Interactive and multimedia contents associated with a system for computer-aided assessment **

Rui C. Paiva [*†‡] Milton S. Ferreira[*§] Ana G. Mendes[*] Augusto M. J. Eusébio[*¶]

Abstract

This paper presents a research study addressing the development, implementation, evaluation and use of Interactive Modules for Online Training (MITO) of mathematics in higher education. This work was carried out in the context of the MITO project, which combined several features of the learning and management system Moodle, the computer-aided assessment for mathematics STACK, the mathematical software GeoGebra, several packages from the type-setting program LaTeX, and tutorial videos.

A total of 1962 students participated in this study. Two groups of students taking a Calculus course were selected for a deeper analysis.

In regard to usability and functionality, the results indicate that MITO scored well in almost all aspects, which is fundamental for their introduction into formal university courses. The analysis of the data reveals that the use of MITO educational contents by students mainly occurs about one week and a half prior the evaluations. Moreover, there is a strong correlation between the results of online assessments on MITO in a continuous assessment model and the final grade on the course.

Keywords: mathematics, computer-aided assessment, e-learning, interactive contents, multimedia contents.

* Departamento de Matemática da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria. Morro do Lena, Alto do Vieiro, 2411-901 Leiria, Portugal. Phone: (+351) 244820300. E-mail addresses: rui.paiva@ipleiria.pt, milton.ferreira@ipleiria.pt, aimendes@ipleiria.pt, augusto.eusebio@ipleiria.pt
†Correspondence author.
‡Centro de Matemática da Universidade do Porto.
§Centro de Investigação e Desenvolvimento em Matemática e Aplicações da Universidade de Aveiro.
¶Centro de Estudos de Gestão do Instituto Superior Técnico.

** Accepted author's manuscript (AAM) published in [Journal of Educational Computing Research, 52 (2) (2015), 224-256] [DOI: 10.1177/0735633115571305]. The final publication is available via url http://jec.sagepub.com/content/52/2/224.short?patientinform-links=yes&legid=spjec;52/2/224.

1 Introduction

The integration of Information and Communication Technologies (ICT) in mathematics education provides essential tools for teaching and learning mathematics enhancing student's learning (NCTM, 2000). Anderson et al. (2001) divides the use of ICT for educational purposes into two categories: technology for instructional or assessment purposes. For instance, the use of interactive and multimedia books as presentation resources in the classroom and as a support tool for textbooks serves instructional purposes, while the use of online quizzes as tools to test student knowledge is related to the assessment of students' skills. There are many studies showing that the use of instructional technologies helps to improve the teaching and learning processes of mathematics (Kaput & Hegedus, 2007; Baki & Güveli, 2008; Lazakidou & Retalis, 2010; Reed, Drijvers & Kirschner, 2010; Andrade Aréchiga, López & López-Morteo, 2012; Kim & Chang (2010)). With respect to the impact of e-assessments in mathematics the majority of the studies show that e-assessments with formative feedback contribute to student learning and improve student performance on the final course exam (Varsavsky, 2004; Roth, Ivanchenko, & Record, 2008; Ruokokoski, 2009; Rasila, Majander, & Malinen, 2010; Angus & Watson, 2009; Huisman & Reedijk, 2012; Rivera, Ochoa, & Perez, 2013.). However, there are also studies showing that online quizzes do not significantly improve student learning. For instance, Steenhuis et al. (2009) found that online quiz scores are related to the grade that students achieve in the course arguing, that both grades could reflect student ability and that the observed relationship is not a good indicator of the added value of online quizzes.

One facet that is always present and important in education is the learning outcomes. These represent the domain of competencies (cognitive and affective) that should be developed during the course of education (Frye, 1999). Assessment is a fundamental and integral part of any curriculum based on student learning outcomes which includes measurement, feedback, reflection, and change. Frequent assessment of students helps them to review and refine concepts and to obtain a deeper understanding, depending on the type of assessment (Frye, 1999). Anderson et al. (2001) identify two different types of assessments of student learning. One type is the summative assessment, which assesses the knowledge and skills acquired by the students at the end of a module or teaching unit. The other type of evaluation is the formative assessment, which is intended to collect information related to the learning progress and to address, in an

opportune way, problems observed during the learning process. The essential difference is that the summative assessment concludes a learning period, while the formative assessment provides intermediate feedback to improve the final result. For instance, a self-assessed quiz and a homework assignment with significant weight on the overall course grade can be regarded as formative, if the examinations cover the same material (Steenhuis, Grinder, & Bruijn, 2009).

The fast development of Computer Algebra Systems (CAS) and web based resources opened a window of opportunities for supporting the math educational process that can take many forms under different learning environments (Kaput & Hegedus, 2007). Currently, there are several commercial and open source Computer-Aided Assessment (CAA) software with varying levels of mathematical and pedagogic sophistication (Sangwin, 2012). Some of the better-known are WeBWorK, STACK, WebAssign, eGrade, MapleT.A., and MyMathLab. Presently, publishers provide online support for textbooks that include exercises with links to automatic evaluations.

In this context, the software STACK, "System for Teaching and Assessment using a Computer algebra Kernel", represents an advanced CAA system for mathematics, with an emphasis on formative assessment. STACK is open source software licensed under the General Public License (GPL, 2004) originally developed by Sangwin (2003) at the University of Birmingham. To make STACK compatible with other software that we use in the construction of our online contents, we made some adjustments and translated it to Portuguese. The first advantage of STACK is the ability to enter the answer in the form of an algebraic expression. After entering an answer, the computer algebra system Maxima (www.maxima.sourceforge.net) is used to evaluate the response, and students receive specific feedback, with hints to guide them when their incorrect submissions are within the expected responses. In this way, feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning (Gibbs & Simpson, 2004). This feature encourages students to continue to search for the correct answer. Another advantage of STACK is its sophisticated method for creating different versions of a question depending on random parameters. Thus, the author may generate as many versions as there are possible variations of the parameters used.

STACK has been used by thousands of students at the universities of Birmingham and Manchester in the United Kingdom, 60,000 students at Aalto University in Finland, 1800 students at Polytechnic Institute of Leiria (IPL) in Portugal, and hundreds of students at Nagoya

University in Japan, among others (Sangwin, 2010). At IPL STACK has been used since 2009 by engineering, economics, science, and biology students for homework and training assignments.

In 2010, we started a project of online support for teaching mathematics in higher education called "Interactive Modules of Online Training" and known by its Portuguese acronym MITO. This project combines a set of tools that we thought to be among the most developed and promising for teaching mathematics. Our main goal is to complement the face-to-face classes and the traditional teaching content through interactive and multimedia online content. The theoretical content is supported by multimedia and interactive books that include videos and Java applets, and the practical content consists of online quizzes in STACK with a high level of interactivity.

The MITO platform (www.mito.ipleiria.pt) results from an adaptation of the source codes of the learning management system Moodle and STACK to integrate static and randomized graphics with high resolution and level of detail, Java applets, interactive books, and STACK questions.

For the realization of the project we have computer support from the Distance Education Unit of the IPL and a dedicated server.

In this paper we present in detail a three-year experience on the development and implementation of MITO. We describe the special features of its design, development, implementation, and evaluation of the tools used in our project and the results of the study conducted with students involved in this project. The study was designed to explore the accessibility and usability issues of MITO in higher education and the extent that web-based practice affects students' mathematics learning and achievement. As described above, there are many studies that describe the use of open-source software for educational purposes and the development of educational contents. Our main contribution to the field is the design, development, implementation and evaluation of instructional materials involving Moodle, STACK, GeoGebra (www.geogebra.org), PSTricks (Voss, 2011), and tutorial videos with useful modifications on the open source code. The following research questions are answered:

- How did teachers and reviewers evaluate MITO?
- How did students evaluate the use of MITO contents?
- How did students use MITO?
- How are related the results of MITO e-assessments and final grade in a continuous assessment model?

This paper is organized in five sections. In Section 2 we present the phases of analysis, design, development, and implementation of the MITO project. In Section 3 we describe the methodology used in our work. Section 4 shows the data analysis and finally, in Section 5 we present the discussion and future plans for the MITO project.

2 MITO project

With the purpose of familiarizing the reader with the educational content used in this research, we present the MITO project. This is a research project running at IPL, developed by teachers from the Department of Mathematics (DMAT) of the School of Technology and Management (ESTG). In the MITO project we used the most common model for creating instructional materials, the ADDIE Model (Piskurich, 2006). This model considers five phases: Analysis, Design, Development, Implementation and Evaluation. We proceed with descriptions of the first four phases. The evaluation of MITO by teachers, reviewers and students is described in Section 3.

2.1 Analysis phase

The inclusion of Portugal in the pan-European Bologna Process in 1999 resulted in a major reorganization of the formative process in Portuguese higher education since 2007 (Veiga & Amaral, 2009). In IPL, the evaluation of knowledge throughout the semester instead of solely during final exams at the end of the semester represented the most significant change in the evaluation system of undergraduate courses. In DMAT, we use continuous assessment models (Biggs, 2003).

