

# Intelligent tutorial system based on personalized system of instruction to teach or remind mathematical concepts \*\*

Rui Paiva – School of Technology and Management, Polytechnic Institute of Leiria and Centro de Matemática da Universidade do Porto, email: rui.paiva@ipleiria.pt.

Milton Ferreira – School of Technology and Management, Polytechnic Institute of Leiria and Center for Research and Development in Mathematics and Applications, University of Aveiro, email: milton.ferreira@ipleiria.pt.

Miguel Frade – School of Technology and Management, Polytechnic Institute of Leiria and Computer Science and Communication Research Centre, email: miguel.frade@ipleiria.pt.

## Abstract

*The growth of the higher education population and different school paths to access an academic degree has increased the heterogeneity of students inside the classroom. Consequently, the effectiveness of traditional teaching methods has reduced. This paper describes the design, development, implementation and evaluation of a tutorial system (TS) to improve student's engagement in higher mathematics. The TS design was based on the Personalized System of Instruction of the Mastery Learning pedagogical approach and can be implemented in any higher education course with mathematics needs. The TS consists on small self-paced modularized units of educational contents, including tutorial videos, notes and formative e-assessment with personalized feedback. The TS ensures that the student is only allowed to proceed to the next unit after he or she achieves the required mastery criterion of the current unit. The TS was implemented in the Quantitative Methods course of an undergraduate degree and received good acceptance from students. It was also recognized that TS contributed to learning and engagement with the discipline. Through an experimental research experience it has been shown that the imposition of restrictions on the advance to the next level by a mastery criterion leads to a significant improvement in student's engagement and performance.*

*Keywords: Mathematics; Mastery Learning; Online assessment; Video; Tutorial system.*

## Introduction

The democratization of the Portuguese higher education in recent years launched a number of challenges in the teaching-learning methodologies of its institutions. The accommodation of students with different school routes has increased the heterogeneity of students of different cycles of training and has reduced the effectiveness of traditional education based on groups (Santos, 2000). The problem of maintaining the scientific requirements when there is a growth of the higher education population is difficult to resolve and motivates an adaptation of the traditional teaching methods (Boavida, 2010).

Mastery Learning (ML) is a teaching-learning model that assumes that the individualized education leads to better learning than usual groups based learning. The concept of ML dates back to 1963, when John Carroll assumed the radical position that fitness, and even more than intelligence, is a measure of the amount of time necessary for a person to learn, suggesting that all students are able to achieve the same degree of learning, if the time and sufficient learning opportunities are provided (Carroll 1963, 1989). It is based on this optimistic view that all students can learn any subject that Benjamin Bloom developed his ML theory (Bloom, 1968). From the perspective of Bloom, the variability of the performance levels of students in "traditional education" is explained by the fact that teachers teach the same content in the same way to all students of a group in a given period of time.

The second key feature of the ML is the alignment of targets (Guskey, 2001, 2007). All three components of education, including learning objectives, teaching and evaluation should focus on the same goal. For example, if a teacher expects his students to learn a certain subject, he should provide a reasonable number of opportunities for them to practice it during the education process. The final evaluation should also evaluate the proficiency of the students. Another unique feature of the ML is frequent testing and the establishment of a mastery criterion, where a student can only advance to the next level if he or she achieves the established mastery criterion (Slavin, 1987).

The ML can be implemented in two different ways: Learning for Mastery (LFM) and Personalized System of Instruction (PSI). Developed by Bloom himself, LFM involves small groups of about 30 students, where the teacher is the main manager of knowledge and time for each step. Although LFM is more widely used because of its adaptability to conventional classrooms where teaching time and curriculum are relatively fixed, it is more difficult to ensure that all students in a group receive sufficient amounts of time and attention (Guskey & Pigott, 1988). The PSI was conceived in 1968 by Fred Keller to mit-

\*\* Accepted author's manuscript (AAM) published in [Journal of Computer Assisted Learning, 33 (4) (2017), 370-381] [DOI: 10.1111/jcal.12186]. The final publication is available via url <http://dx.doi.org/10.1111/jcal.12186>.

igate this problem. The main idea is based on learning through small modules of teaching materials of autonomous study, rather than directly through the teacher (Fox, 2004). The classes are complementary or additional in PSI, because unlike classes, which occur only at certain times, learning materials can be accessed by students the number of times they require. The students perform formative assessments when they consider themselves prepared. Each module has a number of training assessments, similar but not identical, that can be done by students until they reach the mastery criterion (Slavin, 1987). In PSI, the individualized feedback and answering questions of the students is ensured by several teachers or assistants. The main differences between LFM and PSI are shown in Table 1 (Kulik & Bangert-Drowns, 1990).

