# AN EXPERIMENT WITH ONLINE PRACTICE FOR ASSESSMENT

**Isabel Vieira, Ana Paula Lopes, Isabel Cristina Lopes, José Manuel Azevedo, Cristina Torres, Lurdes Babo**

*CEOS.PP / ISCAP / Polytechnic of Porto (PORTUGAL)*

## Abstract

When teaching is a passion, we do whatever we can to pass it over to our students, and it is even more relevant when we teach a subject to students from a different area. As professors in a Higher Education Institution (HEI) we have the responsibility to prepare the students for the future. We have to gain their attention so that they become engaged in the subject taught and have success at the end.

The aim of this paper is to describe and analyse an experiment that took place in the Spring term of 2019, with first year students of a Mathematics course from the undergraduate degree in Marketing. We have developed several questions, related to the course syllabus, on an online platform, and created a Question Pool. Outside the classroom, students at their own pace could solve these exercises. All questions have the feedback and solution step by step, thus the students can realize what is expected gaining insights about the subject. We discuss how the students accepted the initiative and how they used it. The results will be presented, analysed and used to improve the methodology.

Keywords: E-learning, Higher Education, Mathematics, Technology, Assessment.

## 1 INTRODUCTION

When teaching Mathematics with passion we wish to pass it to our students so that we can engage them to learn it. Universities and Polytechnic Institutes a little everywhere in non-Mathematics degrees face increasing difficulties at enrolling students in studies in which mathematics is a substantial component. Being a Mathematics lecturer in a non-Mathematics degree is a challenge to our creativity, imagination, and perseverance and it may be necessary for educators to create or foster motivation [1]. Every day we have to be prepared to produce and develop new tools to motivate students and technology could be a very good help for this purpose. We cannot ignore technology, as we are surrounded by it. The integration of educational technologies into the classroom becomes an essential part of a classroom to enhance 21st century teaching and learning experiences [2]. Therefore, this integration opens new possibilities for the teaching and learning process.

Nowadays a large number of educators use Virtual Learning Environments (VLE), to supplement the traditional face-to-face lecture and improve the teaching process [3]. VLE could be used in all degrees of teaching, especially in large university courses, once VLE can reach a great number of students in different geographical areas and courses [4]. Moodle is one of the most popular 'free' software e-learning platform (Learning Management System), or VLE and was introduced in ISCAP in 2005.

It was thinking in overcoming the difficulties described earlier that since 2006/2007 a team of lecturers from the Porto Accounting and Business School (ISCAP), at the Polytechnic of Porto (P. PORTO) has been developing an Online Mathematics Education Project – *MatActiva* (www.matactiva.com), making use from the interactivity of Moodle (see [5][6]). This project provides innovative resources and diversified materials around themes such as Elementary Mathematics, Calculus, Algebra, Financial Mathematics, and Statistics. The main objective of this project was to innovate the teaching and learning process exploring technologies as a pedagogical resource taking into account the importance of inducing bigger motivation to the students, allowing the student to engage more deeply and committed to learning, improving the rates of success.

The *MatActiva* project is used for blended learning, distance education and flipped classroom, and is structured in eight sections (About us, Mathematics zero, Learning, Tests, Doubts, MathChallenge, Erasmus, Etc), containing diversified materials ranging different students' needs and levels of knowledge, according to the areas mentioned above.

In this particular experience, we made use of the section Tests. In the topic Tests, students can solve, diagnostic, formative and summative tests (multiple-choice, true/false, matching and numerical type questions), that are created from a question bank (with approximately 1700 questions) organized into categories and subcategories according to themes previously specified. Each test allows multiple attempts and for each attempt the questions are immediately corrected and for each wrong answer a feedback solution, step by step, is visualized.

For several authors ([7][8][9][10][11]) the subject of feedback in higher education it is a very important issue to the development of effective learning, yet the number of studies carried out in higher education when compared to other sectors is small [12]. For [13], feedback from tests constitutes a main function for learning. Furthermore, the authors think that it is crucial that this information must be timely and provide specific explanations about mistakes and recommendations to improve student learning.

This paper is organized as follows: section 2 describes the methodology, section 3 presents the results obtained. We conclude with section 4, given the main ideas from this study.

## 2 METHODOLOGY

This paper aims to report an experience that took place in the Porto Accounting and Business School (ISCAP) at the Polytechnic of Porto (P. PORTO) during the Spring term of 2019, in an undergraduate first year course on Mathematics, from the Marketing degree.

