ADeAPTIVE

**Advanced Design of e-learning Applications
Personalizing Teaching to Improve Virtual Education**

Advanced Design of e-Learning Applications Personalizing Teaching to Improve Virtual Education.

Erasmus + Project 2017-1-ES01-KA203-03826

ADeAPTIVE

Project Information:

Project Acronym:
ADeAPTIVE Project

Project Title:
Advanced Design of e-Learning Applications Personalizing Teaching to Improve Virtual Education

Agreement Number:
2017-1-ES01-KA203-03826

SubProgramme or KA:
 KA203 Strategic Partnerships for Higher Education

Project Website:
www.adeaptive.com

Participant Partners:
UCLL, JYU, UAM, Eurecat, EucA, UVT

© Advanced Design of e-Learning Applications
Personalizing Teaching to Improve virtual Education

Funded by the
Erasmus+ Programme
of the European Union

1

IO3: Study of the different learning strategies detected and recommendations

CONTENTS:

A1: create contents and guidelines for the students in order to create a reflective learning diary

A2: create contents and guidelines for teachers to assess reflective learning diaries

A4: methodology and design of adaptive tests that will give real time feedback to students about their learning strategy

A8: report on the what, why, now, how, what and next of the student reflective diaries.

Author: UCCL

SUMMARY

●   For starting this IO and establishing the first steps of content related to learning diaries, UCLL has chosen to recreate courses within the model UAM presented and is described in IO2.

●   These courses were 'an entrepreneurship 2D-game' for refugees wanting to start a business of their own, a networked course for international entrepreneurs and finally a course on social entrepreneurship for teachers either with a refugee background or teaching refugee children.

●   All courses are situated in a non-formal context and are organised for adult learners from different backgrounds, from different countries of origin and perhaps more importantly for ADeAPTIVE with different pedagogical concepts.

●   UCLL opted to work from the perspective of the course developer in co-creation with the target audiences.

●   This means that we have redeveloped courses, modifying them in the UAM-model and later in the gamification techniques of JYU demonstrating the strong interaction between partners and IOs.

ADeAPTIVE

1. Introduction

From a pedagogical perspective, we introduced 'nesting' for the 2D-game, Bloom's taxonomy and Csikszentmihalyi's flow theory for the international entrepreneurs and the pedagogical recommendations of L2LByTe2 for the social entrepreneurship course. Furthermore, we moved towards blended learning as the 2D-game was initially offered as a stand-alone course but this did not work out. In the second year, we worked within the Flipped Class and connectivism approach whereas in the third year, the social entrepreneurship course is embedded in a Community of Practice context introducing mentoring and peer learning.

From the co-creation sessions, we learnt from the refugees that a) they need more digital literacy; b) their pedagogical experience is quite often limited to objectivism where the teacher transfers all knowledge and the learners regurgitate this at the exam; c) in their culture oral communication is highly valued d) on a personal level they need to reconnect to people to feel a sense of belonging again. Merely offering a virtual learning environment clashes with these needs.

Entrepreneurial learning with its fail forward approach was an unknown concept as well and this is probably why our 2D-game which was our way of testing the basic knowledge about setting up a business of their own, was not a success. We discussed the reflective diaries during the meeting at JYU but it was dismissed as our Finnish colleagues had experienced that this was not a very popular testing method amongst learners. We then introduced the digcompedu at the project meeting in Timisoara and agreed that we would ask learners to create a selfie based on this framework. Selfies are part of reflective learning, creating awareness within the learner of his/her learning challenges. Again, this is not a learning path most refugees are familiar with.

Based on these co-creation sessions and the debates within the consortium, we have created a zero test for the social entrepreneurship course for teachers in which we both explain the pedagogical concepts and the levels of digital literacy while asking them to link it to their expertise.

Funded by the
Erasmus+ Programme
of the European Union

Based on the three development cycles transforming textbooks into the ADeAPTIVE model we have formulated some recommendations for educators both as developers as well as facilitators.

2. WORK CARRIED OUT

2D-Game: Opgefrist!

Target audience: refugees

Learning outcomes: at the end of the game learners have acquired a Dutch vocabulary linked to entrepreneurship.

At the end of the game learners have a map of all steps to set up a business in Flanders linked to the relevant entrepreneurship agencies.

The challenges are tame problems as there is always one right answer.

The game was offered as a stand-alone learning environment

The game was developed in php by UCLL. It put the participants in 'authority' addressing them as entrepreneurs through the look and feel of the game (entrepreneurship visualisations). There is a dashboard monitoring their progress, an exit button, a competitors' chart, virtual money to be gained based on their result, a digital badge and links to external resources are provided should participants make a mistake.

Based on the UAM-model we created 6 questions for each of the five chapters. We achieved personalization through allowing learners to work through it anywhere, anytime and anyplace.

All learners could reach the end as mistakes are supported by resources (external links) making sure they could all tick the right answer.

Second year: Fresh Start

Target audience: refugees

Funded by the
Erasmus+ Programme
of the European Union

Learning outcome: at the end of the game participants can start up a business of their own.

The challenges are a combination of tame and wicked problems.

The game is developed by Eurecat in Moodle

The game was offered in a connectivist, flipped class format, together with a rhizome course with further links for deep learning and with a face-to-face demo day for external experts in the field of entrepreneurship.

We introduced Bloom's taxonomy and Csikszentmihalyi's flow to increase both immersion and personalization of the learning trajectory.

Each chapter was subdivided in challenges based on the competences 'remember', 'understand', 'apply', 'analyse', 'evaluate', 'create'.

The model looks as follows. All four chapters- Well-Being (2), Starting up a business (4), Legal (2) and Financial issues (4)- are defined by challenges per Bloom's category: 6 for remember, 5 for understand, 4 for apply, 3 for analyse, 2 for evaluate and 1 for create. Per category the questions are organized from facets to the major task which is the number 1 in the model. All learners start from the first chapter of the course, Well-Being, finding their own pitch for the future. From 'apply' onwards learners are asked to find answers between the entrepreneurial culture of their country of origin and of their host country. Thus, we create intrinsic motivation as it is about their own passions within their own contexts.

Learner A starts from the final task (task 1). Should he fail he has to answer two facets of this category before he can get to the next category. If he does not fail, however, he has a fast trajectory from category to category to the final task in 'create'.

Learner B can follow a more linear path building up to the main challenge per category from the bottom up. He does all the tasks moving from the easier questions to the final task (task 1).

In this way, we ensure the tasks are challenging and they have the skills to be doing them.

Funded by the
Erasmus+ Programme
of the European Union

With the exception of the remember-category, all final tasks involve designing and sharing a file about their own business. These files are given feedback by their peers and by the external experts from the refugee community. Rubrics are given as a link to guide this feedback. As such, the programme guarantees active and engaging tasks with clear factors of success.

3ʳᵈ Year M.Genies, the journey that takes you into the UNE (Unknown Nation of Educators)

Target audience: teachers with a refugee background and teachers of refugees and migrant children/adults

Learning outcome: to become a social entrepreneur (or intrapreneur) addressing learning pains in multicultural educational settings.

The challenges are predominantly wicked problems. Remember and understand are about the concept of social entrepreneurship. From application onwards learners have to bridge the gaps between culture of origin and receiving countries.

The game is offered within the community of practice of participants.

We created 24 questions according to Bloom's taxonomy and included the gamification techniques as developed by JYU:

User types:

Achievement orientation: dashboard (carpet M) to see progress, Deadline is visualized by blurring the contours of the cave, badge

Immersion orientation: storification, humour, avatar but also the immediacy of the others and the choice between the fast track (only the number 1 challenges) or the at ease track exploring all questions. The look and feel of the game is based on the genies entering Aladdin's cave.

Social orientation: living library, cave M (a CoP), M-coach (individual guide), brightness of the magic lamp indicates number of learners working in the virtual learning environment simultaneously, peer feedback, connect to outside experts.

And we have a variety of tasks which is good for all three orientations

User report: length of participation, number of days participated, discontinuity of participation, drop-out alarm.

3. ZERO TEST

How would you define your level of digital skills? Tick one level

O Newcomers (A1) have had very little contact with digital tools and need guidance to expand their repertoire.

O Explorers (A2) have started using digital tools without, however, following a comprehensive or consistent approach. Explorers need insight and inspiration to expand their competences.

O Enthusiasts (B1) experiment with digital tools for a range of purposes, trying to understand which digital strategies work best in which contexts.

O Professionals (B2) use a range of digital tools confidently, creatively and critically to enhance their professional activities. They continuously expand their repertoire of practices.

O Experts (C1) rely on a broad repertoire of flexible, comprehensive and effective digital strategies. They are a source of inspiration for others.

O Pioneers (C2) question the adequacy of contemporary digital and pedagogical practices, of which they themselves are experts. They lead innovation and are a role model for younger teachers

There are many pedagogical approaches in the digital age. Tony Bates has compiled them in his collaboratively written open source book: Teaching in a digital age- Second edition. Guidelines for designing teaching and learning. You can find both the book and the knowledge clips here: https://www.col.org/resources/teaching-digital-age

How would you describe the dominant model of knowledge construction and its teaching

practice in your country of origin? Tick one

O   Objectivism. Knowledge is facts, formulas, terminology, principles, theories. Teaching is transmission of this body of knowledge. Learning is to comprehend, reproduce and add to the knowledge based on empirical evidence and the testing of hypotheses. Tests would require students to find 'right answers' and justify them.

O   Behaviorism. Human behaviour is seen as predictable and controllable. Teaching relies on the stimulus-response theory. Certain behavior can be reinforced through feedback for instance.

Behaviourists attempt to maintain a high degree of objectivity in the way they view human activity, and they generally reject reference to unmeasurable states, such as feelings, attitudes, and consciousness.

O   Cognitive approaches define three domains of learning: cognitive (thinking), affective (emotion) and psycho-motor (doing). Learning consists of comprehension, summary icon, analysis, synthesis, generalization, evaluation, decision-making, problem-solving and creative thinking. Teachers teach learners how to learn, how to develop stronger or new mental processes for future learning, and how to develop deeper and constantly changing understanding of concepts and ideas.

O   Constructivists emphasise the importance of consciousness, free will and social influences on learning. Learners construct new knowledge and create meaning or understanding is achieved by thinking or reflecting on new information. Social constructivists believe that this process works best through discussion and social interaction, allowing us to test and challenge our own understandings with those of others. Reflection, seminars, discussion forums, small group work, and projects are key methods and online collaborative learning, and communities of practice.

O   Connectivism: knowledge is distributed across a network of connections, and therefore, that learning consists of the ability to construct and traverse those networks. It implies a pedagogy that:

(a) seeks to describe 'successful' networks (as identified by their properties, which I have

characterized as diversity, autonomy, openness, and connectivity) and

(b) seeks to describe the practices that lead to such networks, both in the individual and society – which I have characterized as modelling and demonstration (on the part of a teacher) – and practice and reflection (on the part of a learner) (Bates, 99).

In your teaching practice, so far, have you predominantly worked with

O   tame problems, where the answer is fixed, predetermined, or provided by the teacher for simple regurgitation.

O   wicked problems where the data are in flux, simple solutions are hard to come by, and what works one time, doesn't work the next (Hanstedt, 89).

Hanstedt, P. (2018). Creating wicked students. Designing courses for a Complex world. Sterling, Virginia: Stylus.

When playing games (serious or entertainment) I

O   never play games

O   want to keep track of my progress, be the first to complete tasks

O   step into another world

O   connect with other players

My future in the receiving country will be as

O   volunteer in an educational context

O   Co-teacher

O   Language teacher

O   Teacher of Religion

O   Teacher

O   Educational entrepreneur

O   Intercultural mediator

Recommendations Developer

1) Check their pedagogical DNA: how are they used to learn? Check how gamification is perceived in their country. Make sure they can situate their learning habits within what they have to do in the game.

2) Find your ideal achievement for each chapter, cut it down in ideal facets for each competence level and let go of the notion that 'students' should cover all.

3) Create a social context in which your gamers can meet.

4) Avoid failure: provide resources (online library) right from the beginning so that gamers can review what they should know prior to starting the game or check while answering, and check having answered incorrectly.

5) Appoint a game moderator from within the group who takes care of the informal conversations, reminds them to continue, explains.

6) Put them in 'authority': address them as experts.

7) Put a timer on the tasks as in different cultures time and especially deadlines are defined differently.

8) Make sure that the badge is accredited so that gamers feel they haven't been wasting their time.

Facilitator

1) Introduce the game in a face-to-face context

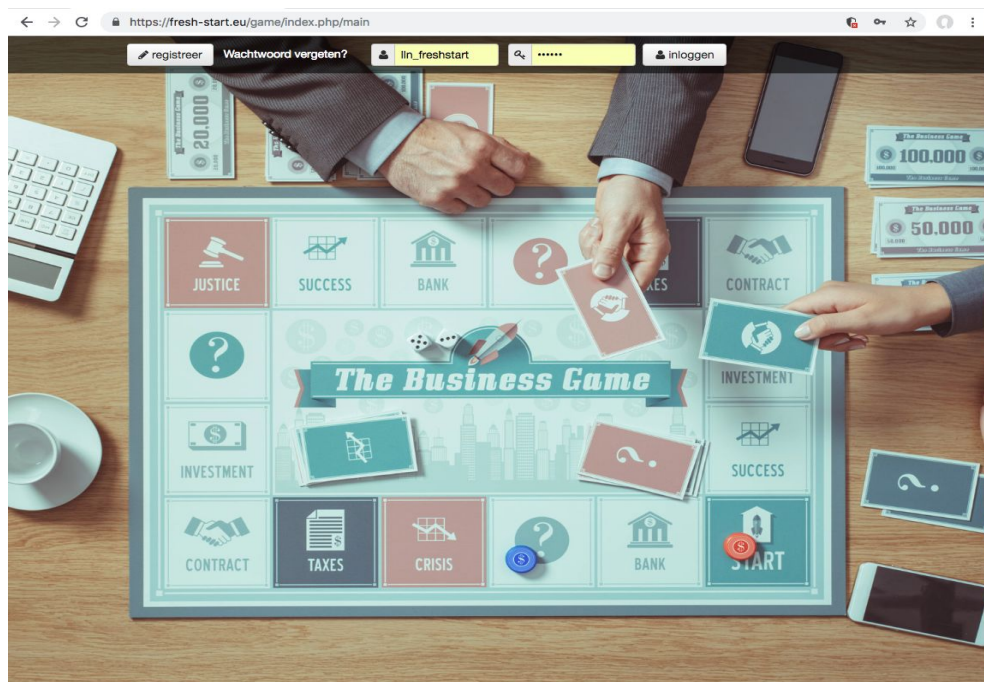2) Provide face-to-face time to discuss what's happening in the game

Funded by the
Erasmus+ Programme
of the European Union

3) Organise group gaming sessions where people work together through the game.

Screenshots

2D-Game:

Year one

Look and feel of a business game

2nd year

Year 2

U: upload

P: peer evaluation

B:

badge

ALL ones lead to the final step. The 'create' level generates an open badge.

CREATE:

Produce new or original work

Design, assemble, construct, conjecture, develop, formulate, author, investigate

EVALUATE:

Justify a stand or a decision

Appraise, argue, defend, judge, select, support, value, critique, weigh

© Advanced Design of e-Learning Applications
Personalizing Teaching to Improve virtual Education

Funded by the
Erasmus+ Programme
of the European Union

17

ANALYSE

Draw connections among ideas

Differentiate, organize, relate, compare, contrast, distinguish, examine, experiment, question, test

APPLY

Use information in new situations

Execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch

UNDERSTAND

Explain ideas or concepts

Classify, describe, discuss, explain, identify, locate, recognize, report, select, translate

REMEMBER

Recall facts and basic concepts

Define, duplicate, list, memorize, repeat, state

Learner's trajectory

Remember

Funded by the
Erasmus+ Programme
of the European Union

ADeAPTIVE

Learner's trajectory

Remember

They start from question six. They can't move on before they have uploaded the link with their result (stop sign). They need to have two questions right (green arrows) before they are given the choice to stay at the same level or move on to the next level (smiley).