The idea of implementing homework in the mathematic disciplines arose when it was found that freshmen students were entering with insufficient work habits in mathematics. In the academic year 2008-09, one solution found in DMAT was the implementation of homework assignments on paper with the uploading of solutions to the e-learning platform. However, this experience did not bring the expected benefits, as we discovered that a significant percentage of students copied the solutions from peers. The enormous amount of work involved in grading homework and the poor improvement in student achievement as a result of cheating, led us to the conclusion that randomized exercises were necessary. Moreover, in order to reduce the demotivation of students with more difficulties, we decided to associate such a system with interactive and multimedia contents that the students could consider more attractive.

In September 2010 the MITO project was formally initiated with the primary goal of promoting mathematics through the development of interactive modules to support traditional classroom teaching and distance learning in higher education.

2.2 Design phase

The online quizzes available in conventional e-learning platforms are limited to multiple choice questions, and their system of randomized questions consists of choosing a question from a group of questions from the question bank. To create a system of randomized quizzes with this method requires too much work and has the disadvantage of being limited to multiple choice questions. In this sense, an intensive internet search led to the conclusion that STACK provides the best solution to all these limitations. Its method of creating different versions of a question is very practical. In this system, each question may depend on random parameters and it is allowed to create as many versions as the possible variations of the parameters. For example, we might choose a question set involving a function $f(x) = mx\sin(nx)$ where $m \in \{1,\ldots,9\}$ and $n \in \{2,3,4\}$. Of course, attention must still be given to the choice of random parameters to prevent the generation of impossible or trivial questions.

The time invested in programming the CAS code may have, as a return, the generation of thousands of different versions for each question. The version of each question of an online quiz is chosen between the versions that the teacher provided in the database versions of the question. The teachers from DMAT verify the versions and ensure its content validity. From a practical standpoint, for a course with about 100 students, it is enough to provide about 50 versions for each question. For example, when a student opens a 10 questions quiz, each question is associated with one of the 50 versions of the database. As a result, the probability that two students have the same quiz is approximately zero and cheating is more difficult.

The high compatibility between STACK and Moodle (Wild, 2009) and the fact that they are open source software supported the decision to opt for these CAA and learning and management systems.

The e-learning platform of MITO is a complex modification of the source code of Moodle and STACK intended to accommodate interactive and multimedia learning contents. We briefly describe these changes throughout this document.

An interactive module of MITO is composed of interactive and multimedia books and training along exercises with online evaluations produced with STACK. The interactive books consist of theoretical notes supported with Java applets, highly detailed graphs, and tutorial videos. The STACK exercises may also include Java applets and random graphs. We proceed with descriptions of each of these components.

2.2.1 Graphs

The graphics used in the interactive modules of MITO are distinguished by their high resolution and the wide range of possibilities for creating graphics. In Figure 1 we present an example.

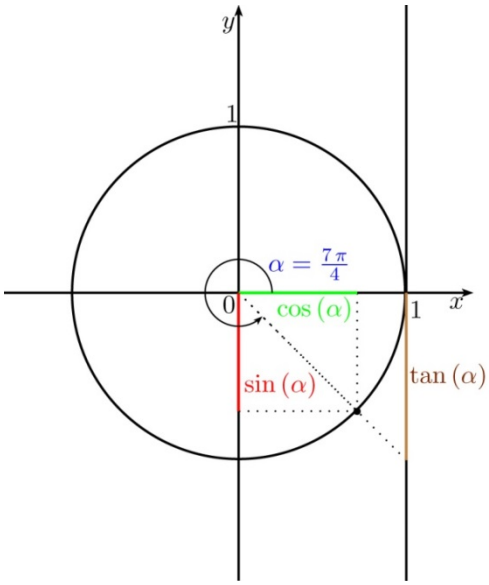

Figure 1: Figure used in a trigonometry question from STACK.

Our graphs are vector graphs constructed using PSTricks code, which is a LaTeX package (Voss, 2011). These graphs do not lose quality when they are enlarged, and it is possible to isolate objects and areas and treat them independently. The e-learning platform of MITO was adapted to recognize this code and generate the correspondent image. The following improvement was to adapt the STACK source code to accept PSTricks code with random parameters. As a result, we have random questions from STACK in which the graphics vary with the parameters. The graph of Figure 1 was implemented as a random graph, in a STACK question, with 24 different versions corresponding to variations of the amplitude of the angle α. We often work with random graphs that can have hundreds of different versions.

2.2.2 Videos

The use of tutorial videos to support teaching is a technique used for some years. Mayer (2003) provides examples in which students who received videos performed better on assessments than students who received text. With the development of ICT, several online video sharing communities have emerged. One of the most popular sources of free videos is the KHAN Academy. For a recent and sophisticated use of tutorial videos in higher education, see Wells et al. (2012), Maxwell, & Angehrn (2010) and Chen (2012). The videos of MITO were produced using LaTeX code (Oetiker, Partl, Hyna, & Schlegl, 2011) and possess the advantages of high resolution images and the ability to be edited. The platform of MITO records when users access the videos. This feature allows monitoring the student's interaction with this educational content and it was also useful to gather data to answer the third research question.

2.2.3 Java Applets

The Java applets used in MITO were produced using GeoGebra (www.geogebra.org). This is an interactive geometry, algebra, and calculus software intended for teachers and students. The platform of MITO has been adapted to incorporate the Java applets into all activities. We then enabled STACK questions with random Java applets. As a final result, Java applets can vary with the parameters of the questions.

In Figure 2, we present an example of an applet embedded in a STACK question.

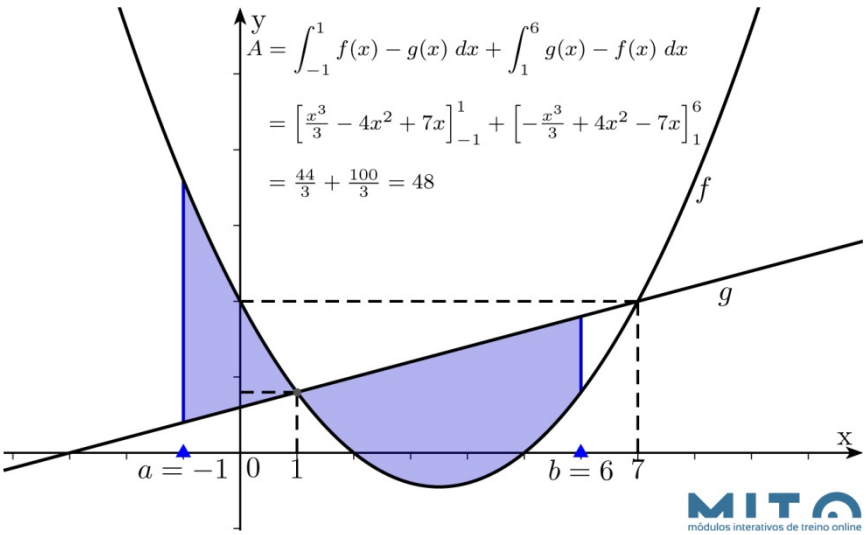

Figure 2: GeoGebra input interactions with a STACK question.

In the applet shown in Figure 2, the student may change the location of points a and b by dragging on the screen. As this is performed, the region and feedback update dynamically. Once

the student finds an example of his version of the question, he may use the same idea to obtain the answer. This question also has a link to a tutorial video. We have also implemented questions with embedded applets that directly provide the answer after appropriate interaction.

2.2.4 Computer-aided assessment

The online quizzes provided by learning management systems such as Blackboard or Moodle are very limited in the type of questions that may be included on their tests, consisting generally of true/false or multiple-choice questions. Questions whose correct answers can be given in many different ways or with a wide variety of correct answers, such as Example 2.1, cannot be implemented using the conventional question types of these platforms.

To give the reader an idea of the added value when compared to conventional online quizzes, we highlight some features of STACK framed by our experience at IPL. For more detailed description, see Sangwin (2007), Sangwin (2012), and Sangwin (2013).

In STACK, students may submit their answers in the form of a mathematical expression. For example, the student might answer a question by entering a polynomial or a matrix. Essentially, STACK asks for mathematical expressions and evaluates these using a computer algebra system. STACK uses the free software Maxima for the manipulation of symbolic and numerical expressions and establishes the mathematical properties of student answers. This feature is particularly useful when the question has several correct answers. The system includes several answer tests to compare expressions, a syntax checker and the abilities to transform the answer entered with an "Answer Preview" in LaTeX typeset format and to generate feedback related to the student answer by the execution of a potential response tree. This last feature contributes to maintaining student motivation and may facilitate the student's reasoning process (Harjula, 2008). Generally, the potential response tree works as follows. Depending on the result of the answer test, either the true or false branch is executed. In each branch, we may adjust the mark and penalty for the attempt, generate and add specific feedback, generate and add a specific hint, proceed to another node, or end the process. In the next example, we present a question that has several possible answers and the formative feedback that accompanies the responses.

