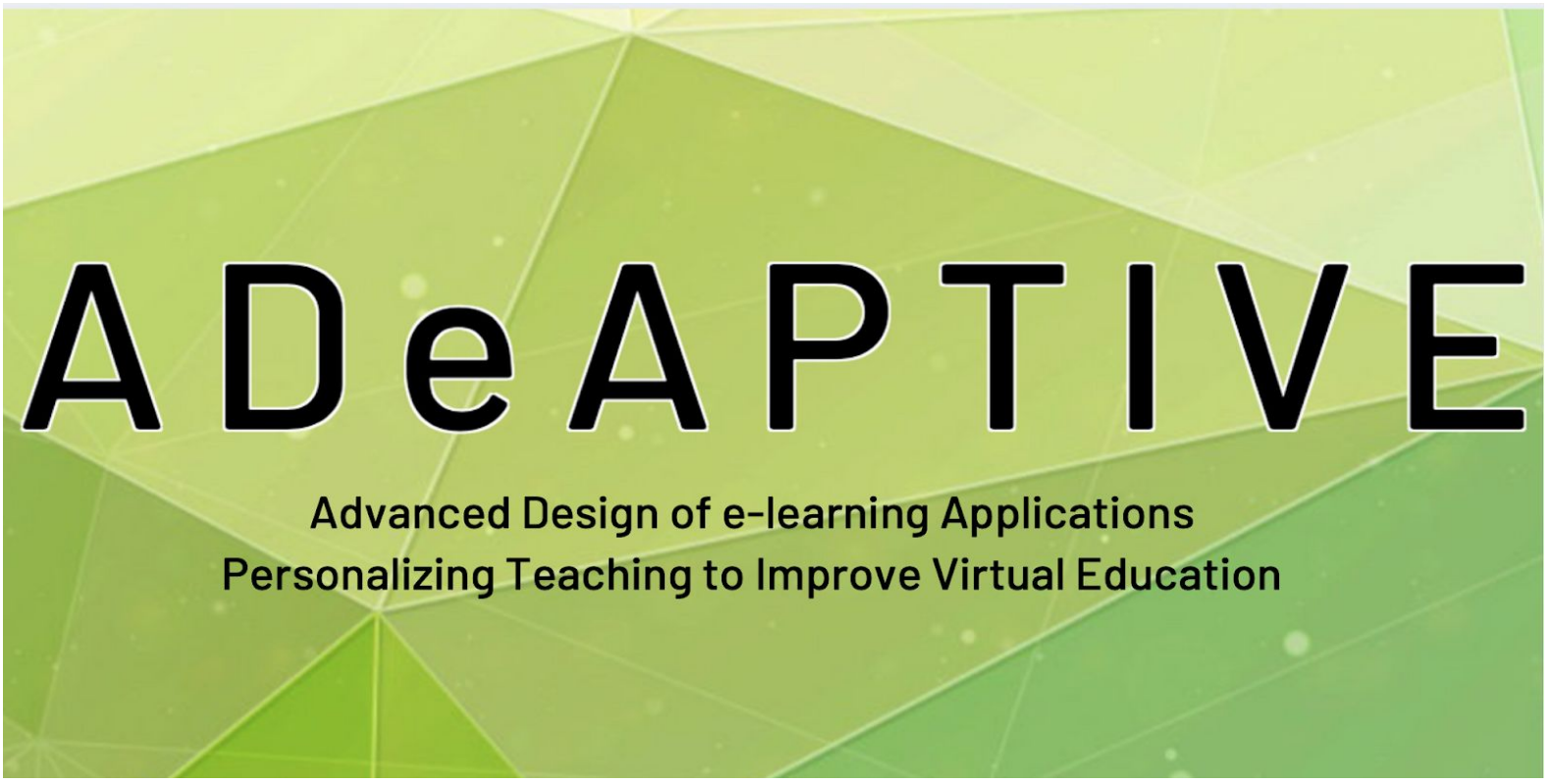


Advanced Design of e-Learning Applications Personalizing Teaching to Improve Virtual Education.

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

Funded by the
Erasmus+ Programme
of the European Union

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

Disclaimer: The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Funded by the Erasmus+ Programme of the European Union

CONTENTS:

Funded by the Erasmus+ Programme of the European Union

AUTHORS: UAM, EURECAT, JYU, UVT, UCLL

All tasks described before are included in different functionalities of the platforms. Since not all the tasks could be effectively included in a single platform, we decided to create three different platforms in order to be sure that all aspects could be correctly developed. This does not mean that a single platform cannot include all the aspects. This means that by developing different platforms we can cover in a more accurate way all the requirements. By doing so, we have made a higher effort than the one expected at the beginning of the project. However, we decided to do it in order to be sure that we offer the best results we can.

- The moodle-based platform focuses more on the wrong use of the application in terms related to tasks 3x (communication, content creation, etc).
- E-valUAM focuses mainly on tasks related to 1x (computer skills). However, this is not exclusive and both platforms are ready to be used in all tasks. It is just an optimization procedure.
- Results from these tasks are the development of platforms that are online though the project web page (www.adeaptive.com) and, in the case of ADeAPTIVE(, in the following links:

www.evaluam.eu

for version 3.0

also available in

e-valuam.ii.uam.es

for version 2.0

- The details of this applications are described in tasks A11, A14, A22
- A tutoring system was used in tasks related to the creation of an analyzer (A2x). The archaeology tutor was in production in August 2019 during the El Salt archaeology campaign. A demo is available under request. Because this software, due to its complex design, is not available for all public without supervision, it is fully described in this document. (A22, A23 and A24)

Funded by the Erasmus+ Programme of the European Union

A11: Creation of general end user computer skills, learning procedures and social media self-management evaluation methods by using adaptive tests.

Author: UAM

SUMMARY

In this document we present:

- A general system of adaptive tests that can be used for all the concepts proposed in this task.
- We have defined sequential adaptive tests to improve the assimilation of contents during learning.
- This system is used to avoid learning strategies where only learning answers in instead of the assimilation and individual reflection of the contents.
- A conclusion of this document is that parameters such as the number of attempts needed to achieve the maximum grade or the first grades obtained in each set of questions are relevant factors in order to predict and correct inappropriate methods in the teaching and learning process.
- This document includes a full analysis based on data collected before the beginning of the project, so it is not included as a pilot experience. It is included in this task as a guide supported by data about how to create the contents required in the task.

1. INTRODUCTION

Computer-Aided Learning (CAL) environments are more present in educational institutions. It is common practice for teachers and students to use computer procedures to support their teaching / learning processes. [1]. In this way, evaluation processes, or self-evaluation, can specially benefit from the use of these practices. The use of new technologies has even made it possible to create Massive

Funded by the Erasmus+ Programme of the European Union

Open Online Courses (MOOC) [2], where the large number of students enrolled means that personalised follow-up by teachers is not feasible. In this scenario, the use of technical and methodological environments from the world of CAL is indispensable.

In the phases of student learning and self-evaluation, systems that are capable of providing feedback to students have very high teaching potential as they provide elements that can be considered substitutes for teacher interaction [3]. Thus, assessment tools can be found for some specific learning domains that are capable of advising students [4][5][6][7]. However, trying to provide information and advice to students for more generalized domains, adaptive tests (Computer Adaptive Testing, CAT) seem to offer advantages that go beyond providing a mere instantaneous score [8].

Environments that make use of adaptive tests take into account the particular abilities of the students in order to adapt the questions presented to them based on their previous answers, often including some type of personalized feedback [9]. Examples of the application of adaptive testing can be found to improve language skills [10], identify learning styles [11], measure chess skills [12] or mathematical ability [13], improve some personality characteristics [14] or assess the health status of participants [15]. In the field of education, different approaches and systems based on CATs have been proposed and used for decades to improve the learning process [16] [17] [18] [19]. This type of tests requires additional work compared to the traditional ones, due to its high complexity. Therefore, a good calibration process is essential [13], with student models sometimes used to improve the quality of the assessment process [20] [21] [22] [22] [23].

One of the most frequent problems found in this environment is the lack of control of the study methodology of the users. Given that this type of tests are usually closed answers (multiple choice), one of the most obvious drawbacks is that the learning method can sometimes lead to the memorization of the correct answers instead of real learning of the content.

In this document we show that the vast majority of students memorize answers when using CATs as a method of study. To solve this problem, we proposed a CATs design where the questions are presented in different groups and they are shown to the students gradually. With our study based on a representative sample of students, we show that this system is very effective in helping students to change their learning strategy towards methods that help them assimilate the contents correctly. The proposed methodology has been implemented in e-valUAM [24], an application that aims to increase the quality of tests by improving the objectivity, robustness, security and relevance of content [25].

Section II describes the proposed sequential adaptive test method to improve efficiency during

Funded by the Erasmus+ Programme of the European Union

autonomous learning. Section III explains the data used during the experiment with real students. Section IV discusses the results obtained and the relevant parameters of the study. Finally, section V summarizes the finds of the work.

2. METHODOLOGY

The proposed method has two main objectives: first is to detect the students who use the adaptive tests incorrectly (memorization of the answers without a correct assimilation of contents), and second to correct this undesirable effect.

For this purpose, adaptive tests will be used where the questions are divided into different levels of difficulty. When students answer a number of questions correctly, the test will begin to show questions of the next level. The modification proposed in this document is to divide the question repositories in 4 groups depending on the difficulty. The first group is shown initially to the students. When a student obtains 3 grades superior to 9 points, the questions from the second group are automatically shown on the next test. In this second group, there are also questions of 3 levels of difficulty. In a similar way, questions from the second group are shown to the student until they demonstrate high performance. In order to present content from the third group of questions the student must answer 2 over 9 questions correctly. In this second step, the difficulty of changing groups has been reduced, assuming that the students reach this point with a higher level of knowledge.

During the learning phase, the questions will no longer change, so that after reaching the third group, the student will continue to answer the same questions in all his or her attempts. The fourth group of questions corresponds to the final exam, and in this case a small number of questions of the most basic level have been kept identical to those of the third group. These questions can be used to control the type of learning the student does. A schematic representation of this methodology is shown in Fig. 1.

3.EXPERIMENT

The experience with real students took place during the 2015/2016 academic year and was monitored in successive courses. There were 44 students involved in the subject "Environment as an Educational

Funded by the Erasmus+ Programme of the European Union

Resource", with the mention of Knowledge of the Environment from Science and Mathematics. This subject was taught through theoretical and practical lessons and is part of the educational plan of the fourth year of the Master's Degree in Infant Education at the Faculty of Teacher Training and Education of the Autonomous University of Madrid (UAM). The work has been possible thanks to collaboration of the Polytechnic School of the UAM and the aforementioned Faculty.

The test that was used in the autonomous learning phase by the students was designed with three levels, questions of each level increased in difficulty as they advanced to higher levels. At lower levels, the questions contained general knowledge, while at the higher levels, advanced knowledge of accuracy was required.

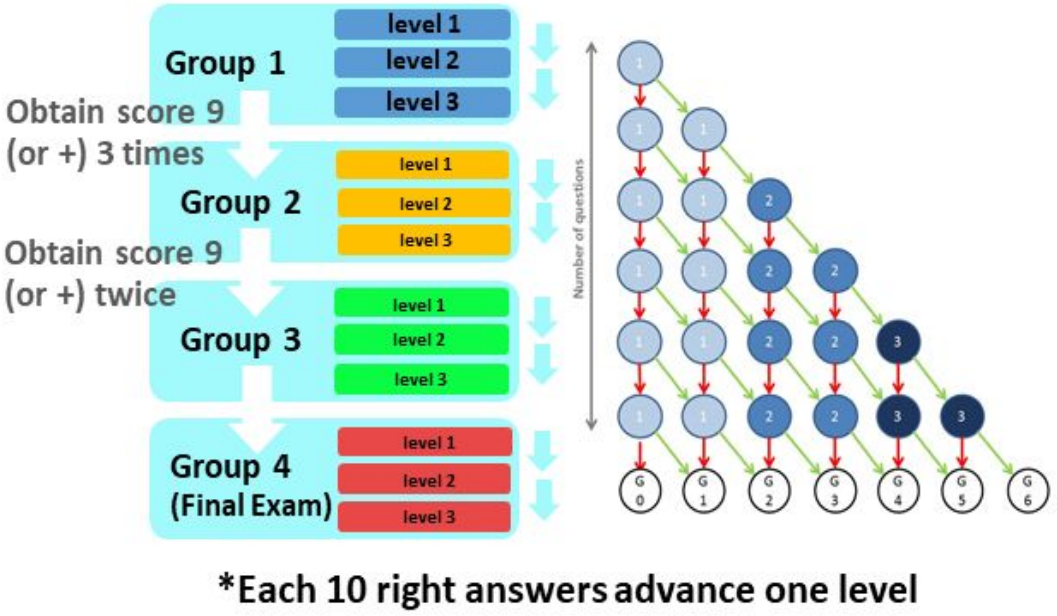


Fig. 1. Representation of the proposed sequential adaptive test model. On the left it is shown the 4-group method, and on the right it is shown model Unidirectional Levels Method (ULM).