Third Year

EMERgenCeS UCLL

Project code: 2019-1-BE02-KA204-060226

ADeAPTIVE

Welcome to M.Genies, the journey that takes you into the UNE (Unknown Nation of Educators)

Dear M. Genie

Welcome to our country: UNE. We are still building it! Like any indefinite document it invites the quest for specific features. And that is why we need you: the M.Genies.

Like in any wonder tale, genies come from another world. Be they a ghost, a spirit, a phantom, they always bring pieces of information that help protagonists or even the whole story to go on. Likewise, brilliant thinkers help societies to evolve into a brighter future. And this brighter future is the creation of UNE: the Unknown Nation of Educators. Here, educators, both as or of refugees, create a knowledge society where all learners (refugees, non-refugees, old, young) step with confidence into the unknown.

In M.Genies, we take you into the unknown and offer you challenges. These challenges allow you to become your own M.Genie who becomes a guide towards UNE.

After you have completed all quests, you will get a magic lamp, a badge certifying your expertise and a key into UNE.

You have four powers during the trajectory

1) For inspiration, you can rely on the 'living library of EMERgenCes': (http://erasmus-emergences.eu/). Here you can find Communities of practice of refugee teachers, literature review, good practices, inspiring examples, links to our social media pages with recent initiatives. We have also included a 'for your information' sheet where you can learn more about the pedagogical principles underlying M.Genies.

2) Cave M.: here you can chat with your peers, find reviewers. Each group has its own closed facebook group of EMERgenCeS. In this group, you introduce your avatar, get graphic presentations about the group's digital skills, knowledge construction, wicked/tame problems, and preferred future profiles. You also upload your tasks to get feedback from your peers.

3) M-coach: a teacher/ intercultural mediator with a refugee background who is already familiar with the educational system of the receiving country. Once registered, you are introduced to the M-coaches. These M-coaches chat about your wellbeing, give feedback, explain and also connect with you should you lag behind.

4) Carpet M.On top of your screen you see your personalized dashboard with the name of your avatar, the steps you have taken and still have to take and the remaining time to do so. At the end, you are awarded a badge that you can link to your Europass.

(Visualisation: number of people participating is seen through the brightness of the magic lamp. Progress is visualized by the movement of a carpet on top of the screen. Deadline is visualized by blurring the contours of the cave).

(Back office information: length of participation, number of days participated, discontinuity of participation, drop-out alarm)

© *Advanced Design of e-Learning Applications*
Personalizing Teaching to Improve virtual Education

Funded by the
Erasmus+ Programme
of the European Union

21

ADeAPTIVE

A3: analysis of students' learning strategies based on the feedback in piloting experience 1.

A5: analysis of students' learning strategies based on the feedback in piloting experience 2.

Author: UAM

SUMMARY

- This document presents a self-learning method based on a system of adaptive tests in which questions are automatically modified while the student is being tested.
- The main objective of this method is to serve as support for student learning and self-evaluation, as a model for study and assessment, and to improve the acquisition of knowledge in higher education.
- The method has been tested with real students in subjects that include both theoretical and practical classes.
- Results from this experiment demonstrate that the tool is useful in most of the cases, with most advantages in theoretical contents.

1. INTRODUCTION

The increased need of monitoring and interacting tasks with students, resulting from the application of new teaching models, can be met using computer-aided learning (CAL) strategies [1]. The large numbers of students enrolled on Massive Open Online Courses (MOOC) make manual monitoring impractical without utilizing the educational potential of computers [2]. As large-scale data processing can be carried out with the power of information technology, MOOCs have become a viable teaching method which requires no direct supervision of teaching staff.

However, the use of computers in teaching poses challenges such as the replacement of student-teacher interaction by automatic methods. [3]. One of the topics that has gathered more interest is the possibility of using an automatic tutoring system to effectively advise and guide learners [4-7]. In this regard, Computer Adaptive Testing (CAT) offers great potential due to its dynamic adaptation to different learner profiles [8].

One of the main disadvantages of utilizing these tests is the requirement of a good calibration method [13] in which learner models improve the evaluation process [19-21]. Both Multistage Adaptive Testing (MAT) [22] and cognitive diagnostic assessment [23] offer advantages in combination with CAT specific feedback.

In this document, we present a series of adaptive tests applied to both the theoretical and practical parts of the same subject with the objective of analyzing their potential positive effects in both kinds of teaching strategies. To demonstrate the positive effects of the method, we will compare results of students that used the adaptive tests with students that did not have the opportunity of doing so (students from previous years and from other parts of the subjects).

In Section II of this document, we propose the self-evaluation system based on adaptive tests. Section III analyses the results of the study conducted in academic years 2016/17 and 2017/18 with students from two different university degrees sharing a common subject. Finally, Section IV summarizes and describes the results of the study.

2. Experimental and computational details

The experiment with real students took place during academic years 2017/2018 and 2016/2017 and involved 2 subjects with similar numbers of students. The first one was Biochemistry which forms part of the second year of the University Degree in Biology at the Faculty of Sciences at the Universidad Autonoma de Madrid (UAM). Of the 140 students enrolled in the subject, 25 to 30 students were enrolled in their compulsory practice classes for a duration of two weeks (4 hours a day). The adaptive tests were included in the e-valUAM platform [24], which was available to students from the first day of the classes, so that they could access the tests throughout the teaching period. At the end, the students took a final exam which consisted of knowledge imparted in class and of the material covered

by the e-valUAM adaptive tests. The second subject included in the study was also called Biochemistry, but in this case, it was taught in the Degree of Chemistry, involving 3 groups of students, 2 of Biochemistry Theory and 1 from Biochemistry Laboratory Practices. The total number of students enrolled was 145 in theoretical classes and 129 in practical classes. The duration of teaching was one school semester.

In this document we have used e-valUAM: a platform designed to deliver adaptive tests that dynamically modify questions presented to students. It is primarily a self-assessment and learning system, which can also be used for final evaluation purposes. The adaptive test strategy used by e-valUAM in this study is represented in Fig. 1. e-valUAM is also optimized for giving detailed and accurate information on students' activity. This fact enables instructors to monitor student development, informs learners of their progress and provides feedback on assimilation of content. This feedback can be used by students to correct their learning deficiencies in a timely manner. An example of real time positive and negative feedback is presented in Fig. 2.

Funded by the
Erasmus+ Programme
of the European Union
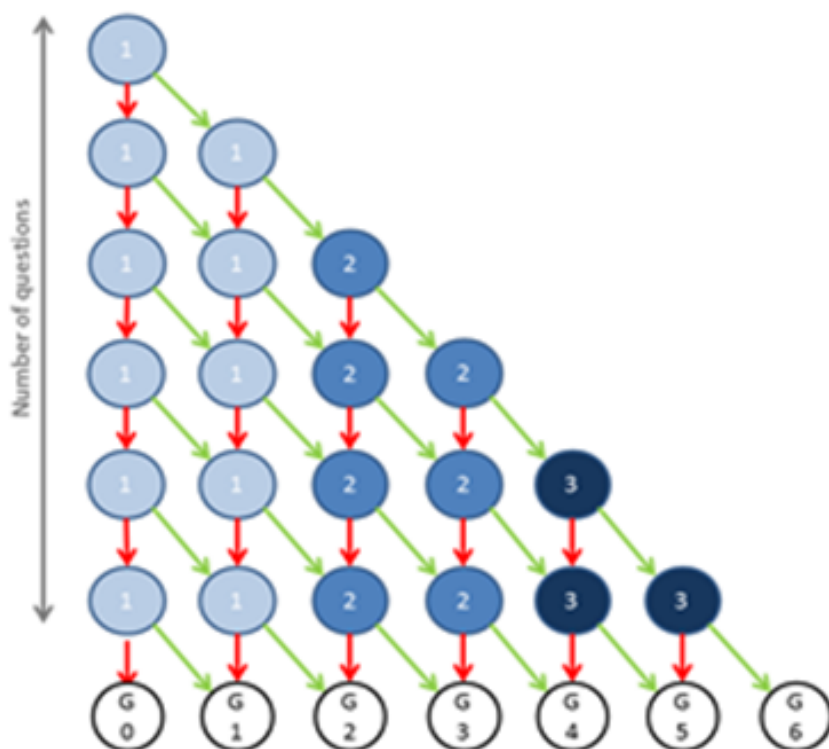
Figure 1: Representation of the proposed adaptive test model: Unidirectional Levels Method (ULM). Green arrows indicate the path students take when they answer a question correctly. Red arrows point the way when a question is answered incorrectly. In all cases, the number of total questions is fixed (6 in the case of the diagram represented). Grades are denoted at the bottom of the figure, being better as you move to the right.

The tests that were used by students in the autonomous learning phase were designed with 4 levels, where questions of each level increase in difficulty. At lower levels, the questions contained general knowledge, while at the higher levels, advanced knowledge and answer accuracy were required.

Students had 3 possible answers to each question, with only one correct answer. The difficulty of the questions would remain at the same level until the minimum number of correct questions was achieved. At that point, students would start answering questions from the following level, as shown in Fig. 1. We developed 3 different tests. The first one corresponded to the theoretical contents (20 questions), and the other ones corresponded with the contents of practical lessons (20 and 16 questions respectively). The length of the tests was adapted to the size of the contents of each part of the subject. We have divided the practical contents in 2 different tests because the practical lessons from the 2 grades involved in the study (Chemistry and Biology) had differing contents. In the case of the theoretical lessons, we only included the subject of Chemistry, so we did not need to develop 2 different tests.

Figure 2: Questions in tests. a) Example of questions to be answered from Biochemistry practice classes in the Chemistry degree. The notification in red indicates that the previous answer was wrong. b) Example of question from Biochemistry theoretical classes of the Biology degree. The notification in green indicates that the previous answer was right.

3. Results and discussion

Graphs in Fig. 3 correspond to results obtained by two students with different behaviors. Fig. 3a corresponds to a student who has used the platform many times in a single day. The score obtained in the theory exam (carried out with different questions and without using

the platform) appears in orange. Fig. 3b contains the results of a student who has used the platform for 3 days. It is worth noting that both exam scores are lower as the ones obtained in the final attempts on the platform since questions in the theory exam did not correspond to the e-valUAM questions.
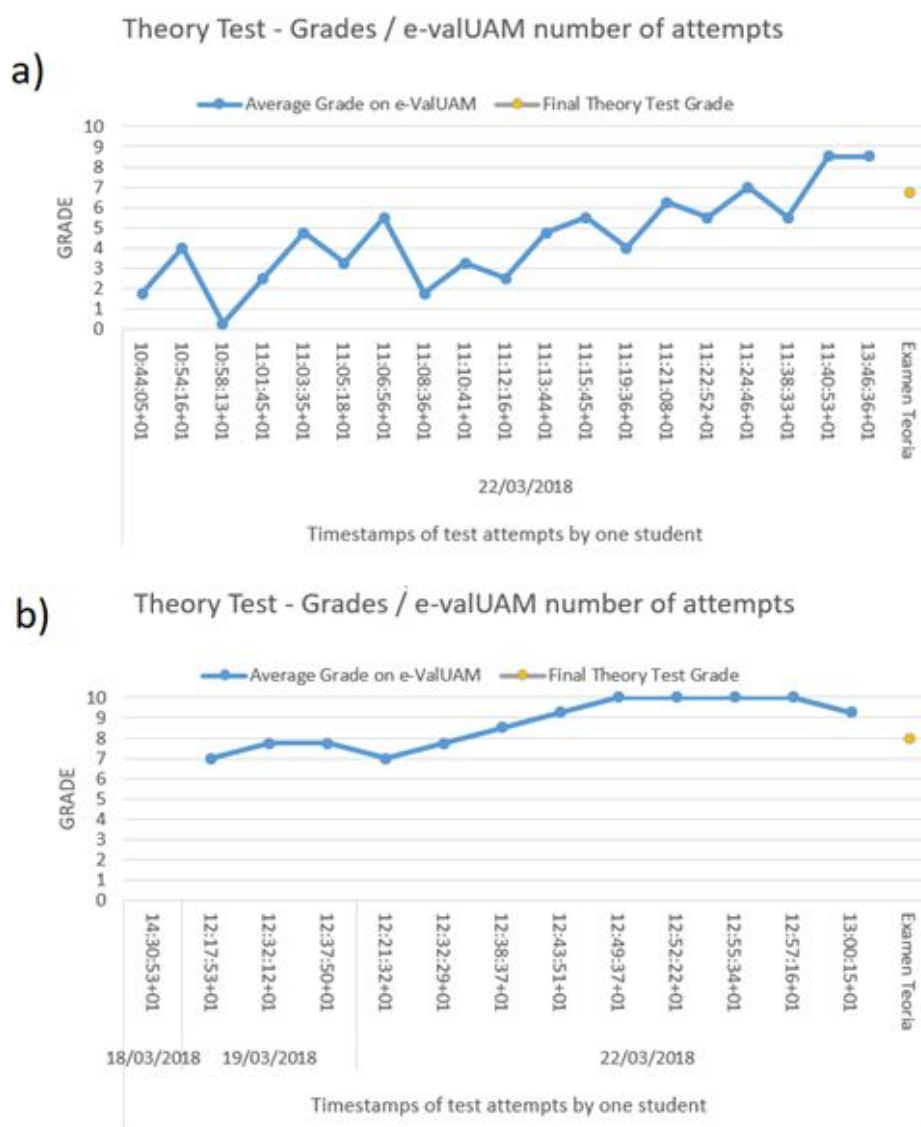
Figure 3: a) Grades of a student using the platform many times over the course of one day. b) Grades of a student using the platform many times over the course of 3 days. Last dot in both graphics represents the final theory test grade. Degree: Chemistry, Subject: Biochemistry, Exam: Practice Theory Test on e-valUAM, Academic Year: 2017/18.

