, flipped learning approach was introduced in the academic year 2019-2020, with the consequent (but slight) modification of the evaluation method. To implement continuous assessment an automated evaluation tool has been developed both for quizzes and for the correction of programming exercises. The result is a continuous work of the student, good grades and satisfaction from both the students and the teacher.

KEYWORDS

Continuous assessment, flipped learning, automated assessment, programming subject, Standards: 8, 11

INTRODUCTION

It has been stated that the assessment method is what most influences what and how students study (Rust et al., 2005). Studies indicate that it is the most important element of the learning process and should be integrated into it. (Crawley et al., 2014)

Several studies show the positive impact of continuous assessment on academic outcomes (Martín-Carrasco et al., 2014; Pérez-Martínez et al., 2009; Walser, 2015; Cole and Spence, 2012). One of the reasons for this is the increased activity of students, who start working earlier and spend more time on the course (Edström et al., 2003). In addition, it facilitates meaningful learning since new concepts or knowledge are related and build on previous knowledge that, thanks to continuous assessment, are well established.

Often this continuous assessment is implemented with 1 or 2 partial tests, which leads to an improvement in performance, but sometimes they become for the students "early final exams" with the consequent disruption in the usual preparation of the semester.

The summative evaluation for obtaining the final grade, must also include the formative evaluation. This formative evaluation provides teachers and students with data to take the necessary measures and reach the desired level of knowledge (Gikandi et al., 2011).

Some studies establish that continuous assessment favors superficial learning (Jordan, 2009; Tan, 1992). In fact, what favors superficial learning is not the frequency of the evaluation but an inadequate way of carrying it out. The solution is to design tests that not only evaluate basic cognitive skills (Draper, 2009; Leung et al., 2008), which are the easiest to implement with automated systems, but other higher cognitive functions such as analyzing, relating, etc.

Another major difficulty in implementing continuous assessment is the increased workload and time required for teachers to prepare, correct, and score different activities (Reina-Paz et al., 2014). A time that is not motivating for the teacher nor is it valued externally. Trotter's (2006) study recognizes the increased teacher dedication, but states that the results in student motivation and learning are excellent.

At Tecnun, the School of Engineering of the University of Navarra, throughout different curricula and with different calendars, and mainly as a result of the European Higher Education Area adaptation process, we have gone from having only one final exam, to also having intermediate exams, and in some cases biweekly exams were introduced. In addition, in order to evaluate the knowledge with which the students joined the university, diagnostic assessments on the fundamental subjects of mathematics and physics were also introduced. This has resulted in a large percentage of the teachers' time devoted to teaching being spent on correction. A task that, as explained, brings a reduced value to the student and the teacher. For this reason, we decided to work on methods that could relieve the teacher of this workload without giving up the advantages provided by a greater number of evaluations.

EXPERIENCE IN A PROGRAMMING SUBJECT

This article describes and analyzes the implementation of a method of continuous assessment in a second-year programming subject of the School of Engineering of the University of Navarra in the academic years 2016-2017 to 2019-2020.

As mentioned above, it was necessary to develop an automatic correction tool that would adapt to the needs of this subject, which was used for the first time in the 2016-2017 academic year. This is an application in which the teacher defined the exercises that the students had to do at each moment and a program checked the validity of the answer. The system showed the viability of automated evaluation and was moved from two partial evaluations to an evaluation every 10 days. Classroom and home exercises were done and evaluated with the same tool. The students' response was very positive, since it was an optional system in which they could objectively verify their learning process and already have a part of the final grade. (Serrano, et al. 2018). In the 2017-2018 school year, the evaluation became weekly.

In the 2019-2020 school year, Flipped-Learning was introduced, along with the need to motivate and verify that students were doing their pre-class work. The following is a description of the course structure as it was developed.

The course is 13 or 14 weeks long depending on the calendar and takes place from the second half of January until the beginning of May, when the final exams of the second semester are held.