Funded by the Erasmus+ Programme of the European Union

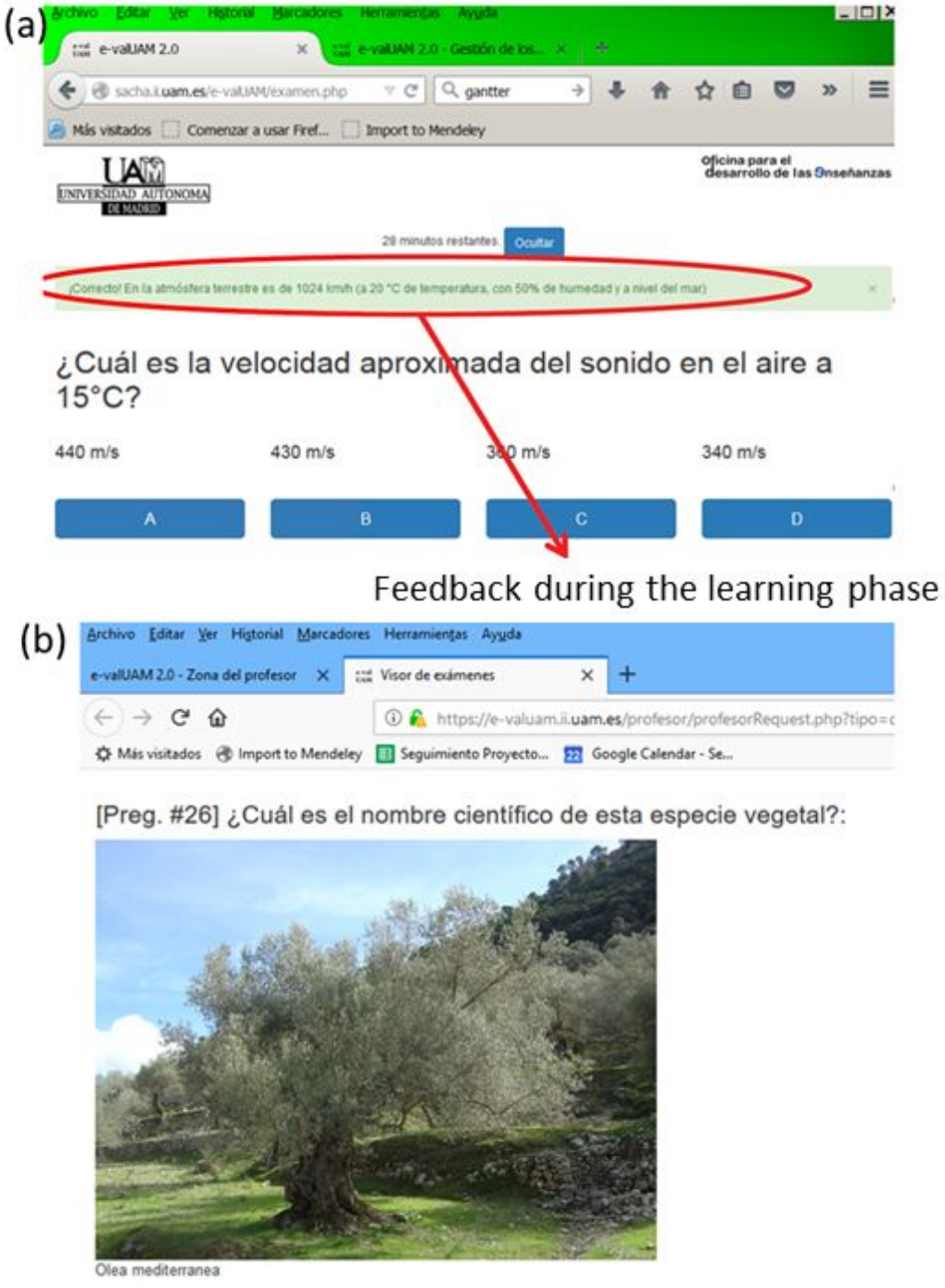


Fig. 2. Examples of questions included in the test. (a) Level 1 question. (b) Level 3 question.

Funded by the Erasmus+ Programme of the European Union

Two test questions are shown in Fig. 2. Fig. 2a shows a first level question with the interface that is shown to students when they answer the questions, and Fig. 2b corresponds to a question of the last level shown with the interface that teachers see in a completed exam. As can be seen, the most advanced questions require knowledge of technical terms not needed at previous levels.

Students have four answers to each question, with only one correct answer. Each question has 4 possible answers, only one of them correct. In order to pass the level, students must answer correctly 10 out of 30 questions. As long as this objective has not yet been achieved, the difficulty of the questions will not change and will remain at the same level until necessary questions are resolved.

44 students participating in this experiment out of a total of 55 students. Further consideration was taken into account to participate in this experiment:

- Students who have not reached the third group have been excluded.
- Only finished attempts. Students who made fewer than 20 attempts during the self-study phase were not used.

4.Analysis

Fig. 3 shows the scores of two students according to the number of attempts made during their study phase with the application. The attempts are ordered chronologically in ascending order, so that the last attempts correspond to those closest to the date of the final test.

This Fig.3 shows a group of points divided into three groups. Each colour change represents a group transition from a level of questions to the following one. In the graph above, a similar trend can be seen in the three groups of points. In the first group, a student has made several attempts in which they or she has gradually increased his or her grade until reaching the number of 9, or more than 9, on three occasions. into the second group, the student has suffered a sharp drop in grades, indicating minimal assimilation of content. This effect is repeated in group 3, which may mean that the student has been more concerned with memorizing than studying efficiently throughout the learning phase. We consider that if this student had acquired the content the number of attempts needed to move from group would decreased, so the drop in his/her grades would be less pronounced, as it is shown in Fig 3b.

Funded by the Erasmus+ Programme of the European Union

To sum up, the indicators that make us think of more efficient learning are: 1) the number of attempts needed to change groups has decreased and 2) the score of the first attempts in the next group has improved.

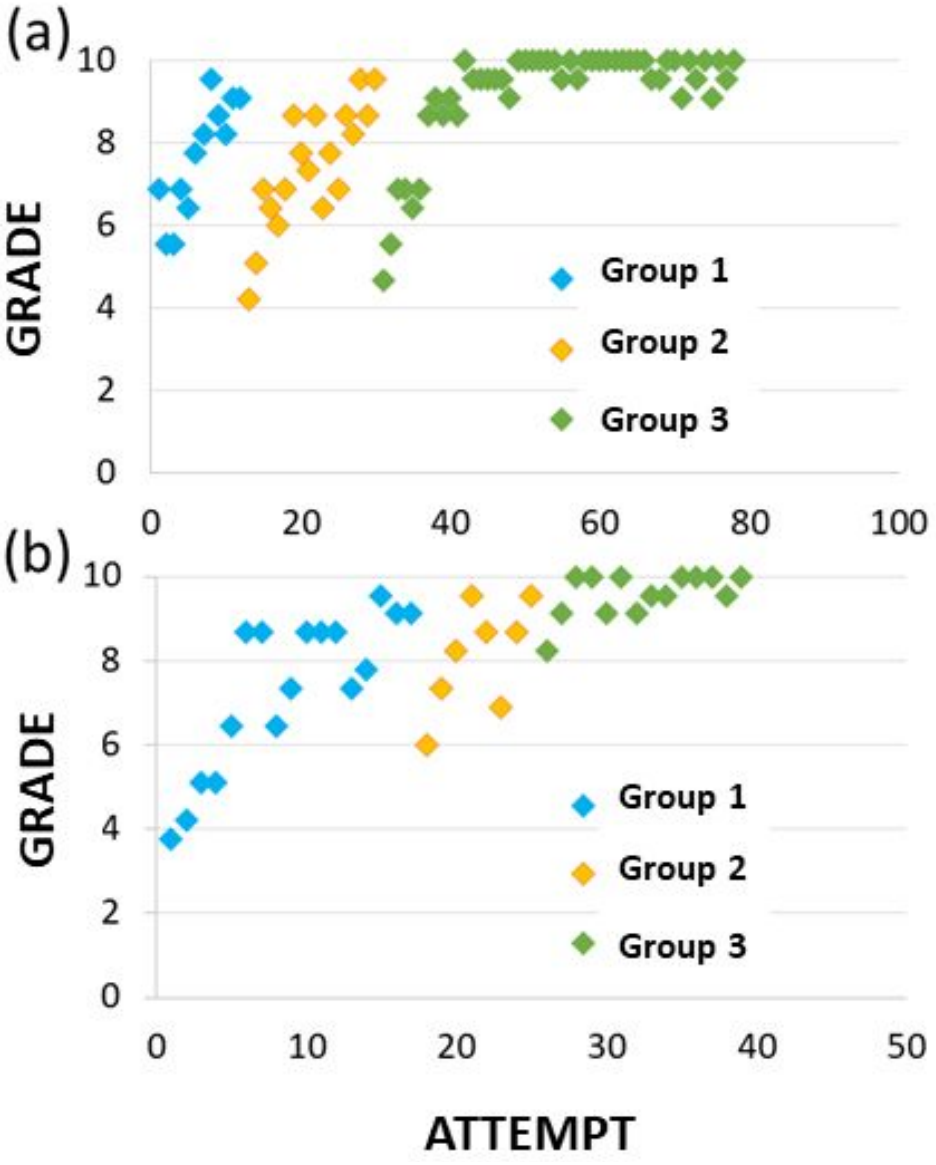


Fig. 3. Grades obtained by students in chronological order during the learning phase. The scores are represented with different colour codes depending on the group from which the questions were extracted. The figure shows the two typical profiles of students, one of them (a) has a very sharp drop in grades when changing groups, and the other (b) has a much smaller drop as they progress through the learning phase.

Funded by the Erasmus+ Programme of the European Union

We first analyse the influence of the number of attempts to overcome a group. Fig. 4 shows the average number of attempts that students have needed to change from a group to another. The data is displayed in groups, where each bar corresponds to students with a similar number of total attempts. It can be seen that, in practically all cases, the number of attempts required to move to the second and third groups decreases with respect to the first group. In the case of students with the highest (37-47) and lowest (10-24) number of attempts, it is observed that they use a greater number of attempts to overcome the second group, but in these cases this effect is also corrected when facing the third group. These data are an indicator of the beneficial effects of dividing the questions into groups, as even students who did not improve in the second group did so in the third.

Another significant effect seen in the Fig.4 is that the number of attempts needed to overcome the first group follows a logarithmic trend, while the trend changes to exponential when the data from the third group are represented.

This effect indicates that students who have intermediate values benefit most of the inclusion of several groups, since the distance between both tendencies is greatest.

Fig. 5 shows the first grade obtained for 4 students in the 3 groups of the learning phase and the final grade of the exam (group 4). It can be seen the effect in the grades when students are first confronted with a new set of questions. As shown in Fig.5, there are different tendencies in these 4 students that, however, have some characteristics in common. In general, in all cases there is an upward trend, which is symptomatic of their learning process. It is difficult to find students who have not learned at least to some extent through this process.

However, to analyse the advantages of including several groups of questions, we must pay attention to the score obtained in the second group, since it is the first time that the questions change. Comparing the grades obtained in group 2 and the subsequent ones, it can be seen that, in general, the ascending trend continues, which is a proof of the beneficial effect of dividing the learning phase into groups. Only students (a) and (b) show no improvement between group 2 and group 4, and this is an effect of the high grades obtained in group 2 (We can consider these students had these students would have a good assimilation of content from the beginning)

Funded by the Erasmus+ Programme of the European Union

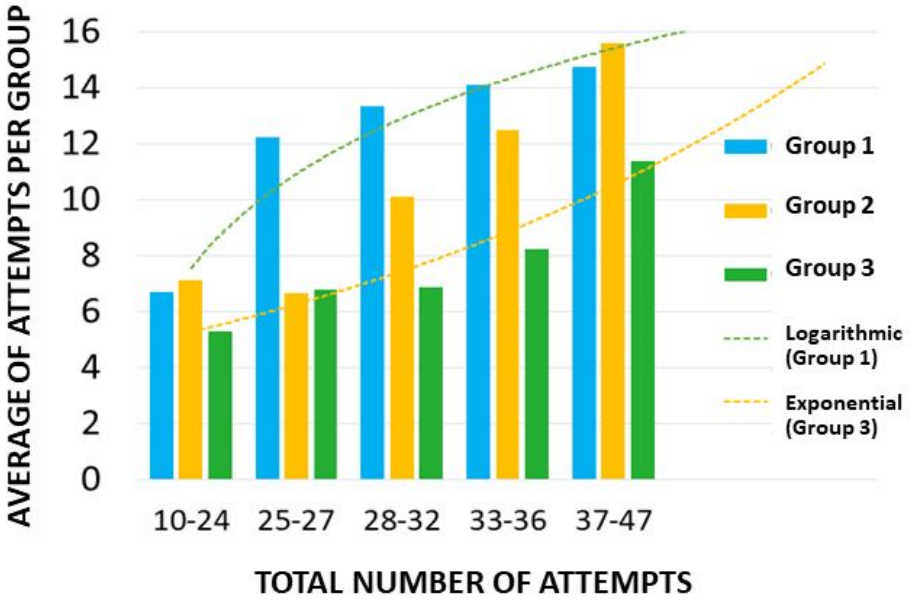


Fig. 4. Average number of attempts students need to overcome each set of questions. Data are displayed for groups of students sorted by the total number of attempts made.

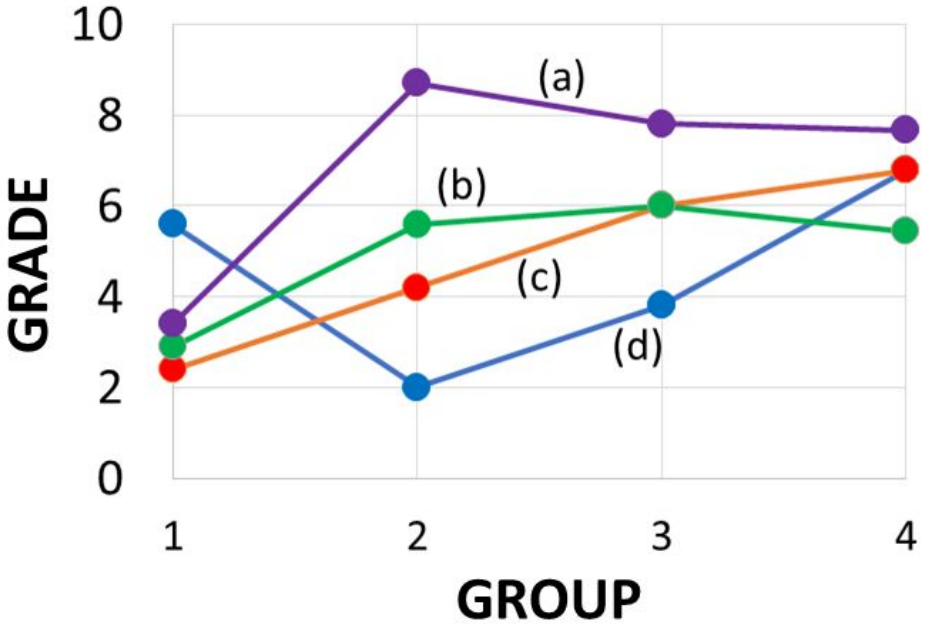


Fig. 5. Qualification of students on their first attempt of each group of questions. The data in group 4 correspond to the grades of the final exam.

Funded by the Erasmus+ Programme of the European Union

If we analyse the data of the 44 students in the study, we obtain that 75% of students have grades better in the final exam (group 4) than they would have gotten grade in group 2 (a grade they would have scored in a hypothetical final exam if the test had not been divided into groups). If the split had been made into 2 groups during the learning phase instead of 3, the number of students who would have benefited are reduced to 63.64%. Therefore, we can conclude that the division into groups significantly improves the students' grades and so their learning. In addition, a greater number of partitions is more beneficial to students. This fact competes with the additional work that requires the realization of a higher number of questions.

5.CONCLUSION

We have proposed a sequential adaptive test model in which students are presented questions divided into groups to maximize performance during their learning process. Using this strategy, we have shown 75% of students' grades improve with our system presented in this document compared to the use of conventional adaptive testing.

The analysis also shows that prolonged use of the application does not lead to adequate assimilation of knowledge, and is not necessarily the most effective method during autonomous learning.

REFERENCES

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

A. McAuley, B. Stewart, D. Cormier, and G. Siemens, "In the Open: The MOOC model for digital practice," SSHRC Application, Knowledge Synthesis for the Digital Economy, 2010.

A. Kumar, "Using online tutors for learning - what do students think? Proceedings of Frontiers in Education Conference (FIE 2004)", IEEE, pp. 524–28, 2004.

C. Bravo, WR. van Joolingen, T. de Jong, "Using co-lab to build system dynamics models: students' actions and on-line tutorial advice". Comput Educ, pp. 243–51, 2009.