The course had 246 students enrolled, but unfortunately, just 53% attend the classes regularly and do the continuous assessment. The so-called continuous assessment consists of three tests, that happen during the classes. This experiment consisted in constructing two tests with multiple-choice random questions in the *MatActiva* platform, similar to those of the second and third assessment tests so that the students could practice, as many times as they want, at their own pace, become aware and reflect on their own difficulties and being better prepared to pass the course. The second test (Test 2) had five questions on Linear Algebra applications: the first on cryptography, the second about Input-Output analysis, the third to calculate the eigenvalues of a matrix, followed by the eigenvectors and the last question was the discussion of a linear system of equations. Each question is worth 1.7 points, except for questions number 3 and 4 with just 1.5 points resulting in the maximum grade of 8.1 for the second test. The reason for this is that the sum of all questions for the three tests adds up to 20, the maximum for the Portuguese scale. All the questions are equivalent, and with the same value, as the ones in the final assessment exam. The third test (Test 3) had just three questions about Linear Programming: the first is to formulate a problem, the second is to solve a linear programming problem with two variables using the graphical method, and the last one is to solve a linear programming problem using the Simplex method. Each of the questions is worth 1.7 points, which adds up to 5.1. All the questions are multiple-choice questions and if the student fails the answer, we subtract 1/3 of the question mark. With this option, for test two we can obtain results in the interval [-2.7, 8.1], while for test three the possible results are in the interval [-1.7, 5.1].

The second assessment test (Test 2) was scheduled for April 8, 2019 and so the students started their attempts on *MatActiva* Test 2 (MA2) on March 30, and the last one was the day of the test. 118 students took the assessment test, but just 16 of them used the *MatActiva*.

The third test (Test 3) was scheduled for May 15, 2019 and so the students started to use the *MatActiva* Test 3 (MA3) on May 12. There were 96 students in the classroom test, as, at this time, some already gave up the course. But the number of students that used the *MatActiva* platform increased to 33, and all completed Test 3. This means that the students considered that they have profited from the previous experience in the second test and wanted to repeat it and spread the word among the colleagues.

Several statistical tools were used to analyse the results of this experiment. Starting with descriptive statistics to obtain a broad picture of the data, the analysis progresses searching for association between each question in *MatActiva* tests and the corresponding one in the assessment tests in class, computing Pearson's correlation. Grades obtained in *MatActiva* tests and in the assessment tests in class are compared using a paired samples t-test and a Wilcoxon signed rank test for paired samples.

Looking for differences between final grades obtained in the course by the students that used the *MatActiva* tests and those that did not use it, the Welch two sample t-test and Mann-Whitney-Wilcoxon rank sum test were performed.

The results obtained are analysed and discussed in detail in the next section.

## 3 RESULTS

Concerning the second test, there were 4 students that logged in *MatActiva* but did not answer any question. Some statistics regarding the results obtained by the students that used the *MatActiva* tests can be seen in Table 1. Each student tried to solve MA2 between 1 and 13 times, spending a total time of 31741 seconds (8 hours and 49 minutes), on average. The classification obtained in the best attempt of each student showed a mean of 6.28 points with a standard deviation of 2.14 points.

Only two of the 16 students that completed the MA2 on *MatActiva* did not use them for the third test. The average number of attempts that these 14 students tried on *MatActiva* was 5. There were some students that just did it once, and there were some students that did it many times. As the tests are random, and with many different questions, the students hardly would get the same test again. As shown in Table 1, the number of trials ranged from 1 to 26 trials, the mean of the total time spent was 7864 seconds (2 hours and 11 minutes), with the best classification achieving a mean of 3.71 points and standard deviation of 1.56 points.

*Table 1. Statistics of the results from both tests.*

| | Count | Minimum | Mean | Median | Maximum | Std. Deviation |
|---|---|---|---|---|---|---|
| *MatActiva* Test 2 Number of Trials | 16 | 1.00 | 5.13 | 4.00 | 13.00 | 3.93 |
| *MatActiva* Test 2 Total Time | 16 | 1574 | 31741 | 12101 | 179357 | 46230 |
| *MatActiva* Test 2 Best | 16 | 1.20 | 6.28 | 6.40 | 8.10 | 2.14 |
| Test 2 Grade | 16 | 3.40 | 6.01 | 6.40 | 8.10 | 1.48 |
| | | | | | | |
| *MatActiva* Test 3 Number of Trials | 33 | 1.00 | 4.97 | 3.00 | 26.00 | 5.21 |
| *MatActiva* Test 3 Total Time | 33 | 143 | 7864 | 3478 | 66171 | 13258 |
| *MatActiva* Test 3 Best | 33 | 0.58 | 3.71 | 3.40 | 5.10 | 1.56 |
| Test 3 Grade | 33 | -1.70 | 2.82 | 2.3 | 5.10 | 1.59 |

