e-Assessment in Mathematics Courses with Multiple-choice Questions Tests

José Manuel Azevedo
ISCAP, Polytechnic Institute of Porto, Rua Jaime Lopes de Amorim, Porto, Portugal

Keywords: E-Assessment, Continuous Assessment, Formative Assessment, Moodle, Multiple-choice Questions, Guidelines for MCQ.

Abstract: With the implementation of the Bologna Process several challenges have been posed to higher education institution, particularly in Portugal. One of the main implications is related to the change of the paradigm of a teacher centered education, to a paradigm that is student centered. This change implies the change of the way to assess courses in higher education institutions. Continuous and formative assessments emerged as the focus, catalyzed by electronic assessment, or e-assessment. This paper presents a case of the implementation of an e-assessment strategy, implemented in order to allow continuous, formative assessment in numerous mathematics classes using multiple-choice questions tests implement in Moodle open-source learning management system. The implementation can be considered a success.

1 INTRODUCTION

The Bologna Process officially started in June 1999 with the Bologna Declaration. This defines a set of steps to be followed by the European higher education systems, in order to build a European area of higher education globally harmonized. After a few years of its implementation in Portugal it revealed as a huge opportunity for the reorganization of higher education in Portugal, being that the polytechnics and universities faced major challenges. The emergence of a new paradigm, valuing the student as the central subject in the construction of their learning requires new pedagogical approaches. However, according to Redecker and Johannessen (2013), changes in teaching practices and learning processes can only happen when assessment also changes. Moreover, according to (Perrenoud 1993), to change assessment is to change school. Historically, assessment in higher education consisted in the application of final exams for each of the courses, the so-called final assessment. The Bologna process points out to another type of assessment, encompassing diverse forms of assessments carried out during the semester/academic year, the so-called continuous assessment.

The mathematic teachers at ISCAP, the school of Accounting and Administration of the Polytechnic Institute of Porto, started to work on the necessary adjustments, related to the Bologna Process, as soon as possible. Nevertheless, several constraints were present. With the necessity to adjust the different courses accordingly to the Bologna Process, began to arrive at ISCAP students who did not attend Mathematics in secondary education. Among these, account a significant number of students entering through the special type of access, named as "Access to Over 23 Years". Many of these students no longer study several years ago, thus have more difficulties. On the other hand, with the restructuring of the programs, the weekly duration aimed at the Mathematics courses has been significantly reduced, classes were numerous and it also became necessary to articulate the mathematics courses with other courses to provide, in a timely manner, the necessary mathematical foundations. Therefore, there was a need to implement new strategies and methodologies to support students, because it was very difficult to implement continuous assessment in such conditions.

Taking into account all these aspects, we began to develop a continuous e-assessment process, which includes the use of Multiple-Choice Questions (MCQ) tests, implemented in Moodle. The tests are randomly generated by Moodle, allowing that each student is presented with a different test. A bank of questions, divided into categories defined to allow

Azevedo J..
e-Assessment in Mathematics Courses with Multiple-choice Questions Tests.
DOI: 10.5220/0005452702600266
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 260-266
ISBN: 978-989-758-108-3

all tests to assess the same learning goals for all students who were making the test, was developed. The existence of a project named as "MatActiva" (Azevedo et al. 2009; Babo et al. 2008; Babo, Azevedo, Torres and A. Lopes 2010; Babo, Azevedo, Torres and A. P. Lopes 2010; Lopes et al. 2010; Lopes et al. 2011; Torres et al. 2009; Torres et al. 2011), whose overall objective is to help students improve their performance in mathematics using the interactivity features of Moodle, and that had our participation, served up as a catalyst.

This paper describes the fisrt phase of this continuous e-assessment process. The main contribution of this paper is to show that it is possible to implement continuous e-assessment in mathematics, having numerous classes, using e-assessment with MCQ. The research work helped understand some important issues related to the use of some MCQ in the area of Mathematics. This also constitutes a novelty, since, as far as our knowledge, there are just a few research works about using MCQ in the area of mathematics.