A. Sai. Sabitha, D. Mehrotra, A. Bansal, Abhay. "An ensemble approach in converging contents of LMS and KMS". Education and information technologies Vol. 22 pp. 1673-1694, 2017

F. Jurado, M. Redondo, M. Ortega. "eLearning standards and automatic assessment in a distributed

Funded by the Erasmus+ Programme of the European Union

Eclipse based environment for learning Computer Programming", Computer Applications in Engineering Education, DOI: 10.1002/cae.21569, 2012.

JJ. Castro-Schez, M.A. Redondo, F. Jurado, "Experience applying language processing techniques to develop educational software that allow active learning methodologies by advising students", Journal of Network and Computer Applications, vol. 41(1), pp. 65-79 (2014) DOI: 10.1016/j.jnca.2013.10.011

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

A.A. Economides, "Personalized feedback in CAT", WSEAS Transactions on Advances in Engineering Education, 2-3, 174-181, 2005.

C.A. Chapelle and D. Douglas, 2006. "Assessing Language Through Computer Technology", Cambridge University Press, 2006.

A. Ortigosa, P. Paredes and P. Rodriguez, "AH-questionnaire: An adaptive hierarchical questionnaire for learning styles", Computers & Education, pp.999-1005, May, 2010.

H.L.J. Van Der Maas, and E-J. Wagenmakers. "A Psychometric Analysis of Chess Expertise", The American Journal of Psychology. Vol. 118, pp. 29-60, 2005.

S. Klinkenberg, M. Straatemeier, H.L.J. Van der Maas, "Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation". Computers & Education, vol. 57, 2 pp. 1813-1824, Sept.2011.

S. Stark, O.S. Chernyshenko, F. Drasgow and L.A. White, "Adaptive Testing with Multidimensional Pairwise Preference Items Improving the Efficiency of Personality and Other Noncognitive Assessments", Organizational Research Methods, pp. 463-487, 2012.

D.A. Revicki, & D.F. Cella, "Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing", Quality of Life Research, vol. 6, 6, pp. 595-600, 1997.

D.J. Weiss, 1982. "Improving measurement quality and efficiency with adaptive testing", Applied Psychological Measurement, 6, 473-492, 1982.

Funded by the Erasmus+ Programme of the European Union

P. Molins-Ruano, C. González-Sacristán, F. Díez, P. Rodriguez and G. M. Sacha. "An Adaptive Model for Computer-Assisted Assessment in Programming Skills". International Journal of Engineering Education. vol 31, pp. 764, 2015.

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

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

M. Antal, and S. Koncz, "Student modeling for a web-based self-assessment system". Expert Systems with Applications, vol. 38, 6, pp.6492-6497, 2011.

M. Virvou, & C. Troussas, "Web-based student modeling for learning multiple languages". In Proceedings of the Conference on Information Society, i-Society. IEEE, pp. 423-428, 2011.

J. Galvez, E. Guzman, R. Conejo, E. Millan, "Student Knowledge Diagnosis Using Item Response Theory and Constraint-Based Modeling. In Proceedings of the Conference on Artificial Intelligence in Education - Building Learning Systems that care: From Knowledge Representation to Affective Modelling", IOS Press, pp.291-299, 2009.

F.A, Dorça, L. V. Lima, M.A Fernandes and C.R. Lopes, "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, vol. 19, 1, pp. 43-58, March, 2013.

Available in e-valuam.ii.uam.es

P. Molins-Ruano, S. Atrio, P. Rodríguez, and S. Gomez-Moñigas "Modelling experts' behavior with e-valUAM to measure computer science skills". Computers in Human Behavior, vol. 61, pp. 378-385, 2016.

Funded by the Erasmus+ Programme of the European Union

A12, A13, A14, A15, A16: Developing of tools and pilot experiences related to Computer Skills.

A17, A25: Third year piloting experiences on digital skills and analysis of students' apparent computer skills.

Author: UAM

Self-assessment method based on questions with variable numerical values

SUMMARY

- This document presents an adaptive test applied to computing skills courses in which the questions are automatically modified each time the students run the application.
- The method presented proves to be effective for the use of the same system in both learning and evaluation phases.
- This homogeneity represents a great advance in the feedback processes between students and teachers as it provides information on the progress made in real time.
- The information acquired during the learning phase can also be used to monitor correct development of the subjects throughout the teaching periods or to detect deficiencies in specific students.
- The results of the proposed method are justified by pilot experiences that include more than 170 students over two academic years.
- Due to COVID-19 pandemic, results of third year A17, A25 were substituted by a different approximation that is divided in two parts.
    - The first one is a general study that involves more disciplines and objectives and is included in IO3.
    - The second one is a statistical study about student preferences in confinement that is included at the end of this section.

1. INTRODUCTION

The requirements of the new teaching models, which increasingly require systems for monitoring and interacting with students, benefit greatly from computer-aided learning (CAL) strategies [1]. A very clear example of the potential of computers in teaching is the Massive Open Online Courses (MOOC)

Funded by the Erasmus+ Programme of the European Union

[2], where the large number of students enrolled makes manual monitoring of the learning/evaluation processes unfeasible. Information technology converts initiatives such as MOOCs into viable practices as massive data processing can be carried out without direct supervision of teaching staff.

One of the biggest challenges of using computers in teaching is the way in which interaction with teachers is replaced by automatic methods [3]. One of the concepts that has aroused most interest is tutoring and the ability to turn an automatic system into an effective tool for advising and guiding students [4][5][6][7]. In this environment, Computer Adaptive Testing (CAT) tests offer a great advantage due to their adaptation to different student profiles [8].

Environments that use adaptive testing are able to adapt the content presented to learners based on their responses, often including feedback to complement training and correct learner deficiencies [9]. There are many settings in which this type of test is commonly used: clinical practice and assessment of patients' quality of life [8], language skills [10], identification of learning styles [11], mathematical ability [12] and computer skills [13]. In education, different approaches and systems based on CAT have been proposed and used to improve the learning process [14][15][16][17]. Among the drawbacks of using these tests is the need for a good calibration process [13], with student models sometimes used to improve the quality of the assessment process [18][19][20]. The advantages of CAT-specific feedback are also combined with multistage adaptive testing (MAT) [21] or cognitive diagnostic assessment [22].

One of the current problems in the evaluation and programming of a subject in higher education is the requirement of continued student work and feedback from teachers. These requirements are a constant challenge for the teaching staff, who are forced to carefully plan their activities. In this document we present a model learning tool for students that can also be used as a final assessment tool. As the same tool is used during the learning and evaluation phases, students receive very precise and concrete feedback on their assimilation of content, which allows them to detect their shortcomings at any time and correct them if necessary. A study carried out with real students during the 2016/2017 and 2017/2018 academic years also demonstrates the effectiveness of this method for teachers, who have the ability to detect problems over the course of the class or in individual student's performance.

In section II we present the proposed method. Section III shows the results of the study carried out during the 2016/2017 and 2017/2018 academic years. Finally, in section IV we describe the conclusions of the work.

Funded by the Erasmus+ Programme of the European Union

2. METHODOLOGY

The system proposed in this document has been developed with a view to providing a system of self-assessment that can also be used as a system for final student assessment. This approach is very dangerous in the sense that students could focus their study strategy on learning only how to pass the assessment test instead of learning the contents correctly. To avoid this effect, the system will have the following characteristics:

1) A large number of questions. The more questions the students have to answer, the less likely they are to focus on learning the answers by heart. In addition, a sufficiently large number of questions would cover the entire agenda, so it would be irrelevant for the student to learn them by heart. In other words, if the student memorized how to answer all the questions, they would have learned the whole content of the subject, so that the objectives would be covered.

2) Even with a large number of questions covering the content of the course, there is the additional problem that students may learn the answers by heart rather than by procedure. To this end, we have designed a dynamic system in which the questions focus on numerical problems in which the parameters of the statement vary in each execution of the problem and, therefore, the answers also vary.

This system has the following additional advantages:

1) Questions are shown to students randomly (except for the restriction of the level of each question outlined below). This fact, together with the different results due to the dynamic parameters of the statement, makes it virtually impossible for students to copy from each other. In order to be able to copy, they would initially have to be lucky that two students located nearby were working on the same question. But, even then, the parameters of each one would be different, so they would have to copy the entire procedure, as well as being able to detect the points where the parameters have influence and modify them accordingly.

2) Being able to use literally the same tool during the learning/self-assessment phase and the formal assessment phase has the great advantage of providing very accurate and continuous feedback to learners about their learning process.

3) Another advantage of using the same tool in both phases is that the students have an exact knowledge of what evaluation system is going to be used, not being able to argue that the evaluation test surprised them, was very difficult or inadequate. These are the typical arguments that are used in

Funded by the Erasmus+ Programme of the European Union

exam reviews and they are eliminated with the use of our system.

4) The tool on which the assessment tests were developed generates a great deal of information on the attempts made by students throughout their learning process, which is very useful for teachers when it comes to detecting possible learner deficiencies and being able to prevent them with sufficient time.

- In our model, the questions will have the following elements:

Statement: This is the text field in which the question is asked. As a special feature the numbered variable parameters need to be entered after the $ sign. The following text would correspond to a statement that includes two variable parameters.

"Write the result of the area of a rectangle of height $1 and width $2."

In this case, both the width and the height of the rectangle will be values that will change between different executions of the application, generating different results.

Maximum and minimum values of the variable parameters: Once the application is told how many variable parameters exist in the statement, limit values must be given for each parameter. The system will generate a random actual number between the two values.

Image: The system additionally accepts the inclusion of images in the statement. This content is optional.

Question level: The proposed method is included in e-valUAM [23], which is an adaptive test application. The model used requires that the questions be divided into levels, where the lowest levels are initially shown to the students and only when a minimum number of correct answers are given, are they moved on to the next levels.

Source code of the solution: Since the numerical solution of the exercises varies in each execution due to the variable parameters, a code must be included to solve the problem and generate the solution. The application has been developed to support mathematical functions programmed in Matlab.

The graphical interface of the application at the time of adding a question is shown in Fig. 1. Format in which all the required parameters should be added can be seen.

Funded by the Erasmus+ Programme of the European Union

Figure 2 shows how students would view the question introduced in Figure 1. As can be seen, the variable parameter represented in the statement in Figure 1 as "$1" already appears with a value generated between the limits 1 and 5 that were specified when the question was created. For students, there is no way to know which parameter will change between one run and another.

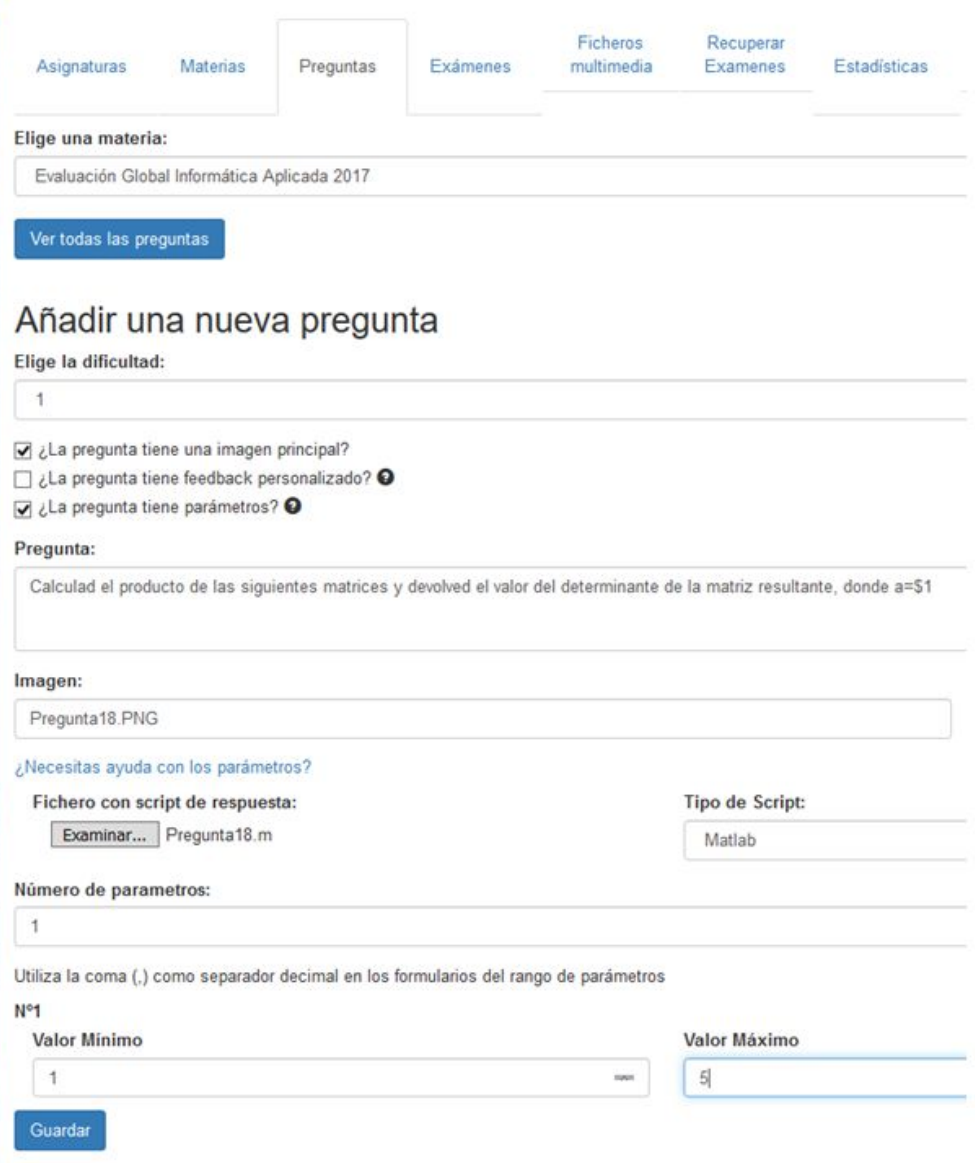


Fig. 1. e-valUAM graphic interface for teachers in the Add new question section.