Example 2.1 Give an example of a real function in *IR* that is continuous and odd but not differentiable at $x = 0$.

$f(x) =$ x+1/x

Your last answer was interpreted as:

$$x + \frac{1}{x}$$

Your answer is partially correct.

$f(x) = x + \frac{1}{x}$ is odd and not differentiable at $x = 0$. However, because it is not defined at $x = 0$ is also not continuous at this point. Consider using an irrational function. Your mark for this attempt is 0.6.

With penalties and previous attempts, this gives 0.6 out of 1.

When authoring STACK questions, it is possible to enter LaTeX code interlinked with parameters and random functions defined by the author. This feature enables the quick generation of large question banks. It is also possible to associate STACK with learning management systems to make use of their authentication systems, quiz classifications, answer attempts and other statistical records. At IPL, we associate STACK with Moodle.

Maxima includes the manipulation of the most common symbolic and numerical expressions and yields high precision numeric results by using exact fractions, arbitrary precision integers, and variable precision floating point numbers. It is possible to use all of these features in STACK. Although Maxima can also plot functions and data in two and three dimensions, the graphs generated by PSTricks code, as described above, clearly have better quality and versatility. Due to the changes made to the source code of STACK, the graphs generated by PSTricks code may also use random parameters.

2.2.5 Interactive and multimedia books

An interactive and multimedia book of MITO consists of a digital book composed of theoretical notes supported by Java applets, detailed graphs and tutorial videos. At the end of each section, there are several multimedia and interactive examples, and the student is encouraged to engage in online quizzes in STACK. Each section can be accessed by clicking on the index located on the left side of each page. Figure 3 presents a screenshot of the interactive book of real functions of two or more real variables.

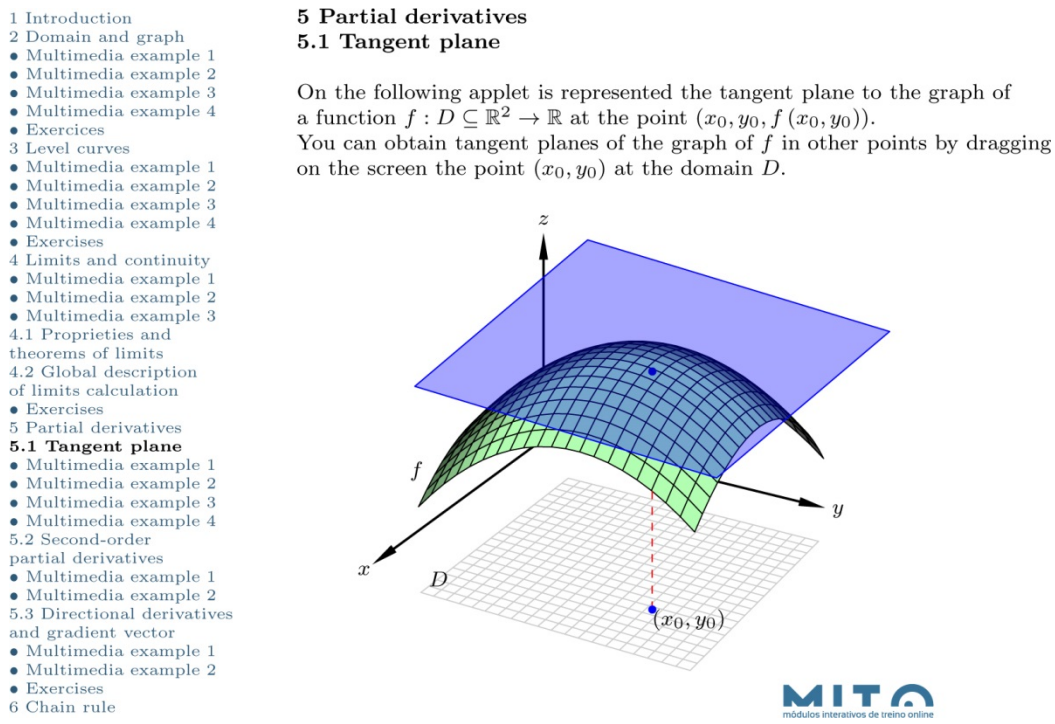

Figure 3: Interactive and multimedia book of real functions of two or more real variables.

2.3 Development and implementation phases

To develop the MITO project, it was necessary to invest much time in finding the appropriate open source software, adapting its code to our goals and solving the bugs that emerged during the three years. It was necessary to dedicate a server exclusively for MITO and often resort to the technical support of Distance Learning Unit of the IPL. After preparing the e-learning platform to build and host the desired educational content, we still had to add a substantial number of working hours to create the educational content. This task must be performed by people who have both the technical skills and the teaching experience required to create the contents. Naturally, the success that we recorded over three years with the students and the support of the IPL was essential to motivate us to proceed with the project.

The experience described in this paper concerns the academic years 2009-10, 2010-11, and 2011-12. In the remainder of this article, we refer to these years as Year One, Year Two, and Year Three. It involved 4 disciplines in Year One, 11 disciplines in Year Two and 11 disciplines in Year Three. These disciplines were:

• Calculus (science/engineering courses/biology/economics courses)

• Linear Algebra (biology/science/engineering/economics courses)

• Statistics (engineering/economics courses)

• Numerical Analysis (science/engineering courses)

In Table 1, we present the numbers of disciplines and participants from our research.

	Year One	Year Two	Year Three
No. of undergraduate courses	4	11	11
No. of participating students	300	778	884
No. of teachers involved	3	9	10

Table 1: Numbers of undergraduate courses and participants from the research.

The data presented in Table1 refers to undergraduate courses involved in online homework assignments for continuous evaluation in mathematic courses. In the three years, more teachers from ESTG recommended that their students use the MITO educational content for learning and training, but they did not require online homework assignments.

The educational contents of MITO were produced during the three-year study. At Year One we build a model of an interactive module of MITO, 55 randomized exercises in Linear Algebra and Calculus, and 23 tutorial videos. These contents were implemented in 4 randomized online homework assignments with a cohort of approximately 300 engineering and biology students. Each homework was supported by online training quizzes with detailed feedback including hints and solutions and with the tutorial videos. After completing each question of the online homework assignment, students received their respective grades and could either rework the answer fields or access the solution. This feature aims to contribute to the student's engagement with the question.

In Year Two, we formally began the MITO project and expanded its use to more courses with more training exercises and tutorial videos. The development of the educational contents proceed with the improvement of the feedback in STACK questions produced in Year One and with the production of about 50 STACK exercises of numerical analysis and statistics associated with Java applets. For three months, we prepared questions with GeoGebra applets embedded that required students to specifically understand the mathematical properties involved. In this type of question, we asked the student to interact with the applet and observe certain mathematical properties. After understanding the property, the student may submit his answer. If it is not correct, he receives feedback with helpful suggestions to interact again with the applet and so on. In the first semester of Year Two, a cohort of 570 students from 8 undergraduate courses in science/engineering and biology had between 3 and 4 online homework assignments with 10 exercises for Linear Algebra and Calculus. In the second semester, a cohort of 208 students from

3 undergraduate courses of numerical analysis and statistics courses of science/engineering completed 4 online homework assignments with approximately 50 STACK exercises with Java applets embedded.

In Year Three, major improvements were introduced with the preparation of the interactive modules of analysis, linear algebra, statistics, and numerical analysis. At this stage, the interactive modules of MITO (IMM) included the following chapters: trigonometry, derivatives, antiderivatives, integrals, real functions of several variables, analytic geometry, matrices, discrete random variables, continuous random variables, nonlinear equations, and polynomial interpolation.

A considerable effort was made to build the several components of the IMM, described in Section 2.2. In total, the module components included 11 interactive and multimedia books with 150 videos and 114 Java applets and 4 question banks with 176 training and evaluation questions created in STACK. The contents were organized in four Moodle disciplines with the users included in independent groups relative to each degree. This feature provides a simpler analysis of the records of the platform.

In the first semester of Year Three, a cohort of 884 students from the undergraduate courses of science/engineering and economy had between 3 and 4 online homework assignments. The main difference relative to previous years was the association with IMM. In some undergraduate courses the interactive and multimedia books were used during face-to-face classes to introduce the theoretical concepts together with the corresponding Java applets. In total, students from 14 undergraduate courses accessed the interactive modules exclusively for learning and training approximately 150,000 times, over the first semester of Year Three.

3 Methodology

3.1 Participants

The three-year research was conducted at ESTG of IPL and involved the undergraduate courses and participants shown in Table 1. Although the research project was carried out over three years, most of the data used answer the research questions was gathered in Year-Three. In order to answer the last two research questions we have focused ourselves on 85 students from two engineering undergraduate courses taking the Calculus course in Year Three. The choice of

the two courses was discussed among researchers until reaching an agreement about the most appropriate ones to provide relevant answers for the last two research questions.