	<b>LFM</b>	<b>PSI</b>
Class setting	Group based	Individual based
Pacing	Teacher-paced	Self-paced
Instructional materials	Teacher-presented	Written
Correctives	Individual/group tutorials	Restudy materials
Duration	2-108 weeks	Usually 1 semester
Mastery criteria	Slightly lower	Higher
Grade level	Mostly pre-college	Postsecondary level

Table 1: Differences between LFM and PSI

The effectiveness of the ML can be influenced by several factors that can even jeopardize its applicability. Among the main adverse factors are the need to spend more time in teaching, self-discipline of students and the inconvenience of the more advanced students have to wait for their peers. One of the main criticisms of the ML is the time required for feedback and corrections of students be substantial, preventing the teaching of the curriculum on time. On the other hand, for all students to achieve the level of mastery provided at each level in time, it requires self-discipline. In this case, the younger students are at a disadvantage. Older age of students in higher education enables the PSI in their courses to be more efficient once they have a greater self-discipline and individual initiative (Klecker & Chapman, 2008). Note, however, that the implementation of the PSI in higher education requires a considerable amount of time and effort required to design and organize the curricula. To mitigate the problem of waiting by the more advanced students, it is recommended to propose them more advanced challenges (Kazu *et al.*, 2005).

Although has been proven the educational effectiveness of PSI in higher education the dominant lecture-tutorial teaching method has never seriously changed and PSI has fallen into disuse (Rae & Samuels, 2011). Fox (2004) states that by 1979 there were close to 3000 publications on PSI, mostly reporting its remarkable success and no less than 5000 PSI courses known to be in existence. Between the reasons for the fall of PSI are the adverse factors described above, the greater time of initial development needed for its courses and the difficulty of adapting self-paced courses to the academic calendar (Elton & Bridge, 1977; Rae & Samuels, 2011).

The Information and Communication Technologies (ICT) have a great potential for any individualised teaching program (Rae & Samuels, 2011). The term Computer Aided PSI (CAPSI) was introduced by Kinsner and Pear (1988) to describe the course they have developed at University of Manitoba in Canada in 1983. It is currently being used at several educational institutions for psychology courses and is referred to webCAPSI ([www.webcapsi.com](http://www.webcapsi.com), accessed 10/03/2016). Their main use of ICT is to deliver and manage the formative testing feature of PSI. The development of the ICT that we had in recent years has facilitated the implementation of the model PSI of ML in other subject areas of higher education (Wong & Kang, 2012). In particular, over the last decade there was a considerable effort to develop software which automatically assesses student's answers to mathematics questions (Sangwin, 2013, 2016). Actually, the transcription of paper-based mathematics examinations into an electronic format is feasible for a significant proportion of the questions as currently assessed (Sangwin, 2016). 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; Pacheco-Venegas, López, & Andrade-Aréchiga, 2015.). Research con-

firmly the positive effects of the implementation of the PSI pedagogy in mathematics and programming higher education by the use of video clips and e-assessments (Rae & Samuels, 2011; Shafie, Shahdan, & Liew, 2010; Kularbphettong, Kedsiribut, & Roonrakwit, 2015).

The MITO - "Online Interactive Training Modules" ([www.mito.ipleiria.pt](http://www.mito.ipleiria.pt)) is a research and development project of the Polytechnic Institute of Leiria (IPL), originating from the Department of Mathematics (DMAT) of the School of Technology and Management (ESTG). It was started in 2010 with the main objective of supporting the traditional teaching and distance learning of mathematics at the higher education level through the construction of interactive educational content. Paiva, Ferreira, Eusébio, & Mendes (2015) present the project design features, its development and implementation, evaluation of the tools used and results of studies with the students involved in the project.