The previous structure of the course consisted of a final exam with 2 partial tests, one in the middle of the course and another one two weeks before the end of the course.

With the incorporation of continuous assessment, weekly evaluations were added, resulting in the 2019-2020 course in 9 evaluated practices, 5 before the first partial test and 4 before the second partial test. Also added, with the incorporation of flipped learning, was a weekly theory test. Therefore, with the exception of the first and last week of the course, every week there was an evaluated practice or a partial test and a theory test, so that continuous assessment was incorporated into the students' routine. Figure 1 shows the calendar for the 2020 course, with the evaluated practices marked in dark green and the partial test in blue. The theory tests are marked in light green.

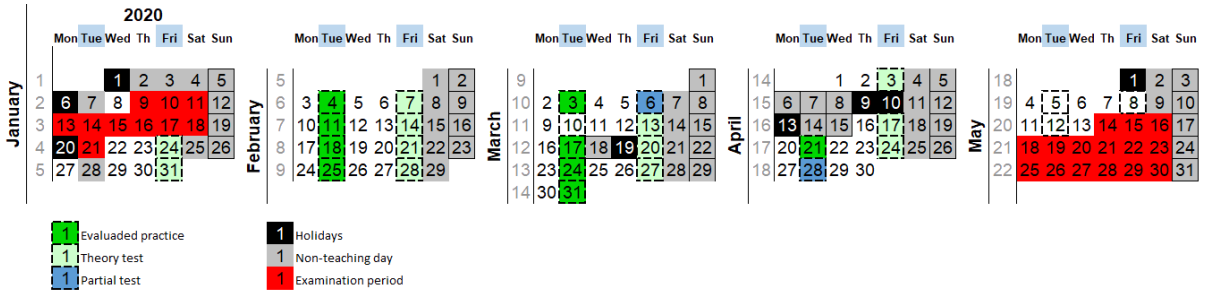


Figure 1. Calendar of the course with the evaluated practices, theory test and partial tests

The course covers 11 or 12 topics, depending on the schedule. These topics are grouped into four blocks: the first of a single topic on computer concepts, the second on the Java language and Object-Oriented Programming which comprises four topics, similar to other courses such as Programming with a Purpose from Princeton University. These two blocks constitute the object of the first partial test. The third block is about the Internet which consists of two topics and the fourth block is about web applications which consists of four or five topics. Block 3 and the first two topics of block 4 are the subject of the second partial tests.

Evaluation mode

Each of the parts has a characteristic type of exercises. The automated evaluator has different procedures for each of these types of exercises. For topic 1 there is a computer processor simulator that is programmed with the instructions written by the students. The evaluator checks that these instructions perform the desired function.

For block 2, the assessor compiles and runs the programs and functions in Java that are written by the students. The automated evaluator checks that the program can be compiled and then executes it taking into account that it may have an error that turns it into an infinite loop, in which case it ends the process and indicates that the solution is incorrect.

In block 3 the evaluator checks that the pages and forms written in HTML language have the elements and characteristics requested in the exercise.

In block 4, where Java applications are written to produce web pages, the evaluator compiles and executes the programs as in block 2. The result is an HTML page that is checked as in block 3.

In addition to the weekly evaluated practices and partial tests there is a final exam and a project in which everything learned in the course is applied. This serves to integrate all the concepts into a larger program and thus the score is not just the sum of evaluations of individual exercises. The final grade is composed of the grade of the first part (25%), the second part (25%), the project (15%) and the final exam (35%). The first part is composed of 5 evaluated practices, the first partial test and 6 theory tests. And the second part is composed of 4 evaluated practices, the second partial test and 6 theory tests.

Automated evaluation

Before the introduction of automated evaluation, the correction of programming exercises was done by viewing and executing the code. The code execution was automated with scripts that made the compilation of the code of each student and the execution with some parameters defined by the teacher to check the correct operation of the program. However, this process was not as direct as described since any error in the execution or a slight variation in the names that the programs should have caused the process to stop.