Relevant information on the students who took part of the study is shown in Table 2.

Group	Undergraduate course	No. of students			Average age
		Female	Male	Total	
A	Electric engineering	6	38	44	19.5
B	Informatics engineering	3	38	41	19.8

Table 2: General information about the 85 students of groups A and B.

The lecturer of electric engineering (group A) was one of the authors of the MITO educational contents used in the study while the lecturer of informatics engineering (group B) did not participate in its construction. In group A the interactive and multimedia books of Calculus described on Section 2 were used during face-to-face classes to introduce the theoretical concepts together with the corresponding Java applets. In Group B, these contents were only used by students outside the class. The two groups followed a continuous assessment model (Biggs, 2003) in the Calculus course. In the next section we will give details about the instruments used in the assessment.

The main reason for this choice of groups was the similar academic programs of the two Calculus courses. The second reason was that two groups have been involved in a variety of similar educational and technical tasks in MITO and were evaluated by a continuous assessment model. The third reason was the relevance of comparing the data associated with an author and a non-author of MITO contents. In this sense, the study was also valuable to gathering information on opinions, use of educational contents, and academic performance across the students from an author and a non-author of MITO contents. Last, the fact that MITO is well known by ESTG students complicates the choice of a control group without access to MITO and with the same teacher. This fact precluded another research question about the influence of MITO on the academic performance of students. In any case as the third and fourth research questions are about how students use MITO and the relationship between e-assessments and final grade the control group would not apply.

3.2 Instruments and Procedures

In the three-year research both qualitative and quantitative data were gathered to answer the research questions. A total of 20 instruments summarized in Table 3 were carefully designed, implemented and conducted. The instruments were evaluated relatively to validity following the content-related evidence method and internal structure method and relatively to reliability using Cronbach's reliability test (Johnson, & Christensen, 2008). In all instruments where the Cronbach's reliability test is applicable, the final result for Cronbach's alpha value was higher than 0.8, which validate them (Johnson, & Christensen, 2008).

Instruments		Year One	Year Two	Year Three	No. of questions	Answer scale	Cronbach's alpha	
							Year Two	Year Three
Learning Object Review Instrument				X	9	Likert (5 options)	Not apply	
Informal conversations		X	X	X		Open opinion	Not apply	
Student survey	Part 1		X	X	9		Not apply	
	Part 2		X	X	22	Likert (5 options)	0.811	0.813
	Part 3		X	X	15	Likert (5 options)	0.823	0.821
	Part 4		X	X	1	Open opinion	Not apply	
Formative and summative assessments	Basic skills test			X	13	Objective (4 options)		0.814
	7 Online homework assignments			X	10 each	Open	Not apply	
	4 Tests			X	Diverse	Open	Not apply	
	4 Exams			X	Diverse	Open	Not apply	
Platform records about use of MITO				X		Open	Not apply	

Table 3: Instruments used to answer the research questions
and the corresponding reliability coefficients

The evaluation of MITO educational contents by teachers and reviewers in the three-year research used both qualitative and quantitative data. The evaluation of the quality, reusability, goal alignment, and other educational and technical aspects of IMM was made in Year Three by 4 reviewers through the Learning Object Review Instrument (LORI version 1.5). LORI rates nine aspects of quality (Nesbit & Belfer, 2005) using a one to five (low to high) point scale. If the item is judged not relevant to the learning object, or if the reviewer does not feel qualified to judge that criterion, then the reviewer may opt out of the item by selecting "not applicable". The qualitative data resource was obtained by informal conversations conducted between the researchers and the teachers that participated in the three-year research.

Measuring how students use and evaluate the influence of MITO in learning, motivation, performance, and other educational aspects involved a well-designed survey conducted in Year

Two and Year Three which gave valuable information about the acceptance of MITO in undergraduate courses (see Student survey in Table 3). It was conducted anonymously at the end of each first semester of Year Two (N=778) and Year Three (N=884) with the students that had online homework assignments and used MITO. The groups A and B described above are also included and studied separately. Table 1 and Table 2 show the numbers of undergraduate courses and participants involved in the survey. To ensure the completion of the survey by all students, the surveys were given to the students in the last two days of each class and were taken via paper and pencil rather than online. The survey has two sections of quantitative type and two of qualitative type and follows the 15 Principles of Questionnaire Construction described in Johnson, & Christensen (2012). The first section collects characteristics about the student's age, gender, academic history and his computer familiarity by 9 closed–ended items. The second part, composed by 22 closed–ended items with answers given in a quantitative scale from 1 to 5 (1 = fully disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = fully agree), gathers opinions about the student's interaction with training exercises and online homework assignments of MITO and the relevance of these in achieving success in the course. The third part is composed by 15 closed–ended items with answers given in a quantitative scale from 1 to 5 (1 = fully disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = fully agree) seeks to assess the usefulness and quality of the interactive and multimedia books of MITO. We employed the factor analysis method (Box, Hunter, & Hunter, 2005) to obtain the 5 dimensions of the second part and the 3 dimensions of the third part which are described below. The Cronbach's alpha value was used to validate them (Johnson, & Christensen, 2012). The categorization of each of the dimensions of parts two and three of the survey was properly discussed among researchers until reaching agreement about the categorization. The most important question categories identified by the researchers for part two were:

Preference — preference of students between online homework on MITO and paper-and-pencil homework.

Learning — learning the mathematical concepts of the discipline due to training exercises and online homework assignments of MITO.

Motivation — increased motivation to study mathematics due to training exercises and online homework assignments of MITO.

Performance — improvement of performance on final grade due to online homework assignments.

Usefulness — recognition of the usefulness of the features of STACK.

The most important categories deemed for part three were the following:

Learning — learning the mathematical concepts due to interactive and multimedia books of MITO.

Usefulness — recognition of the usefulness of interactive and multimedia books of MITO in the learning process.

Replacement — full replacement of the teacher by the learning contents of MITO.

Finally, section four asks the student to write, in an open question, his general opinion about what he liked or disliked about MITO and to suggest improvements. The opinions were gathered and analyzed by categorizing them as inductive themes (Johnson, & Christensen, 2012). Then, the researchers compared the results and obtained agreement about the descriptive details.

Another quantitative data resource used in the study was the platform records of the groups A and B in the first semester of Year Three which held working time period and interaction with educational resources of MITO. These data were used for obtaining statistical results about how students used the educational contents of MITO.

Quantitative information about the student's performance was obtained directly from the grades of an online basic skills test, randomized homework assignments, and written tests and exams of groups A and B. The 16 instruments used (see formative and summative assessments in Table 3) were created by teachers of the Math department of ESTG of IPL. The basic skills test was conducted at the computer labs of ESTG at the beginning of Year Three. The test was made in STACK and required that the students enter all of the multiple choice answers without penalties and received the solutions and final grades after completing the test. This test is used in our research as a pretest. In group A two randomized homework assignments and one written test were given at each half of the semester, covering the first half and the second half of the material. In group B were given three randomized homework assignments during the semester and one written test at each half of the semester. Additionally, after the end of classes, there were two alternative final exams covering all the material in the two engineering courses. These exams

were given for the students that did not succeed in continuous assessment. In the groups A and B, the basic skills test represented 5% in the final grade, the randomized online homework assignments represented 15% and the written tests the remaining 80%.

4 Results

The results in this section provide us with the opinion of teachers, reviewers, and students about MITO, how students use the educational content and how are related the grades of a basic skills test, MITO online homework assignments and of final grade.

The results of our study are presented with four sub-sections according to the research questions.

4.1 How did teachers and reviewers evaluate MITO?

The results of the 4 reviewers that participated in the assessment of the quality and reusability of the educational contents of MITO using LORI show that all the 11 interactive modules scored medium/high in almost all of the aspects. Results are summarized in Table 4 as a set of averaged ratings, one per item, and as a single average.

Item	Mean
Content quality	4.75
Learning Goal Alignment	4.0
Feedback and Adaptation	3.75
Motivation	3.5
Presentation Design	4.0
Interaction Usability	4.75
Accessibility	2
Reusability	4.25
Standards Compliance	4.0
Global mean	4.33

Table 4: Results of the evaluation of educational content of MITO by LORI version 1.5.

The discussion of results is presented below and follows the convergent participation model for collaborative evaluation (Nesbit & Belfer, 2005).

1) Content quality. A mean of 4.75 indicates high rate which means that the content is free of error and presented without bias or omissions that could mislead learners.
2) Learning Goal Alignment. This item obtained a mean of 4 which falls in a medium/high rate. We may conclude that the learning activities, content and assessments provided by the contents of MITO align well with the declared goals. The learning object is enough on its own to enable students to achieve learning objectives.