This paper describes the design, development, implementation and evaluation of an intelligent online tutorial system (Chrysafiadi & Virvou, 2013; Zhang *et al.*, 2014; Kularbphettong *et al.*, 2015) designed to teach or to remind math concepts essential for the integration of students of engineering, economics, management, among others, in mathematics higher education. The tutorial system (TS), which we developed in MITO platform, consists of small self-paced modularized units of educational contents including tutorial videos, notes and formative e-assessments with personalized feedback. It was designed based on the PSI model of ML and can be implemented in any course requiring mathematical bases. In this sense, there is a mastery criterion established in each one of the modules. In addition, we present an experimental research on the impact of the imposition of a mandatory mastery criterion in the students' engagement and performance in a course of Quantitative Methods of an undergraduate course. Our main contribution to the field is the build, the implementation and evaluation of the effectiveness of a TS based on PSI using a set of mathematical software that we thought to be among the most developed and promising for teaching mathematics.

## **Tutorial system of MITO**

In this section we present the stages of analysis, design, development and implementation of the TS for preparation in basic mathematical proficiency. The TS focuses on the theme real functions of one real variable and its design is based on the PSI model of the ML.

### *Analysis*

The main motivation for the creation of the TS was based on the existence of undergraduate courses in ESTG receiving students with very different background in mathematics. This is considered one of the main reasons for the high failure rates and dropout checked in mathematics courses of undergraduate students. To answer this problem it was required a solution based on self-study of mathematics bases, with formative assessment and automatic feedback.

### *Design*

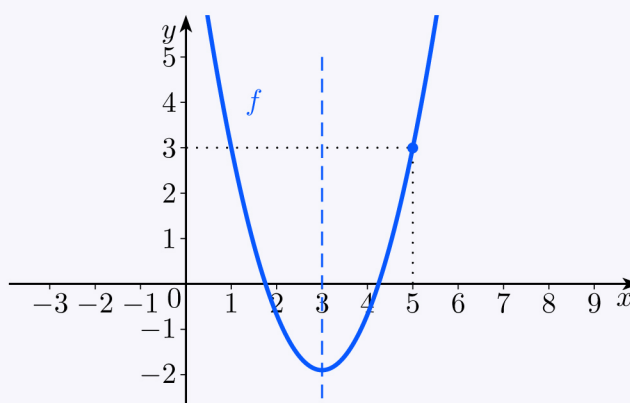
The main methodological idea in TS design was the PSI model of ML, namely the division of the subject into small modules of teaching materials for self-study. The experience in MITO with the use of mathematical software (Paiva *et al.*, 2015) was fundamental to decide that the TS would be constituted by theoretical notes, video lectures, step by step tutorial videos of exercises and parameterized tests with automatic assessment in computer-aided assessment STACK (Sangwin, 2007, 2013, 2016).

The tutorial videos were produced using LaTeX code (Oetiker, Partl, Hyna, & Schlegl, 2011) and possess the advantage of high quality images (Paiva *et al.*, 2015). The theoretical notes consist in a written version of the video lectures. The goal is to allow students to write remarks directly in the theoretical notes when they are watching the video lectures. The step by step tutorial videos of exercises are interspersed with the video lectures.

STACK is a computer aided assessment package for mathematics, which provides a question type for the Moodle quiz. In order to give an idea of its advantages over the conventional tests of teaching and learning platforms like Moodle, we present some of its features. In STACK it is possible to create questions dependent of parameters defined by the author. For example, we can generate a set of questions involving the function  $f(x) = a(x-3)^2 + h$  where  $a \in \{-3, -2, -1, 1, 2, 3\}$  and  $h \in \{0, 1, 2, 4, 5\}$ . Therefore there are 30 similar versions but not identical of this question. When a student opens a test with 5 questions of this