Automated evaluation in a programming course involves compiling and executing the student's code to check its correct operation. This may seem inconvenient since solutions that fail to compile or run would not be considered here. Here it should be mentioned that, being a second-year programming subject, the need to use “incremental development” suggested by many programming professionals is advised to students from the beginning of the course (Larman, 2003; Downey & Mayfield, 2019). Incremental development consists of obtaining in different iterations applications that work correctly, although it still does not accomplish all the requirements that are demanded of it, until reaching the final code. By introducing the automated evaluation and being able the student to test the correctness of their code from the first exercises, the student acquires from the first days the habit of incremental development and always having an application that works. If not, the student knows that the proofreader won't give any points, whereas it will if the student do a partial but working fix. The student also learns the importance of following nomenclature and spell-checking rules, as if the student does not comply, the evaluator indicates this when evaluating the exercise.

Therefore, we can summarize that the automated evaluation process in a programming subject has the advantage that feedback is provided instantaneously, as it would be produced in a real environment, so that it can be analyze what the error is and be able to modify it appropriately or continue with the next iteration of the program until reaching the final solution.

To define how an exercise will be evaluated, the teacher must: decide whether to evaluate the result of a function or what appears in the standard output captured by the evaluator, provide which are the arguments of the program or the function with which test the student code, indicate the expected answer and the value of the correct evaluation. The teacher can define different checks for each exercise, for example with different values of an argument to check that the student's solution is general. The teacher can also define parameters so that each student's statements are different from those of her classmates. In this case, the teacher expresses the solution as a function of the parameter. The teacher also defines how many tries the student can make for each exercise and whether feedback should be provided at that time.

Additionally, question pools can be defined if the teacher wants them to be randomly displayed to each student. Figure 2 shows the definition of a Java exercise by the teacher with how to evaluate it.

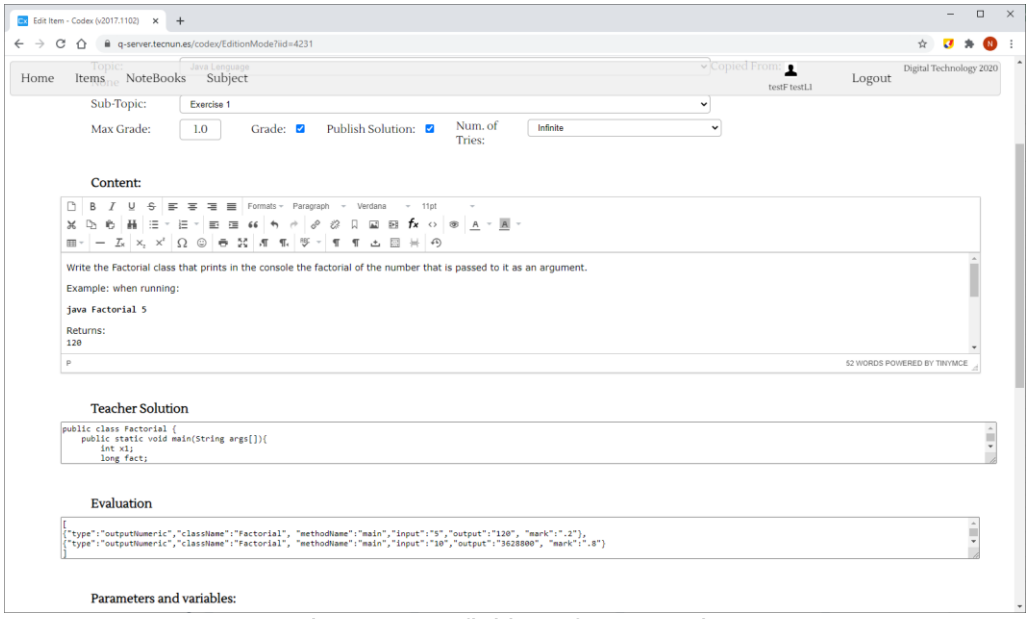


Figure 2. Definition of an exercise

Figure 3 shows the student's view, with the answer that the student introduces and the response of the system when evaluating it.