3) Feedback and Adaptation. On average, this aspect scored 3.75. The learning object has the medium level of ability to adapt instructional activities according to the specific needs and characteristics of the student.
4) Motivation. A mean of 3.5 indicates a medium rate. We may consider that the learning object is motivating in a satisfactory way. MITO content is relevant to the personal goals and interests of the intended learners.
5) Presentation Design. An average of 4 indicates a medium/high rate. This means that the production values and information design enable the user to learn efficiently. The text is legible, the graphs and charts are labeled.
6) Interaction Usability. This item scored 4.75 which is a high rate. This means that the user interface design implicitly informs learners how to interact with the object. The navigation through MITO is easy, intuitive and free of excessive delay.
7) Accessibility. This item scored an average of 2. This means that the educational content of MITO provides a low degree of accommodation for learners with sensory and motor disabilities.
8) Reusability. A mean of 4.25 indicates a medium/high rate. We may conclude that the learning object is a stand-alone resource that can be readily transferred to different courses, learning designs and contexts without modification.
9) Standards Compliance. An average of 4 indicates a medium/high rate. This means that that the learning object adheres to all relevant international standards and specifications. These include the IEEE Learning Object Metadata standards, and technical guidelines developed by IMS, IEEE, SCORM and W3C.

A global mean of 4.33 indicates a medium/high rate of MITO educational contents in quality and reusability aspects considered by LORI.

Through informal conversations with the teachers involved in the three-year research we got opinions and suggestions for improvement of MITO. The general opinion is that homework and training e-assignments contributes for student learning and engagement with the disciplines and that interactive and multimedia books are very useful in the classroom to show parts of the subject that require some dynamic. The suggestions received were related with the creation of more contents and improving accessibility.

4.2 How did students evaluate the use of MITO contents?

To answer this research question we used a survey administered in Year Two (N=778) and Year Three (N=884). Groups A and B are included in Year Three and are also studied separately. The first part of the survey revealed the following information:

−The average age is near 20 (19.9, 20.1).

−Just under half of the students were freshmen males (45.2%, 48.8%).

−Most of the students were day students (75.8%, 79.9%).

−Hours spent on the computer per week: 0-7 (29.5%, 30%); 7-14 (38%, 38.2%); 14-21 (19.9%, 20.6%); > 21 (12.6%; 11.2%).

−Hours spent on the Internet at home per week: 0-7 (27.9%, 28.3%); 7-14 (38.8%, 39%); 14-21 (24.5%, 23.3%); > 21 (8.8%; 9.4%).

In Figure 4 we show the mean values of each of the 5 categories considered in part two of survey (see Section 3.2), about for the Year Two, the Year Three, and the groups A and B. Note that this refers to student's interaction with training exercises and online homework assignments of MITO and the relevance of these in achieving success in the course.

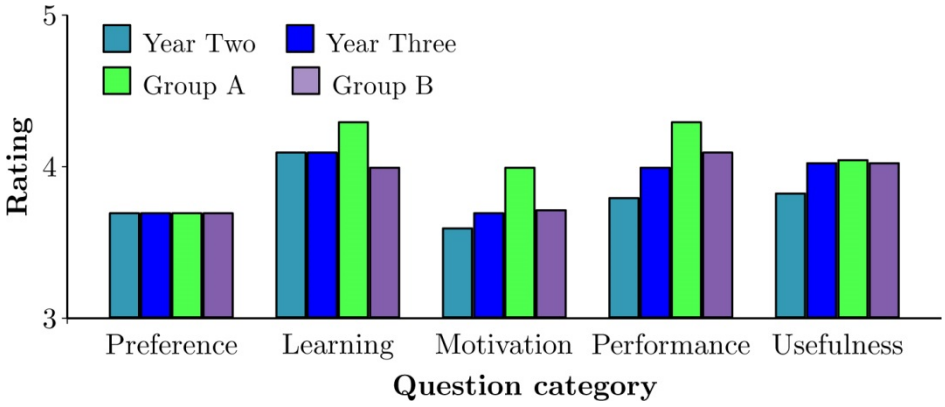

Figure 4: Average student ratings on question categories for the second part of survey. The vertical axis displays the average rank on a 1 to 5 scale; the horizontal axis corresponds to question categories.

In general, the students expressed favorable opinion in the four categories. Regarding the preference of students between online and paper-and-pencil homework, we observe that there are no significant differences between Year Two, Year Three and groups A and B. For the Learning category, there is no significant difference between Year Two and Year Three. It can be observed that Group A had a better opinion than group B. For the categories Motivation, Performance and Usefulness, we observed an improvement in student opinions from Year Two to Year Three and

again, except for Usefulness, that Group A had a slightly better opinion than group B.

In Figure 5, we present the averages for each of the question categories of part three of survey (see Section 3.2). Note that this refers to usefulness and quality of the interactive and multimedia books of MITO.

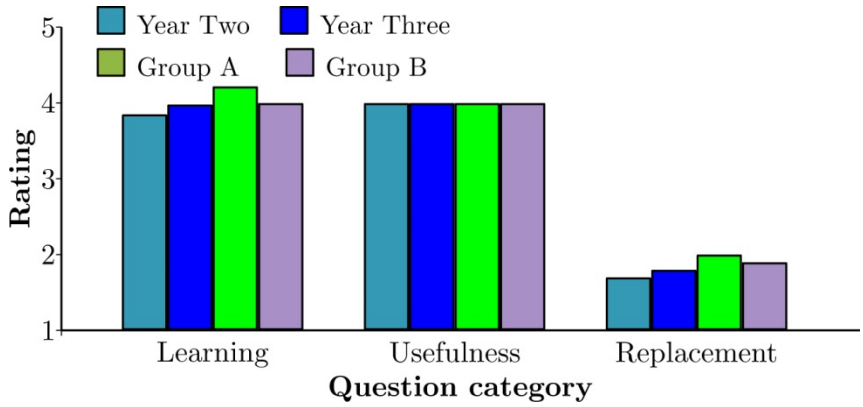

Figure 5: Average student ratings on question categories from the third part of the survey. The vertical axis displays the average rank on a 1 to 5 scale; the horizontal axis corresponds to question categories.

In the Learning and Usefulness categories the students expressed favorable opinion with values near 4 on a 1 to 5 scale. The students' opinions on Replacement category were unfavorable. We observe in the Learning category that there was a small improvement in Year Two relative to Year Three, and the group A had a slightly better opinion than the group B. In the category Usefulness we observe an average equal to 4 in all years and groups. We found that this type of educational content was well-received by the students. In the category Replacement we observe a very low average in all the groups which confirms that students do not feel that the interactive and multimedia contents can fully replace the teacher.

In the last section of the survey, we asked students to write a detailed and frank commentary about the strengths and the weaknesses of MITO with reviews and/or suggestions for future improvement. The positive comments could be categorized in the recognition that MITO is helpful in the study and understanding of the contents, usefulness of the STACK exercises feedback and likes about the online homework assignments. In negative comments we may found requests for more detailed resolutions in training exercises and videos, difficulty entering extensive expressions in Stack questions and requests for more partial credit on the answers given on the computer.

Over the three-year research period, the exchange of views with students in class was

common. Some of the comments received in the last section of the survey were also given in person and via email.

4.3 How did students use MITO?

To evaluate how the students use MITO we use descriptive statistics and informal interviews. Figure 6 and Figure 7 show two graphs with the distribution of the clicks on IMM of the Calculus students of groups A and B over the 23 week semester. These records include only interaction with the interactive and multimedia books and with the online training exercises. The totals of clicks were 41,802 for group A and 26,520 for group B.

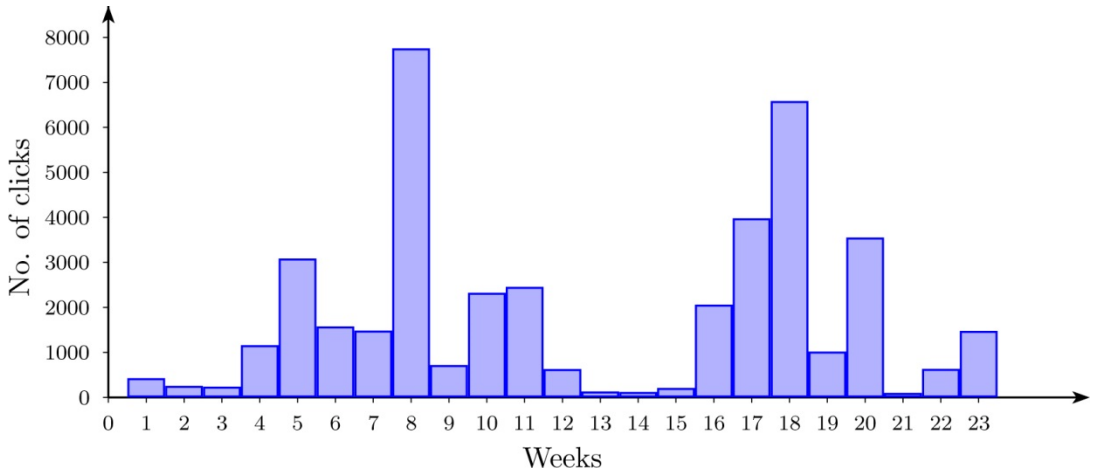

Figure 6: Number of clicks per week on IMM of the students of group A in the first semester of Year Three.