When we look at the time the students spent in the *MatActiva* tests, we conclude that the longer they spent in the platform, the lower is the final grade in the continuous assessment (confirmed by the negative correlations showed in Table 2 and Table 3). This can be strange at first sight, but it makes sense, as the good students made one attempt and succeeded and did not bother to make a second one [14]. So, the more attempts the student makes, the lower is the final grade and the result of the third test.

We have built a table with the best attempt for each student and it showed that everybody got the question on eigenvalues right, even though not all of them got it right on the test.

Another curious fact is that all the students that did the *MatActiva* tests got the Cryptography question right in the classroom test.

We analysed if the results on each question on the *MatActiva* tests correlate to the results on the corresponding question on the classroom tests, but only the question about Input-Output analysis shows a strong positive correlation ($R$=0.7819) and the Simplex question shows a weak correlation ($R$=0.3805). The remaining correlations are near zero, or impossible to compute due to all students having responded it correctly (zero variance).

*Table 2. Pearson Correlations on Test 2.*

| | Total Time (sec) | MA2 Grade | No. Trials |
|---|---|---|---|
| Final Grade | -0.6312 | 0.2896 | -0.3735 |
| Test 2 Grade | -0.1336 | 0.3830 | -0.0617 |

*Table 3. Pearson Correlations on Test 3.*

| | ***Total Time (sec)*** | ***MA3 Grade*** | ***No. Trials*** |
|---|---|---|---|
| Final Grade | -0.2494 | 0.0181 | -0.1218 |
| Test 3 Grade | 0.0697 | 0.0683 | -0.1083 |

We have also tried to compare the best grade obtained on *MatActiva* MA2 with the grade obtained on Test 2, as can be seen in Figure 1. There is a weak positive correlation, with *R*=0.3830, showing that a high grade on Test 2 may somehow be associated with good results on the MA2. The same kind of comparison was done for MA3 test, but the association of both grades was not so evident (*R*=0.0683).

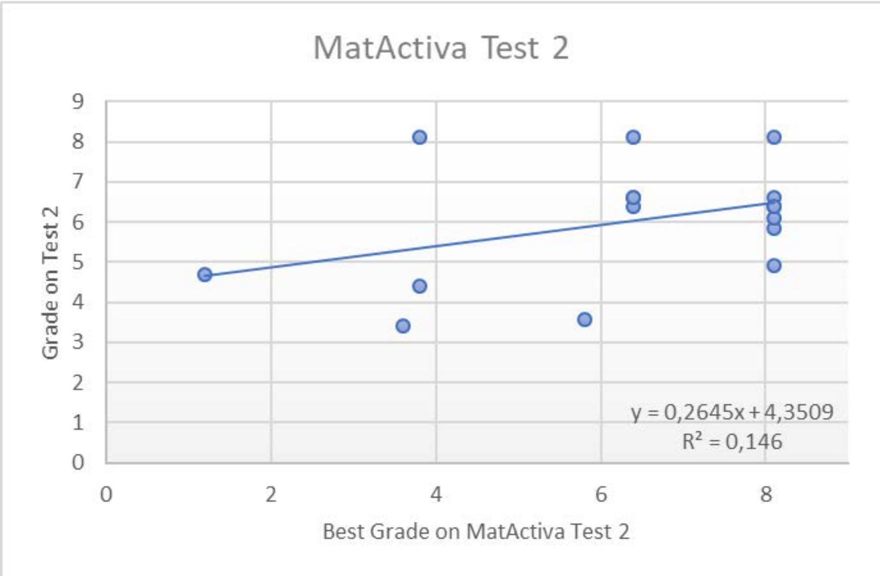


*Figure 1. Scatter plot of grade on Test 2 versus best grade on MatActiva Test 2.*

By converting the grades of both tests on *MatActiva* to a [0, 20] scale, we were able to compare the grades on each attempt, on the actual tests, and the final grade of the course.