The structure for the rest of the paper is the following: firstly related topics are presented, secondly the presentation of the implementation of a continuous e-assessment strategy is introduced, next the construction of the MCQ bank is explained, and the paper ends with the conclusion.

2 RELATED TOPICS

In this section related topics are introduced, namely e-assessment, formative, summative, and diagnostic assessment, Multiple-Choice Questions, and Continuous Assessment.

2.1 e-Assessment

Information and Communication Technologies (ICT) bring up challenges and at the same time offer teachers tools that let them create differentiated learning opportunities for students. Their use is recommended by several European organizations such as the European Parliament and the European Open and Distance Learning Liaison Committee. The use of ICT in the assessment process thus becomes unavoidable. Electronic assessment, or e-assessment, arises. With e-assessment ICT is used throughout all the assessment process from the design of the tests to the storage of the results (Stödberg 2012). One possible approach is the development of specific environments for this purpose (Boticki and Milasinovic 2008; Dascalu and Bodea 2010; Llamas-Nistal et al. 2013). Other authors use the so-called Learning Management Systems (LMS) (Burrow et al. 2005; Salas-Morera et al. 2012). The LMS have the advantage of providing a wide range of tools specifically designed for the implementation of e-assessment. Among these tools we emphasize tests, which can encompass several types of questions, such as multiple-choice, true/false, matching items, short answer, among others.

2.2 Formative, Summative, and Diagnostic Assessment

Regarding its purpose, the assessment can be formative, summative or both, or diagnostic (Jacob et al. 2006; Redecker and Johannessen 2013; Stödberg 2012). We can consider that the summative assessment reflects the paradigm of "Learning to Evaluate" and that the diagnostic and formative assessments reflect the paradigm of "Assess to Learn" (Jacob et al. 2006). We can say that the first paradigm is the most common in the assessment that traditionally is done in higher education, which consists of applying one or more previously scheduled exams in paper written format. It can be said that e-assessment has been serving as a catalyst for a change of this first paradigm to the second one, since it can be found that in relevant scientific studies about e-assessment, the use of formative e-assessment or both types, formative and summative at the same time, is more common than the use of summative e-assessment (Stödberg 2012).

2.3 Multiple-Choice Questions

An important aspect of e-assessment concerns the type of task that is performed. A classification with which we identify ourselves, considering that it results from a careful review of the literature in some of the most important scientific journals in the area and because it corresponds to our practice as a teacher, is presented in (Stödberg 2012). In this study five categories are considered:

- closed questions, such as multiple-choice questions or matching;
- open-ended;
- portfolios;
- products, such as computer programs;
- discussions among students.

In the same study, it is stated that the closed questions are the most used in e-assessment. Among the closed questions, those of multiple-choice are of particular relevance and have some peculiarities,

presenting some advantages and some limitations. These various aspects as well as the comparison of this type of assessment with others, have been studied in scientific research in this area (Bible et al. 2008; Bush 2014; Haladyna et al. 2002; Liu et al. 2011; Lee et al. 2011; Rod et al. 2010; Torres et al. 2009).

In the study presented in (Torres et al. 2009) the following advantages of multiple-choice questions are presented:

- they can be used with diversified contents;
- they can measure a wide range of learning objectives;
- they are adaptable to various levels of cognitive abilities;
- they are very useful for assessment of large classes;
- using computer systems, such as LMS, tests can be graded automatically and statistical analysis can be easily performed;
- they provide the most useful format for comparisons over time due to the objectiveness in grading.

As for the limitations of the multiple-choice questions, the same study shows the following:

- they can be difficult to construct for higher levels of cognitive skills;
- they require good writing skills from teachers, so that the questions are clear;
- they require good reading skills from students, in order to correctly interpret the questions;
- they cannot measure some types of learning objectives, such as the ability to communicate;
- many times it is difficult to find good "distractors" (which corresponds to the incorrect options);
- students can guess the answer.

2.4 Continuous Assessment

As stated in the introduction, the Bologna Process points out to diverse forms of assessment carried out during the semester/academic year, the so-called continuous assessment, while the traditional type of assessment in higher education consisted in the application of final exams, the so-called final assessment.