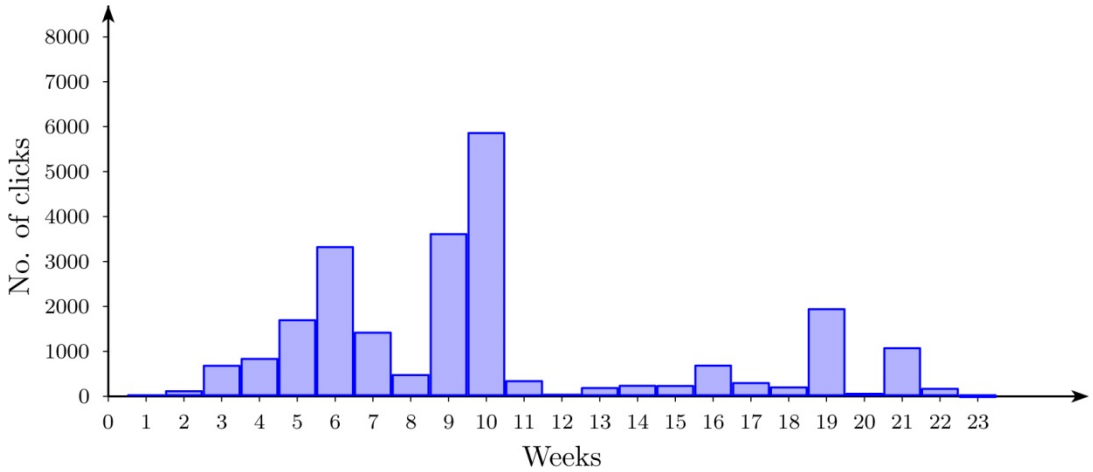

Figure 7: Number of clicks per week on IMM of the students of group B in the first semester of Year Three.

Table 5 presents the weeks in which the assessments took place for groups A and B.

Assessments	**Group A**	**Group B**
Homework assignment no. 1	Beginning of week no. 6	Middle of week no. 6
Homework assignment no. 2	Middle of week no. 8	Beginning of week no. 9
Test no. 1 (first half of material)	Beginning of week no. 9	End of week no. 9
Homework assignment no. 3	Beginning of week no. 12	End of week no. 16
Homework assignment no. 4	Beginning of week no. 18	
Test no. 2 (second half of material)	Middle of week no. 18	End of week no. 19
Final exam no. 1 (all the material)	End of week no. 20	Middle of week no. 20
Final exam no. 2 (all the material)	Middle of week no. 23	Middle of week no. 21

Table 5: Assessment weeks of the Calculus students of groups A and B in the first semester of Year Three.

Figure 6, Figure 7, and Table 5 indicate that the students used the online contents that we have provided for instructional purposes and for testing purposes about a week and a half prior to each one of the assignments. We may also observe that the number of accesses to IMM of the students of group B decreased substantially after the week no. 10 and this tendency was only inverted near the Homework assignment no. 3, Test no. 2 and, Final Exam no. 2. In the case of group A we observe some symmetry between the first nine weeks and between weeks 10 to 19 relatively to the number of accesses. These periods correspond to the first and second half of the semester and had 39% and 45% of the accesses, respectively.

Through informal interviews during the course we found that a small number of students worked in groups while using MITO to study and to solve homework assignments. We believe that in these study groups the students encourage each other to study and that its influence on the data presented in Figure 6 and Figure 7 is negligible. The students also recognized that MITO covers all program contents and that they did not usually use other source of contents to study.

4.4 How are related the results of MITO e-assessments and final grade in a continuous assessment model?

To answer this research question we only need to depict the relationships between online homework and final grades of groups A and B. Nevertheless, in order to better understand the student's performance we start by comparing the grades of groups A and B at the beginning and at the end of the Calculus course.

The results for normal homogeneity on basic skills test, which we use as a pretest, are shown in Table 6. Note that the basic skills test was done before the Calculus course.

Group	N	Mean	Standard deviation	T-value	p-value
A	44	8	4.22	3.813	<0.001
B	41	11.6	4.48		

Table 6: Results for homogeneity test on basic skills test of groups A and B (0-20 scale)

Since the p-value for the homogeneity test is lower than 0.05 groups A and B are not statistically homogeneous with respect to student grades on basic skills test (Box, Hunter, & Hunter, 2005). Moreover, we may consider that the mean of grades in group A is lower than in group B.

We compare now the final grades on the discipline of groups A and B. The results for normal homogeneity are shown in Table 7.

Group	N	Mean	Standard deviation	T-value	p-value
A	44	8.75	4.21	1.94	0.056
B	41	10.54	4.28		

Table 7: Results for homogeneity test on final grades of groups A and B (0-20 scale)

Since the p-value for the homogeneity test is higher than 0.05 we may conclude that groups A and B are statistically homogeneous with respect to student final grades (Box, Hunter, & Hunter, 2005). From the above, we conclude that the initial situation of non-homogeneous groups moved into homogeneous groups. We note, however, that the mean of final grades in group A is lower than in group B.

We proceed with the study of the linear relationship between the grades at the beginning and at the end of the Calculus course. In Figure 8 we show the scores on the basic skills test and on the exams of the students in the current study.

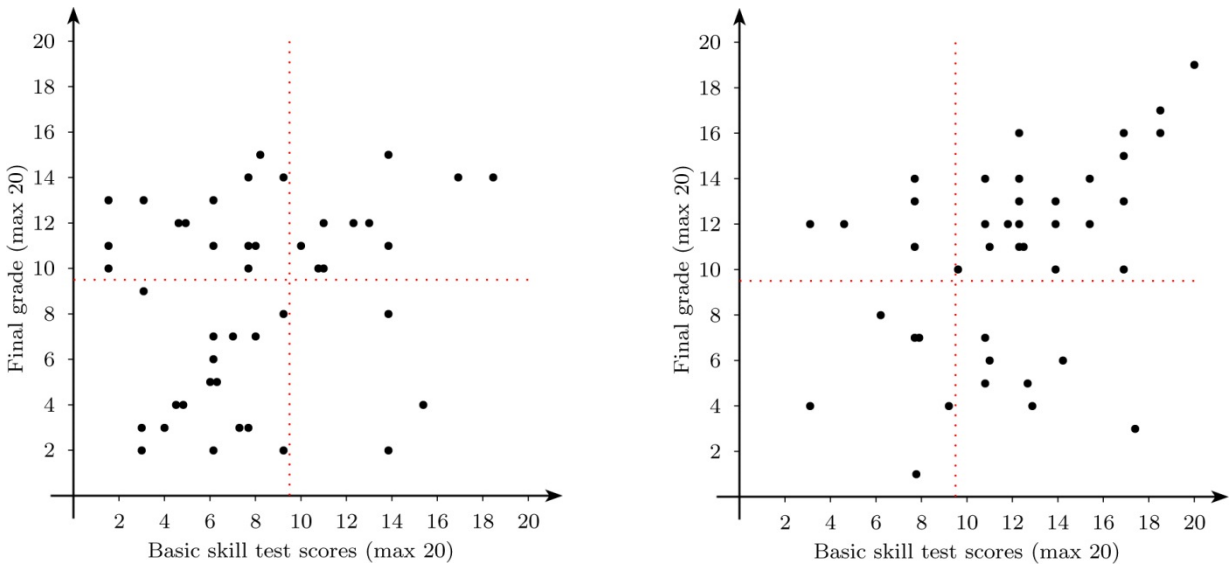

Figure 8: Student scores on basic skills tests and exams from Calculus program in group A (left) and group B (right) in the first semester of Year Three.

We can observe a weak or moderate linear correlation between the two variables. In fact, the Pearson correlation coefficient of these two variables is 0.25 in group A and is 0.43 in group B. Table 8 shows the Anova analysis for the relationship between these two variables and enhances this first diagnosis.

Group	N	F-value	p-value
A	44	2.319	0.135
B	41	0.001	0.009

Table 8: ANOVA results for comparing basic skills test scores and final grade.

The Anova results indicate that in group A the p-value is higher than 0.05 and in group B is lower than 0.05. Therefore, we must preserve the hypothesis of no linear relationship between the two variables in group A and reject this hypothesis for group B (Box, Hunter, & Hunter, 2005).

We also observe in Figure 8 that a significant number of students of each group with negative levels (less than 9.5) on basic skills test obtained approval in the discipline (at least 9.5). The proportion of students in this situation is 32% in group A and 12% in group B. The results of a Z-test for population proportions are shown in Table 9.

Group	N	Proportion	Z-value	p-value
A	44	0.32	2.1699	0.015
B	41	0.12		

Table 9: Z-test for population proportions across groups A and B.