Graphs of the evolution of some sample students are presented in Figure 3 to Figure 6. The first points in each graph are the results obtained by the student in each trial of MA2 on *MatActiva*, the first star is the grade obtained in the classroom test 2 (Test 2). The following points are the results in the trials of MA3 on *MatActiva*, the next star represents the result obtained by the student in the classroom test 3 (Test 3) and the final star is the final grade of the course.

For the 14 students that tried to solve both *MatActiva* tests, we compared the best grade obtained on tests 2 and 3 (see Figure 2). The mean of the best result in MA2 was 15.6 and in MA3 it was 13.3 (in a scale of 0 to 20). Using a paired samples t-test, we found that there are no significant differences between these two grades, given that a t-value of 2.16 and a p-value of 0.26 were obtained. The mean number of trials in MA2 was 5 and in MA3 was 4.5. A paired samples t-test for testing the number of trials in both tests presented a-t value of 2.16 and a p-value of 0.52, meaning that there are no significant differences between the number of trials in each of the tests for these students.

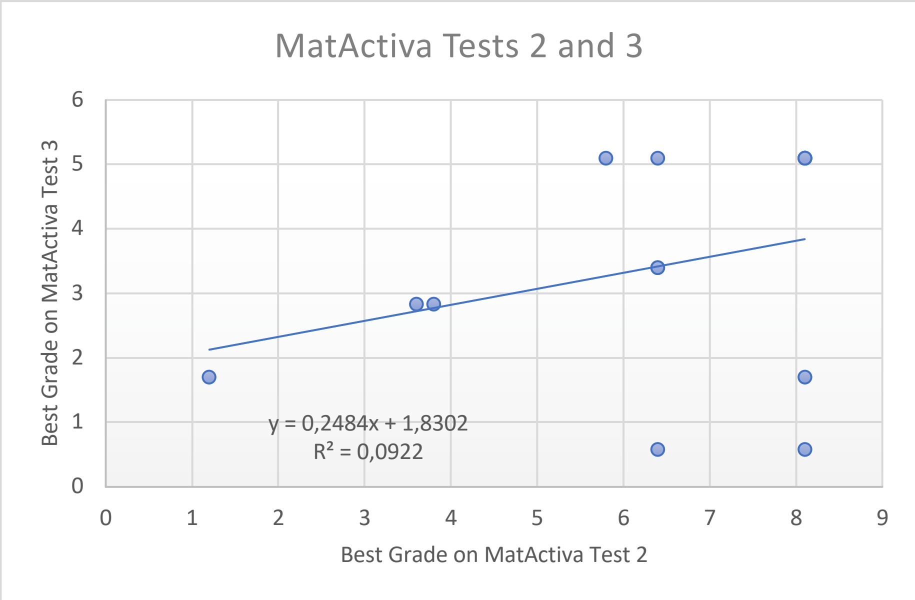


*Figure 2. Scatter plot of grades obtained on MatActiva tests.*

Given the size of the sample of students that used *MatActiva* MA2 test (16 students), we used the Shapiro-Wilk test to check for normality and found that the best grade on MA2 test fails to have a normal distribution (p-value=0.004), forcing to resort to non-parametric tests for this case, but the grade on classroom Test 2 may be considered to follow a normal distribution (p-value=0.174). The normal distribution is often used as an approximation to the distribution of values in a discrete population [15]. Although the number of trials is intrinsically a discrete variable, given its range and the results of the Shapiro-Wilk test (p-value=0.065), we can assume normality also.

Testing for differences between the grades in the *MatActiva* test and the corresponding classroom test showed that, in the second tests, there are no significant differences between the best grade obtained in the *MatActiva* MA2 test and the grade in the classroom Test 2 (Wilcoxon signed rank test p-value=0.489). However, for Test 3, we found that, for a 5% significance level, there are significant differences between the best grade on *MatActiva* test and the classroom Test 3 grade (t=2.037 and p-value=0.024). The mean result for the best trial on *MatActiva* MA3 test was 2.8 and in the classroom Test 3 was 3.7 (on a scale of 0 to 5.1).

