A Case Study of a Probability and Statistics Course: Improving Students' Performance with Formative Assessment

Anna Carolina Finamore[1], Ana Moura Santos[2], Bernardo Pereira Nunes[3,4] and António Pacheco[1]

[1] CEMAT and Dept. of Mathematics, Instituto Superior Técnico, Portugal
[2] CEAFEL and Dept. of Mathematics, Instituto Superior Técnico, Portugal
[3] Dept. of Informatics, PUC-Rio, Rio de Janeiro, RJ, Brazil
[4] Dept. of Applied Informatics - UNIRIO, Rio de Janeiro, RJ, Brazil

anna.couto@ist.utl.pt, ana.moura.santos@tecnico.ulisboa.pt, bnunes@inf.puc-rio.br, apacheco@math.ist.utl.pt

Abstract. This paper presents a case study of a regular course in Mathematics, offered every academic term to about 1,500 STEM students of the *Instituto Superior Técnico* of Lisbon. The main concerns for this course are the low approval and high dropout rates. A new approach, called Online Electronic Quizzes with Random Parameters, based on formative assessment was introduced to run alongside the already adopted summative assessment. The ultimate goal of our work was to map and understand students' weaknesses and then invite professors and staff to discuss and suggest solutions to support students' learning. Preliminary results show that the proposed approach contributed to increase students' performance and approval rates in approximately 11% of the cases.

Keywords: Formative Assessment, Quizzes, Case Study, STEM Education.

1 Introduction

This paper describes a case study of the Probability and Statistics (PS) course at the *Instituto Superior Técnico* of Lisbon (IST)[1]. Briefly, PS is a regular course offered every academic term by the Department of Mathematics (DM) to about 1,500 students from the Departments of Engineering and Architecture and has one of the lowest IST approval rates. Based on internal reports and data provided by the IST Statistics and Prospective Center, the proportion of approved and enrolled students between 2008 and 2014 was only 35%, whereas student's drop out reached over 50% in some academic terms – a very alarming rate of students who either fail or abandon this course.

[1] https://tecnico.ulisboa.pt/en/

Given this negative scenario, both the Pedagogical Council and the DM of IST proposed a new pedagogical strategy to improve students' performance. They noted that the evaluation of PS students has been merely based on summative assessment [3], that is, composed of two exams to evaluate the level of success of students after a set of instructional units. Based on this scenario and motivated by previous experiences in courses such as Linear Algebra and Complex Analysis and Differential Equations at IST, the Department of Mathematics and the Pedagogical Council joined efforts to implement formative assessment in PS [1, 2, 3] as an attempt to improve students' performance. The formative assessment is based on Online Electronic Quizzes with Random Parameters that are graded and stored in the IST Learning Management System Fénix[2]. With the data produced by the students using Fénix, we conducted an analysis [7] to assist PS staff to realign and develop strategies to improve and understand the low approval rates.

This paper presents a case study of the PS course before and after applying formative assessment as an attempt to improve students' approval rates, identify students who need support at an early stage, and enhance the teaching and learning process.

The remainder of this paper is organized as follows. Section 2 reviews existing assessment methodologies. Section 3 presents a case study and the proposed assessment strategy, Section 4 describes preliminary results. Section 5 provides our conclusions.

2 Related Work

Higher Education Institutions (HEIs) have implemented new methodologies in order to enhance teaching and learning, as well as to keep in constant alignment with students' progression and needs. Several methodologies are under investigation to monitor students' development through various methods of assessment, such as diagnostic [8, 11], formative, summative, stealth [9], and game-based [10, 11] assessments. When supported by digital technologies, these methodologies can occur both in synchronous and asynchronous ways [12].

While summative assessment [3] has a qualification purpose, verifying students' achievement after the learning process had occurred, formative assessment has a transformative purpose, providing information during the course about the students' understanding of the taught topic to constantly improve the teaching and learning process.

For this reason, providing and receiving effective feedback is crucial both for students and professors. Learners are made aware of their progress, and instructors identify students' needs and difficulties, and can devise strategies for improving students' performance and their practices.

Formative and summative assessment can also be combined [3] to obtain a balance between the grading and the teaching and learning purposes. Both types of assessment, when combined, foster deeper understanding, since they promote a constant contact with the addressed topics through the frequent assessment moments and feedback. At the same time, this combination promotes the instructor training [3] through the balance

[2] http://fenixedu.org

and timing between different classes – for instance in large courses, where multiple classes are running in parallel.