Funded by the Erasmus+ Programme of the European Union

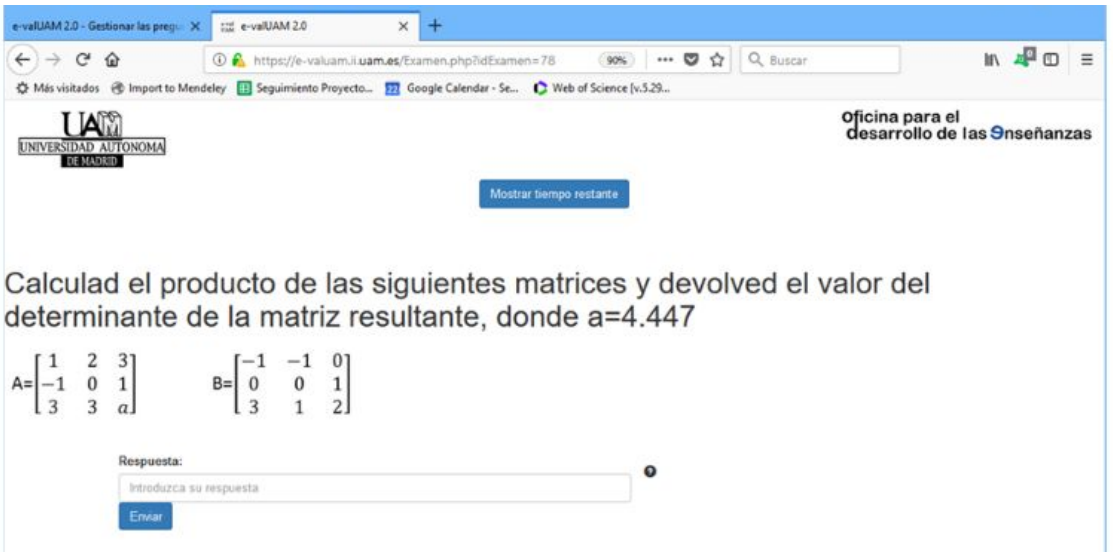


Fig. 2. e-valUAM graphical interface shown to students during a self-assessment test.

3.EXPERIMENT

The experience with real students took place during the 2016/2017 and 2017/2018 academic years. 97 and 73 students were involved respectively in the first year elective course "Applied Computer Science". This subject was taught through theory classes and practical classes in the laboratory. This course corresponds to 6 ECTS and belongs to the Chemical Engineering Degree in the Faculty of Sciences of the Autonomous University of Madrid (UAM).

Figure 3 shows the grades of students in the 2016/2017 academic year as a function of the number of attempts made with the application during their learning phase. The figure also shows the line between a final passing grade and an insufficient grade. The attempts made, shown in this figure, correspond to completed attempts, i.e. where all questions have been answered and a score has been obtained. Students often made incomplete attempts which can be either simple accesses to the application without answering any questions or attempts in which several questions have been answered without reaching the end of the exercise. For the sake of clarity in the study, it has been decided to eliminate all unfinished attempts.

In the figure it can be seen how the grades increase for students with a higher number of attempts. This is the expected result and demonstrates the usefulness of the tool developed. As shown in the figure, it takes only 10 attempts for all students to pass the final exam. In order to better quantify the amount of work required in the subject for a correct assimilation of the contents, Fig. 4 shows the

Funded by the Erasmus+ Programme of the European Union

average scores of students grouped by similar number of attempts. This graph shows that after 16 or more attempts, students' grades do not improve significantly. This data is used to estimate the amount of autonomous work required by the students to assimilate the contents of a subject and to check if this workload corresponds to the assigned ECTS.

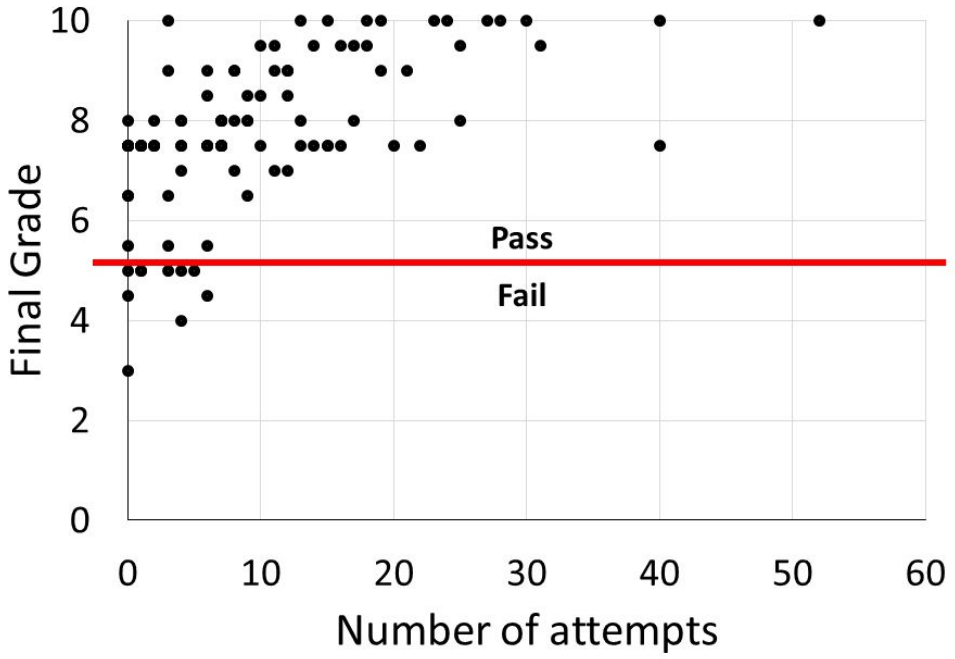


Fig. 3. Final grade ranked by students and number of attempts made during the learning and self-assessment stage. Students have been ordered from the one who tried the least to the one who tried the most.

Funded by the Erasmus+ Programme of the European Union

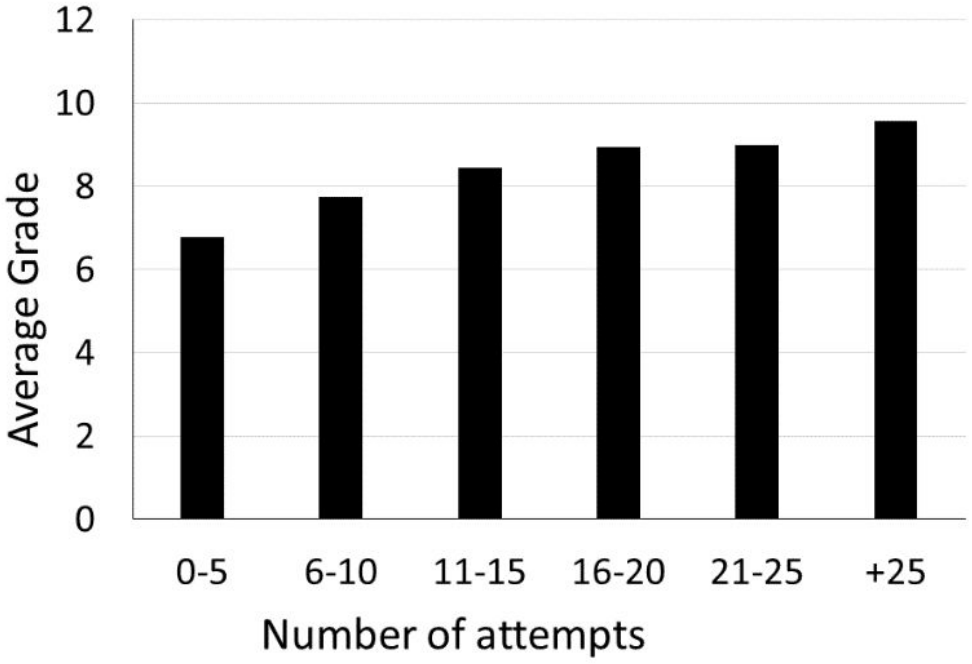


Fig. 4. Average grade on the students' final exam. The x-axis shows the students grouped by the number of complete attempts they made with the application during the learning phase.

Another study that can be done with the application and that would serve as a follow-up tool during the course is the number of times that students use the application during the teaching period. Fig. 5 shows the tests completed by the students in two consecutive courses from the beginning of the course to a specific date. The first course to be analysed is the 2016/2017 course, for which full details are available as it is a course that has already been completed. The second course, at the time of writing this document, still had 4 hours of face-to-face classes to teach, in addition to the ordinary and extraordinary exams. When the results of Fig. 5 are compared for both courses, it is found that there are no significant differences in the amount of autonomous study carried out by the students, which is an indication of the good progress of the course.

Funded by the Erasmus+ Programme of the European Union

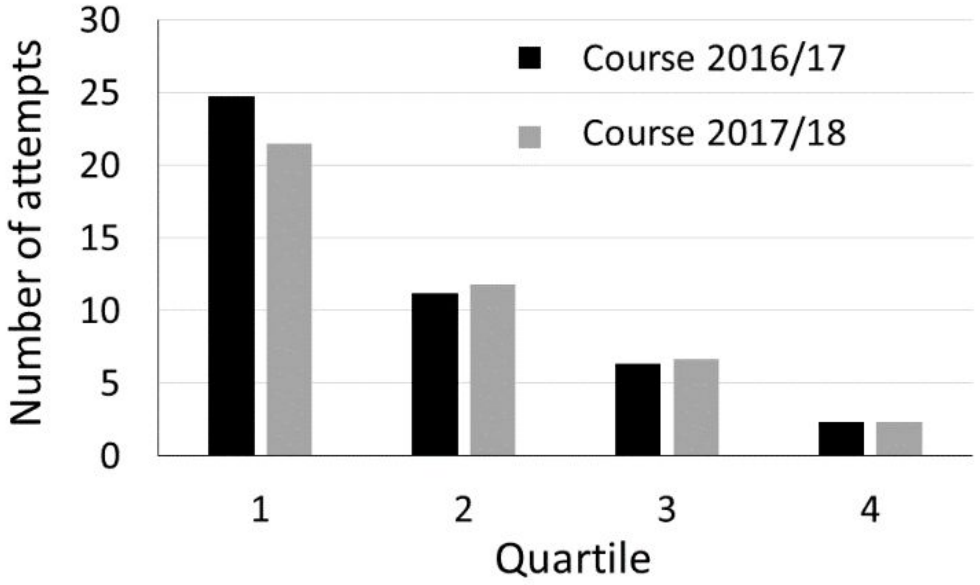


Fig. 5. Comparison of the number of student attempts between two consecutive courses. The students have been grouped into quartiles, where the grouping criterion has been the number of attempts made.

It is important to emphasize that this comparison and these conclusions can only be drawn after obtaining the results of a course that has been generally successful, as is the case of the 2016/2017 course in the subject we are analysing. Such studies cannot be carried out during the first application of this system in a particular subject. On the other hand, Fig. 5 shows the results of the students grouped in quartiles, where the 25% of students who have taken the most tests have been gathered and so on. This study could also be carried out for individual students compared to the rest of their peers and detect or correct any type of deficiency in their learning process.

4.CONCLUSIONS

In this document we show a method of developing adaptive tests that generate questions with numerical solutions in a dynamic way, not repeating the same values between different executions of the application. To ask these questions, the application has a specific format for writing the statements and a programming code that is responsible for generating the answers for each set of numerical values of the statement. Additionally, the questions are divided into levels that are introduced using an adaptive test model in which the questions are displayed from the lowest to the highest levels depending on the number of correct answers given by the student.

Funded by the Erasmus+ Programme of the European Union

The system developed allows the same tool to be used during the learning phase and the final assessment phase. This greatly facilitates feedback between the students and the faculty, which is a requirement in current assessment models.

A study has been carried out in which the test has been developed and used by real students during the 2016/2017 and 2017/2018 academic years. This study demonstrates the effectiveness of the model and its potential to predict student workload together with the correct use of assessment tools dynamically throughout the course.

REFERENCES

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

A. McAuley, B. Stewart, D. Cormier, and G. Siemens, "In the Open: The MOOC model for digital practice", SSHRC Application, Knowledge Synthesis for the Digital Economy, 2010.

A. Kumar, "Using online tutors for learning - what do students think? Proceedings of Frontiers in Education Conference (FIE 2004)", IEEE, pp. 524–28, 2004.

C. Bravo, WR. van Joolingen and T. de Jong, "Using co-lab to build system dynamics models: students' actions and on-line tutorial advice". Comput Educ, pp. 243–51, 2009.

A. Sai. Sabitha, D. Mehrotra and A. Bansal, "An ensemble approach in converging contents of LMS and KMS". Education and information technologies vol. 22 pp. 1673-1694, 2017.

F. Jurado, M. Redondo and M. Ortega. "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, 2012.

JJ. Castro-Schez, M.A. Redondo and F. Jurado, "Experience applying language processing techniques to develop educational software that allow active learning methodologies by advising students", Journal of Network and Computer Applications, vol. 41(1), pp. 65-79, 2014. DOI: 10.1016/j.jnca.2013.10.011

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

Funded by the Erasmus+ Programme of the European Union

S. Narciss, S. Sosnovsky, L. Schnaubert, E. Andres, A. Eichelmann, G. Goguadze and E. Melis. "Exploring feedback and student characteristics relevant for personalizing feedback strategies", Computer and Education, vol. 71, pp. 56-76, 2014.

DOI: 10.1016/j.compedu.2013.09.011

C.A. Chapelle and D. Douglas, 2006, "Assessing Language Through Computer Technology", Cambridge University Press, 2006.

A. Ortigosa, P. Paredes and P. Rodriguez, "AH-questionnaire: An adaptive hierarchical questionnaire for learning styles", Computers & Education, pp.999-1005, May 2010.

S. Klinkenberg, M. Straatemeier, H.L.J. Van der Maas, "Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation". Computers & Education, vol. 57, 2 pp. 1813-1824, Sept.2011.

P. Molins-Ruano, S. Atrio, P. Rodríguez and S. Gómez-Moñivas, "Modelling experts' behavior with e-valUAM to measure computer science skills". Computers in Human Behavior, vol. 61, pp. 378-385, 2016.

CM. Chen and LJ. Duh, "Personalized web-based tutoring system based on fuzzy item response theory", Expert systems with applications, vol. 34, pp. 2298-2315, 2008.

P. Molins-Ruano, C. González-Sacristán, F. Díez, P. Rodriguez and G. M. Sacha. "An Adaptive Model for Computer-Assisted Assessment in Programming Skills". International Journal of Engineering Education. vol 31, pp. 764, 2015.

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

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

M. Antal, and S. Koncz, "Student modeling for a web-based self-assessment system". Expert Systems with Applications, vol. 38, 6, pp.6492-6497, 2011.

Funded by the Erasmus+ Programme of the European Union

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

F.A, Dorça, L. V. Lima, M.A Fernandes and C.R. Lopes, "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, vol. 19, 1, pp. 43-58, March, 2013.

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

M. McGlohe and HH. Chang, "Combining computer adaptive testing technology with cognitively diagnostic assessment", Behavior Research Methods, vol.40, pp. 808-821, nº3, 2008.

Available in e-valuam.ii.uam.es

Funded by the Erasmus+ Programme of the European Union

A22, A23 A24: Design and development of data analyser for e-learning application.

AUTHORS: UAM and EURECAT

SUMMARY

- This work presents a tutoring system which uses three different granularities for helping students to classify animals from bone fragments in zooarchaeology.
- The development of this tutoring system establishes the procedures and methodologies for developing data analyzers for e-Learning application.
- These tasks are also complemented by a pilot experience for demonstrating the usefulness of a prototype that uses the methodology proposed.
- The 3406 bone remains, which have 64 attributes, were obtained from the excavation of the Middle Palaeolithic site of El Salt (Alicante, Spain). The coarse granularity performs a five-class prediction, the medium a twelve-class prediction, and the fine a fifteen-class prediction. In the coarse granularity, the results show that the first 10 most relevant attributes for classification are width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z.
- A user-friendly interface of the tutor has been built in order to train archaeology students to classify new remains using the coarse granularity.
- A pilot has been performed in the 2019 excavation season in Abric del Pastor (Alicante, Spain), where the automatic tutoring system was used by students to classify 51 new remains.
- The pilot experience demonstrated the usefulness of the tutoring system both for students when facing their first classification activities and also for seniors since the tutoring system gives them valuable clues for helping in difficult classification problems.
- As mentioned before the complexity of this algorithm requires supervision at this moment and it is available only on those conditions. For that reason, we have reported all the details about the application here.