kind, each question is chosen between the versions available in the database. As a result, it is possible to provide the student, with relative ease, a reasonable amount of training e-assessments. In STACK, students can submit their answer in the form of a mathematical expression. For example, the student can answer a question by entering a polynomial or a square root. Essentially, the mathematical expressions received are evaluated using the free software Maxima ([www.maxima.sourceforge.net](http://www.maxima.sourceforge.net)). This computer algebra system manipulates symbolic and numerical expressions and allows to establish the mathematical properties of the student answers. This feature is particularly useful when the question has multiple correct answers. STACK includes multiple response tests to compare expressions, a syntax checker, an answer preview and gives feedback related to the student answer. This last feature helps to maintain student motivation and can facilitate the student's reasoning process (Harjula, 2008). In Figure 1, we present a STACK question involving the function  $f(x) = a(x-3)^2 + h$  described above. This question has several correct answers, include a random graph and exemplifies the formative feedback that accompanies the response. The random graph has 24 different versions corresponding to the variations of  $a$  and  $h$ . This feature results from an adaptation of the source code for Moodle and STACK developed in the framework of MITO (Paiva *et al.*, 2015).

In the above figure is part of the graph of a quadratic function  $f$  which has  $x = 3$  as symmetry axis and contains the point  $(5, 3)$ .



Give an example of an expression that defines such function.

$f(x) =$   .

Your last answer was interpreted as:

$$(x - 3)^2 - 2$$

Your answer is partially correct.

The graph of  $f(x) = (x - 3)^2 - 2$  has  $x = 3$  as symmetry axis. However, the graph does not contain the point  $(5, 3)$ .

Your mark for this attempt is 0.6. You may improve the mark if you correct the answer.

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

Figure 1: Example of a STACK question

Once chosen the types of resources that compose the TS, the next step was the choice of the topics to be covered by TS and its division into small modules. An evaluation carried out by teachers, led us to choose the following topics of the chapter of real functions of one real variable: generalities, polynomial functions, module function, odd and even functions, exponential and logarithmic functions, limits, continuity and derivatives. These topics provided a division of the TS in 7 levels with similar working time.

The TS design step became completed with the adaption of the e-learning platform Moodle of MITO to display the contents of each module sequentially with a mastery level e-assessment at the end of each module. Additionally, it was considered essential to ensure the possibility of imposing restrictions on the advance to the next level by a mastery criterion and a good navigability among the contents. The adaptation of the platform was carried out using the modules available in [www.moodle.org](http://www.moodle.org), with adaptations of the respective source code.

### Development

The educational content developed for the 7 TS levels include 30 video lectures associated with 119 step by step tutorial videos of exercises, theoretical notes and 7 STACK formative e-assessment with resolution. Each level is organized sequentially with the video lectures followed by the corresponding tutorial video exercises and ends with a STACK formative e-assessment (see Figure 2). The formative e-assessments include more than 30 similar versions, but not identical, of each question. The student writes the answer in algebraic form and receive immediate feedback with possible suggestions for assistance and video or notes/recommendations. When a student finishes an e-assessment he or she can get a new e-assessment consisting of similar but not identical questions until reaches the desired mastery level. TS has the possibility of providing a task only after the previous one was completed. The passage to the next level can be conditioned by obtainment of a certain rating on the appropriate training test. These features fit with the fundamentals of PSI.



Figure 2: Organization of a TS level

To ensure an intuitive user guidance the progress of a student in each level is shown by a progress bar (see Figure 3).

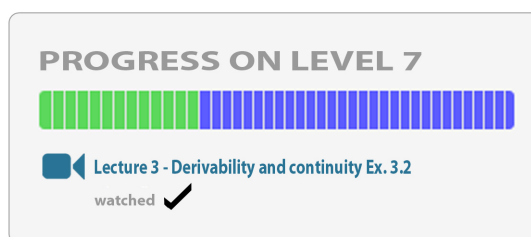


Figure 3: TS progress bar

Additionally, the progress bar provides the teacher with a summary report of the student’s progress in each level. This feature allows the teacher intervention in the tasks where there are more difficulties.