In Fig. 4 we represent the final scores in the theory exam as a function of the number of attempts in the self-evaluation tool. For clarity, we have grouped students with similar numbers of attempts. In the figure, we also represented the standard deviation. As we can see, the average scores only increase slightly when the number of attempts increases. However, the standard deviation reduces significantly. This effect is also shown in Fig. 5, where we have compared the average grades and standard deviation with the number of different days that students used the platform. Both figures show a slight increase in the score, but a considerable decrease in standard deviation. That is, trying more times (or days) does not guarantee that a student's average grade will go up, but the grade is less likely to deviate from the grade average.
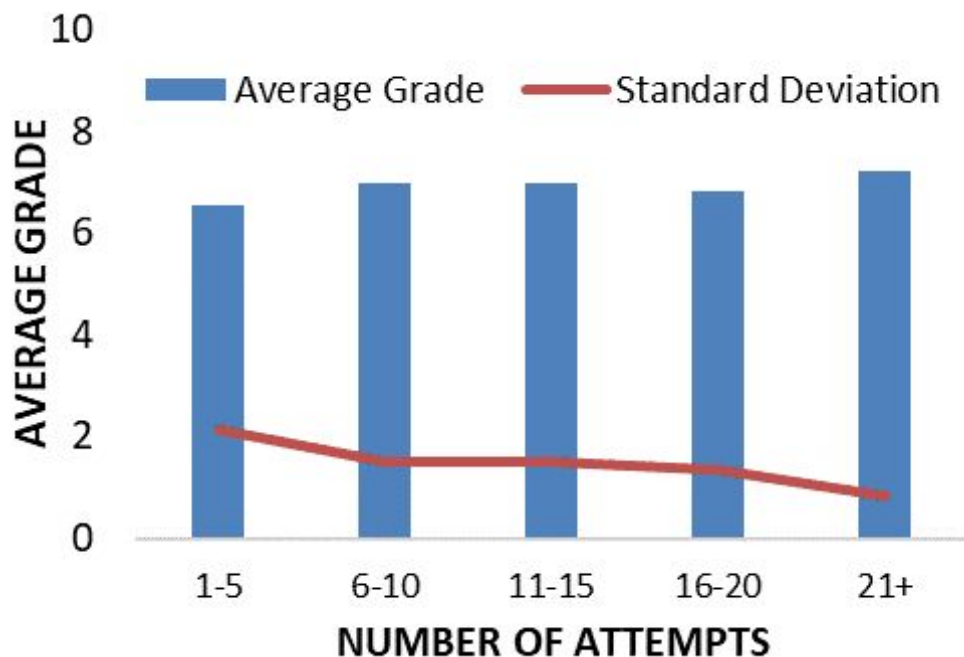
Figure 4: Average grade obtained on e-valUAM summarized by the number of attempts on the platform. The line represents standard deviation of the grades from the mean. Degree:

Funded by the
Erasmus+ Programme
of the European Union

Chemistry, Subject: Biochemistry, Exam: Practice Theory Test on e-valUAM, Academic Year: 2017/18.
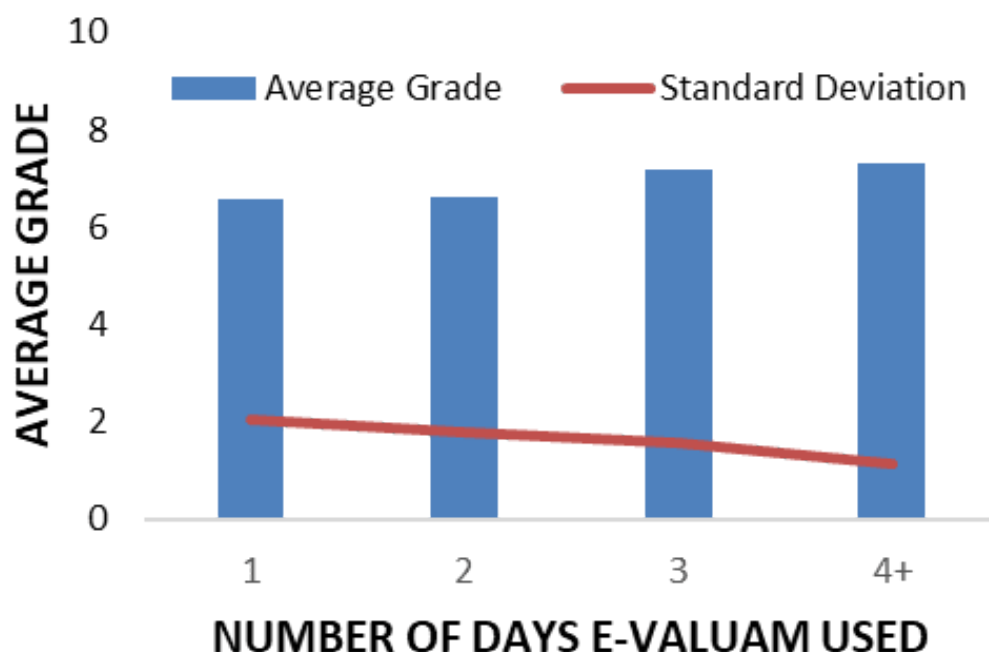
Figure 5: Bars represent the average grade obtained on a progress test on e-valUAM platform. The x-axis shows the students grouped by the number of days the students used the application during the learning phase. The line represents standard deviation of the grades from the mean. Degree: Chemistry, Subject: Biochemistry, Exam: Practice Theory Test on e-valUAM, Academic Year: 2017/18.

In Fig. 6 we can see 2017/2018 student grades split into quartiles. Marked in green, are the average grades of the 14 questions of a test of 20 where the students were able to use e-valUAM and marked in red are the average grades of the remaining 6 of 20 questions related to contents not included in the self-evaluation tool. The objective of this study was to compare the results obtained by the same students, in the same subject and in the same year. For clarity, results have been also divided in quartiles in the figure inset. As we can see, the grades are clearly higher in the questions that were related to contents included in the self-evaluation tool.
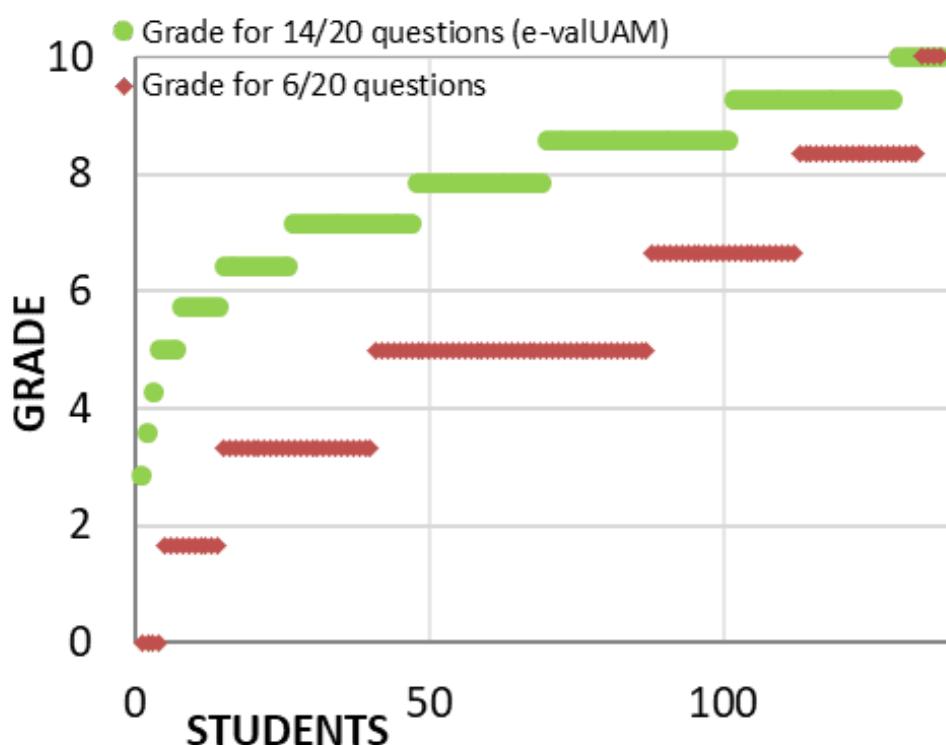
Figure 6: Grades from a 20-question test split into 14 questions (material practiced on e-valUAM) and 6 questions where the contents were not included on e-valUAM. Inset: comparison of average grades in quartiles. Degree: Chemistry, Subject: Biochemistry, Exam: Final Theoretical, Academic Year: 2017/18.

Figs. 7 and 8 are the grades of students, ordered from lowest to highest, including results from different consecutive years. In Fig. 7 we show results of students from the practical lessons in the degree of Biology. On the other hand, Fig. 8 shows results of the practical lessons from the degree of Chemistry. As we can see in both figures, scores are in general higher in cases where students were able to use the self-evaluation tool. There is only one case in Fig. 7 where a group of students (year 2015/16) got scores comparable to the ones obtained when using the tool.

ADeAPTIVE
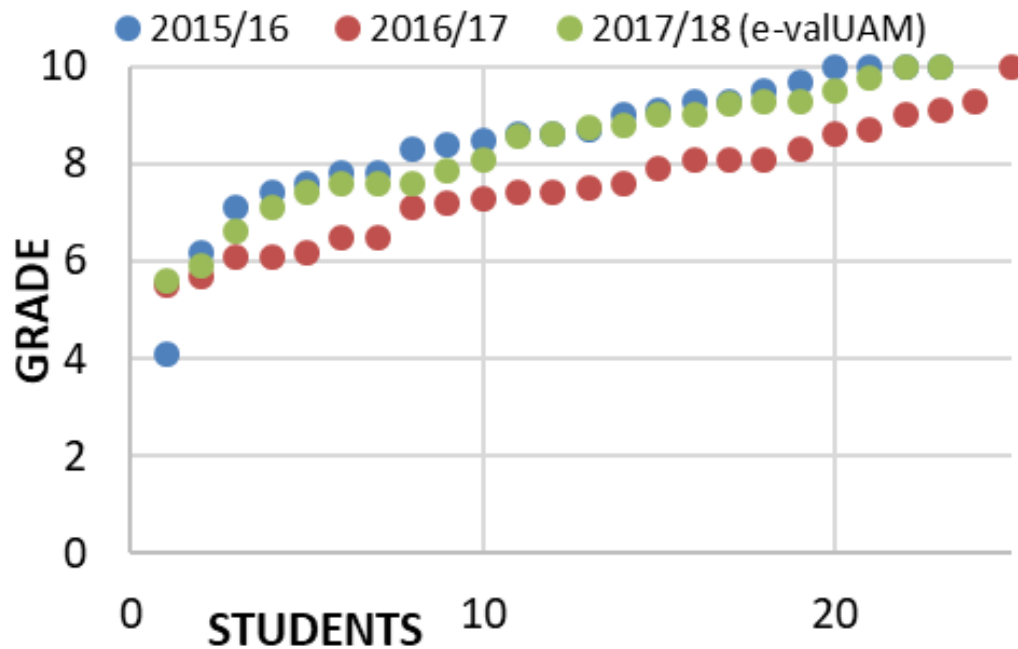
Figure 7: Comparison of student grades, ordered lowest to highest, in 3 consecutive years. Groups in the first 2 years did not use e-valUAM. Inset: comparison of mean grade in quartiles. Degree: Biology, Subject: Biochemistry, Exam: Final, Academic Years: 2015/16, 2016/17 and 2017/18.
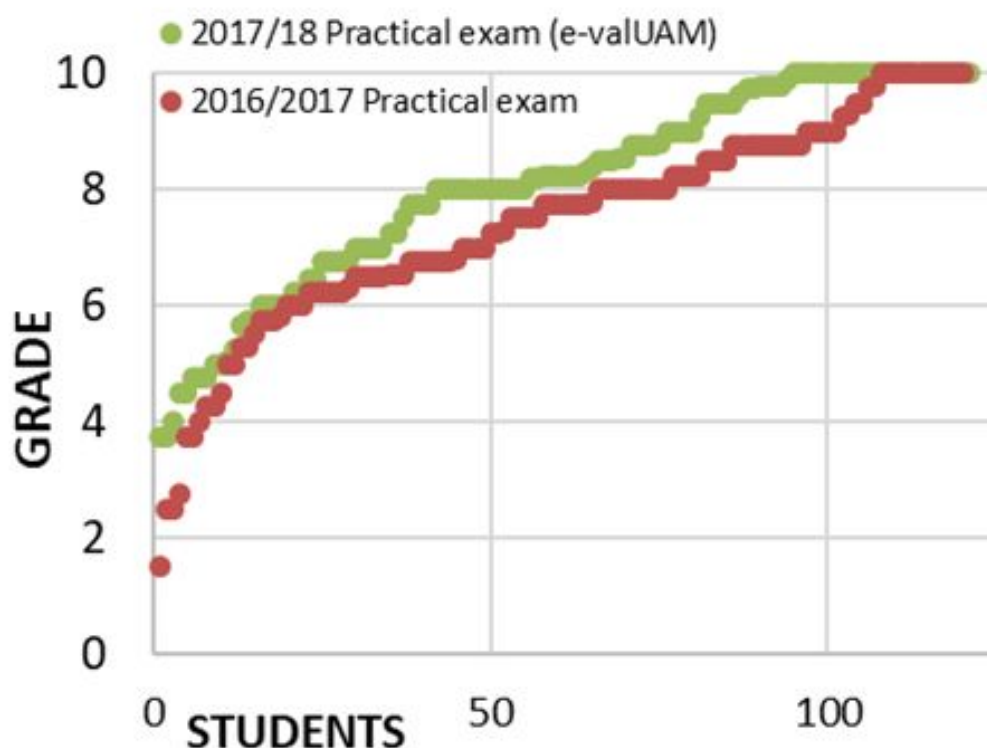
Figure 8: Comparison of student grades in the final practical test, ordered lowest to highest, in 2 consecutive years. Group of 2016/17 did not use e-valUAM and the group in 2017/18 did. Inset: comparison of mean grade in quartiles. Degree: Chemistry, Subject: Biochemistry, Exam: Final Practical, Academic Years: 2016/17 and 2017/18.

4. Conclusions

A system for learning support in Biochemistry subjects based on adaptive tests has been developed. We have proved that it is a useful tool for improving student performance. In general, we can conclude that using the system in a continuous way (extended in time) gives better results. Experience with adaptive testing shows that the students who use the system are more likely to have their grades deviate less from the mean grade. However, repeated use of the application for short periods of time does not guarantee their grade will increase. We

conclude that students risked more uncertainty in their final grades if they made fewer attempts through the online platform.

To test the advantages of using the platform, we have also compared scores of different contents in the same academic year (i.e. done by the same students) and same contents in different academic years (i.e. done by different students). In most of the cases, scores were higher when contents could be also learned on the platform.

Taken into account the results of the experiment with real students, we can conclude that the method is valid in general for subjects that include theoretical and practical contents since we tested it in different subjects with that format. However, the method still requires more analysis in subjects that use, for example, numerical contents. In those cases, there is an additional risk related to potential memorization of numerical answers by the students, making the method useless. An Open Answer format for the items could be a potential future line to extend the results of this document.

REFERENCES

[1] P. Molins-Ruano, C. Sevilla, S. Santini, P. A. Haya, P. Rodríguez and G. M. Sacha. 2014. Designing video games to improve students' motivation. *Computers in Human Behavior*, 31, pp. 571.

[2] A. McAuley, B. Stewart, D. Cormier, and G. Siemens. 2010. The MOOC model for digital practice, SSHRC Application, *Knowledge Synthesis for the Digital Economy*.

[3] A. Kumar. 2004. Using online tutors for learning -what do students think? Proceedings of Frontiers in Education Conference (FIE2004), *IEEE*, 52-28.

[4] C. Bravo, WR. van Joolingen and T. de Jong. 2009. Using co-lab to build system dynamics models: students' actions and on-line tutorial advice. *Computer and Education*, 243–51.

[5] A. Sai. Sabitha, D. Mehrotra and A. Bansal. 2017. An ensemble approach in converging contents of LMS and KMS, *Education and information technologies*, 22, 1673-1694.

[6] F. Jurado, M. Redondo and M. Ortega. 2012. eLearning standards and automatic assessment in a distributed Eclipse based environment for learning Computer Programming, *Computer Applications in Engineering Education*. DOI:10.1002/cae.21569.

[7] JJ. Castro-Schez, M.A. Redondo and F. Jurado. 2014. Experience applying

language processing techniques to develop educational software that allow active learning methodologies by advising students, *Journal of Network and Computer Applications*, 41(1), 65-79.

[8] H. Wainer, N.J. Dorans, D. Eignor, R. Flaugher, R. Green, R.J. Mislevy, L. Steinberg and D. Thissen. 2000. *Computer-Adaptive Testing: A Primer*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

[9] S. Narciss, S. Sosnovsky, L. Schnaubert, E. Andres, A. Eichelmann, G. Goguadze and E. Melis. 2014. Exploring feedback and student characteristics relevant for personalizing feedback strategies, *Computer and Education*, 71, 56-76. DOI: 10.1016/j.compedu.2013.09.011

[10] C.A. Chapelle and D. Douglas. 2006. *Assessing Language Through Computer Technology*, Cambridge University Press.

[11] A. Ortigosa, P. Paredes and P. Rodriguez.2010. AH-questionnaire: An

adaptive hierarchical questionnaire for learning styles, *Computers & Education*, 999-1005.

[12] S. Klinkenberg, M. Straatemeier and H.L.J. Van der Maas.2011. Computer

Adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation, *Computers & Education*, 57 (2), (Sept.2011), 1813-1824.