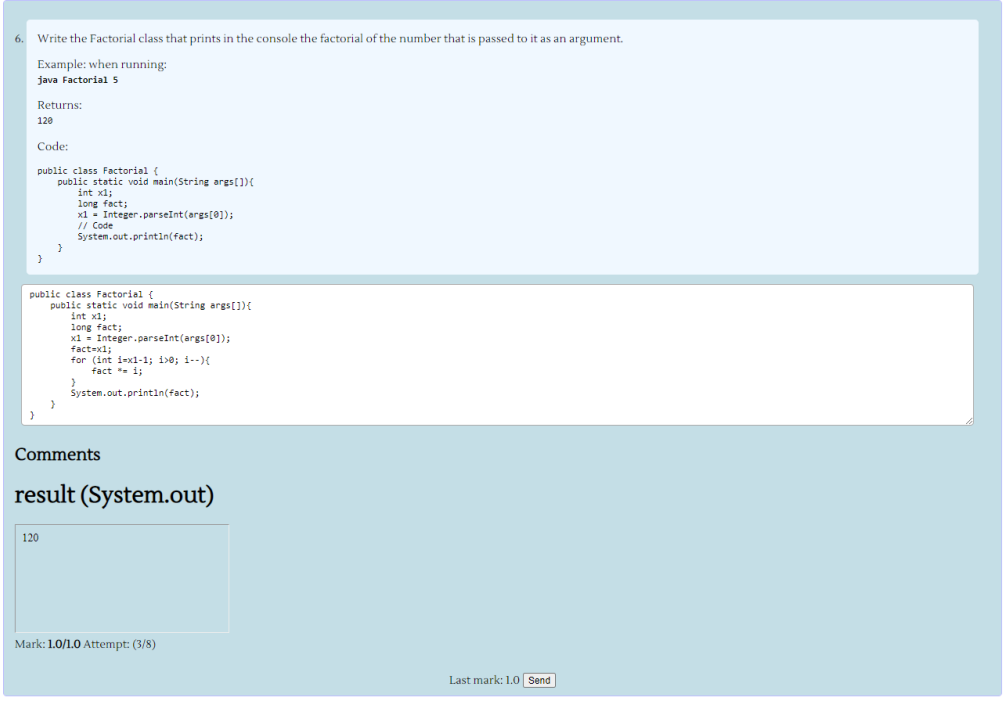


Figure 3. Student view and response of an exercise

Flipped Learning

After the use of the automated continuous assessment in 3 courses, the good results it produced and the good reception among the students were observed. However, the time invested in the theoretical explanations and the evaluated practices left little time for the students to carry out exercises in the classroom. Although they had the automated evaluator to verify the correctness of these exercises, they did not have the possibility to comment them with their classmates, with the teacher or with the assistants of the subject during the class time.

For this reason, Flipped Learning was introduced in the 2019-2020 course, to take most of the theoretical explanations and procedures out of the classroom and to be able to focus the activities on those where the presence of the teacher and students is most valuable. Because the objectives of each class were well defined, having applied continuous weekly evaluation in previous courses, the development of the material, mainly videos, required about 2 or 3 hours of preparation each week for recording and editing, which allowed the flipping of all classes. More than 70 videos are available on the page https://nicolasserrano.github.io/CS/material.

To motivate the visualization of the videos, each week a test was conducted with basic questions from the videos, in which students could take a sheet with their handwritten notes. This has the effect that the evaluation test is done with less tension and the visualization of the videos is done in a more active attitude, and even more comprehensive way than taking notes in class, since the objective is to bring some notes to the next test.