Continuous assessment has got several advantages. According to (Borba and Penteado 2001), the following are some of the advantages of continuous assessment:

- Favours alternative learning itineraries;
- Allows the dynamization of learning;
- Stimulates and supports the advancements, giving priority to the positive elements, instead of the negative ones;
- Develops the reflexion capability, since it gives the student the opportunity to organize himself and understand its course of action and progression.

3 IMPLEMENTING A CONTINUOUS E-ASSESSMENT STRATEGY

With the implementation of the Bologna Process at ISCAP, the school of accounting and administration of the Polytechnic Institute of Porto, it was necessary to implement continuous or, more appropriately, distributed assessment. The word distributed is used in the sense that there are several summative and/or formative assessment moments during the semester. From now on, we will use continuous assessment with this sense of distributed assessment, remarking that there are also some moments of formative assessment, some of them being both formative and summative.

For the Mathematics teachers the implementation of continuous assessment constitutes a hard task, since two antagonistic situations were verified. Firstly, the duration of the courses were substantially reduced while maintaining the necessity to teach almost the same topics, in order to provide, in a timely manner, the necessary mathematical foundations to other courses. Secondly, the number of students per class increased due to budgetary limitations, which conduce to numerous classes. We are talking of about 800 students, distributed among classes of about 40 or more students. Time necessary for teachers to do assessment is proportional to the number of students being assessed. Thus, the existence of numerous classes discourages teachers to implement continuous assessment.

In order to solve this problem, an e-assessment strategy was envisaged. This includes an e-assessment component which consisted in the use of MCQ tests. This was due to the advantages of MCQ referred above (section 2.3). Since the open source LMS Moodle was already available in the institution, it arises as the natural choice to implement the MCQ tests. Moodle has got the advantage of being able to randomly generate tests by selecting a fixed number of questions from some selected categories defined by the teachers, which allows students to be presented with a different test, avoiding the necessity of developing several

different tests. Presenting different tests to adjacent students is important, since it is much easier for students to cheat with MCQ tests.

The tests, implemented with Moodle, were answered by the students as homeworks. This e-assessment component was both of a formative nature and of a summative nature. The summative nature is present because the tests have a weight in the final grades obtained by the students. The formative nature is present because the tests were to be answered by the students 7 to 10 days before the formal tests, to help them ascertain their awareness. These formal tests were the other component of the assessment. It consisted of 3 MCQ tests answered by the students in written format, at the same time for all the students, in a date previously scheduled by the school. It is relevant to say that these tests can also be considered as a type of e-assessment, since the marks were obtained and stored with electronic support. The electronic support consisted of an excel file with adequate formulas. The students answers were converted to electronic format, grades were automatically generated and stored in this format, and statistics were calculated. It is important to say that presenting the students different tests is a problem also in these MCQ tests in written format. At least 8 different versions were necessary for each test.

The implementation was gradually done, in order to carefully test the system. The first step consisted in the development of a bank of MCQ. This bank of MCQ was carefully planned and implemented to allow that the tests were randomly generated by Moodle, allowing that each student is presented with a different test maintaining, even so, uniformity. The construction of the MCQ bank is discussed in the next section.

Next, it was decided to implement 3 tests as homeworks, during specified periods previously defined and communicated to the students. It was intended to do the tests during the lessons, but there were no technical conditions to do it. Due to these circumstances, these MCQ tests implemented with Moodle had initially a weight of 10%, being that the 3 tests MCQ tests in written format had the remaining 90%. The decision for the weights was considered good, since this was the first time that the bank of MCQ was used and some uncontrollable situations could happen. In addition, the tests were optional to the students, and answered out of the lessons environment. As stated before, its purpose was mainly to serve as formative assessment.

This format was maintained during 3 academic years, but with slight changes in format and in the MCQ tests' weights.

4 CONSTRUCTING THE MULTIPLE-CHOICE QUESTIONS BANK

The most important task of the e-assessment was the construction of the MCQ bank. Three aspects were considered namely, defining categories for the questions, building the questions and the tests, and the revision process. These aspects are explained following.