[13] P. Molins-Ruano, S. Atrio, P. Rodríguez and S. Gomez-Moñivas 2016. Modelling experts' behavior with e-valUAM to measure computer science skills, *Computers in Human Behavior*, 61, 378-385.

[14] AJ. Martin and G. Lazendic 2018. Computer-Adaptive Testing: Implications for Students' Achievement, Motivation, Engagement, and Subjective Test Experience, *Journal of Education psychology*, 110, 27-45.

[15] CM. Chen and LJ. Duh. 2008. Personalized web-based tutoring system based on fuzzy item response theory, *Expert systems with applications*, 34, 2298-23.

[16] P. Molins-Ruano, C. González-Sacristán, F. Díez, P. Rodríguez and G. M. Sacha. 2015. An Adaptive Model for Computer-Assisted Assessment in Programming Skills, *International Journal of Engineering Education*, 31, 764.

[17] R. Conejo, E. Guzmán, E. Millán, M. Trella, J.L. Pérez-De-La-Cruz and A. Ríos. 2004. SIETTE: A web-based tool for adaptive testing, *International Journal of Artificial Intelligence in Education*, 14, 1, 29-61.

[18] M. Lilley, A. Pyper, and P. Wernick. 2011. Attitudes to and Usage of CAT in Assessment in Higher Education, *Innovation in Teaching and Learning in Information and Computer Sciences*, 10, 3, 28-37.

[19] M. Antal, and S. Koncz. 2011. Student modeling for a web-based self-assessment system. *Expert Systems with Applications*, 38, (6), 6492-6497.

[20] T. Lehmann, I. Hahnlein and D. Ifenthaler. 2014. Cognitive, metacognitive and motivational perspectives on reflection in self-regulated online learning, *Computers in human behavior*, 32, 313-323.

[21] F.A, Dorça, L. V. Lima, M.A Fernandes and C.R. Lopes. 2013. Automatic student modeling in adaptive educational systems through probabilistic learning style combinations: a qualitative comparison between two innovative stochastic approaches, *Journal of the Brazilian Computer Society*, 19 (1), 43-58.

[22] AJ. Martin and G. Lazendic. 2018. Computer-Adaptive Testing: Implications for Students' Achievement, Motivation, Engagement, and Subjective Test Experience, *Journal of Education psychology*, 110, 27-45.

Funded by the
Erasmus+ Programme
of the European Union

[23] M. McGlohe and HH. Chang 2008. Combining computer adaptive testing technology with cognitively diagnostic assessment, *Behavior Research Methods*, 40, 808-821.

[24] Available in e-valuam.ii.uam.es. https://e-valuam.ii.uam

ADeAPTIVE

A6: Recommendations for serious game-developers based on self-assesment and learning diaries

Author: JYU

SUMMARY

- This activity highlights some key recommendations pertaining to gamification, based on the findings of the pilot studies conducted in the project.
- As described elsewhere, learning diaries were not applied in the project to the extent that was foreseen before the project.
- Therefore, the principal data sources behind the recommendations include the gamification user type studies and open-ended answers from pilot surveys.

1. Take the user type profile of the target group into account.

User type data can help us develop adaptive gamification: with the aid of user profiles, it is possible to recommend a tailored version to meet the needs of the target group. The user type profile helps the developers put emphasis on certain game elements.

Example: Based on our pilots, collaborative opportunities are a priority when developing gamified course platforms intended for teacher students.

Recommendation based on:

IO1 Report

Nousiainen, T., Vesisenaho, M., Ahlström, E., Peltonen, M., Fort, S., & Gómez, S. (2020). Gamifying Teacher Students' Learning Platform: Information and Communication Technology in Teacher Education Courses. In F. J. García-Peñalvo & A. García-Holgado (Eds.), TEEM'20: Proceedings of the Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 688-693). ACM.

2. In addition to game elements, pay attention to task types and contents.

Game elements are not the only factor to consider when planning potential solutions for meeting the users' preferences and tapping into their motivational orientations. In many cases, task types and contents might play a more significant role than actual game elements in shaping the students' experience and in responding to the needs and preferences of a particular user type.

Example: In our pilots with teacher students, especially different creatively oriented tasks supported the motivation of the participants.

Recommendation based on:

IO1 Report

Nousiainen, T., Vesisenaho, M., Ahlström, E., & Peltonen, M. (2020). Moodlea pelillistämässä: kokemuksia opettajaopiskelijoiden tieto- ja viestintäteknologian kurssilta. [Gamifying Moodle: Experiences from an ICT course for teacher students.] Yliopistopedagogiikka, 27(1).

Nousiainen, T., Vesisenaho, M., Ahlström, E., Peltonen, M., Fort, S., & Gómez, S. (2020). Gamifying Teacher Students' Learning Platform: Information and Communication Technology in Teacher Education Courses. In F. J. García-Peñalvo & A. García-Holgado (Eds.), TEEM'20: Proceedings of the Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 688-693). ACM.

3. Gamification of Moodle-based courses should be made easier for teachers.

Gamification of Moodle courses can still be quite burdensome for teachers who are not technically oriented, and organisation-specific restrictions can hinder the use of some useful plugins, limiting the possibilities for developing dynamic game-based elements. Some simple, small-scale game elements should be readily available for teachers building their courses on Moodle (or on other online learning platforms). For example, the options for visualising progress could include more game-like options, such as avatar development, in addition to basic progress bars.

Example: Avatar development was experienced as motivating by the students participating in the ADeAPTIVE pilots. However, the teacher of the course needed some advanced skills in using Moodle in order to build this feature.

Recommendation based on:

IO1 Report

Nousiainen, T., Ahlström, E., & Vesisenaho, M. (2020). School of Digital Wizards: Gamifying teacher students' course platform. Presentation in the ADeAPTIVE Multiplier Event "Gamification Techniques for eLearning Applications" (online), November 19, 2020.

Nousiainen, T., Vesisenaho, M., Ahlström, E., & Peltonen, M. (2020). Moodlea pelillistämässä: kokemuksia opettajaopiskelijoiden tieto- ja viestintäteknologian kurssilta. [Gamifying Moodle: Experiences from an ICT course for teacher students.] Yliopistopedagogiikka, 27(1).

4. Develop collaborative and immersive learning environments with the aid of Extended Reality.

For user types whose motivational orientations are primarily collaborative and social, the use of basic game elements on traditional learning platforms may not be an ideal approach to support their engagement. Therefore, we suggest looking into the affordances provided by Extended Reality (XR). In particular, social interaction and real-time collaboration are facilitated by immersive Virtual Reality environments. Such tools also allow the construction of various different game-based tasks such as virtual escape rooms that are based on collaborative problem solving.

Example: Virtual Reality based solutions would meet many of the preferences and orientations that were important for the participants of our pilots: collaboration, social interaction, exploration, immersion, and creativity.

Recommendation based on:

IO1 Report

Funded by the
Erasmus+ Programme
of the European Union

Nousiainen, T., Ahlström, E., & Vesisenaho, M. (2020). School of Digital Wizards: Gamifying teacher students' course platform. Presentation in the ADeAPTIVE Multiplier Event "Gamification Techniques for eLearning Applications" (online), November 19, 2020.

A7, A9: report and final versions of self-evaluation of students about their level on each competency.

Author: UAM, EURECAT

SUMMARY

● In this IO we suffered a lot the effects of the COVID-19 pandemic because of the following reasons:

○ Learning diaries were programmed to be extensively used in the last year, being the previous ones used for preparation of the platforms and methodologies. With the pandemic, the teaching format was dramatically changed and we could not act in the same way as expected.

○ This IO includes a lot of reports and recommendations that should be done mainly at the end of the project, with a good data set and a complete recompilation of information. The pandemic interrupted many processes in progress that, if finished, would have given extremely useful information for finishing these tasks as expected.

● Partial results related to A7 and A9 topics are included in other tasks and IOs.

● Instead of finishing these tasks as initially proposed, we decided to adapt the contents to the pandemic situation. For that reason, we are including here a full study about self-evaluation of students in confinement that has achieved an extremely high impact, appearing even in some media such as Radio, Newspapers or TV. By far, this work has been much bigger than the original tasks.

● This work also demonstrates the very nice work done in ADeAPTIVE in the previous years because Data analysis developed here has been only possible thanks to the adequate development of all the previous tasks and tools of the project.

1. INTRODUCTION

This study explores the effects of COVID-19 confinement in the students' performance in higher education. Using a field experiment of 458 students from three different subjects in Universidad Autónoma de Madrid (Spain), we study the differences in assessments by dividing students into two groups. The first group (control) corresponds to academic years 2017/2018 and 2018/2019. The second group (experimental) corresponds to students from 2019/2020, which is the group of students that interrupted their face-to-face activities because of the confinement. The results show that there is a significant positive effect of the COVID-19 confinement on students' performance. This effect is also significant in activities that did not change their format when performed after the confinement. We find that this effect is significant both in subjects that increased the number of assessment activities and subjects that did not change the workload of students. Additionally, an analysis of students' learning strategies before confinement shows that students did not study on a continuous basis. Based on these results, we conclude that COVID-19 confinement changed students' learning strategies to a more continuous habit, improving their efficiency. For these reasons, better scores in students' assessment are expected due to COVID-19 confinement that can be explained by an improvement in their learning performance.

The coronavirus COVID-19 outbreak disrupted life around the globe in 2020. As in any other sector, the COVID-19 pandemic affected education in many ways. Governments decisions have followed the common goal of reducing the spread of coronavirus by reducing social contact, with a direct effect in many countries avoiding face-to-face teaching and/or exams and restrictions on immigration affecting Erasmus students [1]. Where possible, classes are being held via books and materials taken from school, through various e-learning platforms enabling interaction between teachers and students, and, in some cases, with the help of national television shows or social media platforms. Some education systems announced exceptional holidays to better prepare for this distance-learning scenario.

© Advanced Design of e-Learning Applications
Personalizing Teaching to Improve virtual Education

Funded by the
Erasmus+ Programme
of the European Union

41

The start and end dates of the academic year, as well as holidays, are different in each country, therefore the situation was not homogeneous. While some countries suspended in-person classes from March/April until further notice, others were less restrictive, and universities were only asked (without imposing by the government) to reduce face-to-face teaching and replace it with online solutions as much as possible. Depending on the academic calendar, in other cases it was possible to postpone the start of summer semester. [2].

Assessment is probably the most challenging concept to be adapted to this new educational system, and many universities are concerned about how to manage a correct evaluation of students' skills and knowledge. The consequences of university closure potentially extending to the end of the academic year arises questions concerning grading and valuation of progress that rapidly became a significant policy challenge. An additional problem is how to determine if students in their final school year can study adequately or be assessed fairly in terms of accessing higher education programs.

Fortunately, there are many tools available to face the challenge of distant learning imposed by the COVID-19 pandemic [3]. For these reasons, the modification of contents that were previously taught face-to-face is easily conceivable. Unfortunately, there are other important tasks in the learning process, such as assessment or autonomous learning, that can be still challenging without the direct supervision of teachers.

Related to the assessment process, online evaluation has become one of the most concerning concepts in the COVID-19 pandemic because of two reasons. First, teachers should redesign their on-site evaluation in order to meet distance evaluation requirements. Second, it is unclear how to ensure that students follow the instructions and do not use not-allowed additional material in their evaluation tests without the direct supervision of teachers.

All these arguments end in a common topic: How to ensure the assessment's adequacy and correctly measure students' knowledge. Thus, how can teachers read students' results if they differ from previous years? Namely, if students achieve higher scores than previous years, we could think that can be linked with cheats in their online exams or with changes in the common format of the evaluation tools. On the contrary, students can obtain lower grades

which, again, could be related to the evaluation format or to a less effective autonomous learning due to the change in teaching method.

The objective of this document is to reduce the uncertainty in the assessment process in higher education during the COVID-19 pandemic. To achieve this goal, we analyze students' learning strategies before and after confinement. Altogether, our data indicates that autonomous learning in this scenario has increased students' performance and higher scores should be expected. We also discuss the reasons underneath this effect. We present a study that involves more than 450 students enrolled in 3 subjects from different degrees from the Universidad Autónoma de Madrid (Spain) during three academic years, including data obtained in the 2019/2020 academic year, where the restrictions due to the COVID-19 pandemic has been applied.

2. BACKGROUND

E-learning has experienced significant change due to the exponential growth of the internet and information technology [3]. New e-learning platforms are being developed for tutors to facilitate assessments and for learners to participate in lectures [4-5]. Both assessment processes and self-evaluation have been proven to benefit from technological advancement. Even courses that include all the contents online such as Massive Open Online Courses (MOOCs) [6-7] have also become popular. The inclusion of e-Learning tools in higher education implies that a greater amount of information can be analyzed, improving teaching quality [8-9-10]. In recent years, many studies have been performed analyzing the advantages and challenges of massive data analysis in higher education [11]. For example, a study of Gasevic et al [12] indicates that time management tactics had significant correlations with academic performance. Jovanovic et al also demonstrated that assisting students in their management of learning resources is critical for a correct management of their learning strategies in terms of regularity [13].

Related to autonomous learning, many studies have been performed regarding the concept of self-regulated learning (SRL), in which students are active and responsible for their own learning process [14-15] as well as being knowledgeable, self-aware and able to select their own approach to learning [16-17]. Some studies indicated that SRL significantly affected students' academic achievement and learning performance [18-19-20]. Researchers

indicated that students with strongly developed SRL skills were more likely to be successful both in classrooms [21] and online learning [22]]. These studies and the development of adequate tools for evaluation and self-evaluation of learners have become especially necessary in the COVID-19 pandemic in order to guarantee good performance in e-learning environments [23].

Linear tests, which require all students to take the same assessment in terms of the number and order of items during a test session, are among the most common tools used in computer-based testing. Computer adaptive test (CAT), based on item response theory, was formally proposed by Lord in 1980 [24-25-26] to overcome the shortfalls of the linear test. CAT allows dynamic changes for each test item based on previous answers of the student [27]. More advanced CAT platforms use personalization to individual learners' characteristics by adapting questions and providing tailored feedback [28]. Research contains numerous examples of assessment tools that can guide students [29-30-31] and many advances have been also developed in the theoretical background of CAT [32] . In this aspect, advantages offered by CAT go beyond simply providing a snapshot score [33], as is the case with linear testing. Some platforms couple the advantages of CAT-specific feedback with multistage adaptive testing [34]. The use of CAT is also increasingly being promoted in clinical practice to improve patient quality of life. Over decades, different systems and approaches based on CAT have been used in the educational space to enhance the learning process [35-36]. Considering the usage of CAT as a learning tool, establishing the knowledge of the learner is crucial for personalizing subsequent question difficulty. CAT does have some negative aspects such as continued test item exposure, which allows learners to memorize the test answers and share them with their peers [37-38]. As a solution to limit test item exposure, a large question bank has been suggested. This solution is infeasible in most cases, since the majority of the CAT models already require more items than comparable linear testing [39].

3. PURPOSE

The aim of this study is to identify the effect of COVID-19 confinement on students' performance. This main objective leads to the first hypothesis of this study which can be formulated as H1: COVID-19 confinement has a significant effect in students' performance. The confirmation of this hypothesis should be done discarding any potential side effects such as students cheating in their assessment process related to the distant learning.

Moreover, a further analysis should be done to investigate what factors of COVID-19 confinement are responsible for the change.

A second hypothesis is H2: COVID-19 confinement has a significant effect in the assessment process. The aim of the project was therefore to investigate the following questions:

1.   Is there any effect (positive or negative) of COVID-19 confinement in students' performance?

2.   Is it possible to be sure that COVID-19 confinement is the origin of the different performance (if any)?

3.   What are the reasons for the differences (if any) in students' performance?

4.   What are the expected effects of the differences in students' performance (if any) in the assessment process?

4. MATERIALS AND METHODS

4.1 Measurement Instruments

We have used two online platforms. The first one is e-valUAM [40], an online platform that aims to increase the quality of tests by improving the objectivity, robustness, security and relevance of content. e-valUAM implements all the CAT tests described in the following sections. The second online platform used in this study is the Moodle platform provided by the Biochemistry Department from Universidad Autónoma de Madrid, where all the tests that do not use adaptive questions are implemented. Adaptive tests have been used in the subjects "Applied Computing" and "Design of Water Treatment Facilities". Traditional tests have been used in the subject "Metabolism".