The tests consisted of multiple-choice questions with a correct answer. They were made with the same evaluation tool to have all the marks unified in the same system, but they could be made with the LMS of the subject or another system like Socrative or Kahoot.

It is necessary to mention that since this subject was taught in spring 2020, it coincided with the change to online teaching from March 2020. The classes continued in the same way from that date, the theoretical part continued to be learned through the Flipped Learning videos and the practical classes continued to be the laboratory practices with resolution of doubts. What did occur is a delay in the evaluation tests of both the practices evaluated and the theory tests due to the university's indication not to carry out evaluated tests in the first days of the confinement. When, after two weeks, methods were available to verify the identity and correct performance of the tests, an update of the calendar was published with the dates for the performance of these tests in the following weeks, with the test and the evaluated practice being performed on the same day during the following two weeks.

Since the school wanted to test the proctoring tools in advance before the final exam dates in May, and all the material was available, students of this course were given the option of preparing the material in advance, also advancing the evaluated practices and the tests on the videos in order to take the exam during the class period. It was expected that a dozen students would take this modality, but it turned out that most of them signed up for it. So the school was able to take a most complete test of the proctoring tool.

RESULTS

The result of applying the continuous assessment, as already mentioned in the article on its application in previous courses, has been very positive both in the aspect of learning and student satisfaction. Figure 4 shows the results of the survey carried out with students in February 2018, regarding which exercises were to be maintained with automated evaluation. Only 1 student out of 23 did not want the weekly evaluation, while 95.6% wanted to continue it in the second part of the course. And 86.9% wanted the practice exercises during the week to be evaluated. 100% wanted some kind of early evaluation.

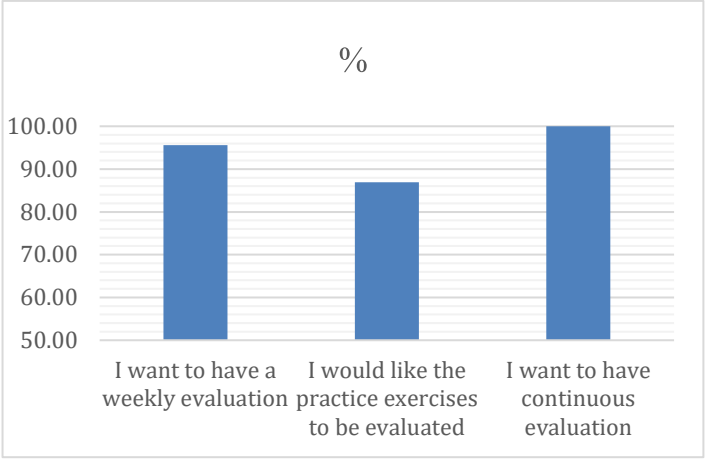
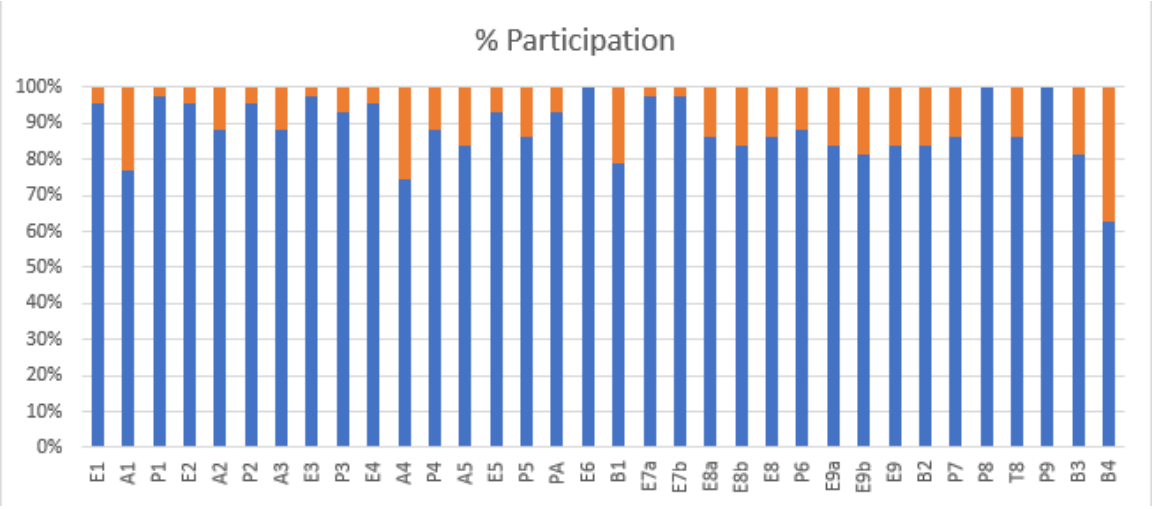