### Implementation and evaluation

The TS was implemented in the 1st semester of the academic year 2014/15 in the discipline Quantitative Methods of the undergraduate course of Marketing at ESTG. With the goal of preparation in basic mathematical proficiency, the students used the TS combined with 2 e-assessments conducted in person (see Figure 4) and a written exam. We will give more details about this topic and present the evaluation of TS in the Methodology and Results sections.



Figure 4: e-assessment conducted in person

## Methodology

The research experience took place in ESTG during the 1st half of the school year 2014/15 with 72 students of Quantitative Methods of an undergraduate course of Marketing. The TS was used by these students as a self-study tool and as a complement to regular classes. According to the fundamentals of PSI, one task should only be made available after the student achieve the desired level of mastery in the previous task. This is known as mastery criterion. To evaluate the effectiveness of this rule, the 72 students were divided into two similar groups through a random assignment technique (Creswell, 2008): experimental group (EG) and control group (CG). Table 2 presents relevant information on the participants.

Group	No. of students			Average age
	Female	Male	Total	
EG	18	17	35	18.5
CG	19	18	37	18.4

Table 2: General information about the 72 students of EG and CG

The mastery criterion established for EG in each of the seven levels of TS was that a student could only advance to the next level after getting a grade not less than 60% in the formative test of the current level. Otherwise, he or she was unable to advance and receives instructions to do some more work and try again a similar but different formative test. The CG students did not have any restrictions on the access to content, but received the same study recommendations and were recommended to obtain a grade not less than 60% in the formative tests at the end of the levels. In all of the seven levels of the TS, the students from EG and CG received deadline recommendations for achieving the mastery criterion.

During one semester both qualitative and quantitative data were gathered to evaluate the students' opinion, engagement with the TS and the influence of the mandatory mastery criterion on the students' performance. A total of 15 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		No. of questions	Answer scale	Cronbach's alpha
Metacognitive Awareness Inventory		52	Dichotomic	0.814
Informal conversations			Open opinion	Not apply
Platform records				Not apply
Student survey no. 1		9		Not apply
Student survey no. 2	Part 1	3		Not apply
	Part 2	13	Likert (5 options)	0.822
	Part 3	1	Open opinion	Not apply
Formative and summative assessments	7 formative e-assessments	Diverse	Open	Not apply
	2 summative e-assessments	7 and 8	Open	Not apply
	1 written exam	12	Open	Not apply

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

The data used to obtain the EG and CG involved a well-designed student survey and the Metacognitive Awareness Inventory (MAI). The survey (see student survey no. 1 in Table 3) was conducted before the beginning of the experience and measures the student level on math (SLM) and demographic characteristics. At the same time was carried out an inventory of metacognitive perceptions using the Metacognitive Awareness Inventory (MAI) instrument proposed by Schraw and Dennison (1994) in order to identify students' perception on their learning skills and abilities, study habits, awareness of learning and cognitive processes among others.

Another quantitative data resource used in the study was the platform records of EG and CG which held the interaction with the educational resources of TS. These data were used to obtain statistical results about students' engagement with TS and how much the students of the two groups used it.

Quantitative information about the student's performance was obtained directly from the grades of two summative e-assessments and one written exam (see Table 3). In order to motivate the self-discipline, the summative e-assessments were conducted on the class one day after the proposed dates for the completion of levels 3 and 7 and evaluated respectively the three first and last four levels of TS. The written exam covered all the material of the 7 levels. Table 4 presents the formative and summative assessments in their chronological order.

Item	Local	Weight
Formative e-assessment of level 1	Outside the class	-
Formative e-assessment of level 2	Outside the class	-
Formative e-assessment of level 3	Outside the class	-
Summative e-assessment 1	Class	15%
Formative e-assessment of level 4	Outside the class	-
Formative e-assessment of level 5	Outside the class	-
Formative e-assessment of level 6	Outside the class	-
Formative e-assessment of level 7	Outside the class	-
Summative e-assessment 2	Class	15%
Written exam	Class	70%

Table 4: Formative and summative assessments

Measuring how students use and evaluate the influence of the TS in learning, motivation, performance, and other educational aspects involved a well-designed survey conducted at the end of the experience (see student survey no. 2 in Table 3). A qualitative data resource was obtained by informal conversations conducted between the teachers and the students that participated in the experience.