4.1.1 CAT theoretical model

Let us consider a test composed by N_0 items. In the most general form, the normalized grade S_j obtained by a student in the j-attempt will be a function of the weights of all the questions α and the normalized scores ψ ($S_j=S_j(\alpha, \psi)$), and can be defined as:

$$S_j\left(\vec{\alpha}, \vec{\varphi}\right) = \left(\sum_{i=1}^{N_Q} \alpha_i\, \varphi_i\right) \tag{1}$$

where the ψ_i is defined as

$$\varphi_i = \delta_{A_i R_i} \tag{2}$$

where δ is the Kronecker delta, A_i the correct answer and R_i the student's answer to the i-question. By using this definition, we limit ψ_i to only two possible values: 1 and 0; $\psi_i = 1$ when the student's answer is correct and $\psi_i = 0$ when the student gives a wrong value. This definition is valid for both open answer and multiple-choice tests. In the case of multiple-choice tests with N_R possible answers, ψ_i can be reduced to consider the random effect. In this case:

$$\varphi_i = \delta_{A_i R_i} - \frac{1 - \delta_{A_i R_i}}{(N_R - 1)} \tag{3}$$

Independently of using equation 2 or 3, to be sure that $S_j(\alpha, \psi)$ is normalized (i.e. $0 < S_j(\alpha, \psi) < 1$), we must impose the following additional condition on α:

$$\sum_{i=1}^{N_Q} \alpha_i = 1 \tag{4}$$

In the context of needing a final grade (FG) between 0 and a certain value M, which typically takes values such as 10 or 100, we just need to rescale the $S_j(\alpha, \psi)$ value obtained in our model by a factor K, i.e. $FG_j = K\, S_j(\alpha, \psi)$.

We will now include the option of having questions with an additional parameter L, which will be related to the level of relevance of the question. L is a number that we will assign to all the questions included in the repository of the test (i.e. the pool of questions from where the questions of a j-test will be selected). The concept of relevance can take different significance depending on the context and the opinion of the teachers. In our model, the questions with lower L values will be shown initially to the students, when the students answer correctly a certain number of questions with the lower L value, the system starts

proposing questions from the next L value. By defining N_L as the number of possible L values, the L value that must be obtained in the k-question of the j-test can be defined as:

$$L_k = trunc\left[\left(\frac{N_L \sum\limits_{i=1}^{k-1} \varphi_i}{N_Q}\right) + 1\right]$$

(5)

where *trunc* means the truncation of the value between brackets. It is worth noting that L_k is proportional to the sum of the student's answers to all the previous questions in the test. This fact means that, in our model, the L_k depends on the full history of answers given by the student. L_k is inversely proportional to N_Q, which means that it takes a higher number of correct answers to increase L_k. Once L_k is defined, a randomly selected question is shown to the student. Another important fact that implies the use of equation 5 in the adaptive test is that we will never have $L_k < L_{k-1}$. In other words, once a student starts answering questions from a certain L value, they will never go back to the previous one. In Fig 1 we show a simple example of the multiple possible paths that a student can follow when facing a test that uses our model. In the example shown in Fig 1 we have used $N_Q = 6$ and $N_L = 3$. In this very straightforward example, the L value changes every time a student gives 2 correct answers (following diagonal arrows on the diagram). Wrong answers imply a vertical drop on the diagram. In this case, the L value will not change ($L_k = L_{k-1}$). The final student's grade G is shown on the lower part of the diagram. The lowest grade (G=0) corresponds to a test where the student has failed all the N_Q questions. In this case, L=1 for all the possible j values. The highest grade (G=6) implies that the student has faced questions from all the possible L values (L={1,2,3} in Fig 1).
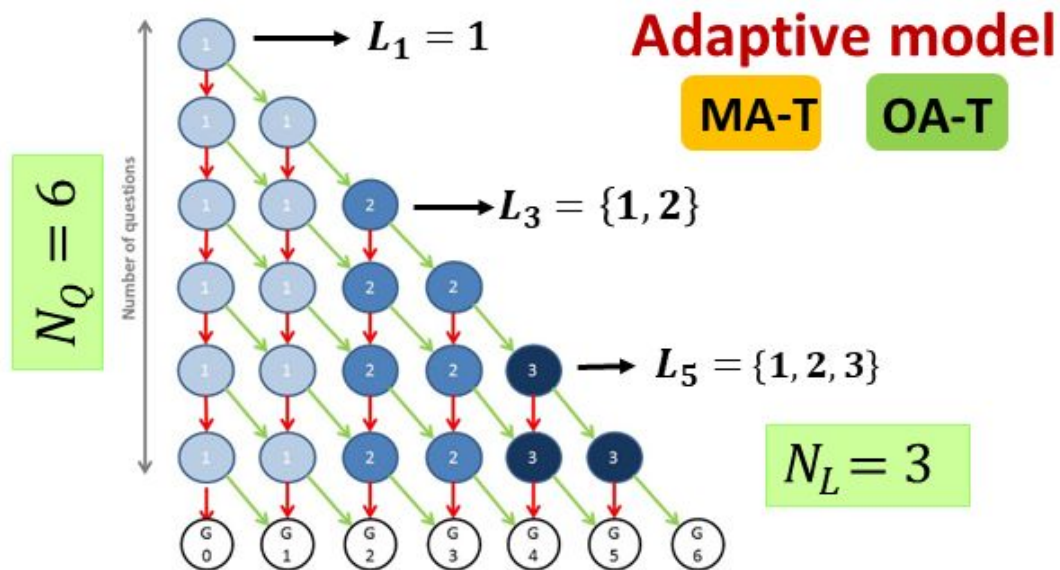
Fig 1. Representation of the CAT model.

Green arrows indicate the path students take when they answer a question correctly. Students follow red arrows when an incorrect answer is given. Grades are represented at the bottom of the figure, improving with each step to the right.

4.1.2 Multiple Answer Test (MA-T)

MA-Ts are used for evaluation and self-evaluation of theoretical contents. In this case, we use a standard format where a single correct answer must be chosen from a pool of possible answers shown by the system. Since students can sometimes answer correctly by a random selection, equation 3 must be used when evaluating the score of any item. The format of the MA-T in our CAT method requires the following elements:

-       Statement

-       Correct answer

-       Rest of possible answers (wrong ones)

-       Level of the item

In addition, the e-valUAM platform, where the method proposed in this document has been implemented, allows optional use of images or sounds for both the statement and the answers. e-valUAM shows the statement and the possible answers. All other information relating to the CAT method such as the level of the item is hidden to the student. The interface also shows optional feedback information about the performance of the learner in the previous item. Finally, it shows the time remaining to finish the test.

4.1.3 Open Answer Test (OA-T)

To train the contents of numerical problems, an OA-T was developed in which the statements include at least one parameter that will change its value with each execution of the application. This kind of question requires the following elements to be created:

- Statement with explicit indication of the modifiable parameter(s).

- Minimum and maximum values of each modifiable parameter.

- Programming code (Matlab in e-valUAM) that calculates the solution to the problem.

- Level of the item.

As with MA-T, there is also a possibility of including multimedia files in the item. However, in this type of question, the multimedia option is only available in the statement as the answer is numerical and included by the user in all cases. In this case the interface has a space for entering the numerical answer. In this figure we also show the feedback of the application when an incorrect previous answer has been introduced.

4.1.4 Traditional tests

Traditional tests have been used in the subject "Metabolism" from the Human Nutrition and Dietetics Degree. The course contents are divided in 6 parts corresponding with different aspects of human metabolism. Each part is taught by face-to-face lectures followed by different on-line activities that the students perform in the Moodle platform and that are later discussed between them and with the professor in face-to-face workshops. The on-line

Funded by the
Erasmus+ Programme
of the European Union

activities in each part of the course performed after the corresponding lecture are, in this order: "Exercises Workshop", "Discussion Workshop", "Self-assessment" and "Test". Except for the Discussion Workshops, discussed below, all activities are on-line questionnaires (multiple-choice and cloze questions) that the students must perform in Moodle, achieving a score automatically calculated and shown that will be part of their continuous assessment. Exercises, Workshops and Self-assessment are done by the students without supervision of the tutor, whereas the Tests are done on-line but under controlled conditions, just in the same way as a regular examination. The students should perform a total of 15 on-line activities: 5 Exercises Workshops, 6 Self-assessment activities and 4 Tests. Discussion Workshops are practical cases that must be discussed by the students and sent to the tutor by Moodle. As these activities do not obtain an automatic score but require a correction by the professor, they have not been considered in this study.

4.2 Design of the experiment

4.2.1 Control group

The control group of our study is formed by the students of "Applied Computing" and "Metabolism" from academic years 2017/2018 and 2018/2019 and by students of "Design of Water Treatment Facilities" from academic year 2017/2018. In the case of "Design of Water Treatment Facilities", a longitudinal study has been performed in the academic year 2017/2018 to analyse the effect of rewards in the students' learning strategies, especially those related to time management.

4.2.2 Experimental group

The experimental group of our study are students of "Applied Computing" and "Metabolism" from the academic year 2019/2020. In the longitudinal study of "Design of Water Treatment Facilities", an experimental group corresponds to the third stage of the study.

4.2.3 Study of autonomous learning strategies in the control group

To answer our research questions, we first set up an experiment that obtained accurate measurements of the autonomous learning activities both in the control and experimental group. The high accuracy in the measurements in the autonomous learning activities is

achieved by using the adaptive tools previously described both in learning and assessment. In our experiment, students know from the beginning the evaluation format (and e-Learning tools), available for their autonomous learning. This fact is extremely motivating for students to use the tool and perform the tests. In Fig 2 we show the procedure used in our experiments to measure autonomous learning. There are two kinds of contents in the subjects included in the experiment: theory and numerical problems. In the case of theory, students can use the e-valUAM platform through MA-T. The final evaluation is performed by OA-T. In the case of numerical problems, both learning and evaluation use OA-T. In this case, the motivation of students for using the platform is even higher because their final scores will be obtained by exactly the same tool they use in their autonomous learning.
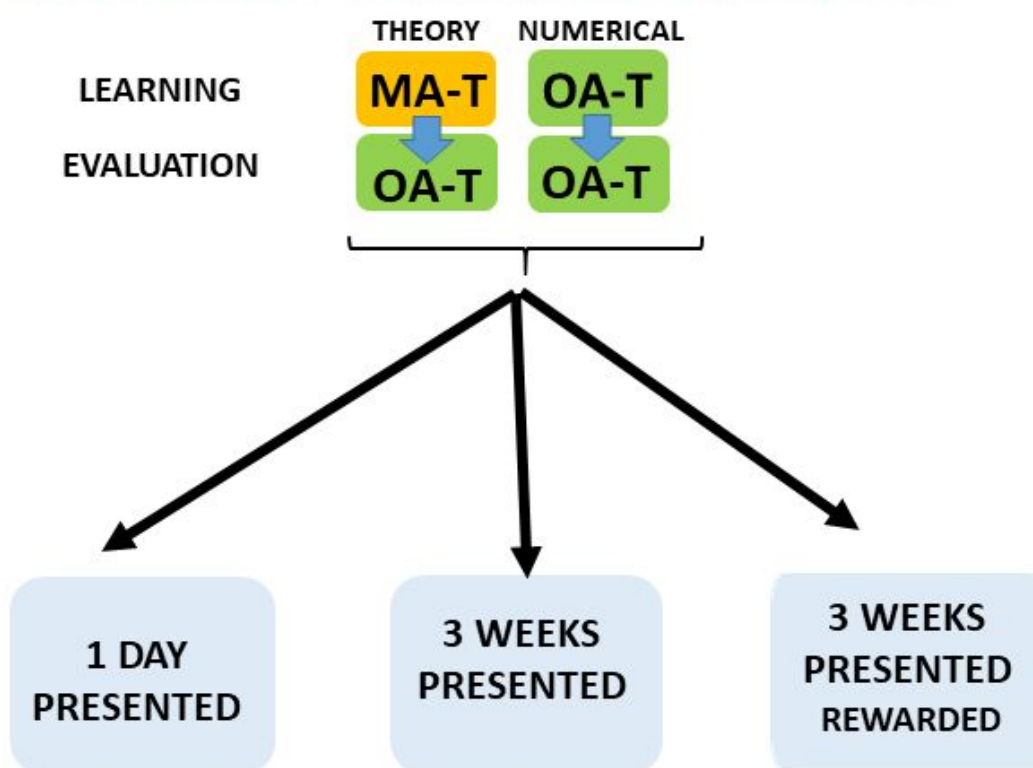
Fig 2. Scheme of the strategy followed to accurately measure the autonomous work of students and the influence of rewards.

This system is applied in all the experiments related to analyse the autonomous learning activity.

In the case of "Design of Water Treatment Facilities", an additional longitudinal study has been performed as is shown in Fig 2. This study has been performed in three stages. In the first stage, the self-evaluation material was presented to the students right before the evaluation test. In the second stage, the self-evaluation material was available 3 weeks before. To manage the theory contents, a MA-T with three possible answers to each question and only one correct, has been created as well. In the first stage, with the test available only the day before the exam, students had at their disposal a MA-T composed of 16 questions and with a time limit of 15 minutes. In the second stage, the time available was increased to 30 minutes and the question number to 18. These changes are related to the nature of the subject contents and do not have a significant impact on the study. The OA-T tests had 6 and 9 questions in the first and second stage, respectively. In order to maintain consistency in the subsequent study, an additional effort has been made to keep the difficulty of the questions as similar as possible.

Stage 3 had no new material added, instead, it used all the material created in stages 1 and 2. Stage 3 corresponds to the time window between the end of classes and the last exam. In this stage, all the material was available from the beginning. The protocol used in this stage only varied from stage 2 in the inclusion of a reward for students who used the application 3 or more different days. The reward was related to a bonus in the final grades.

4.2.4 Effect of COVID-19 confinement

Once the correct measure of the autonomous learning is ensured, as explained in the previous section, we must develop an experiment to describe the effect of COVID-19 confinement. This experiment compares results of control and experimental groups in all the assessment processes. There are two stages that must be considered to determine the effect of the COVID-19 confinement. The first one corresponds to the period without confinement in the course 2019/2020 (before March 11), when all the measurable activities are performed in

Funded by the
Erasmus+ Programme
of the European Union

similar conditions for experimental and control groups. The second stage corresponds to the period of COVID-19 confinement (after March 11), where some measurable activities are performed in a different format and statistical differences can be found by comparing experimental and control groups.
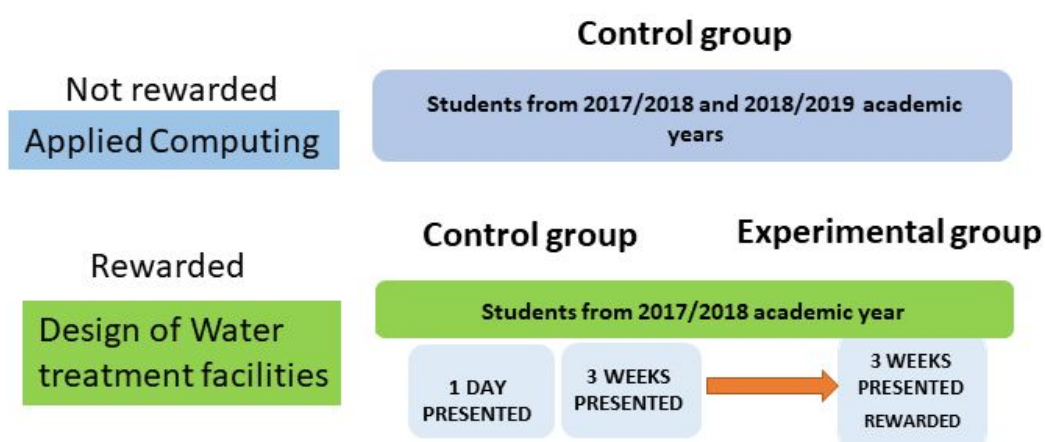
In the case of "Applied Computing", students use the e-valUAM application on a continuous basis as it is available from the beginning of the course. They use the same OA-T for training and for their final exam, therefore the students use the application often from the beginning of the course. For this reason, in "Applied Computing", data analysis is performed over a continuous self-evaluation process over a single test along the whole academic year. The test performed by the students did not change its format and questions in the whole three years under study.