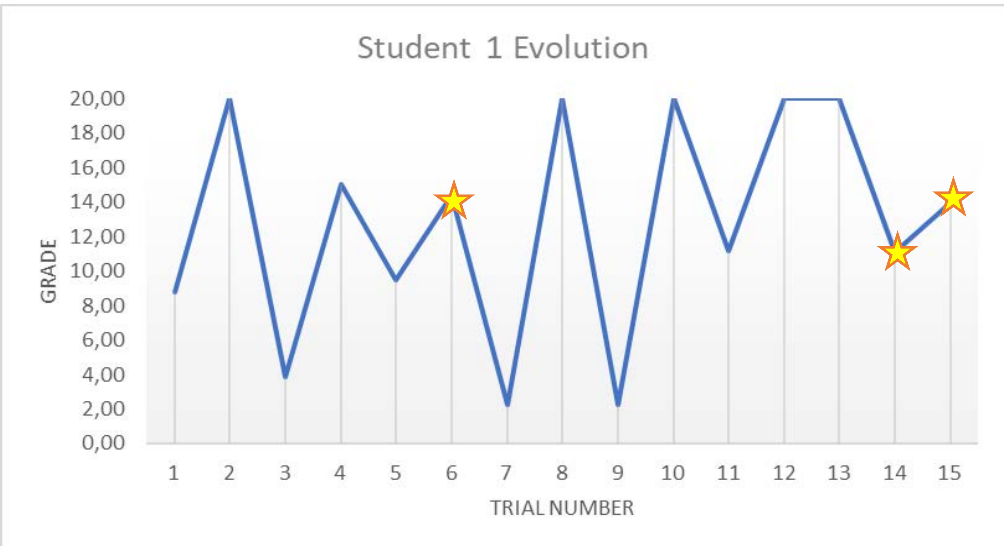


*Figure 3. Student 1 evolution.*

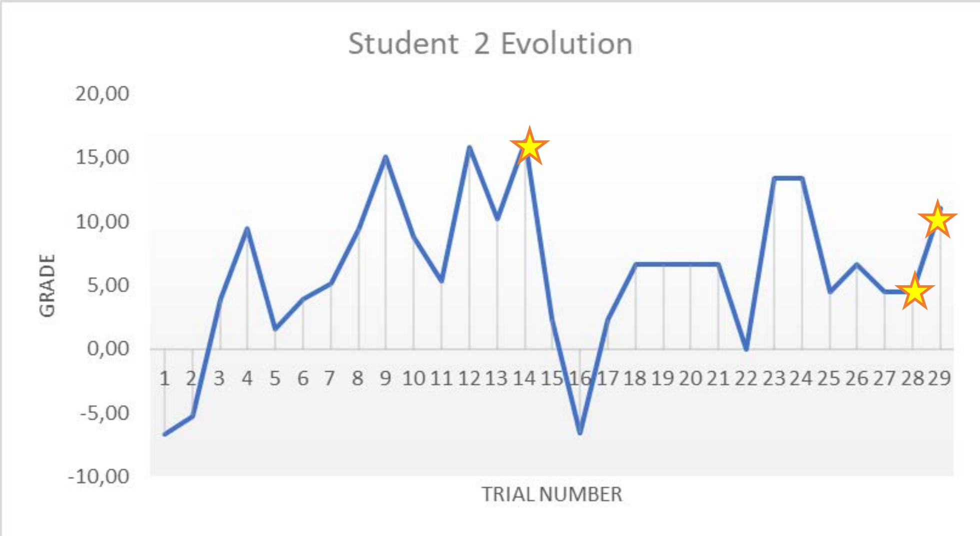


*Figure 4. Student 2 evolution.*

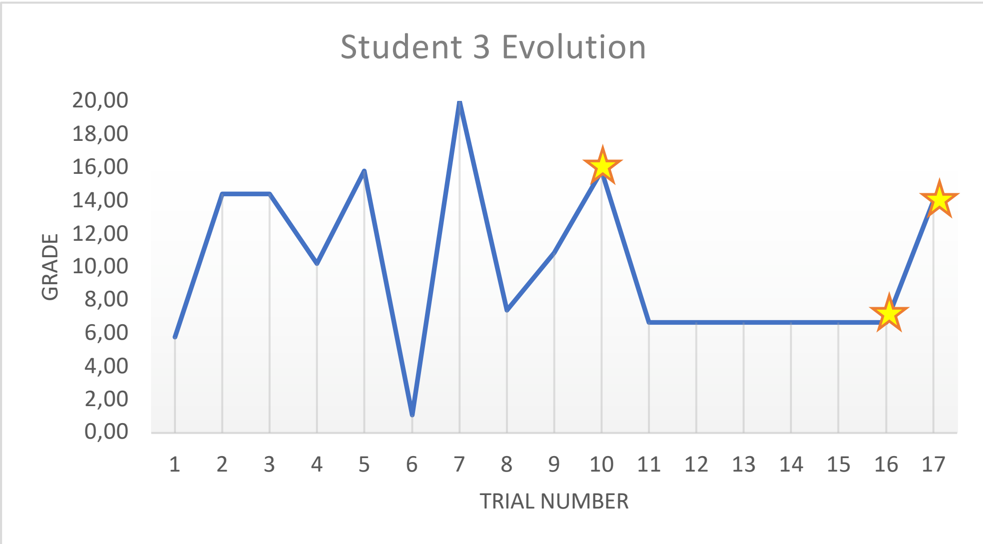


*Figure 5. Student 3 evolution.*

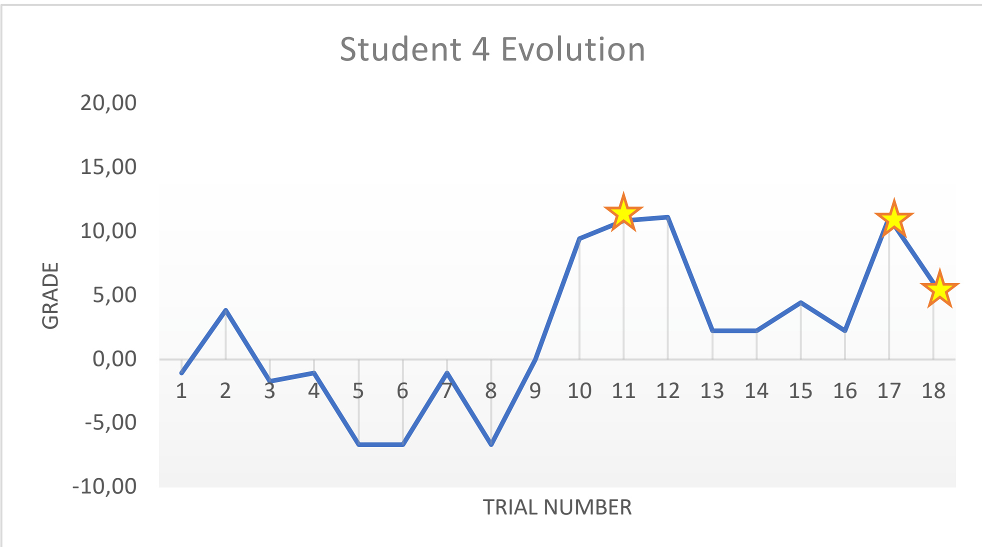


*Figure 6. Student 4 evolution.*

In Figure 7, a comparison is made between the final grades of the course achieved by students that used *MatActiva* tests and students that did not use them. The grades of the students that used *MatActiva* tests are generally higher and more concentrated around a positive classification (approved in the course), while students that did not use *MatActiva* show a higher variety of grades but with a very low median grades (No *MatActiva* Test2 Q2=6; No *MatActiva* Test3 Q2=3; No *MatActiva* Both Tests Q2=6.5), which correspond to negative classifications (more than 50% of these students fail the course). The first quartile of these boxplots (*MatActiva* Test2 Q1=12.75; *MatActiva* Test3 Q1=10; *MatActiva* Both Tests Q1=12.25) also shows that 75% (or more) of the students that used *MatActiva* were approved in the course.

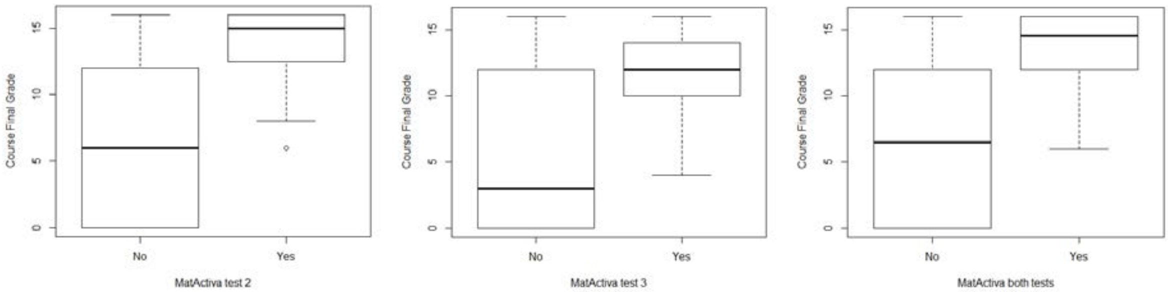


*Figure 7. Boxplots of the course final grade according to the groups of students that used the MatActiva tests 2, 3 and both tests.*