Since the p-value in the proportions test is lower than 0.05 there is statistically significant difference between the proportion of students with negative levels on basic skills test that obtained approval in the discipline in groups A and B (Box, Hunter, & Hunter, 2005). Moreover, the proportion on group A is higher than in group B.

In order to answer the last research question, we proceed with depicting the relationship between online homework, given during the semester, and final grade. Figure 9 shows the scatter plots relating these two variables in groups A and B.

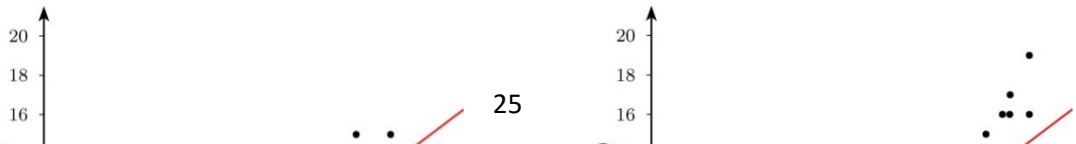

Figure 9: Student scores mean on homework and final grade from Calculus program in group A (left) and group B (right) in the first semester of Year Three.

The ANOVA results shown in Table 10 indicate that the p-value is lower than 0.05 in the two groups. Therefore, the hypothesis of no relationship between the two variables may be rejected.

Group	N	F-value	p-value
A	44	138.369	<0.001
B	41	48.641	<0.001

Table 10: ANOVA results for comparing homework scores and final grade.

The Pearson correlation coefficient of these two variables is 0.88 in group A and 0.81 in group B. In both cases the Pearson correlation coefficient is higher than 0.8 which indicates a high degree of positive linear correlation between the two variables. The determination coefficient is 0.77 in group A and 0.66 in group B. We can thus consider that students with better ratings on the final exams were those who had better scores on the homework and that the observed outcomes are well replicated by the linear model (Box, Hunter, & Hunter, 2005). These results are consistent with the results of the category Performance on part two of the survey (see Figure 4). Note that this category evaluates the students' opinion about the improvement of performance on final grade due to online homework.

5 Discussion and future work

This paper focused on the use of interactive and multimedia content associated with a system for computer-aided assessment. The interactive and multimedia books were used for instructional purposes, as they served as presentation resources in the classroom and as an interactive support to textbooks. The purpose of using a system for computer-aided assessment was to provide a formative assessment instrument that would allow students to study in a progressive

way, to learn from quiz feedback and to improve their academic performance. Therefore, self-assessed quizzes, homework assignments, and interactive and multimedia books enable us to combine the technology for instructional and assessment purposes. The study was conducted in a technology and management business school with 300 students in Year One, 778 students in Year Two and 884 students in Year Three. In Year Three a deeper study was conducted with two groups of Calculus students from engineering about several educational aspects in the learning process. The description of the creation process of MITO project and four research goals were proposed.

The presentation of the analysis, design, development and implementation shows that a project of creating instructional material requires a well-planned model and a great investment of time. Moreover, developing a project of this kind involves technical issues and advanced programming tasks. The construction of the educational contents had to be performed by people who have both technical skills and the teaching experience required for creating the contents. It was necessary to dedicate a server exclusively to MITO and often resorting to the technical support of the Distance Learning Unit of IPL. Our experience has shown that benefits compensate the costs at least at the long run.

The first research question was: How did teachers and reviewers evaluate MITO? The evaluation of MITO by reviewers and teachers through LORI indicates positive opinions about quality and reusability of MITO educational contents. In general, with exception of accessibility, the interactive modules of MITO scored well in all quality and reusability aspects of LORI. The item accessibility needs further consideration. The LORI results allows us to consider that the interactive modules of MITO are an e-learning resource with quality, are aligned with the goals, present good features of reusability, durability, and navigability and are capable to give useful feedback to the learner (Nesbit & Belfer, 2005). The teacher opinions gathered through informal conversations indicate a good acceptance of MITO and that is aligned among learning goals and learner characteristics. Throughout the three-year experience we felt the support and interest of colleagues with respect to MITO. We frequently received suggestions for improvement and development of new content. We note at this moment some interest from some colleagues in building interactive modules for Physics.

The second research question was: How did students evaluate the use of MITO contents? To

answer this question we resort to a user survey for two years presenting good features of validity and reliability (Johnson, & Christensen, 2012). The second part of survey, designed to gather opinions about the student's interaction with online training exercises and online homework assignments and the relevance of the exercises in achieving success in the course, presented good outcomes. In each one of the years, these were confirmed by student's claims. Relatively to learning, performance and usefulness, the students' opinions are good. Significant positive results were also obtained regarding the students' perceptions of the effect of these educational contents on motivation. The third part of the survey, related to the evaluation of the interactive and multimedia books of MITO, also gave us positive feedback. The students recognize the usefulness and the contribution of these contents to learning. An interesting point in the category of Replacement confirms that students do not feel that the interactive and multimedia contents can fully replace the teacher. Parts two and three of survey showed in each of the categories of Year Two and Year Three an overall improvement of opinion from Year Two to Year Three and that group A achieved better opinions than Group B. We believe that this was due to the improvement of the quality of the MITO contents from Year Two to Year Three and because of the different framework of lecturer with MITO in each group. The fact of the lecturer of group A being an author of the educational content of MITO and having used it in classroom may have influenced the students' opinions. We thus consider that although the classifications obtained in LORI indicate that the navigation through MITO is intuitive, a guided tour of MITO taken in the classroom by the teacher can improve this point. In the last part of the survey we collected general opinions about what students liked or disliked and suggestions for future improvements. We obtained an overall acceptance of the platform which is in agreement with the remainder of the survey. The negative comments gathered were particularly useful for future improvement.

The third research question was: How did students use MITO? To answer this question we considered two distinct groups in real-world settings: the same discipline in different courses and teachers. The two groups followed a continuous assessment model including mandatory randomized homework assignments in MITO: four in group A and three in group B. We concluded that the students of each group concentrated the use of MITO for instructional and for testing purposes at least ten days prior to each one of the assignments. Groups A and B had in the first half of semester a similar number of accesses. On the second half of semester, group A acceded

quite more to the educational contents of MITO than group B. We think that the main reasons were that the group A had one more homework assessment than group B and the use of interactive books as presentation resources in the classes of group A. By indication of the teacher, the students from group A may have considered the IMM as their first reference to study using the interactive and multimedia books together with online training quizzes. On the contrary, platform records indicate that the interaction of students from group B with MITO relied mostly on the online training quizzes.

From previous experience, we know that before the introduction of homework assignments the students' study was concentrated just before the exams. It is not surprising that students will accept more easily a regular homework regimen if they know that it will have a significant impact on the final grade (Croft, Danson, Dawson, & Ward, 2001). In our courses, homework contributed approximately 20% to the final mark. From the above, we believe that the use of a greater number of homework assignments with a greater weight can be easily accepted by the students and may increase student engagement with the contents of MITO. This could help the teacher to get good insight into the learning processes of students and to make grounded choices in the construction of his lessons.

The last research question was: How are related the results of MITO e-assessments and final grade in a continuous assessment model? The relations between the grades of a basic skills test, of MITO online homework assignments, and of final grade used to answer this question were carefully validated for the groups A and B. It was found that the scores of online homework assignments and final grade are strongly correlated in any of the groups. This result is in agreement with the existing literature (Varsavsky, 2004; Roth, Ivanchenko, & Record, 2008; Ruokokoski, 2009; Rasila, Majander, & Malinen, 2010; Angus & Watson, 2009; Huisman & Reedijk, 2012; Rivera, Ochoa, & Perez, 2013) and with the students' opinions gathered in survey. Certainly, if the correlation was weak, the teachers and the students would not have motivation to accept this kind of evaluation component in their courses. The extension of this finding to other groups and areas can be made based on this experience.

The data analyzed to answer the last research question gave us other findings that we may use as exploratory research for future work. It was found that despite the group A had a statistically lower basic Math skills level than the group B, the final grades of the two groups were

statistically homogeneous. On the other hand, it was found that the basic math skills scores and the final score on course were weakly correlated in group A and moderately correlated in group B. Additionally, it was verified that the proportion of students in group A that passed from a negative level in basic skills test for a positive final grade was significantly higher in group A than in group B. If we take into account that the group A had a lower basic Math skills level than the group B we can thus conclude that the results in terms of performance improvement were better in group A. In this experiment, groups A and B were different in teacher and undergraduate course, in the use of interactive books in the classroom, in the number of homework assignments and in the initially level of basic skills on math. This is clearly a real-world setting that may be found in many undergraduate courses. These differences between groups do not permit to conclude that the better performance improvement in group A was due to the greater use of MITO. Moreover, we must also consider the possibility that students had resorted to external sources of educational content like textbooks or tutorial videos. Despite this, all of the results obtained together with the literature cited lead us to believe in the positive effects of MITO on different aspects of the learning process.