In the case of "Metabolism", the confinement started just before the beginning of Part IV of the subject. Therefore, in this second stage the scheduled in-person lecturers from Part IV where recorded on video by the tutor and provided to the students Afterwards, they had the opportunity to perform the corresponding on-line continuous assessment activities, just in the same way as they would have done in normal conditions (as these activities were always not face-to-face). As this subject already has an important number of on-line, not face-to-face, activities programmed previously to the cancellation of face-to-face teaching, it is an extremely valuable tool to analyse the effect of the confinement on students' performance.

In Fig 3, we show the full experiment described in this section. First, the two studies performed in years prior to COVID-19 confinement. In these first studies, we focus on the learning strategies of the students and their response against reward scenarios. Information from these first studies will be useful for the later discussion about the reasons for the changes due to the COVID-19 confinement. The second part of our experiment is related to the effect of COVID-19 confinement, where we compare results of similar assessment activities between control and experimental group. In Fig 3 we include information about all control and experimental groups in the different interventions.

ADeAPTIVE

Learning strategies previous to COVID-19 confinement

Control group

Not rewarded

Applied Computing

Students from 2017/2018 and 2018/2019 academic years

Control group   **Experimental group**

Rewarded

Design of Water treatment facilities

Students from 2017/2018 academic year

1 DAY PRESENTED   3 WEEKS PRESENTED   →   3 WEEKS PRESENTED REWARDED

Effect of COVID-19 confinement

Control group   **Experimental group**

Additional activities

Applied Computing

Students from 2017/2018 and 2018/2019 academic years → Students from 2019/2020 academic year

Without additional activities

Metabolism

Students from 2017/2018 and 2018/2019 academic years → Students from 2019/2020 academic year

Fig 3. Description of the experiment.

4.3 Statistical analysis

4.3.1 Applied Computing

Funded by the
Erasmus+ Programme
of the European Union

In this subject we have analyzed students' performance expressed as their score (1-10) obtained in all the tests performed in self-evaluation. Those tests (OA-T format) were available from the beginning of the academic year and students used them on a continuous basis. First, we tested if data followed a Gaussian distribution by a D'Agostino and Pearson omnibus normality test [41]   As the data did not pass the normality test, we used non-parametric statistical tests to compare the data from the 3 academic years. Means were compared by a Kruskal-Wallis test [42] and, when differences were found, by a Mann-Whitney post hoc test [43] to determine which pairs of means were different. All the statistical analysis was performed using GraphPad Prism 6 (GraphPad Software, La Jolla, California, USA). Data is presented as mean±SD and statistical significance was set at $p<0.05$.

4.3.2 Metabolism

Here, we have analysed 2 variables, the score (0-10) obtained by the students in the different activities and the proportion of students that pass the activity (score≥5), from 2019/2020 and the 2 previous academic years. Firstly, we checked if the results from the previous 2 years (2017/2018 and 2018/2019) were similar between them, thus allowing us to compare them with the results from present year. To perform the analysis, we tested the score data for normality by a D'Agostino and Pearson omnibus normality test [41] , comparing them later with an unpaired 2-tailed t-test, if they were normal, or with a non-parametric Mann-Whitney test, if they were not normal. Afterwards, we proceeded similarly to compare the results from the present year with those from the 2 previous. When data from each activity followed a normal distribution, we compared the 3 means by a 1-way ANOVA followed by a an unpaired 2-tailed t-test; when data did not follow a normal distribution, we compared the 3 means with a Kruskal-Wallis test followed by a Mann-Whitney post hoc test. In order to compare the proportion of students that pass the activities (score≥5) in the different years, we performed a z-test (after confirming that our data met the central limit theorem). All the statistical analysis was performed using GraphPad Prism 6. Data is presented as mean±SD and statistical significance was set at $p<0.05$.

4.3.3 Design of Water Treatment Facilities

In order to substantiate these related samples of results, we performed a Wilcoxon Signed Rank test to compare two related samples that cannot be assumed to be normally distributed.

In the experiments related to this subject, the same students were involved in the two samples. Data is presented as mean±SD and statistical significance was set at p<0.05.

4.4 Participants

4.4.1 Applied Computing

This study with real students took place during the 2016-2017, 2017-2018 and 2019-2020 academic years in the first-year subject "Applied Computing". 97, 73 and 91 students were involved each academic year, respectively. This subject was taught through theory lessons and practical classes in the computer laboratory. This course corresponds to 6 ECTS and belongs to the Chemical Engineering Degree in the Faculty of Sciences from Universidad Autónoma de Madrid, Spain. Due to COVID-19 pandemic, the face-to-face teaching was cancelled on March 11, having a strong influence in the 2019/2020 course since it started in the first week of February and lasts until mid-May.

4.4.2 Metabolism

The data for this study has been obtained from real students enrolled in the "Metabolism" course from the Human Nutrition and Dietetics Degree from Universidad Autónoma de Madrid, Spain. This is a 6 ECTS compulsory subject taught during the second semester of the first year of the degree. The study comprises data from 2017/2018 and 2018/2019 academic years, with 64 and 63 enrolled students, respectively, and from the present one, 2019/2020 (47 students), which has been strongly affected by the mobility restrictions caused by the COVID-19 pandemic. This year, the course started on January 28 and finished on April 30. Face-to-face teaching cancellation and beginning of confinement occurred when the subject was approximately at its half.

4.4.3 Design of Water Treatment Facilities

The experience with real students took place during the 2017-2018 academic year. 23 students were involved in the fourth-year optative subject "Design of Water Treatment Facilities". This subject was taught through theoretical and practical classes in the classroom and laboratory, respectively. Our study focused on the theoretical part of the subject. This

course corresponds to 6 ECTS and belongs to the Chemical Engineering Degree in the Faculty of Sciences of the Universidad Autónoma de Madrid, Spain.

5. RESULTS

5.1 Autonomous learning strategies in the control group

5.1.1 Continuous working of students in a non-rewarded scenario

In this section, we analyze students' self-learning strategies in the years that were not influenced by COVID-19 (2017/2018 and 2018/2019). However, since data is very similar in both courses, we will focus on 2017/2018 for clarity. All statistical results can be also applied to 2018/2019. In those years, theoretical lessons were taught in the classroom and practical lessons in the computer laboratories with the physical presence of a teacher. Practical lessons always used the e-valUAM platform and the adaptive test described previously.

In Fig 4a, we show the scores obtained by all the students in the course 2017/2018 in all the attempts made with the adaptive tests. Attempts are ordered chronologically. A score of -1 in the figure implies that the student did not finish the test and did not take a numerical score. Those tests have been included in the figure as a part of student's autonomous work and must be considered as time of learning; however, will be excluded from score data analysis. There are some effects in the figure that must be considered. First, there are certain scores more repeated than others (2.5, 5 and 7.5), which correspond to a jump in the level (L) (level 1: 0-2.5, level 2: 2.5-5, level 3: 5-7.5 and, level 4: 7.5-10), meaning that students are not able to answer more difficult questions yet. Another effect that can be easily seen in the figure is a region where a huge number of tests had a score -1 (between 600 and 700 approximately). This effect appears because, in those days, teachers asked students to find the most difficult questions in the test. For that reason, students interrupted the tests when they found those questions.
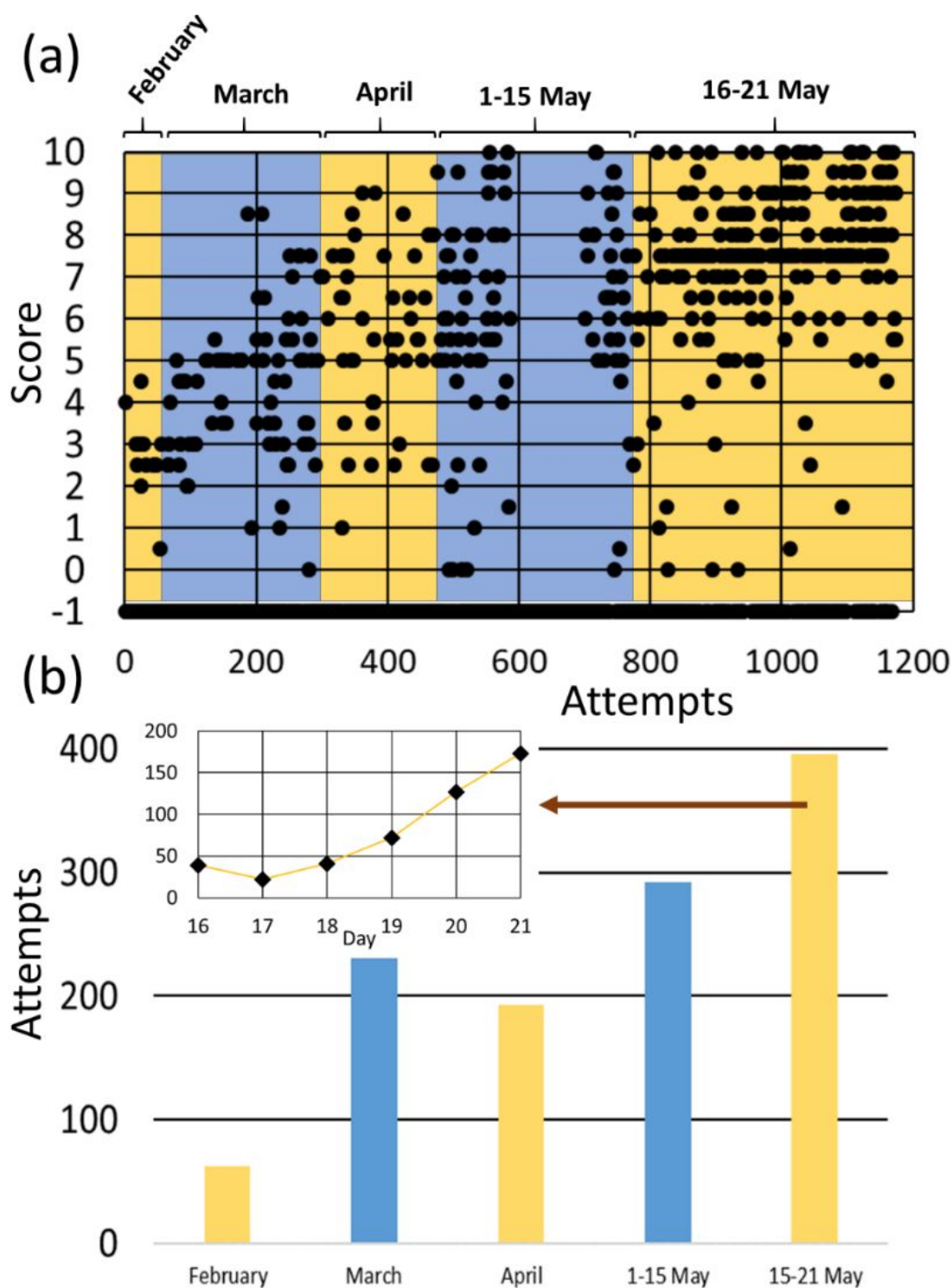
However, the most interesting effect, which can be observed in Fig 4a, is the distribution of the tests in time, with a much bigger number of tests done in May, the last month of the academic year for this subject. Students used the platform 688 times in total in May and 486 times in the period between February and April. It is worth noting that the academic year for this subject ended on May 21 (final exam). In Fig 4b we show the distribution of attempts

between February and May. We found 62 in February, which is a low but reasonable number since students do not have yet enough knowledge to face tests properly. They usually start using it at the middle of the month. In March, there is an increase in the number of attempts followed by a slight decrease in April. This difference can be justified by the Easter holidays (one week) in April. The number of attempts strongly increased in May. Even when counting only the first two weeks of the month, we have more attempts than in previous months (292 from May 1-15). In the period of May 16-21, we have recorded 396 attempts (469 if we count the attempts in the final exam). In the inset of Fig 4b we show the distribution in the last period (May 16-21). As we can see, students used the platform 127 times the day before the exam and 173 times the day of the final exam (95 of studying and 78 in the exam itself). More than 33% of tests were performed the last 6 days before the final exam (and more that 50% of those tests were performed the day before the exam).

ADeAPTIVE

© Advanced Design of e-Learning Applications
Personalizing Teaching to Improve virtual Education

Funded by the
Erasmus+ Programme
of the European Union

59

Fig 4. Distribution of self-evaluation activities in "Applied Computing".

(a) Individual scores of all the students in the academic year 2017/2018 (ordered chronologically). Each dot corresponds to a single attempt of a student. (b) Number of attempts in the academic year 2017/2018 distributed in time periods. Inset shows the distribution in the days between May 16 and May 21.

5.1.2 Continuous working of students in a rewarded scenario

To study the effects of a rewarded scenario, we have developed an experiment in the academic year 2017/2018 for students from the subject "Design of Water Treatment Facilities". We have used 3 stages (fully described in Materials and methods section): stage 1, where the self-evaluation tests have been made available for 1 day; stage 2, where the self-evaluation tests have been made available for 3 weeks; and stage 3, where self-evaluation tests were available for a long time with a reward related to their use. Fig 5 shows the attempted distribution in all the stages. Focusing on the MA-T, 13/20 students have made more attempts in the first stage than in the second, which is striking considering that in the second stage application was available for longer. Three students have not made any attempt in either of the 2 stages.

As for the group of students who have used the application for more than one day, we noted only 3 individuals working for at least 3 days. Considering that the application has been available for 3 weeks in this second stage, these results show the very low willingness of the students to work continuously. Focusing on the OA-T, this effect is even more significant, as only two cases have been recorded in which students have used the application for more than one day. Increasing student motivation and analyzing the results of a more persistent use of the tool is the objective of stage 3.