In the case study described below, we applied formative assessment to improve the teaching and learning process [1, 2]. The main reason for this choice was our previous positive experience of applying this method in the context of other IST's courses, carried out alongside a shift in the pedagogical focus from the instructor-centered to the student-centered approach [4].

3 Case Study: Introducing Formative Assessment in PS

Course Overview

PS is a regular course, offered every semester[3] by the Department of Mathematics at the *Instituto Superior Técnico* of Lisbon. The course is composed of 6 ECTS[4] credits, divided into three classes per week of 1.5 hours each, and one class is devoted to discuss and solve only theoretical exercises. Moreover, each professor – 10 in total for around 1,500 students per semester – offers 1.5 extra hours per week (as their office hours) to answer individual students' questions. Students can attend all the office hours they want to. This is approximately 15 additional hours dedicated to a more personal contact with the PS professors.

As previously mentioned, the PS course approval rate has been low for several consecutive years. Figure 1 and 2 present respectively the rate of approved and enrolled students for 6 different degrees as well as their final status. The proportion of approved and enrolled students between 2008 and 2014 was only 35% (Figure 1) whereas students drop out rates reached over 50% in some academic terms (Figure 2). The figures show the analysis for the following 6 degrees: *Industrial Engineering and Management (IEM)*, *Applied Mathematics and Computation (AMC)*, *Computer Science (CS)* which is available in both campuses: *Alameda (CS-A)* and *TagusPark (CS-T)*, *Biomedical Engineering (BE)* and *Electrical Engineering (EE)*.

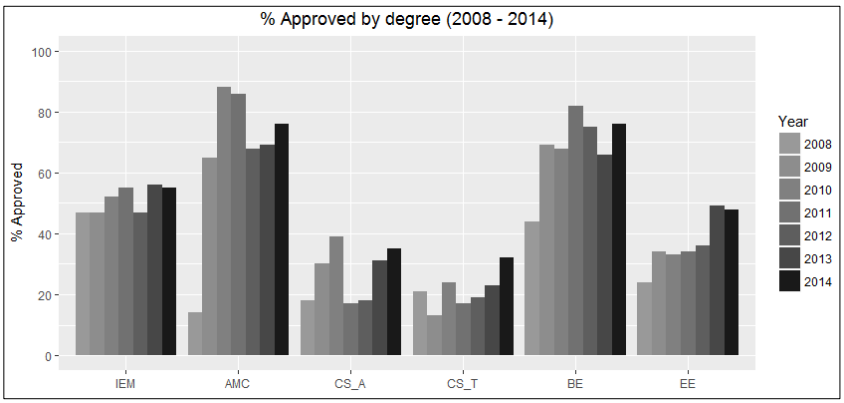

Figure 1. Rates of approved vs enrolled students.

[3] https://fenix.tecnico.ulisboa.pt/cursos/lmac/disciplina-curricular/1529008374337

[4] http://ec.europa.eu/education/resources/european-credit-transfer-accumulation-system_en

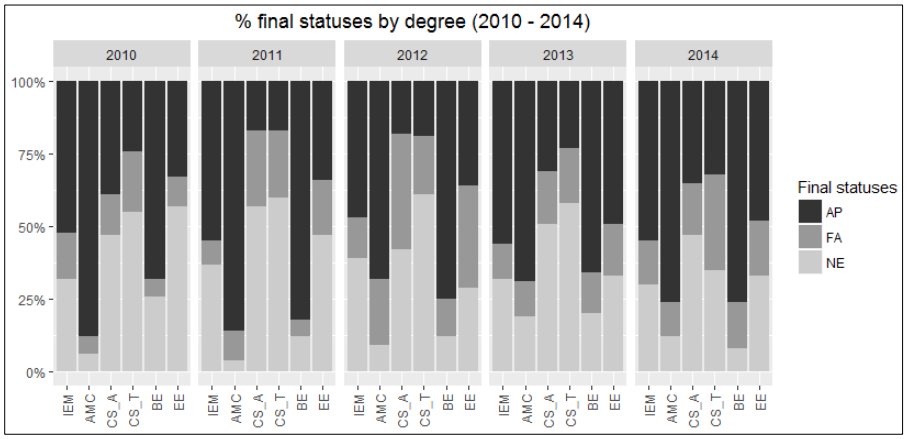

Figure 2. Rates of students' final statuses: Approved (AP), Failed (FA) and Not Evaluated (NE).