4.1 Defining Categories for the Questions

Moodle can generate tests randomly by selecting a fixed number of questions from selected categories, which render a different test for each student. This poses two important questions:

- how to guarantee that the tests assess the same topics?
- how to guarantee that the tests are uniform in difficulty for all the students?

It was defined that the guarantee that the same topics are assessed could be achieved with the definition of categories in which to classify each of the elaborated questions, each category corresponding to a learning outcome. The learning outcomes were carefully defined by the group of teachers based in the learning outcomes of the courses being taught. These were defined at the beginning of each semester by the group of teachers, based on the students' necessities, but are somehow uniform along the years. It was detected that if the tests include more than one question from a category, Moodle can select the same question two times (at least), which is somehow common in Moodle randomly generated tests. Thus, to avoid this issue, the tests randomly generated by Moodle presented to each student, includes only one question per category, in order to avoid that the same may question appear more than once in the test. The categories to include in each test, and consequently the learning outcomes that are meant to be assessed, are defined by the teachers for each test.

As for the guarantee that the tests are uniform in difficulty for all the students, it was decided that the teachers should develop questions with low to medium difficulty level. The questions should also be uniform in format: for instance it is not acceptable to have a question with 3 options, and another question with 7 options, since it is more difficult for the students to analyze the late case.

4.2 Building the Questions and the Tests

As mentioned in the previous section, it is important to define a format to be followed by the teachers in the design of the questions. It was defined that all the questions will have 4 options: 1 correct and 3 distractors. A penalty of 33% was introduced for the distractors, in order to try to avoid students guessing the answer to the questions. It was also decided that the first option should be the correct one, in order to facilitate later revisions that, eventually, became necessary. This is not an issue for the students, since Moodle shuffle the various options before presenting the question in the generated tests.

Special care was taken when generating the tests in Moodle:

- Generating a different test for each class, defining duration, date, time the test became available, and time test stop being available;
- Only one attempt was allowed for the student to solve the test;
- Tests randomly generated by Moodle through the selection of 1 question from each of the predefined categories (each category corresponds to a predetermined learning outcome);
- The options in each of the MCQ were randomly shuffled for each generated test

4.3 The Revision Process

Producing questions without errors is crucial for building trust in the assessment process in any case, but is more difficult to ensure when building MCQ. A careful revision process was designed, so that errors could be minimized. The process consisted of the following seven steps:

1. Groups of two teachers were assigned with the responsibility of preparing a specified number of questions for each of the defined categories.
2. The course coordinator reviewed the questions and suggested changes.
3. The same team concretized the changes and prepared a detailed resolution of the questions.
4. A second team of two teachers analyzed the questions in detail and the respective resolution and proposed changes considering, for instance, the time required to solve them, the difficulty level, and the encountered errors.
5. The first team performed the changes.
6. The course coordinator analyzed the final version of the questions and proposed changes that, at this stage were minimal.
7. The last version of the questions was verified by the whole group and the final agreement was given.

It is relevant to say that a similar revision process was followed for the MCQ tests in written format, including the necessity of generating several versions. Later, after the students answering the MCQ tests in written format, the resolution (not just the answers) for one of the versions was made available to students.

Today the process is more streamlined. Taking into account the experience previously gained, it was possible to eliminate steps 3, 4 and 6.

This revision process revealed to be effective, since no errors have been found in the tests so far.

5 RESULTS

The implementation of the bank of questions can be considered a success, due to the hard work of the teachers, along the 3 academic years of the project. Two courses were included in the process, one in the first semester of the academic year, and another one in the second semester of the academic year. For the courses of the first semester, at the end of the process, there were 17 main categories in the bank of questions, being that 7 of these categories were subdivided in subcategories. Thus there were 30 categories that included questions. The total number of questions was 699. This gives a mean of 23.3, despite the categories not having the same number of questions. For the course of the second semester, at the end of the process, there were 21 main categories in the bank of questions, being that only 1 of these categories was subdivided in subcategories. Thus, there were 24 categories that included questions. The total number of questions was 661. This gives a mean of 27.5, despite the categories not having the same number of questions.