Students that used the *MatActiva* tests achieved a final grade in the course significantly higher than the students that did not use *MatActiva*. This is the conclusion drawn from Table 4, where the results of the F test for comparing variances in two samples and the Welch two sample t-test along with the Wilcoxon rank sum test for comparing differences in the course final grades of the two groups (students that used *MatActiva* tests and students that did not use it). The results of the Shapiro Wilk normality test show that only the group of students that responded to *MatActiva* MA3 test can be considered to have course final grades with normal distribution. Given the large sample of students that did not use *MatActiva* tests, the parametric F test and t-test may be used for this case. In the two remaining cases, the non-parametric Wilcoxon rank sum test should be considered instead of the t-test. In the F test, the null hypothesis is rejected for a 5% significance level, meaning that there can be significant differences between the variances in the two groups of students who responded to *MatActiva* MA3 test. Therefore, we proceed using the Welch t-test for two samples with different variances. In the Welch t-test performed, the null hypothesis is also rejected for a 1% significance level, meaning that there are significant differences in the mean of the course final grade for students that used the *MatActiva* MA3 test and for students that did not use it. Analogously, the results of the Wilcoxon test for the remaining two cases present p-values lower than 5%, meaning that there are significant differences between the median final course grades of students who used *MatActiva* MA2 test (or both tests) and the students that did not use it.

*Table 4. Comparison of the course final grades in the groups of students that used MatActiva.*

| | | ***MatActiva Test2*** | | ***MatActiva Test3*** | | ***Both MatActiva Tests*** | |
|---|---|---|---|---|---|---|---|
| | | ***No*** | ***Yes*** | ***No*** | ***Yes*** | ***No*** | ***Yes*** |
| Mean of the course final grade | | 6.384 | 13.625 | 5.921 | 11.848 | 6.500 | 13.357 |
| Standard deviation of the course final grade | | 5.908 | 3.008 | 6.055 | 3.114 | 5.959 | 3.128 |
| Number of students | | 156 | 16 | 139 | 33 | 158 | 14 |
| Shapiro Wilk test | p-value | $7.5*10^{-12}$ | 0.002 | $5.3*10^{-12}$ | 0.089 | $6.7*10^{-12}$ | 0.008 |
| Wilcoxon rank sum test | W | 369 | | 1095.5 | | 371.5 | |
| | p-value | $2.5*10^{-6}$ | | $2.2*10^{-6}$ | | $2.9*10^{-5}$ | |
| F test | F | 3.857 | | 3.782 | | 3.629 | |
| | p-value | 0.005 | | $4.5*10^{-5}$ | | 0.011 | |
| Welch t-test | t | -8.149 | | -7.938 | | -7.135 | |
| | p-value | $5.8*10^{-9}$ | | $3.6*10^{-12}$ | | $3.3*10^{-7}$ | |

Using a comparison with t-tests to identify causality, in an observational study such as this, must be made with caution. As stated by Devore [15], although statistical analysis may indicate a significant difference in response between the two groups, the difference may be due to some underlying factors that had not been controlled, rather than to any difference in treatments. Our results may mean that the use of *MatActiva* prepared the students better for the classroom tests, but it may also be due to a factor we could not control: are the students that choose to use *MatActiva* the students that usually study more, and therefore achieve better results? As the tests have students' names, and as we know the students, we can confirm that some of the students that have used *MatActiva* were, in fact, the students that are always trying to find more things to help them succeeding, but there was also a group of students that have difficulties with Maths that have used the platform and gave us positive feedback about their experience.

## 4 CONCLUSIONS

The present study described and analysed an experiment that took place in the context of a math course from the Marketing degree. Two tests (MA2 and MA3) were constructed with random multiple-choice questions supported by *MatActiva* web site. The MA2 test had questions on Linear Algebra and the MA3 test had questions on Linear Programming and the students could practice, as many times as they want, at their own pace, to be prepared for the assessment. All questions have the solution step by step, thus the students can realize what is expected to do. These two online tests are similar to those from continuous assessment performed in class (Test 2 and Test 3).

The statistical analysis performed allows concluding that, when investigating if the results on each question on the *MatActiva* tests correlate to the results on the corresponding question on the classroom tests, only the question about Input-Output analysis shows a strong positive correlation. Also, no significant differences were found when comparing grades obtained in MA2 and Test 2. Instead, significant differences were found in the case of MA3 and Test 3 where the students achieved better marks in Test 3 than in MA3. It is important to mention that, although a small number of students completed the *MatActiva* tests, this number increased from MA2 to MA3, showing that the students considered that they have profited from the experience in the second test and wanted to repeat it and spread the word among the colleagues.