## Results

The Portuguese higher education system uses a grade scale from 0 to 20, which is also used throughout this paper. The results of the survey no. 1 and of the MAI show that CG and EG are statistically similar. In Table 5 are shown the results for normal homogeneity of the SLM of EG and CG.

Group	N	Score Mean	Standard deviation	p-value
EG	35	12.75	2.4	0.9328
CG	37	12.8	2.6	

Table 5: Results for homogeneity test on SLM of CG and EG (0-20 scale)

Since the Normal distributions requirements for the homogeneity test are verified and its p-value is higher than 0.05 we may conclude that CG and EG are statistically homogeneous with respect to SLM (Box, Hunter, & Hunter, 2005). The random assignment technique used in the selection of the elements of each group ensures the homogeneity in the other characteristics that may be relevant to the experience (Creswell, 2008).

The averages on the 52 items of the MAI instrument (Schraw & Dennison, 1994) shown in Table 6 indicate that an overall statistical uniformity exists in the two groups in students' perception of their own learning process and study strategies (Box, Hunter, & Hunter, 2005).

Group	N	Score Mean	Standard deviation	Anova
EG	35	64.2	11.3	F = 0.44, p=0.51
CG	37	65.4	11.1	

Table 6: Results on the metacognitive awareness inventory instrument for each group.

To evaluate the students' engagement on TS we use the number of clicks on the contents and the grades of the 7 formative e-assessments of TS. Each one of the variables verify the requirements on normality for the tests below. Table 7 indicates that there is not statistical uniformity in the number of clicks on the TS in the two groups ( $p < 0.0001$ ). The mean of accesses was statistically higher in EG than in CG.

Group	N	Mean	Standard deviation	p-value
EG	35	1002.36	123.4	$< 0.0001$
CG	37	757.92	131.2	

Table 7: Results on the number of clicks for each group.

The average grade of the 7 formative e-assessments are shown in Table 8 together with results of the normal homogeneity test.

Group	N	Mean	Standard deviation	p-value
EG	35	15.42	3.1	$< 0.0001$
CG	37	10.73	3.4	

Table 8: Results for homogeneity test on the grades of the 7 formative e-assessments of CG and EG (0-20 scale)

Since the p-value for the homogeneity test is much lower than 0.05 we may conclude by conventional criteria that EG and EC are not statistically homogeneous on the grades of the 7 formative e-assessments. The EG had an average grade significantly higher than CG.

We compare now the grades on summative assessments of EG and CG. The results for normal homogeneity are shown in Table 9. Each one of the variables verify the requirements on normality for the tests below.

	Group	Score Mean	Standard deviation	p-value
e-assessment 1	EG	14.1	2.3	0.0219
	CG	12.8	2.4	
e-assessment 2	EG	13.2	2.4	0.0044
	CG	11.5	2.5	
Written exam	EG	13.4	2.4	0.0085
	CG	11.8	2.6	
Final grade	EG	13.5	1.8	0.0005
	CG	11.89	1.9	



Table 9: Results for homogeneity test on summative e-assessments and written exam of CG and EG (0-20 scale)

Since the p-value for the homogeneity test is lower than 0.05 in all assessments and also in the final grade we may consider that EG and CG are not statistically homogeneous with respect to student's grades on the summative assessments and in the final grade (Box, Hunter, & Hunter, 2005). Moreover, we may consider that the final average grade in CG is lower than in EG ( $p < 0.05$ ). From the above, we conclude that the initial situation of homogeneous groups moved into non-homogeneous groups with the final grade of EG statistically higher than in CG.

The mean values of each of the four aspects considered in the user survey applied to the 72 students from the two groups are shown in Table 10. In general, the users expressed favourable opinions in the four aspects of the TS considered in the survey.