Formative Assessment Strategy

Until the first academic term of 2016/2017, only summative assessment had been used as a method to evaluate students in PS courses. Once the pattern of low students' grades and high dropouts' rates (for different cohorts for consecutive years) were observed, formative assessment was implemented in an attempt to identify specific topics students were struggling with. This was done in an effort to improve PS teaching and learning strategies. Formative assessment was introduced in the PS course using Online Electronic Quizzes with Random Parameters or, in short, Quizzes.

Development of the Quizzes started in 2015. They were based on a corpus of written exams produced in the five previous years, and using the Mathematica[5] programming language [5]. Currently, 100 exercises covering 23 different topics are available for professors who want to create Quizzes for educational purposes. Professors can create their own Quizzes by selecting the topics to be covered, as well as the difficulty level of the questions to be included: this is determined based on the average time taken by the students to complete the task, and is validated by a small group of experts in probability and statistics. The levels range from "very easy" to "very difficult". Moreover, each professor takes pedagogical actions they judge necessary according to the performance of the students in the quizzes.

Although the Quizzes are not mandatory, each student receives a unique and randomly generated quiz automatically distributed by Fénix [5,6]. The students have one week to complete the quiz, and can resubmit their answers as long as the due date has not passed; only the last submission is considered for grading. After the due date, the correct answers are displayed to the students with their original answers (correct or not). Additionally, a grade is awarded based on the percentage of correct answers.

[5] https://www.wolfram.com/mathematica/

Figure 3 illustrates an example of a randomly generated exercise. The red boxes indicate the random parameters. Each parameter and its range are predetermined by the exercise creator and automatically generated by the system.

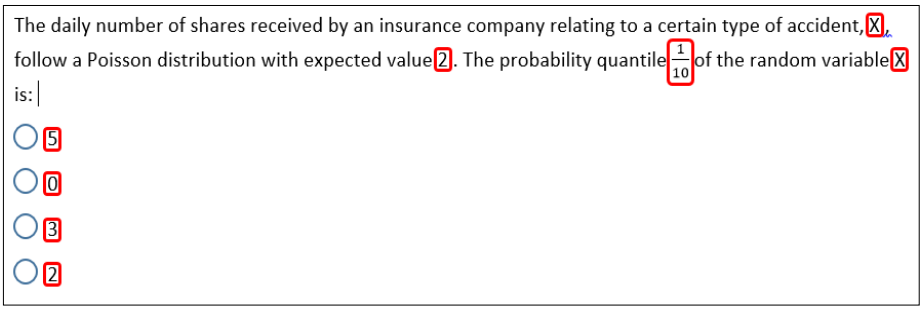

Figure 3. Example of an online PS exercise (in red, the random parameters).

Approval Criterion

The approval criterion before the introduction of the Quizzes was based on the average of two exams (T1 and T2) following a 20-point grading. To be approved, students had to obtain a grade higher than 7.5 in each test, and an overall average (equally weighted) higher than 10. Optionally, students could retake a test, or both tests (RT1 and RT2), to improve their grades. The latest tests are only considered if they effectively improve the test grade. So, the final grade (FG) is calculated by the average between the two best achieved grades:

$$FG = AVG(MAX(T1, RT1) ; MAX(T2, RT2)) \qquad (1)$$

Note that, if a student misses one of the tests, and the respective retake test (T1/RT1 or T2/RT2), his/her final status is considered as "Not Evaluated (NE)"; if the minimum approval criterion is not achieved, then the final status will be "Failed (FA)". In both cases, the student will need to fully retake the course in a subsequent academic term.

After introducing the Quizzes, the approval criterion was slightly adapted to include it in the final grade. However, following the same strategy of the retake tests, the Quizzes are only considered if the final grade improves. Moreover, only the *n-1* Quizzes with better grades are considered in the final grade and no penalty is given to students who did not take the Quizzes. The final grade taking into account the Quizzes is calculated as:

$$FG' = MAX(FG , (0.9*FG + 0.1*QuizzesG)) \qquad (2)$$

where QuizzesG is the average of the *n-1* quizzes with better grades taken by the student.

The intention behind the new approval criterion (FG') is that students use it as an additional source of study and a safe way to improve their final grade. Moreover, professors can use the results of the Quizzes to early identify learning issues and the most difficult topics to their students.

4 Analysis and Results

This section presents an analysis of two editions of the PS course, one using summative assessment and the other using both formative and summative assessment. The goal of this analysis is to verify whether or not the inclusion of formative assessment in the PS course improves the proportion of students' performance. In total, considering the two academic terms, 2,444 students from 18 degrees, where PS is mandatory, were considered in this analysis.