Even being optional, tests implemented with Moodle had good response rates. Table 1 presents the number of students that answered the 3 tests along the 3 academic years(AY)/semesters(S) of the implementation of this project. During the first semester of AY 3 it was decided to implement 2 tests instead of 3. The number of students decreases from the first test to the third test, because several students give up from continuous assessment. This is more evident during the first semester of the first academic year of the implementation of this project.

The author considers that this is due the fact that the students and the teachers were not aware of the bureaucratic implications of continuous assessment at the beginning of the semester. In the second semester of the third academic year the number of students doing the tests is smaller, because continuous assessment was not implemented due to technical issues.

Table 1: Number of students per test.

	AY 1		AY 2		AY 3	
	S1	S2	S1	S2	S1	S2
Nº of enrolled students	**620**	**818**	**772**	**908**	**821**	**1070**
Test 1	589	536	624	588	573	546
Test 2	468	478	598	524	506	466
Test 3	287	388	554	466	-	366

Table 2 resumes the grades obtained by the students in the academic year, first semester before the project being implemented, and in the 3 years that the academic project was implemented. Table 3 resumes the grades obtained by the students in the academic year, second semester before the project being implemented, and in the 3 years that the academic project was implemented. The percentage of approved students is always low, but tends to decrease during the implementation of the project. Nevertheless, the author considers that this is due mainly to the changes introduced by the Bologna Process. The project continued to be developed after the period described in this paper, and the grades have now an ascending tendency.

Table 2: Resume of the grades obtained by the students being assessed in the first semester.

	AY before	AY 1	AY 2	AY 3
Mean	7,3	6	5,5	4,7
Max	15	20	18	19
Min	0	0	0	0
Std	3,6	4,7	4,4	4,3
% Aprov.	31%	38%	36%	29%

Table 3: Resume of the grades obtained by the students being assessed in the second semester.

	AY before	AY 1	AY 2	AY 3
Mean	6,9	5,4	4,7	5,2
Max	16	20	18	18
Min	0	0	0	0
Std	3,7	4,5	4,5	4,2
% +	40%	34%	31%	34%

6 CONCLUSIONS

This paper presents the e-assessment strategy that was envisaged in order to solve some constraints that resulted from the necessity of applying continuous assessment, accordingly to the Bologna Process. The e-assessment strategy consisted in two main components, which comprises the use of MCQ tests. The necessary steps are described along the paper. The paper relates a very early stage of the process.

The lessons learned in the process resulted in some guidelines that can be used by teachers willing to use MCQ tests to assess their students. This is one of the paper's contributions. Another contribution is to demonstrate that it is possible to use MCQ in Mathematics to assess their students. In addition, this research helped to understand some important issues related with the use of MCQ to assess students, more specifically in Mathematics courses.

Despite the success achieved, some limitations were identified. One limitation is related with the guarantee that tests, randomly generated by Moodle, are uniform in difficulty for all the students, since some problems were identified during the process. In the future, it is intended to regularly use statistical measures in order to better ascertain the difficulty level of the tests. Another limitation is the lack of technical conditions to implement the e-assessment during the classes. Future research directions also include the development of technical conditions to make it possible.

REFERENCES

Azevedo, J. et al., 2009. Enhancing Math Skills with Moodle. In *Proceedings of ICERI 2009 - International Conference of Education, Research and Innovation*. pp. 2367–2377.

Babo, L., Azevedo, J., Torres, C. and Lopes, A.P., 2010. Moodle and Multiple-Choice Tests. In *Proceedings of INTED 2010 - 4th International Technology, Education and Development Conference*. pp. 296–303.

Babo, L., Azevedo, J., Torres, C. and Lopes, A., 2010. New Challenges in Mathematics for the European Higher Education. In *Proceedings of ICERI 2010 - International Conference of Education, Research and Innovation*. pp. 4971–4980.