Figure 4. Student survey results

Regarding the learning process, it is observed that from the first week the students have an active attitude, which is shown in the results of the exercises they have as practice, the evaluated practices, and the participation in the subject's forum, where they ask questions about exercises of the subject and their classmates answer most of the times to them, if not the teachers and assistants of the subject answer them. In the survey, 73.9% of the students said that if there were no exercises evaluated, they would not practice weekly, compared to 26.1% who said they would practice anyway, so it can be concluded that the method especially favors those who need additional motivation for the work of the subject.

Figure 5 indicates the participation in the different activities of the continuous assessment (A: test of part A, B: test of part B, E: exercises in class and outside the classroom, P: weekly evaluated practice). It should be noted that the test exercises and evaluated practices were able to carry out only in the classroom. With the total data, all the students who took the final exam had taken more than half of the continuous assessment tests (out of 30 tests), 27.9% took 100% and 81.4% took more than 83% of the tests.



E: Homework exercises
A: Theory test part A
B: Theory test part B
P: Evaluated practice
PA: Partial test A

Figure 5. Percentage of students who participated in each continuous assessment activity (in blue)

In Figure 6 each row corresponds to all the exercises of a student and each red point to an exercise. So one row represents the performance of a student and a column how each student has answered an exercise. In the last two blocks the answers are produced alternatively in one block or the other corresponding to that they chose the anticipated or the normal evaluation. This table shows the general situation of the class.

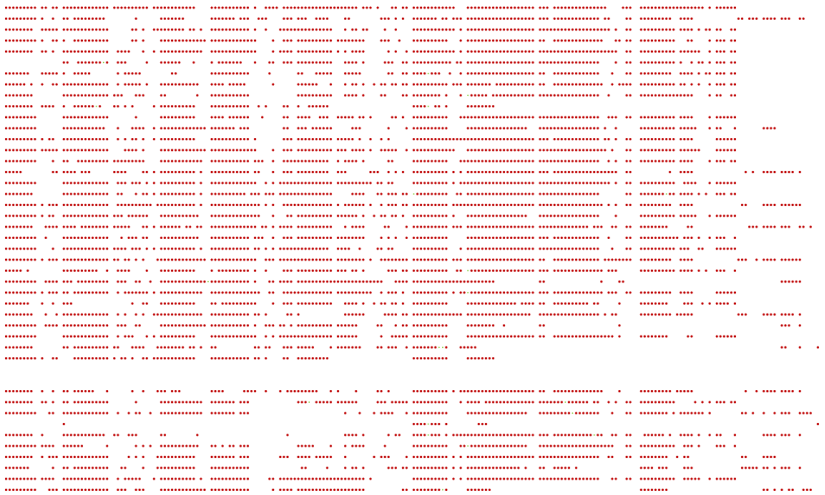


Figure 6. Overview of student exercises

As for the recommendations collected from students in the February 2018 survey, they request:
- More exercises. In fact, they have other exercises, but they prefer to have them in the automated evaluation platform, because if not, they do not even consider them as part of the subject, even though they are within the material part of the subject.
- They prefer videos made by the teacher; along with the 70 videos, 5 were external videos and they prefer that they be made by the teacher because they consider them more applied to the exercises they have to do.
- They prefer videos to documents, even though only 3 documents have been provided, they find them heavier.