Analysis of Students' Performance

The analysis of students' performance was based on their final status: Approved (AP), Not Evaluated (NE) and Failed (FA). Table 1 indicates that formative assessment contributed to a slight improvement of the final status for a proportion of PS students. The number of approved students increased from 46% to 51% – an improvement of approximately 11%. Consequently, fewer students failed or were not evaluated in PS. The number of students who failed PS decreased from 14% to 12% – an improvement of approximately 15% – and students who were not evaluated (i.e. abandoned PS course) decreased from 40% to 37% – an improvement of almost 8%.

An hypothesis test of proportion was conducted (H_0: $p_{2015/2016} \leq p_{2016/2017}$; H_1: $p_{2015/2016} > p_{2016/2017}$) and the p-values were calculated: NE (p-value = 0.0654) and FA (p-value = 0.0736). Another hypothesis test of proportion for the AP was also conducted (H_0: $p_{2015/2016} \geq p_{2016/2017}$; H_1: $p_{2015/2016} < p_{2016/2017}$) and the p-value was 0.0070. So, for the final statuses NE and FA, with 10% of significance level both statuses in 2015/2016 are higher than both statuses in 2016/2017. For the AP status, we rejected H_0 for all the usual levels of significance 1%, 5% and 10%.

Table 1 – Proportion of the final status of PS students.

Semester	Assessment Type	Approved	Not Evaluated	Failed
1st sem. 2015/2016	Summative	46%	40%	14%
1st sem. 2016/2017	Summative + Formative	51%	37%	12%

Analysis of the Quiz Adoption

Due to the noncompulsory nature of the Quizzes, an analysis of its adoption by the students was carried out. The main goal was to validate students' acceptance of the new proposed assessment strategy, as well as to understand students' frequency of study and how this information can assist professors on improving their teaching methodology.

Table 2 shows the percentage of students who voluntarily submitted the Quizzes. In some courses, such as *Applied Mathematics and Computation* (AMC), the adoption is very high, reaching 83%. In others, such as Computer Science - *Alameda* (CS-A), the adoption rate was low, reaching only 11%-13%. Examining the first three Quizzes related to the first test (T1), we see a range of 83% to 32%, a very different result compared with the last three Quizzes related to the second test (T2), which ranged from 77% to 11%. The number of quizzes was determined by the number of topics addressed in each test, in our case, there were six set of topics resulting in the creation of six

quizzes. A comparison of the median grades obtained by the students who participated in all Quizzes and the students who did not, is shown in Figure 4. Note that we use median instead of average as it is less sensitive to outliers. It is clear that students who participate in all quizzes obtained better grades in the tests.

Table 2 – % of submitted responses to quizzes during the 1st semester, 2016/2017.

	T1			T2		
Degree	**Submitted Quiz 1**	**Submitted Quiz 2**	**Submitted Quiz 3**	**Submitted Quiz 4**	**Submitted Quiz 5**	**Submitted Quiz 6**
IEM	80%	65%	66%	62%	47%	38%
AMC	83%	77%	74%	77%	74%	51%
CS-A	47%	31%	32%	23%	11%	13%
CS-T	67%	39%	35%	22%	13%	13%
BE	79%	74%	58%	31%	18%	17%
EE	49%	42%	35%	53%	32%	37%

There are a few possible reasons that can explain the low adoption by some students. One possible reason is the approval criterion. As the lowest grade is always discarded, the last Quiz is usually overlooked, as it will not contribute to students who obtained the highest grades in the first *n-1* quizzes. Another possible reason is the high demand for extra class activities by other courses. This is usually the case for Computer Science, where most of the courses are based on large projects and by the end of the academic term, these students prefer to be fully dedicated to their career-specific courses than to those offered by another department (in this case the Department of Mathematics).

Although there are many other reasons for the low adoption of the Quizzes, the importance of this analysis lies on the possibility for gaining an understanding of students' behavior, and identifying the topics they find most challenging throughout the academic term. For instance, the low submission rate may indicate that the Quiz was difficult or that students did not need them to improve their grade. Therefore, formative assessment can serve as a mean to inform professors about students' behavior and also to assist them in preparing and improving their teaching methodologies.