Funded by the Erasmus+ Programme of the European Union

1. Introduction

The use of technology opens new frontiers in learning and improves data mining from different sources in order to improve students' learning processes [1]. One of the biggest challenges of including communication technologies in learning is the way in which interaction between teachers and students is simulated by automatic methods [2,3]. In this context, tutoring and the possibility of turning an automatic system into an effective instrument for counseling students are stirring a great amount of interest in many areas [4,5]. The interest in this concept is demonstrated by the many initiatives related to intelligent tutoring systems (ITS). For example, Graesser et al. [6] incorporated natural language processing (NLP) to the tutor in order to enhance the student's learning process. In other studies, authors used gamification techniques to enhance the student's learning process [7,8] or integrate affect sensors [9]. Other more recent ITS guided students to choose their learning path according to their profile [10].

The use of ITS in archaeology is a very interesting topic because of the great amount of data that is available for analysis. In this discipline, data analysis has been performed by many statistical methods [11], including artificial intelligence (AI) in both supervised and unsupervised forms [12–14]. In [15], authors focused on computer vision and gave a visual analysis in archaeology. Random forest [16] has also been used as a supervised learning method to classify archaeological remains. There have also been initiatives [17] which use artificial neural networks in the archaeological domain claiming that they are suitable for the sparse data commonly found in this discipline. Augmented reality and serious games have also been used as tools for improving learning in related subjects such as history or archaeology [18,19]. Virtual reconstructions of ancient cities have also been developed in many different formats [20,21].

Unfortunately, there is a lack of references of ITS in more technical tasks that could help students in more specialized techniques, such as zooarchaeology, which is the topic included in this work. In archaeological sites, there are many different specialized tasks that must be performed by well-trained students [22]. In the case of zooarchaeology, the identification and characterization of animals or species from bone fragments is a crucial task that requires the analysis of many parameters obtained from the characteristics of the remains [16]. First, students require good training in the taxonomical and anatomical characteristics. Later, they must identify the animals and all the information related to their use by humans (trapping, butchering process, discard, technological use, etc.). This data analysis is a great effort and an ITS would be of great help. Moreover, implementing and defining an effective protocol for implementing an intelligent tutor in bone fragment characterization will be also useful in other similar disciplines that also need the analysis of huge amounts of data in archaeology such as the identification of constructive styles [23]

Funded by the Erasmus+ Programme of the European Union

or reconstruction of ceramics from small fragments [24].

In this work, we show the steps for creating an ITS, that have been developed from the application of AI classification algorithms to a data set that includes bone fragments correctly identified in the archaeological site of El Salt. A comparison between different granularities and algorithms is also shown. By choosing the most effective combination, we have developed a user-friendly interface of the ITS with the objective of testing it in the 2019 season in the archaeological site. Finally, we show the results and conclusions of the pilot experience in the archaeological site during August 2019.

This work is structured as follows: Section 2 explains the materials and methods used in the study, Section 3 presents the results, Section 4 discusses the results, and finally Section 5 details the conclusions and future work.

2. Materials and Methods

This section is divided into materials, where the dataset of the study is explained, and methods, where the following methodology to build the tutoring system and to evaluate it is explained.

a. Materials

The used dataset had 3406 instances of archaeological remains and 64 attributes (the last one is the predicted class) extracted from faunal assemblage of Stratigraphic Unit Xa of El Salt (Figures 1 and 2) [25–27]. The excavation of "El Salt" can be seen in Figure 1 and examples of the archaeological remains of "El Salt" can be seen in Figures 2 and 3.

The Salt is one of the fundamental sites of the Middle Palaeolithic in the western Mediterranean region due to the entity of its archaeological sequence, for its extraordinary state of conservation, which affects even organic matter and for the integrated and multidisciplinary nature of the research process which takes place in it, with the confluence of different Hispanic, European, and North American Universities and Research Centers. This enclave has been located at the head of the Serpis River, in the locality of Alcoy (Alicante, Spain). It is presented as an outdoor site of about 300 m^2 located at the foot of a large wall that rises up to 38 m tall. The space inhabited at the foot of the wall was protected by a large roof, which at times of maximum development came to serve as protection for almost the entire surface. Highlights are the strategic condition of its location, in the middle of various biotopes of plain, mountains, river valley, and lake-palm environment, etc. and immersed in a territory of the mountains of Alicante, very rich in diversified resources.

Research at the El Salt site attempts to deepen the knowledge of its paleolithic record, from an

Funded by the Erasmus+ Programme of the European Union

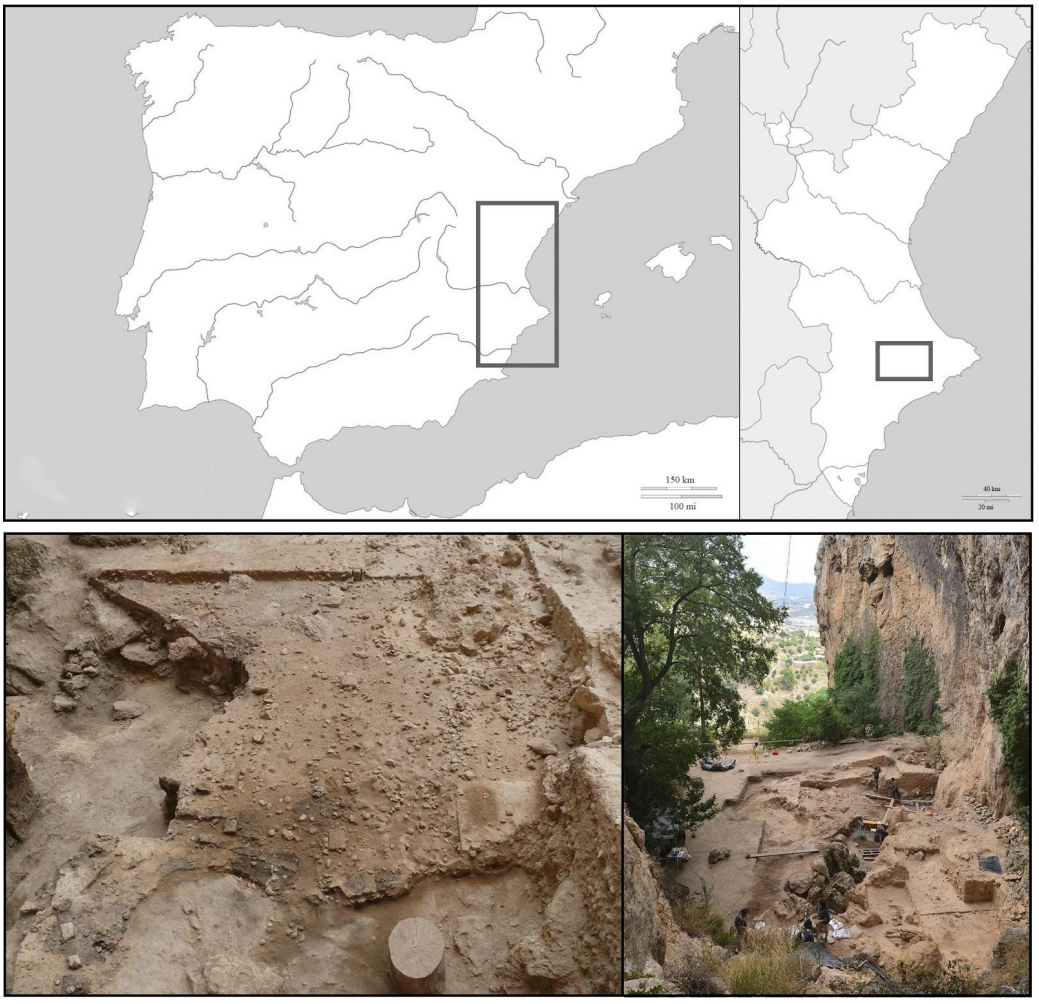


Figure SEQ Figure * ARABIC 1 Geographical location of "El Salt", site overview and current excavation surface.

integrative and multidisciplinary perspective. This record does not stop growing by virtue of the application of increasingly sophisticated excavation techniques and high resolution analytical procedures. To the traditional material record of this type of deposits, consisting of lithic remains, fauna, anthracological material, etc. Currently, a microscopic and even molecular registry is being added that is decisively contributing to enrich the information and improve its quality. These facts give us the opportunity of obtaining a complete and accurate data-set in zooarchaeology, among others, which is a key step for training classification algorithms as the ones needed for developing the ITS presented here.

Funded by the Erasmus+ Programme of the European Union

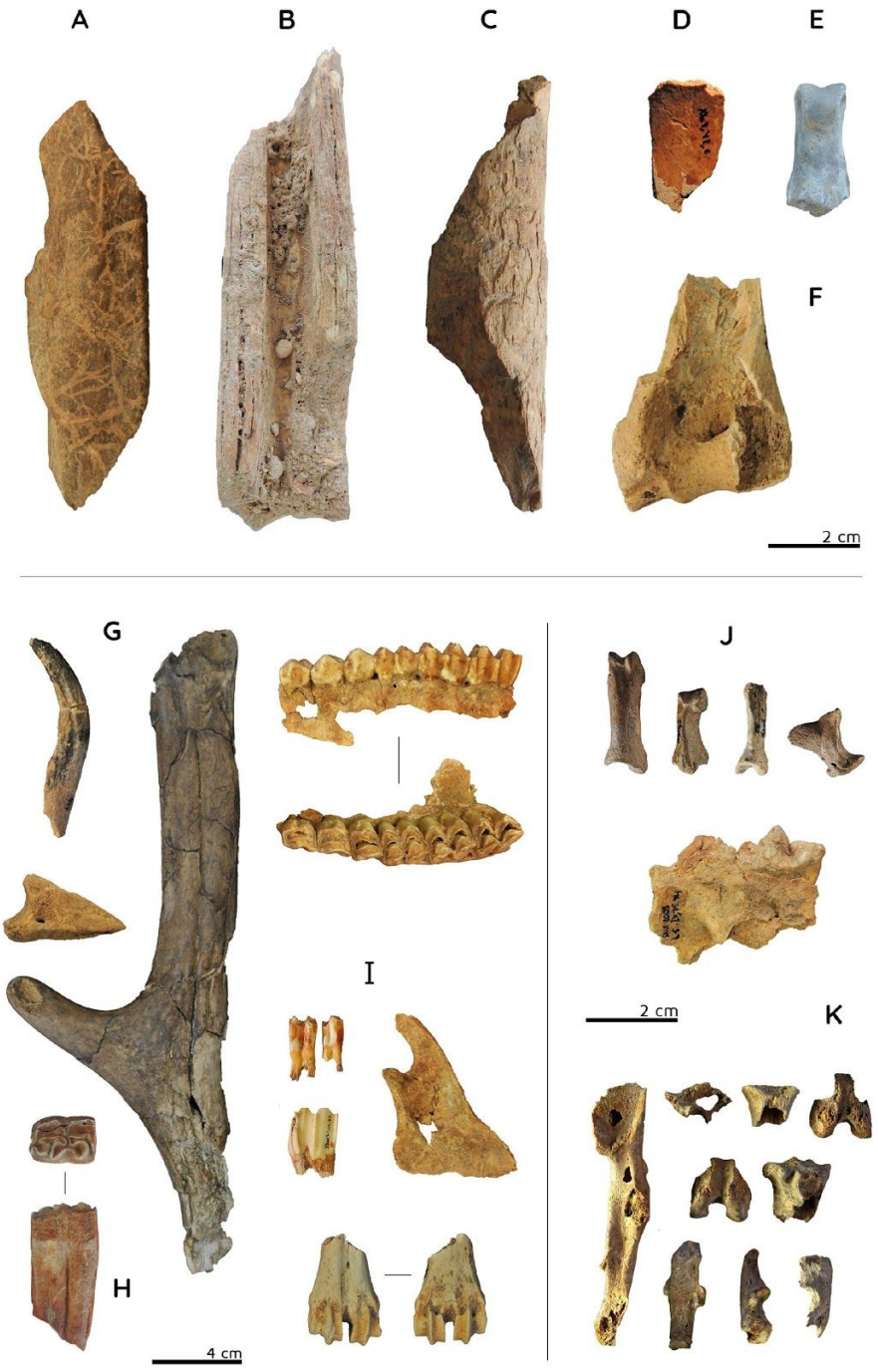


Figure 2 Taphonomic damage and bone remains recovered examples: manganese (A), concreteness (B), root-marks (C), bone flake (D), burned bone (E), erosion (F), bones, teeth and deer antler (G), bones and wild goat teeth (H), horse tooth (I), phalanges and lynx maxillary (J), and rabbit bones (K).