CONCLUSIONS

The students' responses and their performance on the course show that continuous assessment is of interest to the students, which, although it requires continuous work, compensates them for having an early view of their status and distributing their workload.

In terms of student satisfaction, the point that might seem more debatable is the high number of tests assessed, but when asked for recommendations to improve in the subject, only one mentions it, along with the results of surveys that want to keep the tests assessed. This indicates that the continuous assessment, at least in the way it has been implemented in the subject, allows to integrate it in the subject's routine by evaluating all the activities both inside and outside the classroom.

Student performance has improved. It can be seen in the quality of the projects executed by the students and by the number of them performing the exercises in the final exam, although with the change with the anticipated evaluation it cannot currently be shown with a comparative numerical value.

Therefore, the change of evaluation system to a system of continuous assessment has had an effect on improving student participation and academic performance, so it can be inferred that training has improved. From the first day the students were working so they have been able to be more active during the course allowing the teacher to make the class more active and increase the interaction of the students among themselves and with the teacher.

The automatic correction tool favors the use of the incremental development model, a recommended programming method to which students are not usually used to, but which they adapt to because it is a simple system, and it is the one that guides their way of working.

Regarding the workload of the teacher, the first years has increased because they have to prepare the class exercises and evaluation in the tool, but once they are already worked is a basis that reduces considerably the time of preparation as it is to be modified gradually. The same happens with the theoretical videos of flipped learning. It should be noted that the correction time is zero.

In summary, the students' perception of the subject has improved considerably. They consider that they have learned more that the workload is well distributed and that their transversal skills have improved.

REFERENCES

Ausubel, D. P. (1963) *The psychology of meaningful verbal learning.*, *The psychology of meaningful verbal learning.* Oxford, England: Grune & Stratton.

Cole, J. S. and Spence, S. W. T. (2012) 'Using continuous assessment to promote student engagement in a large class', *European Journal of Engineering Education*. Taylor & Francis, 37(5), pp. 508–525. doi: 10.1080/03043797.2012.719002.

Crawley E.F., Malmqvist J., Östlund S., Brodeur D.R., Edström K. (2014) Student Learning Assessment. In: Rethinking Engineering Education. Springer, Cham. https://doi.org/10.1007/978-3-319-05561-9_7

Downey, A. B. and Mayfield, C. (2019) *Think Java*. 2nd edn. Sebastopol, United States: O'Reilly Media, Inc.

Draper, S. W. (2009) 'Catalytic assessment: Understanding how MCQs and EVS can foster deep learning', *British Journal of Educational Technology*, 40(2), pp. 285–293. doi: 10.1111/j.1467-8535.2008.00920.x.

Edström, K., Törnevik, J., Engström, M. and Wiklund, Å. (2003) 'Student involvement in principled change: Understanding the student experience' *Proceedings of the 2003 11th Improving Student Learning,* OCSLD, Oxford.

Gikandi, J. W., Morrow, D. and Davis, N. E. (2011) 'Online formative assessment in higher education: A review of the literature', *Computers and Education*. Elsevier Ltd, 57(4), pp. 2333–2351. doi: 10.1016/j.compedu.2011.06.004.

Jordan, S. (2009) *Assessment for learning: pushing the boundaries of computer-based assessment*, *Cumbria*.

Larman, C. and Basili, V. R. (2003) 'Iterative and incremental development: A brief history', *Computer*, 36(6), pp. 47–56. doi: 10.1109/MC.2003.1204375.

Leung, S. F., Mok, E. and Wong, D. (2008) 'The impact of assessment methods on the learning of nursing students', *Nurse Education Today*. Nurse Educ Today, 28(6), pp. 711–719. doi: 10.1016/j.nedt.2007.11.004.