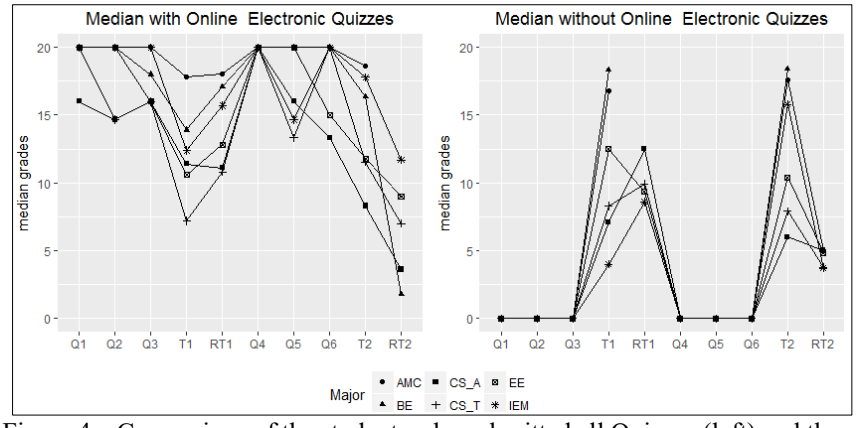

Figure 4 – Comparison of the students who submitted all Quizzes (left) and the ones who did not (right).

Students Questionnaire Analysis

A questionnaire of 14 closed-end and 5 open-ended questions was conducted in order to receive feedback from the students regarding the inclusion of the Quizzes in the PS course. In total, 168 students responded to the questionnaire. We report here only the questions related to the use and acceptance of the Quizzes:

- Did the Quizzes contribute to my exam preparation?
  - From the students who completed the questionnaire, 71% answered this question, and 75% said yes.
- Would you recommend the Quizzes to your friends in the next semester?
  - From the students who completed the questionnaire, 71% answered this question, and 88% said yes.

Finally, the open-ended question, asking "What was the perceived importance of the Quizzes for you?". Answers were classified into four groups: 89% of the students mentioned that the quizzes were useful for regular practice/study topics taught in classroom; 17% mentioned that it was a good opportunity to improve their grade; 3% mentioned that they liked the feedback obtained; and 2% mentioned that online exercises can be done remotely.

Analysis of the PS Grades

The final analysis conducted in this study aimed at measuring the impact of the Quizzes in the final grades of the students. That is, if the students' grades improved after the proposed assessment strategy has been applied. Figure 5 shows the average grade obtained by the students attending the Probability and Statistics (PS) course.

Students who participated in all Quizzes obtained better grades than the ones who did not participate. It is clear that there was a shift in the grades. After applying formative assessment, the number of students who obtained better grades increased. For instance, there was an increase of almost 35% in the number of students who achieved an average grade of 18. Moreover, the number of students who obtained lower grades (10-14) decreased. This is a clear indicator that regular activities and the right form of assessment may improve students' performance.

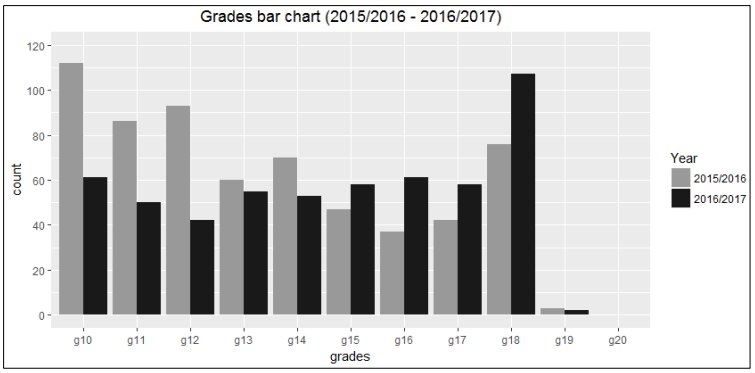

Figure 5 – Students' average grades (x-axis: grades, y-axis: number of students).

5 Conclusions

This paper presents a case study in the Probability and Statistics (PS) course at the *Instituto Superior Técnico* of Lisbon, with a concerning rate of passing and failing students. To tackle the problem of students' drop-out, a new strategy based on formative assessment, called Online Electronic Quizzes with Random Parameters, was introduced to improve students' performance. The proposed formative assessment approach helps students to maintain regular contact with the topics addressed in the specific course (PS). Moreover, the random parameters of the online exercises guarantee a large volume of different exercises, that is, a large volume of unique instances of the Quizzes for each student. This helps professors and staff to individually identify students' difficulties at an early stage, and also contribute to a collaboration purpose, since each exercise is different from student to student, and they can discuss the exercises and construct their knowledge together.