User's opinions on	No. of items	EG mean	CG mean
Positive aspects on learning	4	4.2	4.1
Positive aspects on motivation	4	3.9	3.7
Positive aspects on performance	3	4.1	3.9
Web design	2	3.7	3.7
Totals	13	4	3.9

Table 10: Results of the user survey instrument on different aspects of TS (0-5 scale)

Through informal conversations with the students involved in the research we got opinions and suggestions for improvement of TS. The general opinion is that TS contributes for student learning and engagement with the discipline and that interactive and multimedia content is very useful as self-study tool. The suggestions received were related with the creation of more contents and the web design improving.

### Conclusions and future work

This paper focused on the description of ML teaching-learning methodology and in the analysis, design, development, implementation and evaluation of a TS designed for preparation in basic mathematical proficiency. The TS consists of 7 levels constituted by notes, tutorial videos and formative STACK e-assessments at the end of each level. Its design is based on the PSI model of the ML. One of the foundations of this teaching-learning model is that a student only should move to the next level after achieving the mastery criterion established at the current level. For this reason, TS was developed with the feature of conditioning the advance to the next level by the classification in a STACK e-assessment. To assess the impact of this imposition in the student engagement and performance we developed an experimental research. The study was conducted in a technology and management business school with 72 students of Quantitative Methods of an undergraduate course of Marketing. At the beginning of the experience we obtained two similar groups EG ( $n=35$ ) and CG ( $n=37$ ) using the random assignment technique.

The presentation of the analysis, design, development and implementation of TS shows that a project of creating instructional material to implement ML in higher education 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 data analysed to measure the engagement in self-study by TS were the platform records and the grades of the 7 formative e-assessments of TS. We found by the platform records that the number of accesses to contents of the students of EG is significantly higher than in CG. Relatively to the 7 formative e-assessments it was found that the EG had an average

grade significantly higher than CG. Since the only difference between EG and CG was the restriction on the advance to the next level by the grade not less than 60% in the formative test, we may conclude that this feature of TS promotes the student's engagement. It confirms the effectiveness of the ML rule that the student must achieve the desired level of mastery before progressing to the next level.

To assess the impact of the mandatory mastery criterion on the performance of the students we used the grades on summative e-assessments and on a written exam. It was found that the grades of CG are statistically lower than of EG in each of the two summative e-assessments, in the written exam and in final grade. As the initial situation of homogeneous groups moved into grades of EG higher than in CG we may conclude that the grade restriction feature of ST in the formative tests has positive effects in student's performance. This is true not only on the summative e-assessments but also on written exams.

The student's opinions about TS were gathered with a survey conducted at the end of the experience. Relatively to learning, motivation and performance, the students' opinions are favourable. The item web design needs further consideration. The survey results and the informal conversations allows us to consider that students think that ST contributes positively to learning, motivation and performance on the discipline of Quantitative Methods.

We can conclude that this paper describes a successful experience of implementation of PSI as complement to regular classes in teaching mathematics at the higher education level. The PSI inconveniences of the incremented time needs, self-discipline of students and that the more advanced students have to wait for colleagues have been successfully contoured by the features of the TS, by being online and through the articulation of the TS with two e-assessments and a written exam. The TS was well received by students and it was found that the positive effects in student's engagement and performance are boosted by the imposition of mandatory mastery criteria on the advance to the next level. We can thus conclude that this feature allows a better use of the same educational contents. The experience, the results and the tutorial system design 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.

Future work includes the web design improvement of TS, its evaluation by reviewers and the implementation in other undergraduate courses and distance learning. 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. This and other exciting features called "reasoning by equivalence" will be available in the next STACK version (Sangwin, 2016). At that time we will have much more to investigate about the gains that these will provide for the PSI.

## **Acknowledgments**

The authors thank the Department of Mathematics of ESTG for the support given to the experience.

We also thank to the Distance Education Unit of the Polytechnic Institute of Leiria, for providing us with computer support.

The second author acknowledges support by Portuguese funds through the CIDMA – Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (“FCT - Fundação para a Ciência e a Tecnologia”), within project UID/MAT/04106/2013.