Funded by the Erasmus+ Programme of the European Union

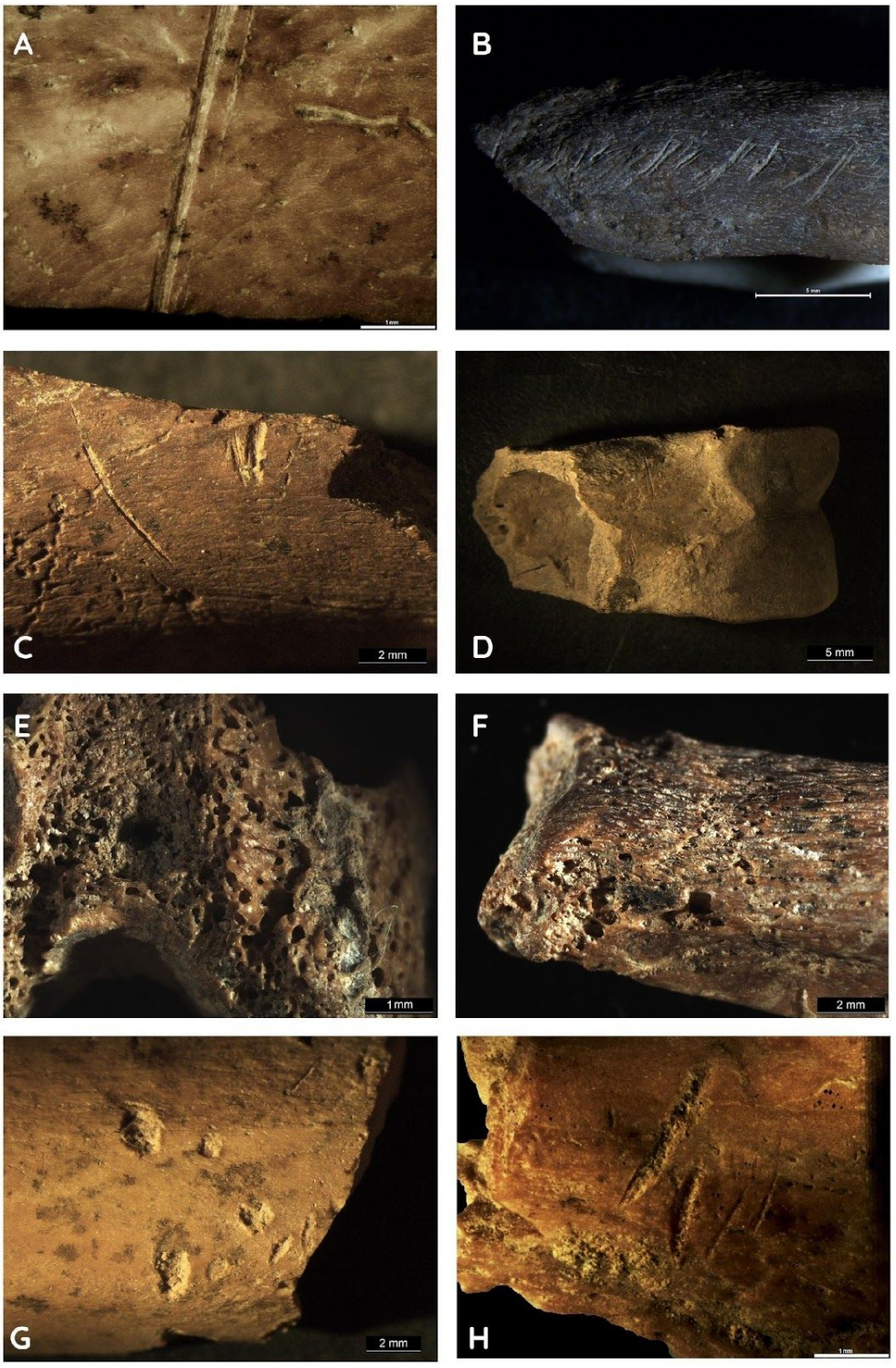


Figure 3 Bioestratinomic damage examples. There are anthropogenic marks: slicing and scraping marks (A–D), Predator non-anthropogenic marks: digestion (E,F), punctures (G), and scores (H).

Funded by the Erasmus+ Programme of the European Union

In Figure 2, we can observe taphonomic damage and bone remains recovered examples. In Figure 3, we can observe bioestratinomic damage examples.

In Table 1, a descriptive analysis of the sample is shown. There are two columns in the table. In the first one, we have listed the most relevant parameters (according to the classification algorithm with best performance in the coarse, medium, and fine granularities, which will be explained in the following section) and the predicted class 'Family' (taxonomic rank between Order and Genus) both in coarse, medium, and fine granularities. Following the name of each parameter, we have included a description of the possible values that it can take. For instance, "Bone" can take the listed 93 unique values, while the anatomical group can take only 11 values. It is worth noting that, in the case of the prediction classes, the coarse granularity can take five values, the medium granularity 12, and the fine granularity 15. The second column of the table briefly describes the attribute. If the attribute is numeric, count (number of not null values), mean, standard deviation (std), minimum (min), first quartile (Q1), second quartile (Q2), third quartile (Q3), and maximum (max) are given. If the attribute is categorical, count (number of not null values), unique (number of categories), top (top category), and frequency (freq, which is the frequency of the top category) are given. In the categorical granularity families, the number of each category is given because it is relevant for the predicting classifier that will be explained in Section 2.2. The figure after the table shows an example of the available data.

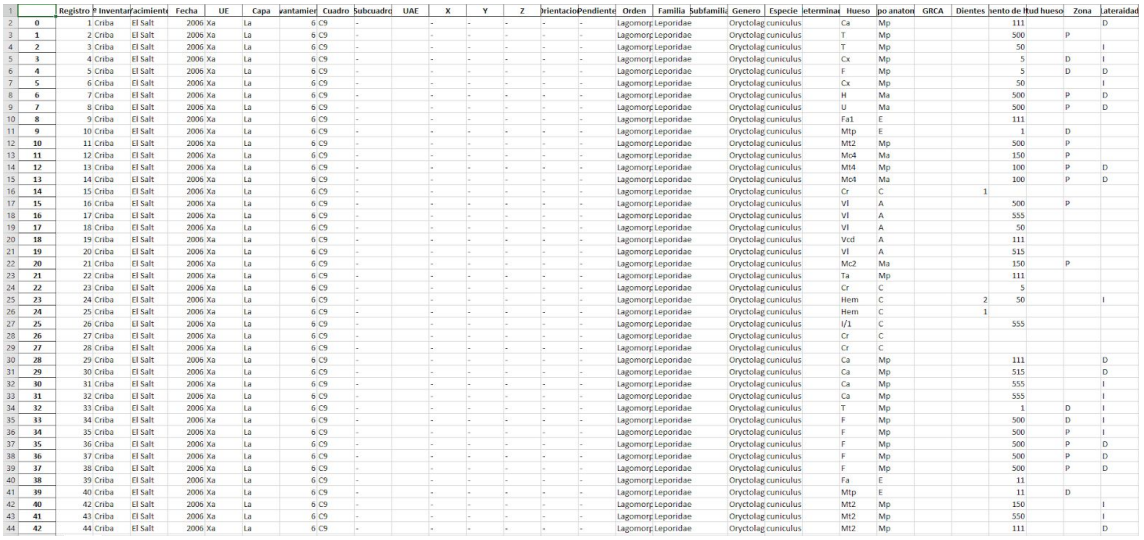
Funded by the Erasmus+ Programme of the European Union

Attribute	Description
Width	Count: 2036, Mean: 14.822829, Std: 8.973862, Min: 0.99, Q1: 8.457500, Q2: 13.43, Q3: 19.32, Max: 71.01
Bone: Ca, T, Cx, F, H, U, Fa1, Mtp, Mt2, Mc4, Mt4, Cr, Vl, Vcd, Mc2, Ta, Hem, I/1, Fa, Mt3, Vc, I1/, Mx, As, Pa, R, Es, Fa2, Fa3, Mt5, nan, Mc5, Ct, Vs, Vt, Mc3, V, Das, Dai, Da, Mt, M3/, M1/, Hy, Mc, P2/, P4/-M1/, M/1, Lt2, Art, Cc, PT, M2/, In, M/P, M/1-M/2, P4/, M3, M1/-M2/, P/2, I, M/X, Pl, P/M, R/, Co, L, Lt1, Lt3, P/4, P/3-P/4, Se, dp2/, P3/, M/x, P4/-P3/, M/3, PTPer8, PTPle, PTPL, P/3, M/, M/2, M, Fa1-V, Fa2-II, Fa2-IV, I/C, Fi, Asta, I/3, d/3, Mc2-3, I/2	Count: 3403, Unique: 93, Top: T, Freq: 374
Thickness	Count: 714, Mean: 4.887451, Std: 2.873843, Min: 0.460000, Q1: 3.072500, Q2: 4.595000, Q3: 6.4, Max: 16.52
Length	Count: 2492, Mean: 36.25859, Std: 28.28524, Min: 2.42000, Q1: 15.94500, Q2: 27.20000, Q3: 47.80000, Max: 284.96000
Bone fragment	Count: 3242, Mean: 189.635102, Std: 204.909877, Min: 1, Q1: 50, Q2: 111, Q3: 500, Max: 555
Anatomical group: Mp, Ma, E, C, A, nan, In, ES, Cr, PT, PTPL, PTPl	Count: 3399, Unique: 11, Top: Mp, Freq: 1275
Long bone circumference	Count: 1378, Mean: 2.526851, Std: 1.107021, Min: 1, Q1: 2, Q2: 2, Q3: 4, Max: 5
X	Count: 1.187000×10^3, Mean: 1.864528×10^3, Std: 6.292228×10^4, Min: 0, Q1: 8.519905, Q2: 2.900000×10^1, Q3: 6.775000×10^1, Max: 2.167891×10^6
Y	Count: 1186, Mean: 1272.859123, Std: 30223.002072, Min: 0, Q1: 11.809257, Q2: 24.500000, Q3: 63, Max: 803982

Funded by the Erasmus+ Programme of the European Union

Table 1. *Cont.*

Attribute	Description
Z	Count: 1232, Mean: 403.047662, Std: 9767.169421, Min: −143.8, Q1: −0.156130, Q2: 166.2, Q3: 192.6, Max: 342938
Manganeso	Count: 1416, Mean: 1.592514, Std: 0.835553, Min: 1, Q1: 1, Q2: 1, Q3: 2, Max: 5
Concretion	Count: 1177, Mean: 1.806287, Std: 1.033386, Min: 1, Q1: 1, Q2: 1, Q3: 2, Max: 5
Coarse granularity family: Bovidae, Cervidae, Equidae, Leporidae, unknown	Bovidae: 420 Cervidae: 516, Equidae: 240, Leporidae: 2164, Unknown: 66, Total: 3406
Medium granularity family: Bovidae, Canidae, Cervidae, Corvidae, Equidae, Felidae, Leporidae, Phasianidae, Rhinocerotidae, Suidae, Testudinidae, unknown	Bovidae: 420 Canidae: 8, Cervidae: 516, Corvidae: 5, Equidae: 240, Felidae: 24, Leporidae: 2164, Phasianidae: 7, Rhinocerotidae: 2, Suidae: 7, Testudinidae: 9, Unknown: 4, Total: 3406
Fine granularity family: Anatidae, Bovidae, Bufonidae, Canidae, Cervidae, Corvidae, Equidae, Erinaceidae, Felidae, Leporidae, Phasianidae, Rhinocerotidae, Suidae, Testudinidae, Ursidae	Anatidae: 1 Bovidae: 420, Bufonidae: 1, Canidae: 8, Cervidae: 516, Corvidae: 5, Equidae: 240, Erinaceidae: 1, Felidae: 24, Leporidae: 2164, Phasianidae: 7, Rhinocerotidae: 2, Suidae: 7, Testudinidae: 9, Ursidae: 1, Total: 3406



		Registro	Inventar	acimient	Fecha	UE	Capa	antamie	Cuadro	Subcuadro	UAE	X	Y	Z	rientacio	Pendiente	Orden	Familia	Subfamili	Genero	Especie	eterminа	Hueso	po anator	GRCA	Dientes	ento de	ttud hueso	Zona	ateraidad
2	0	1	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ca	Mp			111			D
3	1	2	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		T	Mp			500		P	
4	2	3	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		T	Mp			50			I
5	3	4	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cx	Mp			5		D	I
6	4	5	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			5		D	D
7	5	6	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cx	Mp			50			I
8	6	7	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		H	Ma			500		P	D
9	7	8	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		U	Ma			500		P	D
10	8	9	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Fa1	E			111			
11	9	10	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mtp	E			1		D	
12	10	11	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mt2	Mp			500		P	
13	11	12	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mc4	Ma			150		P	
14	12	13	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mt4	Mp			100		P	D
15	13	14	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mc4	Ma			100		P	D
16	14	15	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cr	C		1				
17	15	16	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Vl	A			500		P	
18	16	17	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Vl	A			555			
19	17	18	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Vl	A			50			
20	18	19	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Vcd	A			111			
21	19	20	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Vl	A			515			
22	20	21	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mc2	Ma			150		P	
23	21	22	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ta	Mp			111			
24	22	23	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cr	C			5			
25	23	24	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Hem	C		2	50			I
26	24	25	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Hem	C		1				
27	25	26	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		I/1	C			555			
28	26	27	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cr	C						
29	27	28	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Cr	C						
30	28	29	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ca	Mp			111			D
31	29	30	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ca	Mp			515			D
32	30	31	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ca	Mp			555			I
33	31	32	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Ca	Mp			555			I
34	32	33	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		T	Mp			1		D	I
35	33	34	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			500		D	I
36	34	35	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			500		P	I
37	35	36	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			500		P	D
38	36	37	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			500		P	D
39	37	38	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		F	Mp			500		P	D
40	38	39	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Fa	E			11			
41	39	40	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mtp	E			11		D	
42	40	42	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mt2	Mp			150			I
43	41	43	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mt2	Mp			550			I
44	42	44	Criba	El Salt	2006	Xa	La	6	C9	-		-	-	-	-	-	Lagomorp	Leporidae		Oryctolag	cuniculus		Mt2	Mp			111			D

Figure 4 Screenshot of the available archaeological data

2.2 Methods

We used 33% of the remains for testing and 67% for training. A pipeline with a standard scaler with a SMOTE (synthetic minority over-sampling technique) or ADASYN (adaptive synthetic) method was used, together with 10-fold cross-validation. The parameter that we want to maximize is f1-score_macro. We have chosen it because F1-score is the harmonic mean of precision and recall, and the macro option calculates metrics for each label, and finds their unweighted mean, and this

Funded by the Erasmus+ Programme of the European Union

does not take label imbalance into account. The used algorithms with their parameters have been:

- Random forest with parameters:
  - classifier n_estimators (number of trees in the forest): [100, 500, 700]
  - classifier max_features (number of features to consider when looking for the best split): ['auto', 'sqrt', 'log2']
  - classifier class_weight (weights associated with classes; if not given, all classes are supposed to have weight one; the balanced mode uses the values to automatically adjust weights inversely proportional to class frequencies in the input data): ['balanced',None]

Random forest (rf) classifiers are a collection of decision trees, each of which is trained on a random subset of the training data and only allowed to use some random subset of the features. There is no coordination in the randomization (a particular data point or feature could randomly get plugged into all the trees, none of the trees, or anything in between). The final classification score for a point is the average of the scores from all the trees. The hope is that the different trees will pick up on different patterns and each one will only give confident guesses when its pattern is present. When it comes time to classify a point, several of the trees will classify it correctly and strongly while the other trees give answers that are on the fence, meaning that the overall classifier slouches toward the right answer. The individual trees in a random forest are subject to overfitting, but they tend to be randomly overfitted in different ways. These largely cancel each other out, yielding a robust classifier.

Random forests follow the following steps. Repeat: Take a bootstrap sample from the data, fit a classification regression tree. At each node: (1) Select m variables at random out of all M possible variables (independently at each node) (2) Find the best split on the selected m variables and (3) Grow the trees big.

We can combine it by voting (classification) or averaging (regression). Random forest have several properties: same idea for regression and classification; handle

categorical prediction naturally; quick to fit, even for large problems; no formal distributional assumptions; automatically fits highly non-linear interactions; automatic variable selection; handle missing values through proximities; not very easy to interpret if the tree is small; the terminal nodes does not suggest a natural clustering; the picture cannot give valuable insight into which variable are important and where.