Another important conclusion refers to the final grades of the students who used the *MatActiva* tests. In general, these students' grades are higher and more concentrated around a positive classification (approved in the course), while students that did not use *MatActiva* show a higher variety of grades but with a very low median grades which correspond to negative classifications (more than 50% of these students fail the course). Also, 75% (or more) of the students that used *MatActiva* were approved in the course. It seems that, by using the *MatActiva* tests, students were better prepared, being able to get better results than those students that did not, but it may also be due to a factor we could not control: are the students that choose to use *MatActiva* the students that usually study more for the tests, and therefore achieve better results? Are these students more motivated and committed to mathematics and learning? These are pertinent questions that we will address in future work.

## REFERENCES


[1] T. Benadé, J Liebenberg, "The relevance of a Mathematics course for computer science students" in *Swan Delta Proceedings: The 12th Delta Conference on the Teaching and Learning of Undergraduate Mathematics and Statistics*, pp. 2-10, Australia: Fremantle, 2019.

[2] K.Luu, J. G. Freeman, "An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia". *Computers & Education*, vol. 56, no.4, pp.1072-1082, 2011.

[3] M.Koskela, P.Kiltti, I.Vilpola and J.Tervonen, "Suitability of a Virtual Learning Environment for Higher Education" *The Electronic Journal of e-Learning,* vol. 3, no. 1, pp. 23-32, 2005. Retrieved from www.ejel.org

[4] S. Ko and S. Rossen, *Teaching Online. A Practical Guide (4th)*. New York, NY, US: Routledge/Taylor & Francis Group, 2017.

[5] J. M. Azevedo, C. Torres, A. P. Lopes, and L. Babo, "Learning Analytics : A Way to Monitoring and Improving Students' Learning," in *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017)*, vol. 1, pp. 641–648, 2017.

[6] C. Torres, A. P. Lopes, L. Babo, J. Azevedo. "*MatActiva* Project – A mathematical dynamic environment to engage students in the learning process". In *The e-Learning Excellence Awards 2017: An Anthology of Case Histories* (D. Remenyi ed.) pp. 125-138. UK: Academic Conferences and Publishing International Limited, 2017.

[7] A. Espasa and J. Meneses, "Analysing feedback processes in an online teaching and learning environment: An exploratory study," *High. Educ.*, vol. 59, no. 3, pp. 277–292, 2010.

[8] C. Evans, "Making Sense of Assessment Feedback in Higher Education," *Rev. Educ. Res*., vol. 83, no. 1, pp. 70–120, 2013.

[9] G. Ion, A. Barrera-Corominas, and M. Tomàs-Folch, "Written peer-feedback to enhance students' current and future learning," *Int. J. Educ. Technol. High. Educ*., vol. 13, no. 1, 2016.

[10] T. Lunt and J. Curran, "'Are you listening please?' The advantages of electronic audio feedback compared to written feedback," *Assess. Eval. High. Educ*., vol. 35, no. 7, pp. 759–769, 2010.

[11] G. Singh, & G. Hardaker, "Teaching in Higher Education Change levers for unifying top-down and bottom- up approaches to the adoption and diffusion of e- learning in higher education". *Teaching in Higher Education*, vol. 22. no.6, pp. 736–748, 2017. Retrieved from https://doi.org/ 10.1080/13562517.2017.1289508.

[12] D. R. Sadler, "Beyond feedback: Developing student capability in complex appraisal," *Assess. Eval. High. Educ*., vol. 35, no. 5, pp. 535–550, 2010.

[13] A.P. Lopes, L. Babo, J.Azevedo, C.Torres, "Rethinking feedback in a connected age" in *INTED2018 Proceedings*, pp. 5273–5282, Valencia, Spain: IATED, 2018. Retrieved from http://doi.org/10.21125/inted.2018.1245.

[14] J. Kaminskiene and D. Rimkuviene, "Sharing the Experience of Teaching Mathematics with Moodle" in *18th International Conference Teaching Mathematics: Retrospective and Perspectives*, pp. 30–36, 2018.

[15] J.L. Devore. *Probability and Statistics for Engineering and the Sciences*. 8th edition. Boston, MA: Cengage Learning; 2011.