By introducing the formative assessment in conjunction with the summative assessment, we observed improvements. The first improvement is the students' general performance in the course. The approval rate increased of 11%, the not evaluated rate decreased almost of 8% and the failure rate decreased almost of 15%. Given the hypothesis tests conducted, we can conclude that relating to the approved (AP) status we have an impact by introducing the Quizzes. The second improvement shows the Quizzes adoption by the students, ranging from 83% to 11% according each degree. The difference regarding the adoption rates by some students/degrees can be explained by multiple factors such as: (i) students' drop out after T1, (ii) grades achieved in previous Quizzes are insufficient to improve the final weighted average (FG'), (iii) other specifics career courses' demanding activities, (iv) the difficulty level of the quiz (which indicates that students are having problems to understand some topics), etc. Punctual low adoption rate can be also explained by the exclusion of the lowest grade for the average of the Quizzes. The third improvement illustrates the shift occurred in the final grades. Before the Quizzes' implementation, the high concentration was mainly around grades ranging from 10 to 12 and, after the Quizzes' implementation, this concentration was shifted to grades ranging from 15 to 18 – we remind that IST's grades follow a 20-point grading. The growth, for instance, of the students who achieved 18 as the final grade was approximately 35%. We observed also that better grades are directly associated to the student participation in the Quizzes, which is conform with our idea that the implementation of regular study and assessment contributes both to map students' weaknesses and the improvement of their performances.

As future work, we intend to conduct a large-scale experiment to include formative assessment in all degrees at IST. To increase the approval rates, we also want to experiment other approval criteria, for instance, making the online quizzes mandatory and change the final contribution to the final grade. We also intend to collect and share professors' experiences and actions, based in the past outcomes of the quizzes, to enhance teaching and learning process in future academic terms. With these changes, we expect that students' participation increases and therefore their grades and final statuses.

ACKNOWLEDGEMENTS

The authors would like to thank the Centre of Statistics and Prospective and the Pedagogical Council from Técnico Lisboa by the technical transfer of data published in this paper. This work is partially supported by the Portuguese Foundation "Fundação para a Ciência e a Tecnologia", through the projects UID/Multi/04621/2013 and UID/MAT/04721/2013.

References

1. Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. Assessment & Evaluation in Higher Education, 37(6), 747-760.
2. Wheatley, L., McInch, A., Fleming, S., & Lord, R. (2015). Feeding back to feed forward: Formative assessment as a platform for effective learning. Kentucky Journal of Higher Education Policy and Practice 3 (2).
3. Glazer, N. (2014). Formative plus Summative Assessment in Large Undergraduate Courses: Why Both? International Journal of Teaching and Learning in Higher Education, 26(2), 276-286.
4. Ifenthaler, D. (2010). Learning and instruction in the digital age. In Learning and Instruction in the Digital Age. Springer US.
5. Finamore, A.C., Moura Santos, A., & Ribeiro, P. (2016) Fostering Stem Formative Assessment for Lifelong Learners. In ICERI2016 Proc. ISBN: 978-84-617- 5895-1, DOI: 10.21125/iceri.2016.1031.
6. Finamore, A.C., Moura Santos, A., & Pacheco, A. (2016). Probabilidades e Estatística: Como Conseguir uma Experiência de Aprendizagem Gratificante? In CNaPPES 16 Proceedings.
7. Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. Educational technology & society, 15(3), 42-57.
8. Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: A review of research in assessment and evaluation in higher education. Assessment & Evaluation in Higher Education, 41(7), 1008-1032.
9. Gibson, D., Coleman, K., & Irving, L. (2016). Learning Journeys in Higher Education: Designing Digital Pathways Badges for Learning, Motivation and Assessment. In Foundation of Digital Badges and Micro-Credentials (pp. 115-138). Springer International Publishing.
10. Bodnar, C. A., Anastasio, D., Enszer, J. A., & Burkey, D. D. (2016). Engineers at Play: Games as Teaching Tools for Undergraduate Engineering Students. Journal of Engineering Education, 105(1), 147-200.
11. Nunes, B. P., Nurmikko-Fuller, T., Lopes, G. R., Siqueira, S. W., de Campos, G. H., & Casanova, M. A. (2016). Treasure Explorers - A Game as a Diagnostic Assessment Tool. In Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on (pp. 217-221). IEEE.
12. Brierton, S., Wilson, E., Kistler, M., Flowers, J., & Jones, D. (2016). A Comparison of Higher Order Thinking Skills Demonstrated in Synchronous and Asynchronous Online College Discussion Posts. NACTA Journal, 60(1), 14.