- Support vector machine (SVM) with parameters:
  - classifier kernel (specifies the kernel type to be used in the algorithm): ['linear','rbf']

Funded by the Erasmus+ Programme of the European Union

  - classifier gamma (kernel coefficient for 'rbf', 'poly', and 'sigmoid'): $[1 \times 10^{-3}, 1 \times 10^{-4}]$
  - classifier C (penalty parameter C of the error term): [1, 10, 100]
- Naive bayes.
- Neural networks with parameters:
  - classifier_ solver (the solver for weight optimization. 'lbfgs' is an optimizer in the family of quasi-Newton methods. 'sgd' refers to stochastic gradient descent. 'adam' refers to a stochastic gradient-based optimizer): ['lbfgs', 'sgd', 'adam']
  - classifier alpha (L2 penalty (regularization term) parameter): $[1 \times 10^{-4}, 1 \times 10^{-5}]$
- k-nearest neighbors (KNN) with parameters:
  - classifier n_neighbors (number of neighbors to use): [3, 5, 7, 9]

More information about the algorithms can be found in the literature [28]. These algorithms have been implemented using the Python library scikit-learn and the ITS has been implemented using Flask. The code of the main parts class of the Tutor is the following:

Form:

Funded by the Erasmus+ Programme of the European Union

```
from flask import Flask, render_template, flash, request
from wtforms import Form, TextField, TextAreaField, validators, StringField, SubmitField

# App config.
DEBUG = True
app = Flask(__name__)
app.config.from_object(__name__)
app.config['SECRET_KEY'] = '7d441f27d441f27567d441f2b6176a'

class ReusableForm(Form):
    name = TextField('Name:', validators=[validators.required()])
    @app.route("/", methods=['GET', 'POST'])
    def hello():
        form = ReusableForm(request.form)
        print form.errors
        if request.method == 'POST':
            name=request.form['name']
            surname=request.form['surname']
            i=1
            unarray=np.zeros(72)
            while i < 73:
                if (request.form['at'+str(i)] not in ''):
                    unarray[i-1]=float(request.form['at'+str(i)])
                i=i+1
            print name
            print surname
            i=1
            while i<73:
                print(unarray[i-1])
                i=i+1
            unarray = unarray.reshape(1, -1)
            prediction=clf.predict(unarray)
            print("Prediction: ",prediction[0])
            finalprediction=['Bovidae','Cervidae','Equidae','Leporidae', 'unknown']
            print(prediction)
            surname=surname +". Your prediction is: "+str(finalprediction[prediction[0]])+ "."

        if form.validate():
    # Save the comment here.
            flash('Thank you ' + name +" "+ surname)
        else:
            flash('Warning: The name and surname form fields are required. ')
        return render_template('hello.html', form=form)
```

Figure 5 Code of the form

Preparation of the dataset:

Funded by the Erasmus+ Programme of the European Union

```
# Prepare the dataset
data_frame = pd.read_excel('output.xlsx')
#SC
data_frame = data_frame.replace('-', np.NaN)
print(data_frame.columns)
#data_frame = data_frame.drop(['Fecha', 'Registro', 'Registro Diente', 'Dientes'], axis=1)
data_frame = data_frame.drop(['Fecha', 'Registro'], axis=1)
column_to_int(["Hueso", "Grupo anatomico", "Color 1", "Color 2"], data_frame)

# Dataset just with numerical columns
numerics = ['int16', 'int32', 'int64', 'float16', 'float32', 'float64']
data_frame_num = data_frame.select_dtypes(include=numerics)
data_frame_num = data_frame_num.fillna(0)  # Replace Nan values with zero
#print(data_frame_num.columns)  # Debug to check the numeric columns

# Change class labels for classes with less than 100 members to 'unknown'
familias = ['Anatidae', 'Bufonidae', 'Canidae', 'Corvidae', 'Erinaceidae', 'Felidae',
            'Phasianidae', 'Rhinocerotidae', 'Suidae', 'Testudinidae', 'Ursidae']
artiodactyla = ['Bovidae', 'Cervidae', 'Equidae']
labels_familia = (data_frame[data_frame['Familia'].notnull()].Familia).as_matrix()
# print(np.unique(label_familia_matrix), np.unique(labels_familia))
mask_unknown = np.isin(labels_familia, familias)
labels_unknow = labels_familia[mask_unknown]
labels_familia[mask_unknown] = 'unknown'
mask_artiodactyla = np.isin(labels_familia, artiodactyla)
labels_artiodactyla = labels_familia[mask_artiodactyla]

# Generate label matrix
le_familia = preprocessing.LabelEncoder()
label_familia_matrix = le_familia.fit_transform(labels_familia)
le_artiodactyla = preprocessing.LabelEncoder()
label_artiodactyla_matrix = le_artiodactyla.fit_transform(labels_artiodactyla)
# Debug purpose
print(sorted(collections.Counter(labels_familia).items()))
print(np.unique(label_familia_matrix), np.unique(labels_familia))

# Generate data matrix
train_val_familia_df = data_frame_num[data_frame['Familia'].notnull()]
train_val_familia_matrix = train_val_familia_df.as_matrix()
# Generate data matrix for the second classifier
train_val_artiodactyla_df = train_val_familia_df[mask_artiodactyla]
train_val_artiodactyla_matrix = train_val_artiodactyla_df.as_matrix()
print(np.shape(train_val_artiodactyla_matrix))

# Generate train and validation dataset for the main classifier
X_train_familia_aux, X_test_familia_aux, y_train_familia_aux, y_test_familia_aux = train_test_split(train_val_familia_matrix,
                                                                                                    label_familia_matrix,
                                                                                                    test_size=0.33,
                                                                                                    random_state=42)

# Generate train and validation dataset for the second classifier
X_train_artiodactyla_aux, X_test_artiodactyla_aux, y_train_artiodactyla_aux, y_test_artiodactyla_aux = train_test_split(train_val
                                                                                                    label_artiodactyla_matrix,
                                                                                                    test_size=0.33,
                                                                                                    random_state=42)
```

Figure 6 Code of the preparation of the dataset

Random Forest algorithm:

Funded by the Erasmus+ Programme of the European Union

```
#Random Forest
from sklearn.tree import DecisionTreeClassifier, export_graphviz
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
import itertools
from sklearn import tree
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
from sklearn.metrics import classification_report
from sklearn.ensemble import RandomForestClassifier as RF
from sklearn.metrics import accuracy_score
from imblearn.over_sampling import SMOTE
from imblearn.pipeline import Pipeline
from sklearn import preprocessing

#rfc = RandomForestClassifier(n_jobs=-1,max_features= 'sqrt' ,n_estimators=2, oob_score = True)

rf_pipeline = Pipeline([
                        ('std', preprocessing.StandardScaler(copy=True, with_mean=True, with_std=True)),
                        ('smote', SMOTE(n_jobs=-1, random_state=42,kind='regular')),
                        ('classifier', RF())
                       ])
#pipeline = Pipeline([
#     ('smote', SMOTE(n_jobs=-1, random_state=42,kind='regular')),
#     ('normal',StandardScaler()),
#     ('clf',RandomForestClassifier(n_jobs=-1,max_features= 'sqrt' ,n_estimators=2, oob_score = True))
#])

# Set the parameters by cross-validation
#'min_samples_leaf': [1,5,10,50,100,200,500],

tuned_parameters = {
    'classifier__n_estimators': [500],
    'classifier__max_features': ['auto'],
    'classifier__class_weight':[None]
}
#{'classifier__max_features': 'auto', 'classifier__class_weight': None, 'classifier__n_estimators': 700
#tuned_parameters = {
#    }

#scores = ['precision', 'recall', 'f1']
#scores = [ 'f1_weighted','f1_micro','f1_macro','accuracy','roc_auc']
scores=['f1_macro']

for score in scores:
    print("# Tuning hyper-parameters for %s" % score)
    print()
    #clf = GridSearchCV(GaussianNB(), tuned_parameters, cv=10,scoring='%s_macro' % score)
    #clf = GridSearchCV(estimator=rfc, param_grid=tuned_parameters, cv=5,scoring='%s_weighted' % score,n_jobs= -1)
    clf = GridSearchCV(estimator=rf_pipeline , param_grid=tuned_parameters, cv=10,scoring='%s' % score,n_jobs= -1)
    clf.fit(X_train_familia_aux, y_train_familia_aux)
#app.run(debug=False)
app.run(host= '172.20.49.55', port=5000, debug=False)
```

Figure 7

Code of the Random Forest algorithm

Funded by the Erasmus+ Programme of the European Union

Fragment of HTML code:

```
<title>Intelligent Tutor System</title>
            <link rel="stylesheet" media="screen" href="static/bootstrap.min.css">
            <link rel="stylesheet" href="static/bootstrap-theme.min.css">
        <meta name="viewport" content="width=device-width, initial-scale=1.0">
<div class="container">
<img src="static/Logo.jpg" alt="Logo" width="500">
<h2>Intelligent Tutor System for archeological sites</h2>

{% with messages = get_flashed_messages(with_categories=true) %}
{% if messages %}

{% for message in messages %}
{% if "Error" not in message[1]: %}
<div class="alert alert-info">
            {{ message[1] }}</div>
{% endif %}

{% if "Error" in message[1]: %}
<div class="alert alert-warning">
            {{ message[1] }}</div>
{% endif %}
{% endfor %}
{% endif %}
{% endwith %}


<form action="" method="post" role="form">
    {{ form.csrf }}
<div class="form-group">
<label for="name">Name*:</label>
<input type="text" class="form-control" id="name" name="name" placeholder="What's your name?"></div>
 <label for="surnamename">Surname*:</label>
<input type="text" class="form-control" id="surname" name="surname" placeholder="What's your surname?"></div>
<p style='margin-left: 13.5em;margin-right: 13.5em'>
<label for="at1">X</label>
<input type="text" class="form-control" id="at1" name="at1" placeholder="Insert value"></div>
<label for="at2">Y</label>
<input type="text" class="form-control" id="at2" name="at2" placeholder="Insert value"></div>
<label for="at3">Z</label>
<input type="text" class="form-control" id="at3" name="at3" placeholder="Insert value"></div>
<label for="at4">Hueso: {0:'Ca', 1:'T', 2:'Cx', 3:'F', 4:'H', 5:'U', 6:'Fal', 7:'Mtp', 8:'Mt2', 9:'Mc4', 10:'Mt4', 11:'Cr', 12:'V1', 13:'Vod', 14:'Mc2', 15:'Ta', 16:'Hem', 17:'I/1', 18:'Fa', 19:'Mt3', 20:'Vo', 21:'I1/', 22:'Mx',
<input type="text" class="form-control" id="at4" name="at4" placeholder="Insert value"></div>
<label for="at5">Grupo anatomico: {0:'Mp', 1:'Ma', 2:'E', 3:'C', 4:'A', 5: nan, 6:'In', 7:'ES', 8:'Cr', 9:'PT', 10:'PTPL', 11:'PTP1'}</label>
<input type="text" class="form-control" id="at5" name="at5" placeholder="Insert value"></div>
<label for="at6">GRCA</label>
<input type="text" class="form-control" id="at6" name="at6" placeholder="Insert value"></div>
<label for="at7">Dientes</label>
<input type="text" class="form-control" id="at7" name="at7" placeholder="Insert value"></div>
<label for="at8">Fragmento de hueso</label>
<input type="text" class="form-control" id="at8" name="at8" placeholder="Insert value"></div>
<label for="at9">Longitud hueso largo</label>
<input type="text" class="form-control" id="at9" name="at9" placeholder="Insert value"></div>
<label for="at10">Circunferencia hueso largo</label>
<input type="text" class="form-control" id="at10" name="at10" placeholder="Insert value"></div>
<label for="at11">Longitud</label>
<input type="text" class="form-control" id="at11" name="at11" placeholder="Insert value"></div>
<label for="at12">Anchura</label>
<input type="text" class="form-control" id="at12" name="at12" placeholder="Insert value"></div>
<label for="at13">Espesor</label>
<input type="text" class="form-control" id="at13" name="at13" placeholder="Insert value"></div>
```

Figure 8 Fragment of the HTML code

Regarding the criteria used to distinguish the different taphonomic alterations, zooarchaeological and taphonomic analysis are performed using already established standard methods [29,30]. Faunal remains are taxonomically and anatomically identified whenever possible, while non-identified specimens with insufficient information were classified into three bone categories (long, flat, or articular) and associated with a weight-size category based on bone density, circumference, and thickness of the cortical surface: large-sized >300 kg, medium-sized 100–300 kg, small-sized 5–100 kg, and very small-sized <5 kg [31–33]. On the other hand, all remains were analysed taphonomically using macroscopic and microscopic techniques to identify biostratinomic and diagenetic modifications. In the fracture analysis, we classified all of the fragments by fracture type (recent, old fresh, old dry, or indeterminate) following the criteria established by [34] and the morphotypes created by Real [35,36]. All dimensions were measured for each bone fragment. Also, bone surface modifications were observed and quantified to identify different damage caused by anthropogenic activity (percussion, butchering marks, or thermal alteration) or predator action (tooth marks, or digestion), as well as the diverse diagenetic processes that produce alterations on bone surfaces as erosion, sediment concreteness, roots marks, weathering, pigmentation, and trampling [37–44]. Damage caused after the prey's death and previously their deposit and sedimentation, the

Funded by the Erasmus+ Programme of the European Union

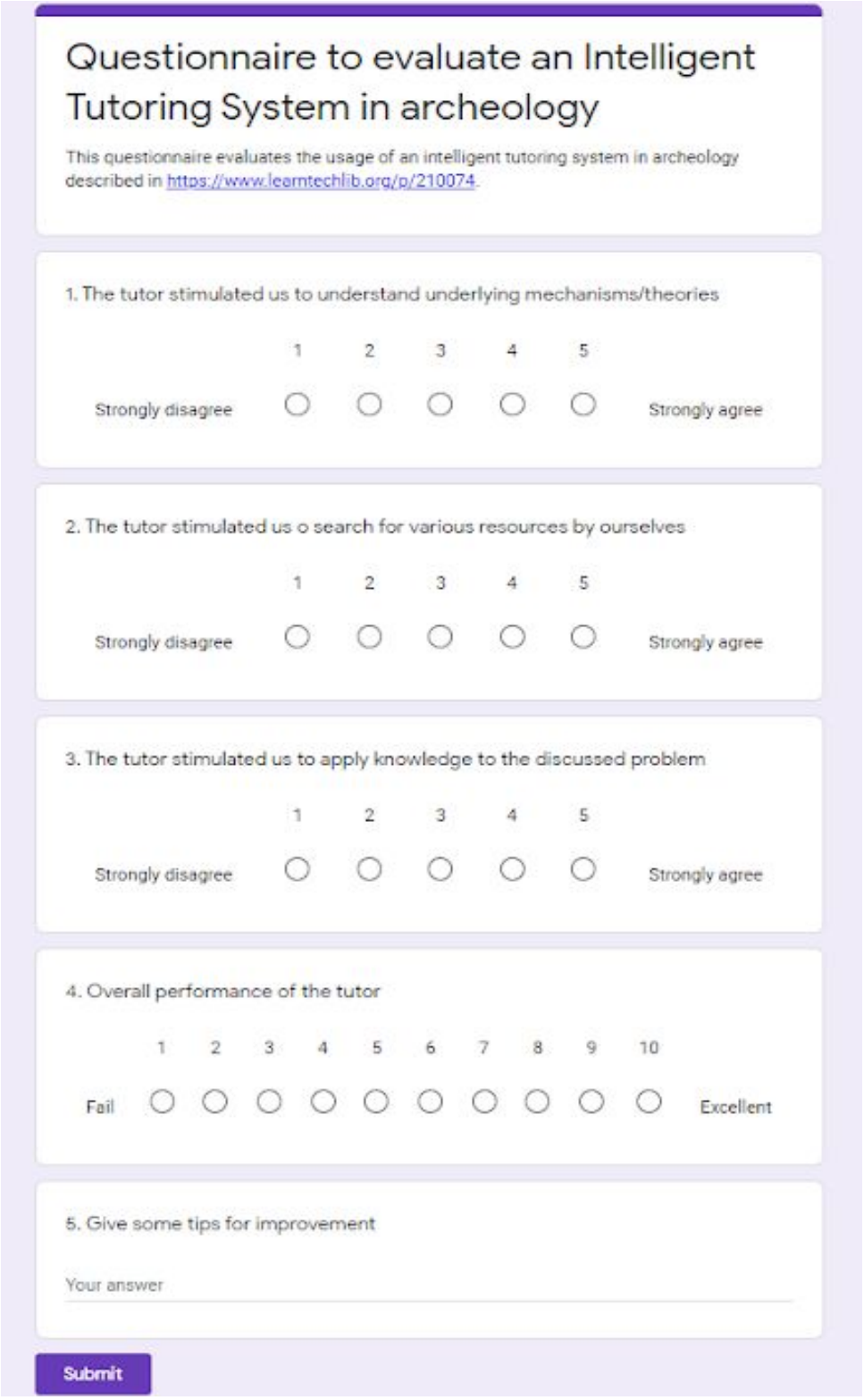
biostratinomic process, was classified and analysed individually in a specific database according to the type of damage, origin, agent, location, morphology, distribution, direction, intensity, quantity, and dimensions. Diagenetic processes were recorded based on their presence and the degree of alteration. In the specific case of butchering marks, different activities were established based on the ethnoarchaeological literature [37,45–47].