In Fig 5 we can see the distribution of attempts made by the students in the whole study (3 stages). There is a fourth stage in the figure that corresponds to an additional exam made by the students that did not pass the previous ones. Since only a few of them were involved in this fourth stage, it was excluded from the analysis. As we can see in the figure, attempts from stage 1 were limited to a couple of days, which corresponds to the short period when the test was available. Some attempts can be found in a wider window in stage 2. However, a

much higher number of attempts extended in time can be found in stage 3, which corresponds to the rewarded stage.
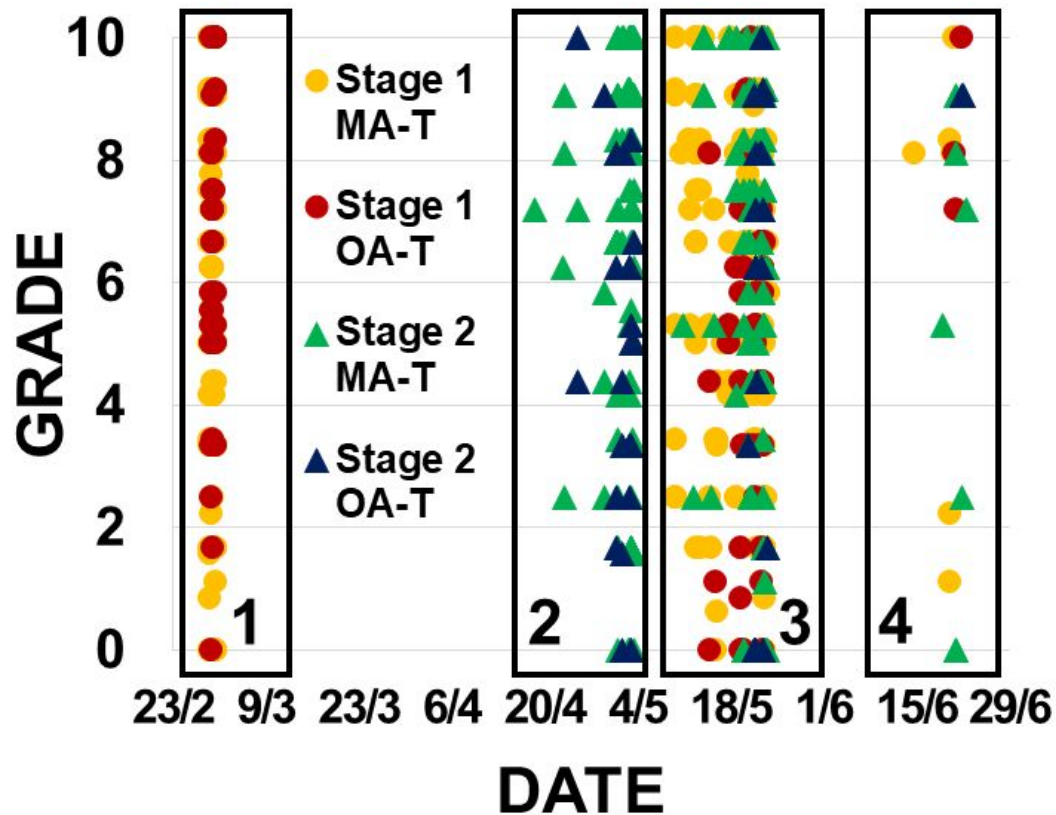

Fig 5. Distribution of attempts in the self-evaluation/learning phase in "Design of Water Treatment Facilities".

Horizontal axis corresponds to the dates and vertical axis to the grades obtained in every individual attempt. The attempts made by all the students in the 3 stages are presented.

In order to measure the use of the tool over time, we summarized the use of the system by students in stage 2 and 3 by comparing the number of days the students attempted the stage 2 MA-T and OA-T tests during the 3 weeks leading up to the exams given after each stage. It is interesting to note that, during stage 3, six more students utilized the system than in stage 2; however, the total attempts reduced between the 2 stages from 123 to 102. This translates to

Funded by the
Erasmus+ Programme
of the European Union

a 23% increase in the average use of the application over time from 1.9 days in stage 2 to 2.4 days in stage 3 combined with a reduction of attempts per student from 9.5 in stage 2, to 5.4 in stage 3. In the earlier stage, fewer students used the application intensively by making multiple attempts a day. Students who logged in at the later stage were more likely to use the application once or twice a day over several days because of the reward related to this behavior. Out of a total sample of 19 students who used the tool in stage 3, 5 were eliminated by the test as they accessed the tests the same number of days in both stages. Then, for the use of application by the remaining 14 students the W-value was of 15.5, with the critical value of W=25 at p.0.05. Therefore, the result is significant.

5.2 Effect of COVID-19 confinement

We have studied two cases related to two different and common scenarios in COVID-19 distant learning. The first one is related to subjects where teachers increased the number of tasks that students should perform autonomously. The second one is related to subjects where all the autonomous tasks are just the same as in previous courses. In this case, only face-to-face sessions have been replaced by distant learning activities such as online sessions or recorded classes.

5.2.1 Additional autonomous activities during confinement

In Fig 6 we show the scores obtained by all the students in the 3 courses analysed in all the e-valUAM tests performed in the subject "Applied Computing". We indicate in the figure the period of the COVID-19 confinement during the 2019/2020 course (from March 11 to March 30). For clarity, we also indicate the equivalent period in the two previous years. As we can see in the figure, the number of tests performed before confinement were similar (225 and 246) in the two previous years (control group), but there is a remarkable increment in the number of tests (328) performed in the experimental group. These differences can be easily explained normalizing these numbers to the number of students each year (97, 73 and 91 respectively). By doing that, we obtain 2.31, 3.37 and 3.6 test/student, respectively. Thus, differences between courses 2018/2019 and 2019/2020 are reduced, although there is still some difference with respect students from course 2017/2018, that can be explained by the fact that some of these students were retaking the subject and, therefore, did not use the e-valUAM application often until the contents were more difficult. If we count the number of

tests finished until the end of the academic year in the courses 2017/2018 and 2018/2019, we find 1345 and 1175 tests respectively. Normalizing with the number of students, we find 13.87 and 16.1 test/student respectively. This is a difference of 2.23 tests/students in a period of 3.5 months, which means a difference of less than 1 test/student each month.

The conditions after the COVID-19 confinement radically changed in the course 2019/2020 because students were required to perform a higher number of tests each week. As we can see in Fig 6, there are still small differences between courses 2017/2018 and 2018/2019. However, the amount of tests performed in 2019/2020 is almost 5 times higher. In the inset of Fig 6 we show a histogram with the data of course 2019/2020. The adaptive tests used in this experiment induce the effect of higher number of attempts at scores 2.5, 5 and 7.5, which are the points where the level of the questions increases. For that reason, the curve is not normal. This fact is statistically tested in the following analysis.

Funded by the
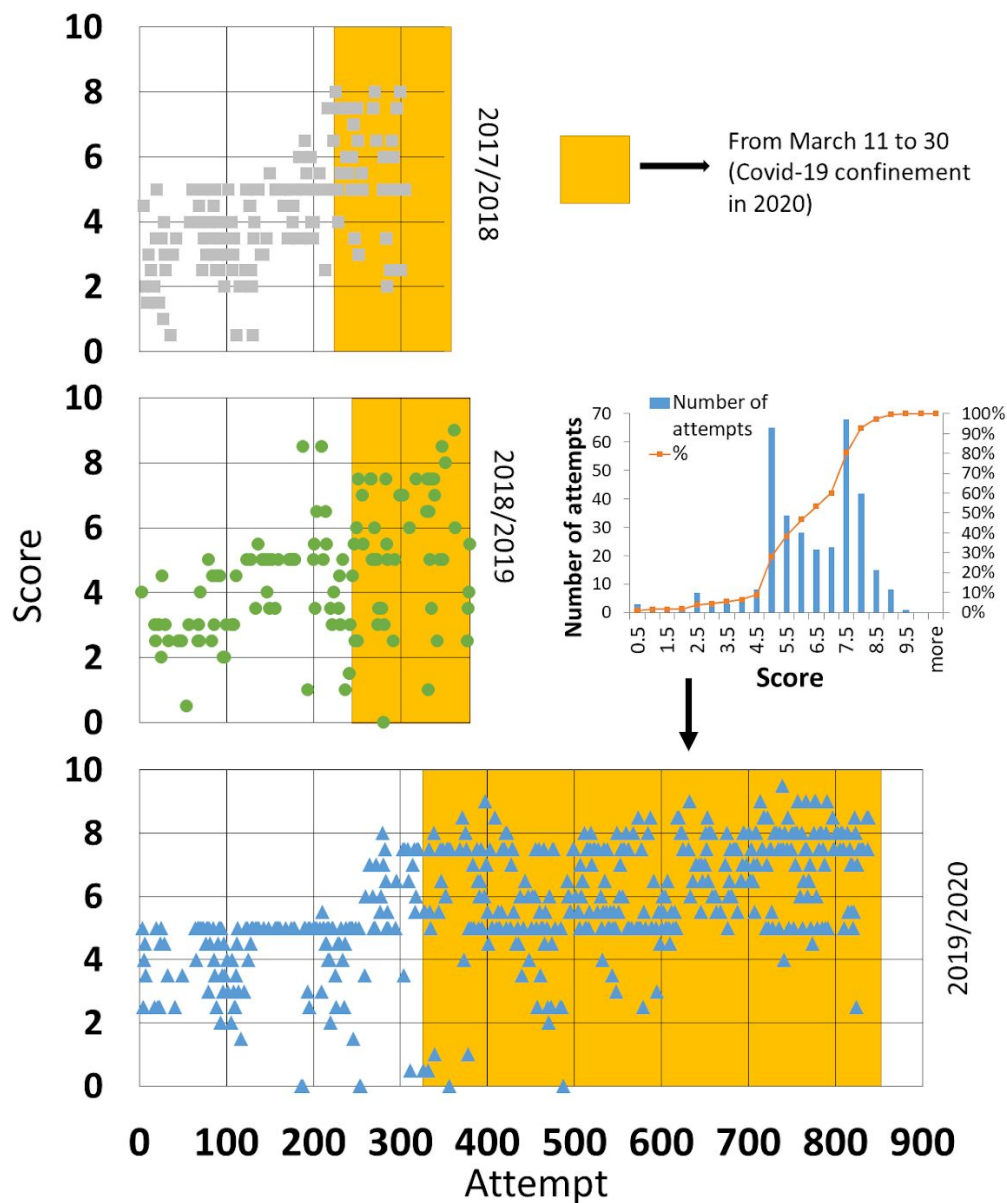Erasmus+ Programme
of the European Union

Fig 6. Attempts vs. Scores ordered chronologically for the three academic years under study in "Applied Computing".

When comparing the mean scores from the period before confinement (Fig 7), we found that they were statistically different between the 3 academic years , differences being between

the mean score from the 2019/2020 (experimental group) and both the other 2 years (control group): 4.5±1.6 in experimental group, vs. 3.9±1.5 in 2017/2018, p=0.0003, and vs. 3.9±1.5 in 2018/2019, p=<0.002), with no differences between 2017/2018 and 2018/2019 (p=0.997). Similarly, in the period between the beginning of confinement and March 30, there were statistically significant differences between the mean of present course 2019/2020 and the previous 2 years (6.3±1.6 in experimental group, vs. 5.6±1.8 in 2017/2018, p=0.0168, and vs. 5.2±2.1 in 2018/2019, p=0.0002), with no differences between 2017/2018 and 2018/2019 (p=0.4451). The mean scores are significantly different both in the control and in the experimental groups between both periods (before and after confinement). We have also seen an increase in the mean score differences after the confinement.
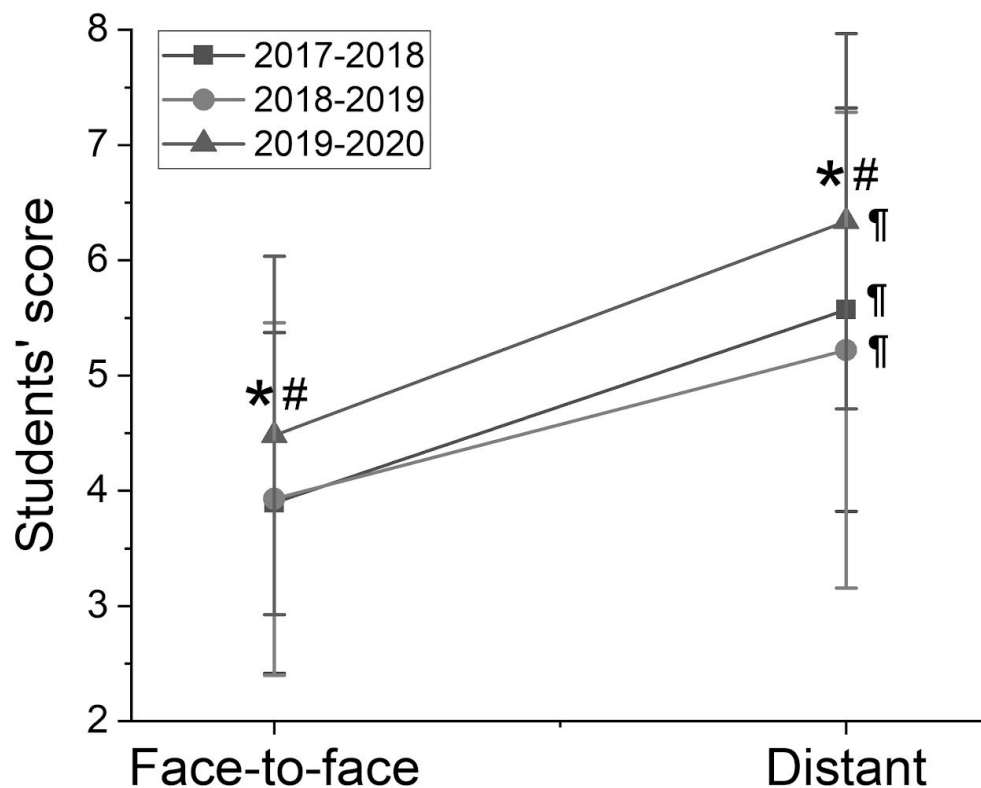

Fig 7. Students' scores in the face-to-face period (before confinement) and in the distant teaching period (during confinement).

Data is presented for the 3 academic courses (2017/2018 and 2018/2019 being the control group and 2019/2020 the experimental group) as mean±SD. Symbols indicate statistically significant differences (p<0.05) between course 2019/2020 and 2017/2018 (*), 2019/2020 and 2018/2019 (#) or between the period before and after the confinement during the same course.

5.2.2 Not additional autonomous activities in confinement

This study has been performed with students from the subject "Metabolism", which is a subject where no additional tasks have been imposed on students because of COVID-19 confinement. We have analysed students' score and proportion of students that pass the activity (score≥5). As it can be observed in Fig 8a, the scores obtained by the students in the different activities along the course were consistently similar in both years of the control group, with the only exception of the second activity where there was a slight decrease in year 2018/2019 (p<0.05). Similarly, the proportion of students with a score≥5 was very similar in the control group (p>0.05) (Fig 8b). Therefore, students' performance in the previous 2 years (control group) were not significantly different, allowing us to compare with them the results from 2019/2020. Scores obtained by the students in the 2019/2020 academic year before confinement were similar to those obtained by students from the control group, although some differences were found at the beginning of the course (activities number 2, 3 and 4, p<0.05). We think that this could be due to the adaptation of the students to the course and the new assessment methodology, being afterwards similar again.