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

- 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.
- Bloom, B. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1-5.
- Boavida, J. (2010). Ensino superior o novo século. Problema o solução? *Revista Portuguesa de Pedagogia*, 44(1), 9-36.
- Box, G. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd Edition). New Jersey: John Wiley & sons.
- Carroll, J. B. (1963). A model of school learning. *Teaching College Record*, 64(8), 723-733.
- Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26-31.
- 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.
- Elton, L., & Bridge, W. (1977). *Individual study in undergraduate science*. London: Heinemann Educational.
- Fox, E.J. (2004). The Personalized System of Instruction: Instruction: A flexible and effective approach to mastery learning. In D. J. Moran, & R.W. Malott (Eds.), *Evidence-based educational methods* (pp. 201-221). San Diego: Elsevier Academic Press
- Guskey, T. R., & Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *The Journal of Educational Research*, 81(4), 197-216.
- Guskey, T. R. (2001). Mastery learning. In J. S. Neil & B. B. Paul (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 9372-9377). Oxford: Pergamon.
- Guskey, T.R. (2007). Closing Achievement Gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery", *Journal of Advanced Academics*, 19, 8-31.
- Harjula, M (2008). *Mathematics exercise system with automatic assessment* (MSc). Finland: Helsinki University of Technology.
- 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.
- Johnson, B., & Christensen, L. (2008). *Educational Research: Quantitative, Qualitative, and Mixed Approaches* (fourth edition). USA: Sage Publications.
- Kazu, I. Y., Kazu, H., & Ozdemir, O. (2005). The effects of mastery learning model on the success of the students who attended "usage of basic information technologies" course. *Journal of Educational Technology & Society*, 8(4), 233-243.
- Kinsner, W., & Pear, J. J. (1988). Computer-aided personalized system of instruction for the virtual classroom. *Canadian Journal of Educational Communication*, 17(1), 21-26.
- Klecker, B. M., & Chapman, A. (2008). Advocating the implementation of mastery learning in higher education to increase student learning and retention. Paper presented at the Meeting of the Mid-South Educational Research Association Knoxville, TN.
- Kulik, C.L.C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265-299.
- Kularbphetong, K. & Kedsiribut, P., Roonrakwit, P. (2015). Developing an Adaptive Web-Based Intelligent Tutoring System using Mastery Learning Technique. *Procedia – Social and Behavioral Sciences*, 191, 686-691.
- 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>.
- Pacheco-Venegas, N.D. & López, G. & Andrade-Aréchiga, M. (2015). Conceptualization, development and implementation of a web-based system for automatic evaluation of mathematical expressions. *Computers & Education*, 88, 15-28.
- Paiva, R.C. & Ferreira, M. S., Mendes, A. G., & Eusébio, A. M. J. (2015). Interactive and multimedia contents associated with a system for computer-aided assessment. *Journal of Educational Computing Research*, 52(2).
- Rae, A. & Samuels, P. (2011). Web-based Personalised System of Instruction: An effective approach for diverse cohorts with virtual learning environments? *Computers & Education*, 57, 2423-2431.
- 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.
- 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. (2016). Automation of mathematics examinations. *Computers & Education*, 94, 215-227.
- Sangwin, C. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press.
- Santos, S. M. (2000). As responsabilidades da Universidade no acesso ao Ensino Superior, Em A. P. Soares, A. Osório, J. V. Capela, L. S. Almeida, R. M. Vasconcelos & S. M. Caíres (orgs.), *Transição para o Ensino Superior*. (pp 69-78). Braga: Universidade do Minho.
- Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460-475.
- Shafie, N., Shasdan, T., & Liew, M (2010). Mastery learning Assessment Model (MLAM) in Teaching and Learning Mathematics. *Procedia Social and Behavioral Sciences*, 8, 294-298.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175-213.
- 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.
- Wong, B. S., & Kang, L. (2012). Mastery learning in the context of university education. *Journal of the NUS Teaching Academy*, 2(4), 206-222.
- Zhang, L., VanLehn, K., Girard, S., Burlison, W., Chavez, H., Gonzalez-Sanchez, & C., Hidalgo-Pontet, Y. (2014). Evaluation of a meta-tutor for constructing models of dynamic systems. *Computers & Education*, 75, 196-217.