Martín-Carrasco, F. J., Granados, A., Santillán, D. and Mediero, L. (2014) 'Continuous assessment in civil engineering education: Yes, but with some conditions', *CSEDU 2014 - Proceedings of the 6th International Conference on Computer Supported Education*, 2(January), pp. 103–109. doi: 10.5220/0004797401030109.

Pérez-Martínez, J. E., Jesús García-García, M., Perdomo, W. H. and Jesús Villamide-Díaz, M. (2009) *Analysis of the results of the continuous assessment in the adaptation of the Universidad Politécnica de Madrid to the European Higher Education Area*. Available at: http://mgie.eui.upm.es (Accessed: 8 January 2021).

Reina-Paz, M. D., Rodriguez-Oromendia, A. and Sevilla-Sevilla, C. (2014) 'Effect Of Continuous Assessment Tests On Overall Student Performance In The Case Of The Spanish National Distance Education University (UNED)', *Journal of International Education Research (JIER)*, 10(1.5), pp. 61–68. doi: 10.19030/jier.v10i1.5.8460.

Rust, C., O'Donovan, B. and Price, M. (2005) 'A social constructivist assessment process model: How the research literature shows us this could be best practice', *Assessment and Evaluation in Higher Education*. Taylor and Francis Ltd , 30(3), pp. 231–240. doi: 10.1080/02602930500063819.

Serrano, N., Blanco, C., Carias, F. and Reina, E. (2018) 'Information from Automated Evaluation in an Engineering School'. doi: 10.4995/head18.2018.8132.

Tan, C. M. (1992) 'An evaluation of the use of continuous assessment in the teaching of physiology', *Higher Education*. Kluwer Academic Publishers, 23(3), pp. 255–272. doi: 10.1007/BF00145016.

Trotter, E. (2006) 'Student perceptions of continuous summative assessment', *Assessment and Evaluation in Higher Education*. Routledge , 31(5), pp. 505–521. doi: 10.1080/02602930600679506.

Walser, T. M. (2015) 'Evaluability Assessment in Higher Education: Supporting Continuous Improvement, Accountability, and a Culture of Assessment', *Journal of Assessment and Institutional Effectiveness*. Penn State University Press, 5(1), pp. 58–77. doi: 10.5325/jasseinsteffe.5.1.0058.

BIOGRAPHICAL INFORMATION

Nicolás Serrano is a professor of Languages and Computer Systems at the School of Engineering of the University of Navarra, in San Sebastian. He has worked at the CEIT research center and in industrial and service companies. His experience covers the areas Computer Science, Software Engineering, Digital Technology and Information Systems. He has directed and developed several information systems such as the academic management system of the University of Navarra, an ERP in web platform for business management and applications for technological innovation.

Carmen Blanco is an Associate Professor of Mathematics, Engineering School of the University of Navarra (Tecnun), Spain. Currently, her research is focused on the analysis, development and application of the new technologies and methodologies for the improvement of the learning process.

Kevin Calderón is a PhD student graduated in Telecommunication Engineering in the Engineering School of the University of Navarra, Spain. His research is focused on the development and application of data analytics techniques in the educational and learning area.

Iñigo Gutierrez is Professor of Electronic Technology and Director of Development at the School of Engineering of the University of Navarra (Tecnun), Spain. Currently, his research is focused on the analysis and application of new teaching technologies and methodologies.

María Serrano is an industrial engineer and has completed the master's degree project at the School of Engineering of the University of Navarra, Spain. His area of interest focuses on information systems, data analysis and machine learning.

Corresponding author

Nicolás Serrano
University of Navarra, TECNUN School of Engineering, San Sebastián, Spain
Paseo de Manuel Lardizabal, Nº 13
20.018, Donostia - San Sebastián
Spain
nserrano@tecnun.es