However, after the end of face-to-face teaching and beginning of confinement, students' scores were significantly higher than in the previous academic years (Fig 8a). The most remarkable differences are evident in activities 10 and 11, that were always on-line activities. Thus, in activity 10, mean score in the present year (2019/2020) was 8.1±0.2 vs. 6.5±0.2 (p<0.0001) and vs. 6.7±0.3 (p=0.0005) in 2017/2o18 and 2018/2019, respectively. Similarly, in activity 11, the mean score of the experimental group was 7.8±0.2 vs. 6.8±0.2 (p=0.0014) and vs. 6.1±0.3 (p<0.0001) in 2017-2o18 and 2018/2019, respectively. Better performance of students was strongly evident when analysing the proportion of students that pass these 2 activities (Fig. 8b). Thus, 95.6% of students passed activity 10 (p<0.0047 vs. 2017/2018 and p=0.0025 vs. 2018/2019) and 97.7% activity 11 (p<0.0068 vs. 2017/2018 and p=0.0111 vs. 2018/2019). Moreover, we also found higher scores and an increase in students that pass the

activity number 9, which is a test (Figs 8a and 8b). In the present academic course (2019/2020), this test has been performed by the students confined at their homes, in contrast to previous years where it was performed in the classroom under the supervision of the tutor. In addition, we have analysed the proportion of students that perform the activities in the different courses (Fig 8c), finding that this proportion is very similar, with only a difference in one activity.
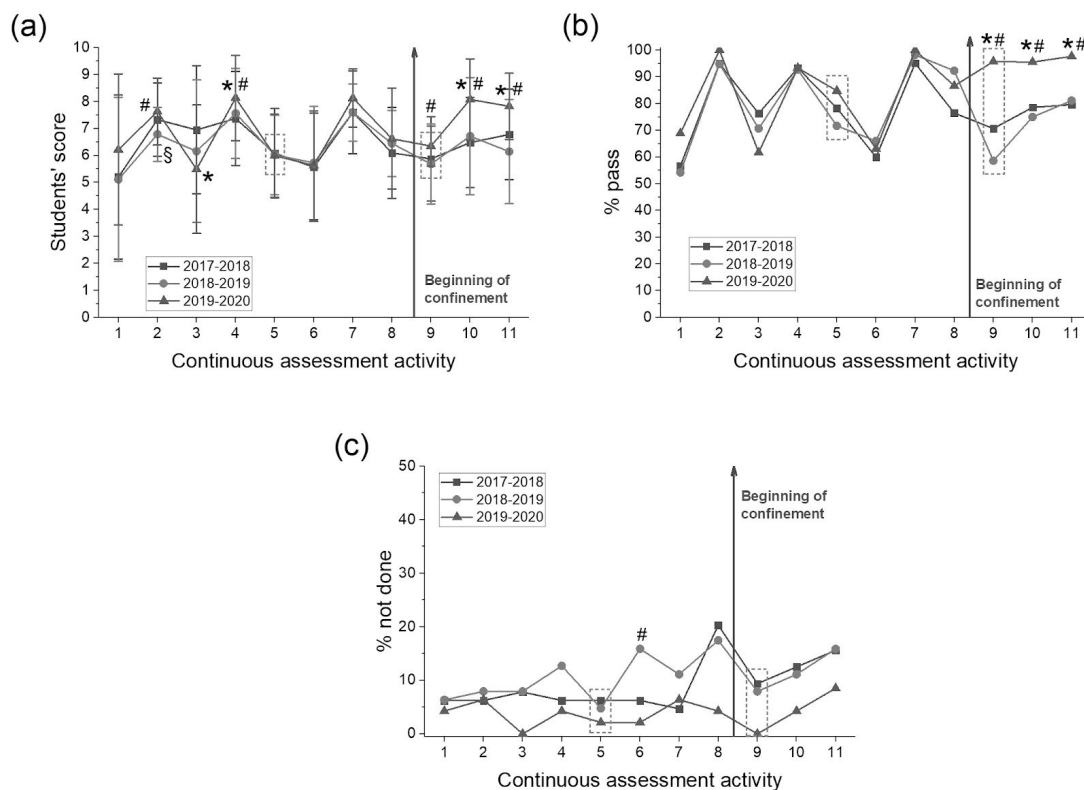

Fig 8. Results of students enrolled in Metabolism during the last 3 academic years.

(a) The score obtained by the students in the different activities of the continuous assessment is represented. (b) Proportion of students that pass the continuous assessment activities (score≥5). (c) Proportion of students that did not carry out the continuous assessment activities. From the 11 activities, only 2 are in-person activities under normal conditions (number 5 and 9, highlighted with a box), the other activities were always on-line.

The end of in-person teaching and beginning of confinement is indicated with a red arrow. Data is presented as mean±SD. Symbols indicate statistically significant differences (p<0.05) between courses 2017/2018 and 2018/2019 (§), 2017/2018 and 2019/2020 (*) and 2018/2019 and 2019/2020 (#).

6. DISCUSSION AND CONCLUSIONS

The aim of this study was to identify the effect of COVID-19 confinement on students' performance. Therefore, we conducted an experiment among 450 students from three subjects in different degrees of higher education at Universidad Autónoma de Madrid, Spain. The results of this study answer our 4 research questions:

1.   Is there any effect (positive or negative) of COVID-19 confinement in students' performance?

2.   Is it possible to be sure that COVID-19 confinement is the origin of the different performance (if any)?

3.   What are the reasons for the differences (if any) in students' performance?

4.   What are the expected effects of the differences in students' performance (if any) in the assessment process?

With respect to research question 1, the results show that there is a significant positive effect of COVID-19 confinement on students' performance. The results indicate that students obtain better scores in all kinds of tests that are performed after the beginning of confinement. Different sources of error have been removed from our study by including only subjects that fit the following requirements for the last three academic years:

-       Same teaching methodology and teachers all the years.

-       Same assessment process all the years, including both distant and on-site activities.

-       Tests performed both before and after the COVID-19 confinement in the present academic year.

ADeAPTIVE

To be sure that this positive effect is due to the COVID-19 confinement (research question 2), we have compared results of students from the present 2019/2020 academic year (experimental group) with results of students from the last two academic years 2018/2019 and 2017/2018 (control group). Students from the two previous academic years had similar scenarios from the beginning to the end of their course. When compared to students of the 2019/2020 academic year, we have two periods with different conditions. The first period is taught in the same conditions as in previous years (before confinement). In the second period, those conditions dramatically changed, and all the teaching and learning activities are limited to distant learning. The results of our studies clearly indicate that there are significant differences in students' performance after the confinement that cannot be found before in the same year or when comparing to the previous academic years.

At this point, it is clear that the variable that correlates with the change in students' performance is the beginning of confinement. However, we cannot establish yet if the difference is due to the:

- New learning methodology.

- New assessment process.

The problem with confinement is that not only learning and teaching strategies should be modified, but also the assessment process as it cannot be done face-to-face. There are some concerns related to these new assessment methods such as the opportunity of cheating by the students. This is the reason why we have chosen only subjects that include several tests that have not been modified because of the confinement. The results of our study show that students have a significant improvement in their scores also in tests that were performed in distant format also in previous years. Moreover, this improvement is only significant when comparing data after the confinement (i.e. there are not significant differences in tests in distant format that were performed before the confinement). These findings reveal that the new assessment process cannot be the reason for the improvement in students' performance because they also get better scores when the format of the assessment does not change. For these reasons, we establish that the new learning methodology is the main reason for the changes detected in students' performance after the confinement.

Funded by the
Erasmus+ Programme
of the European Union

Now, we shall discuss research question 3 (What are the reasons for the differences (if any) in students' performance?). We have proven that something has changed in students' learning methodology. The question is which is the common element in those methodologies that impulse the improvement in the learning process.

We have analysed data from two different subjects that used two very different learning strategies after the confinement. In one of them, additional e-Learning tasks were imposed on the students. Theoretical lessons were replaced by writing documents. In the other one, no additional e-Learning tasks were imposed, and theoretical lessons were replaced by multimedia classes. In both cases, we have found a significant increment in students' performance in the evaluation tests after the confinement. It seems that students' performance is increased independently of the learning strategies followed by teachers. Since we have established that the assessment process cannot be responsible for the differences, and we cannot find any common element in the learning methodologies, we must think about a general change in the autonomous learning process.

We have also analyzed data from previous years in two subjects that demonstrate that students do not work on a continuous basis. In both subjects, students work hard only the last days before the final exams. In one subject, we have found that more than 33% of the autonomous work was done the previous 5 days to the final exam. In the other subject, students only used the e-Learning material the last 2 days before the final exam, even when they were provided three weeks in advance. In this study, we have also found that students easily change their learning behavior and study continuously when a reward is offered. This extra motivation dramatically changed their learning strategy and students worked in a much wider time window, increasing their performance. An increase in the students' performance due to an adequate time management in the learning is well-established in the literature [12 – 13].

In the present COVID-19 confinement, students can find by themselves many different motivations (rewards) to work on a continuous basis. First, confinement is a new scenario that has never been faced by the students. For these reasons, students do not have any previous experience to use as a reference in their learning process. Without any previous

Funded by the
Erasmus+ Programme
of the European Union

reference, students should be confident that they are following the course correctly and therefore, work continuously to be sure that they do not miss any important content. Another interpretation is that they are afraid of missing the academic year because of the COVID-19 confinement and they work harder to overcome any difficulty. Finally, students may be motivated by their intrinsic responsibility in a very confused situation and work hard to contribute as much as they can to solve the problems that higher education is facing. Most probably, different students will find different motivations in this new scenario (probably a combination of many). We conclude that there is a real and measurable improvement in the students' learning performance that we believe can guarantee the good progress of this academic year despite the COVID-19 confinement.

Answering question 4, we have demonstrated that students get better grades in activities that did not change their format after the COVID-19 confinement. Moreover, we have demonstrated that there is an improvement in their learning performance. In conclusion, higher scores are expected due to the COVID-19 confinement that can be directly related to a real improvement in students' learning.

REFERENCES

[ ] https://eacea.ec.europa.eu/national-policies/eurydice/content/how-covid-19-affecting-schools-europe_en

[2] https://www.study.eu/document/impact-of-covid-19-on-studying-abroad-in-europe-overview

[3] Rodrigues H, Almeida F, Figueiredo V, Lopes SL. Tracking e-learning through published documents: A systematic review. Computers & Education. 2019;136;87-98.

[4] Biasutti M. A comparative analysis of forums and wikis as tools for online collaborative learning. Computers & Education. 2017. 111;158-171.

[5] Molins-Ruano P, Sevilla C, Santini S, Haya PA, Rodríguez P, Sacha GM. Designing video games to improve students' motivation. Computers in Human Behavior. 2014;31;571. https://doi.org/10.1016/j.chb.2013.06.013.

[6] Rizvi S, Rienties B, Rogaten J, Kizilcec RF. Investigating variation in learning processes in a FutureLearn MOOC. Journal of computing in higher education. 2020;32(1).

[7] McAuley A, Stewart B, Cormier D, Siemens G. The MOOC model for digital practice. SSHRC Application, Knowledge Synthesis for the Digital Economy. 2010.

[8] Caballero-Hernandez JA, Balderas A, Palomo-Duarte M, Delatorre P, Reinoso AJ, Dodero JM. Teamwork Assessment in Collaborative Projects Through Process Mining Techniques. International Journal Of Engineering Education. 2020;36;470-482.

[9] Palomo-Duarte M, Dodero JM, Medina-Bulo I, Rodriguez-Posada EJ, Ruiz-Rube I. Assessment of collaborative learning experiences by graphical analysis of wiki contributions. Interactive Learning Environments. 2014;22;444-466.

[0] Maldonado-Mahauad J, Perez-Sanagustin M, Kizilcec RF, Morales N, Munoz-Gama J. Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. Computers in Human Behavior. 2018;80;179-196.

[1] Tsai YS, Poquet O, Gašević D, Dawson S, Pardo A. Complexity leadership in learning analytics: Drivers, challenges and opportunities. British Journal of Educational Technology. 2019;50(6); 2839-2854

[2] Gašević D, Matcha W, Jovanović J, Pardo A. Analytics of time management strategies in a flipped classroom. Journal of Computer Assisted Learning. 2019;36(1);70-88

[3] Jovanovic J, Mirriahi N, Gasevic D, Dawson S, Pardo A. Predictive power of regularity of pre-class activities in a flipped classroom. Computers & Education. 2019;134;156-168.

[4] Zimmerman BJ, Kitsantas A. Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. Contemporary Educational Psychology. 2014,39(2);145-155.

[5] Zimmerman BJ. Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. American Educational Research Journal. 2008;45(1);166-183.

[6] Corno L. The metacognitive control components of self-regulated learning. Contemporary Educational Psychology. 1986;11(4);333-346. https://doi.org/10.1016/0361-476X(86)90029-9.

[7] Zimmerman BJ, Martinez-Pons M. Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. Journal of Educational Psychology. 1990;82(1);51.

[8] Panadero E, Broadbent J, Boud D, Lodge JM. Using formative assessment to influence self- and co-regulated learning: the role of evaluative judgement. European Journal of Psychology of Education. 2019;34(3);535-557.

[9] Winne PH, Theorizing and researching levels of processing in self-regulated learning. British Journal of Educational Psychology. 2018;88(1);9-20.

[20] Wang TH. Developing Web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-Learning environment. Computers & Education. 2011;57(2);1801-1812.

[22] Bellhäuser H, Lösch T, Winter C, Schmitz B. Applying a web-based training to foster self-regulated Learning-Effects of an intervention for large numbers of participants. The Internet and Higher Education. 2016;31;87-100.

[22] Broadbent J. Comparing online and blended learner's self-regulated learning strategies and academic performance. The Internet and Higher Education. 2017;33;24-32.

[23] Özyurt Ö, Özyurt H. Learning style based individualized adaptive e-learning environments: Content analysis of the documents published from 2005 to 2014. Computers in Human Behavior. 2015;52;349-358. https://doi.org/10.1016/j.chb.2015.06.020.

Funded by the
Erasmus+ Programme
of the European Union

[24] Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory; 1991.

[25] Lord FM. Applications of item response theory to practical problems. Lawrence Erlbaum Associates, Hillsdale, NJ; 1980.

[26] Wainer H. Computerized adaptive testing: a primer. Hillsdale, NJ: LEA; 1990.

[27] van der Linden WJ, van Krimpen-Stoop EM. Using response times to detect aberrant responses in computerized adaptive testing. Psychometrika. 2003;68;251-265.

[28] Narciss S, Sosnovsky S, Schnaubert L, Andrès E, Eichelmann A, Goguadze G, et al. Exploring feedback and student characteristics relevant for personalizing feedback strategies. Computers & Education. 2014;71;56-76.

[29] de la Rubia MA, Sacha GM. Adaptive Tests as a Tool for Evaluating Work Groups in Engineering. International Journal of Engineering Education. 2020;36;422-429.

[30] Sai Sabitha A, Mehrotra D, Bansal A. An ensemble approach in converging contents of LMS and KMS. Education and Information Technologies. 2017;22(4);1673-1694. DOI:10.1007/s10639-016-9516-7.

[3] Castro-Schez JJ, Redondo MA, Jurado F, Albusac J. Experience applying Language Processing techniques to develop educational software that allow active learning methodologies by advising students. Journal of Network and Computer Applications. 2014;41;65-79.

[32] Veldkamp BP, Avetisyan M, Weissman A, Fox JP. Stochastic programming for individualized test assembly with mixture response time models. Computers in Human Behavior. 2017;76;693-702.

[34] Wainer H, Dorans NJ, Eignor D, Flaugher R, Green BF, Mislevy RJ, Steinberg L, Thissen D. Computer Adaptive Testing: A Primer (2nd ed.). Mahwah, NJ:Lawrence Erlbaum Associates; 2000.

[35] Martin AJ, Lazendic G. Computer-Adaptive Testing: Implications for Students' Achievement, Motivation, Engagement, and Subjective Test Experience. Journal of Educational Psychology. 2018;110(1);27–45.

[36] Lu H, Hu YP, Gao JJ, Kinshuk. The effects of computer self-efficacy, training satisfaction and test anxiety on attitude and performance in computerized adaptive testing. Computers & Education. 2016;100;45-55.

[37] Molins-Ruano P, González-Sacristán C, Díez F, Rodriguez P, GM Sacha. Adaptive Model for Computer-Assisted Assessment in Programming Skills. International Journal of Engineering Education. 2015;31;764.

[38] Boyd AM. Strategies for controlling testlet exposure rates in computerized adaptive testing systems. Unpublished PhD thesis. The University of Texas at Austin; 2003.

[39] Eggen TJHM. Overexposure and underexposure of items in computerized adaptive testing (Measurement and Research Department Reports 2001–1. Citogroep, Arnhem, The Netherlands: 2001.

[40] Molins Ruano, P, Rodríguez P, Atrio Cerezo S, GM Sacha. (2016). Modelling experts' behavior with e-valUAM to measure computer science skills. Computers in. Human Behavior. 2016;61;378-385. Available in e-valuam.ii.uam.es

[4] D'Agostino R, Pearson E. Tests for Departure from Normality. Empirical Results for the Distributions of b2 and $\sqrt{}$ b1. Biometrika. 1973;60(3), 613-622. doi:10.2307/2335012

[42] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association. 1952;47(260);583–621.

[43] Mann HB, Whitney DR. On a test whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 1947;18;50–60.