The experience, results and project design presented in this study can be useful for similar projects, be generalized to populations of students from other institutions and work across different settings. The use of open source software or freeware in our work facilitates its adoption by teachers and interested researchers.

During our investigation, we received several requests to provide a virtual keyboard for entering answers on STACK. Although we recognize that this feature would facilitate the input of answers and reduce the frustration, we also see some disadvantages. Most students will be using computers in other areas of their work and will need to learn to adapt to the rigors of computer input.

The potential response tree of STACK provides useful feedback to students. However, we recognize that the implementation of exercises in which the number of fields is defined by the student may provide an important improvement in the interaction for some types of questions. The next step may be to adapt STACK and the platform of MITO to this feature.

We believe that teaching based in stages represents an effective way of promoting learning outcomes in mathematics and may reduce the achievement gap between groups of students

(Guskey, 2007). For this reason, we intend to implement the MITO e-learning platform to such a model and investigate its use in the continuous evaluation of our courses. The idea is to create a kind of activity that includes theoretical and practical concepts and an online quiz in STACK at each stage. Depending on the student's score on the quiz, he will either progress to the next stage or go back to a previous stage. This future work is based on mastery learning (Guskey, 2007) and aims to be a big step for MITO into the intelligent tutoring systems (Chrysafiadi, & Virvou, 2013; Zhang *et al.*, 2014).

Acknowledgments

The authors thank Ana Cristina Lemos, Ana Felgueiras, Ana Santiago, Carlos Campos, Eulália Mota, Helena Silva, Jorge Fatelo, Nelson Vieira, Paula Faria, Svilen Valtchev and Teresa Coimbra for collaborating in the research.

We also thank Vítor Rodrigues, engineer of the Distance Education Unit of the Polytechnic Institute of Leiria, for providing computer support.

We thank UED, CTC/OTIC and INDEA for their encouragement and constant availability.

Special thanks go to the Polytechnic Institute of Leiria, School of Technology and Management and Mathematics Department for their support and for the purchase of a dedicated server for the MITO project.

References

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Abridged ed.). New York, NY: Longman.

Andrade-Aréchiga, M., López, G., & López-Morteo, G. (2012). Assessing effectiveness of learning units under the teaching unit model in an undergraduate mathematics course. *Computers & Education*, 59(2), 594-606.

Angus, S.D. & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large dataset. *British Journal of Educational Technology,* 40(2), 255-272.

Baki, A., & Güveli, E. (2008). Evaluation of a web based mathematics teaching material on the subject of functions. *Computers & Education*, 51(2), 854-863.

Biggs, J. (2003). *Teaching for Quality Learning at University* (2nd ed.). Buckingham: The Society for Research into Higher Education/Open University Press.

Box, G. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2rd Edition). New Jersey: John Wiley & sons.

Chen, Y. (2012). A Study of Learning Effects on e-Learning with Interactive Thematic Video. *Journal of Educational Computing Research*, 47(3), 279-292.

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade, *Expert Systems with Applications*, 40(11), 4715-4729.

Creswell, J. W. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Second Edition). USA: Sage Publications.

Croft, A. C., Danson, M., Dawson, B. R., & Ward, J. P. (2001). Experiences of using computer assisted assessment in engineering mathematics. *Computers & Education,* 37, 53-66.

Frye, R. (1999). Assessment, accountability, and student learning outcomes [Electronic version]. Dialogue, 2, 1-11.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports student's learning. *Learning and Teaching in Higher Education,* 1, 3-22.

GNU general public license (2004). Free Software Foundation.

Guskey, T.R. (2007). Closing Achievement Gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery", *Journal of Advanced Academics*, 19, 8-31.

Huisman, R. & Reedijk, H. E. (2012). The impact of individual online tests in addition to group assignments on student learning. *Proceedings of the International Conference on Information Communication Technologies in Education,* 1, 654-667.

Harjula, M (2008). *Mathematics exercise system with automatic assessment* (MSc). Finland: Helsinki University of Technology.

Johnson, B., & Christensen, L. (2008). *Educational Research: Quantitative, Qualitative, and Mixed Approaches* (fourth edition). USA: Sage Publications.

Kaput, J., & Hegedus, S. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh , E. Hamilton, & J. Kaput (Eds.), *Proceedings of the foundations for the future in mathematics and science* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum.

Kim, S., & Chang, M. (2010). Does Computer Use Promote the Mathematical Proficiency of ELL Students? *Journal of Educational Computing Research*, 42(3), 385-305.

Lazakidou, G., & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics, *Computers & Education*, 54(1), 3-13.

Mayer, R. E. (2003). Elements of a science of e-learning. *Journal of Educational Computing Research*, 29(3), 297-313.

Maxwell, K., & Angehrn, A.A. (2010). Lessons learned from deploying a video-based Web 2.0 platform in an executive education context. *Communications in Computer and Information Science* 73, 195-201.

Nesbit, J., & Belfer, K. (2005). Collaborative evaluation of learning objects. In R. McGreal (Ed.), *Online education using learning objects* (pp. 138–153). London: Routledge/Falmer.

NCTM, (2000). Principles and standards for school mathematics. VA Reston.

Oetiker, T., Partl, H., Hyna, I., & Schlegl, E. (2011). *The Not So Short Introduction to* LaTeX 2ε (version 5.01). Retrieved from https://tobi.oetiker.ch/lshort/lshort.pdf.

Piskurich, G.M. (2006). *Rapid Instructional Design: Learning ID fast and right* (Second Edition). San Francisco: Pfeiffer.

Rasila, A., Havola, L., Majander, H., & Malinen, J. (2010). Automatic assessment in engineering mathematics: evaluation of the impact. In *Reflektori 2010: Symposium of engineering education. Teaching and Learning Development Unit.*

Reed, H., Drijvers, P., & Kirschner, P. (2010). Effects of attitudes and behaviors on learning mathematics with computer tools. *Computers & Education*, 55(1), 1-15.

Rivera, Z., Ochoa, R., & Perez, B. (2013, October 23-26). *Improving Student Results in a Statics Course using a Computer-based Training and Assessment System*. Paper presented at IEEE: Frontiers in Education Conference. doi: 10.1109/FIE.2013.6685165.

Roth, V., Ivanchenko, V., & Record, N. (2008). Evaluating student response to WeBWorK, a web-based homework delivery and grading system. *Computers & Education,* 50, 1462-1482.

Ruokokoski, J. (2009). *Automatic assessment in university-level mathematics* (MSc). Finland: Helsinki University of Technology.

Sangwin, C. (2007). Assessing elementary algebra with STACK. *International Journal for Mathematical Education in Science and Technology,* 38(8), 987-1002.

Sangwin, C. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press.

Sangwin, C. (2012). *Computer Aided Assessment of Mathematics Using STACK*. Paper presented at the 12th International Congress on Mathematical Education (ICME12), Seoul, Korea. Retrieved from University of Birmingham website: http://web.mat.bham.ac.uk/C.J.Sangwin/Publications/RL_Sangwin.docx.

Sangwin, C. (2003). Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications,* 23(1), 1-14.

Sangwin, C. (2010). *Who uses STACK? A report on the use of the STACK CAA system (International Technical Report).* Retrieved from University of Birmingham website: http://web.mat.bham.ac.uk/C.J.Sangwin/Publications/2010-3-1-STACK.pdf.

Steenhuis, H., Grinder, B., & Bruijn, E. J. (2009). The use(lessness) of online quizzes for

achieving student learning. *Int. J. Information and Operations Management Education,* 3(2), 119-148.

Varsavsky, C. (2004). Can online weekly quizzes contribute to learning in mathematics? IN S. C., Chu, W.C., Yang, T.D., Alwis, & K.C., Ang (Ed.), *Proceedings of the 9th Asian Technology Conference in Mathematics* (pp. 161-168). Singapore: ATCM Inc.

Veiga, A., & Amaral, A. (2009). Survey on the implementation of the Bologna process in Portugal, *Higher Education*, 57, 57–69.

Velichová, D. (2011). Dynamic Tools in Mathematical Education. In: *Procedings of 14th International Conference on Interactive Collaborative Learning,* 24-29.

Voss, H. (2011). *PSTricks graphics for TeX and LaTeX* (1st ed.). Cambridge, England: UIT.

Wells, J., Barry, R., & Spence, A. (2012). Using Video Tutorials as a Carrot-and-Stick Approach to Learning. *IEEE Transactions on Education,* 99.

Wild, I. (2009). *Moodle 1.9 Math*. Birmingham: Packt Publishing.

Zhang, L., VanLehn, K., Girard, S., Burleson, W., Chavez-Echeagaray, M., Gonzalez-Sanchez, J., & Hidalgo-Pontet, Y. (2014). Evaluation of a meta-tutor for constructing models of dynamic systems. *Computers & Education*, 75, 196-217.