Babo, L., Azevedo, J. and Lopes, A., 2008. The Active Mathematics Project at ISCAP. In *Proceedings of ENMA2008 - International Conference on Engineering and Mathematics*. pp. 27–34.

Bible, L., Simkin, M.G. and Kuechler, W.L., 2008. Using Multiple-Choice Tests to Evaluate Students'

Understanding of Accounting. *Accounting Education: An International Journal*, 17(S1), p.s55–s–68.

Borba, M.C. and Penteado, M.G., 2001. *Informática e educação matemática*, Belo Horizonte: Autêntica.

Boticki, I. and Milasinovic, B., 2008. Knowledge assessment at the Faculty of Electrical Engineering and Computing. In *Proceedings of the ITI 2008 - 30th Int. Conf. on Information Techonology Interfaces*. Cavtat, Croatia, pp. 111–116.

Burrow, M. et al., 2005. Developing formative assessment for postgraduate students in engineering. *European Journal of Engineering Education*, 30(2), pp.255–263.

Bush, M., 2014. Reducing the meed for gueesswork in multiple-choice tests. *Assessment and Evaluation in Higher Education*, cvb.

Dascalu, M. and Bodea, C., 2010. Challenges in Building e-Assessment Services from Project Management Knowledge Perspective. *International Journal of Global Management Studies Professional*, 2(1), pp.35–50.

Haladyna, T.M., Downing, S.M. and Rodriguez, M.C., 2002. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 5(3), pp.309–334.

Jacob, S.M., Issac, B. and Sebastian, Y., 2006. Impact on Student Learning From Traditional Continuous Assessment and an E-Assessment Proposal. In *Proceedings of the PACIS 2006 - The 10th Pacific Asia Conference on Information Systems*. pp. 1482–1496.

Lee, H.-S., Liu, L. and Linn, M.C., 2011. Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Applied Measurement in Education in Education*, 24(2), pp.115–136.

Liu, O.L., Lee, H.-S. and Linn, M.C., 2011. An Investigation of Explanation Multiple-Choice Items in Science Assessment. *Educational Assessment*, 16(3), pp.164–184.

Llamas-Nistal, M. et al., 2013. Blended e-assessment: Migrating classical exams to the digital world. *Computers and Education*, 62(1), pp.72–87.

Lopes, A.P. et al., 2011. Innovating Mathematics in the European Higher Education. In *Proceedings of INTED 2011 - 5th International Technology, Education and Development Conference*. pp. 1215–1222.

Lopes, A.P. et al., 2010. Multiple-Choice Tests - A Tool in the Assessing Knowledge. In *Proceedings of INTED 2010 - 4th International Technology, Education and Development Conference*. pp. 256–265.

Perrenoud, P., 1993. Não mexam na minha avaliação! Para uma abordagem sistêmica da mudança pedagógica. In A. Estrela and A. Nóvoa, eds. *Avaliações em Educação: Novas Perspectivas*. Porto, Portugal: Porto Editora, pp. 171–191.

Redecker, C. and Johannessen, O., 2013. Changing Assessment - Towards a New Assessment Paradigm Using ICT. *European Journal of Education Research, Development ans Policy*, 48(1), pp.79–96.

Rod, J.K., Eiksund, S. and Fjaer, O., 2010. Assessment Based on Exercise Work and Multiple-Choice Tests. *Journal of Geography in Higher Education*, 34(1), pp.141–153.

Salas-Morera, L. et al., 2012. Effective Use of E-Learning for Improving Students' Skills. In R. Babo and A. Azevedo, eds. *Higher Education Institutions and Learning Management Systems: Adoption and Standardization*. Hershey, PA: IGI Global, pp. 292–314.

Stödberg, 2012. A research review of e-Assessment. *Assessment and Evaluation in Higher Education*, 37(5), pp.591–604.

Torres, C. et al., 2009. Developing Multiple-Choice Questions in mathematics. In *Proceedings of ICERI 2009 - International Conference of Education, Research and Innovation*. pp. 6218–6229.

Torres, C. et al., 2011. Improving Multiple - Choice Questions. *US-China Education Review*, B(1), pp.1–11.