The questionnaire used to test the experience of the tutor in the pilot of El Salt 2019 season followed partially the methodology explained in [48]. The following areas were tested:

1. Constructive/active learning: The tutor stimulated us to understand underlying mechanisms/theories.
2. Self-directed learning: The tutor stimulated us to search for various resources by ourselves.
3. Contextual learning: The tutor stimulated us to apply knowledge to the discussed problem.
4. Global score: Overall performance of the tutor.
5. Open answer: Give some tips for improvement.

A rating between 1 (strongly disagree) and 5 (strongly agree) was given to questions 1–4.

Funded by the Erasmus+ Programme of the European Union



Questionnaire to evaluate an Intelligent Tutoring System in archeology

This questionnaire evaluates the usage of an intelligent tutoring system in archeology described in https://www.learntechlib.org/p/210074.

1. The tutor stimulated us to understand underlying mechanisms/theories

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

2. The tutor stimulated us o search for various resources by ourselves

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

3. The tutor stimulated us to apply knowledge to the discussed problem

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

4. Overall performance of the tutor

	1	2	3	4	5	6	7	8	9	10	
Fail	○	○	○	○	○	○	○	○	○	○	Excellent

5. Give some tips for improvement

Your answer

Submit

Funded by the Erasmus+ Programme of the European Union

Figure 9 Evaluation questionnaire

3. Results

Results are divided into the experiments for the prediction of animals in the coarse, medium and fine granularities, and the ITS development and use. The performance of the different methods in the coarse granularity can be seen in Table 2, and the confusion matrices of the coarse, medium, and fine granularities using the random forest method can be seen in Figures 10–13. In the fine granularity, classes with less than three instances do not appear because a stratified split between training and test has been applied.

In the three granularities, the 10 more relevant attributes for the classification have been (by this order):

- Coarse granularity: width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z.
- Medium granularity: bone, width, length, anatomic group, bone fragment, Y, X, Z, thickness and manganese.
- Fine granularity: width, bone, length, bone fragment, anatomic group, Z, Y, X, thickness and concretion.

The best configuration parameters for the random forest were:

- Coarse granularity: 'classifier max_features': 'auto', 'classifier class_weight': 'balanced', 'classifier n_estimators': 500.
- Medium granularity: 'classifier max_features': 'sqrt', 'classifier class_weight': 'balanced', 'classifier n_estimators': 700.
- Fine granularity: 'classifier max_features': 'auto', 'classifier_ class_weight': 'balanced', 'classifier n_estimators': 700

Funded by the Erasmus+ Programme of the European Union

Table 2 Performance of the different methods in the coarse granularity. Synthetic minority

Method (Parameters)	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)
Random forest, SMOTE (classifier class weight: balanced, classifier max features: auto, classifier n estimators: 500)	0.86	0.86	0.86	0.86
SVM, SMOTE (classifier C: 100, classifier gamma: 0.001, classifier kernel: rbf)	0.74	0.81	0.74	0.77
Naive Bayes, SMOTE	0.68	0.76	0.68	0.66
Neural Networks, SMOTE (classifier solver: lbfgs, classifier alpha: 1×10^{-5})	0.67	0.75	0.67	0.71
KNN, SMOTE (classifier n neighbors: 3)	0.75	0.79	0.75	0.77
Random Forest, ADASYN (classifier class weight: balanced, classifier max features: auto, classifier n estimators: 100)	0.86	0.85	0.86	0.86
SVM, ADASYN (classifier C: 100, classifier gamma: 0.001, classifier kernel: rbf)	0.72	0.80	0.72	0.75
Naive Bayes, ADASYN	0.68	0.78	0.68	0.65
Neural Networks, ADASYN (classifier solver: adam, classifier alpha: 1×10^{-5})	0.67	0.79	0.67	0.72
KNN, ADASYN (classifier n neighbors: 3)	0.74	0.79	0.74	0.76

In Figure 13, we show the interface of the ITS implemented for being used in the season. As we can see in the interface, there is a first region where the user must be identified. Right below this first section, the user should start filling the different characteristics and parameters that have been obtained from the bone fragment. The ITS also gives information about the different options of a parameter when applicable. This is very important since the ITS could be used by archaeologists from different sites, and we must be sure that they agree in the format of the data.

In Figure 14, we show the output screen of the ITS right after a user has finished the characterization of a fragment and decides to ask for a prediction. In this case, the prediction is shown at the beginning of the interface. As soon as the prediction is done, the ITS is ready for another prediction. For this reason, the fields below the prediction do not have any value. It is worth noting that all the information filled by the user has been stored in the ITS server and is available for users upon request. Figure 14 is an example of a table used by students in the archaeological site to compare their own predictions and the ITS answers. We have censored information about the fragment identification and the identity of the students.

Funded by the Erasmus+ Programme of the European Union

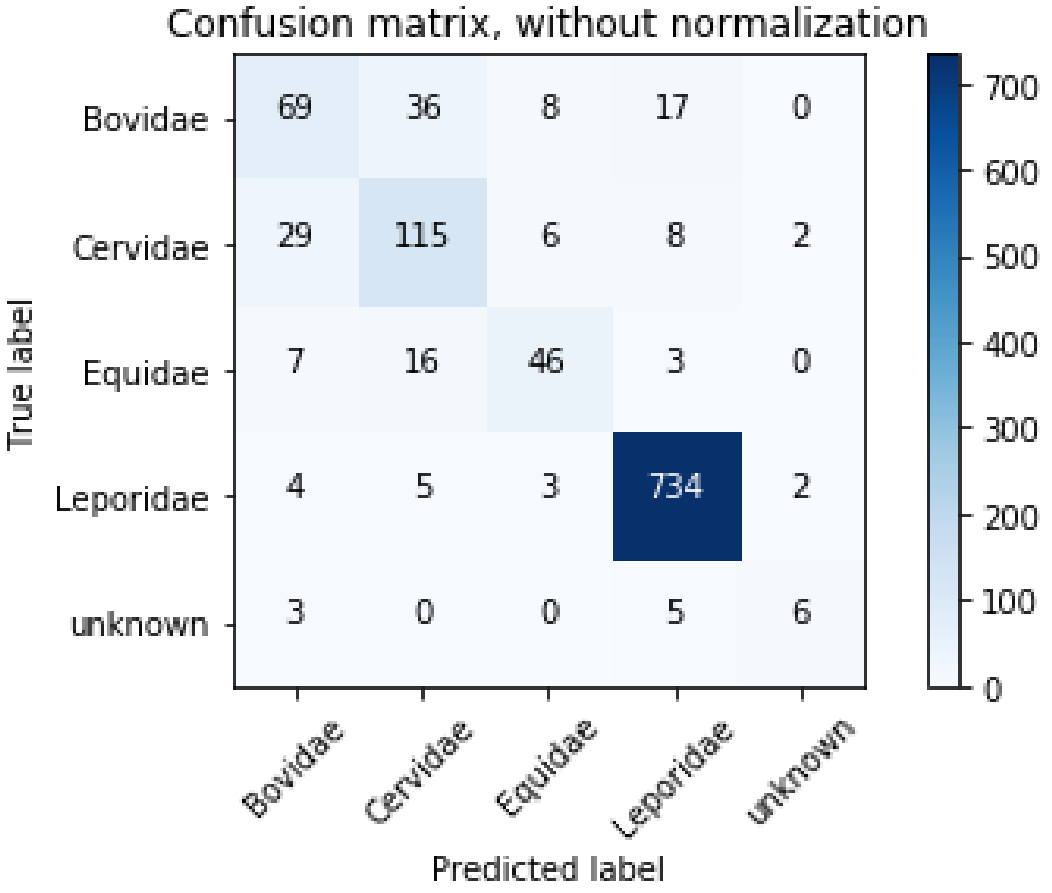


Figure 10 Confusion matrix of the coarse granularity.

Funded by the Erasmus+ Programme of the European Union

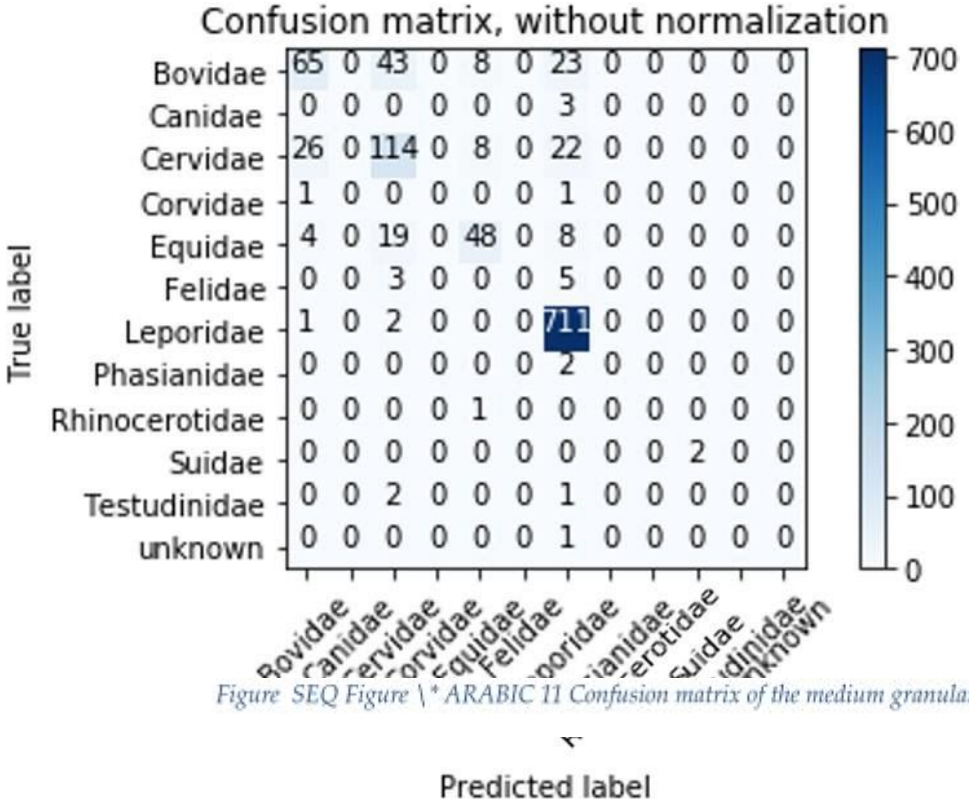


*Figure SEQ Figure * ARABIC 11 Confusion matrix of the medium granularity.*

Funded by the Erasmus+ Programme of the European Union

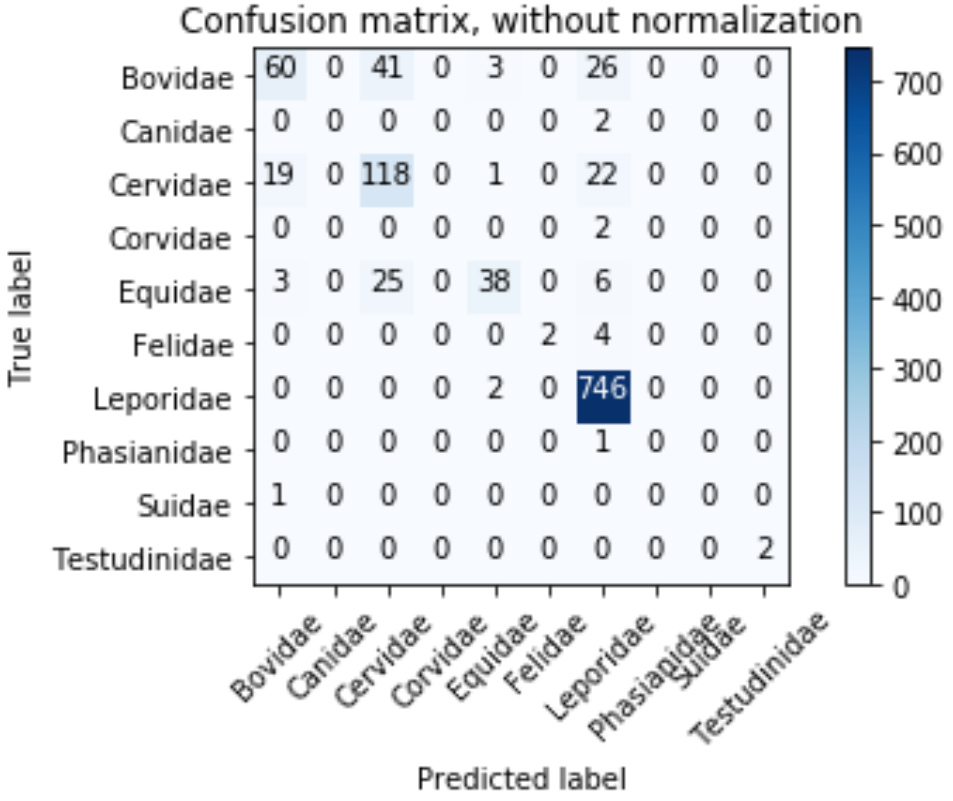


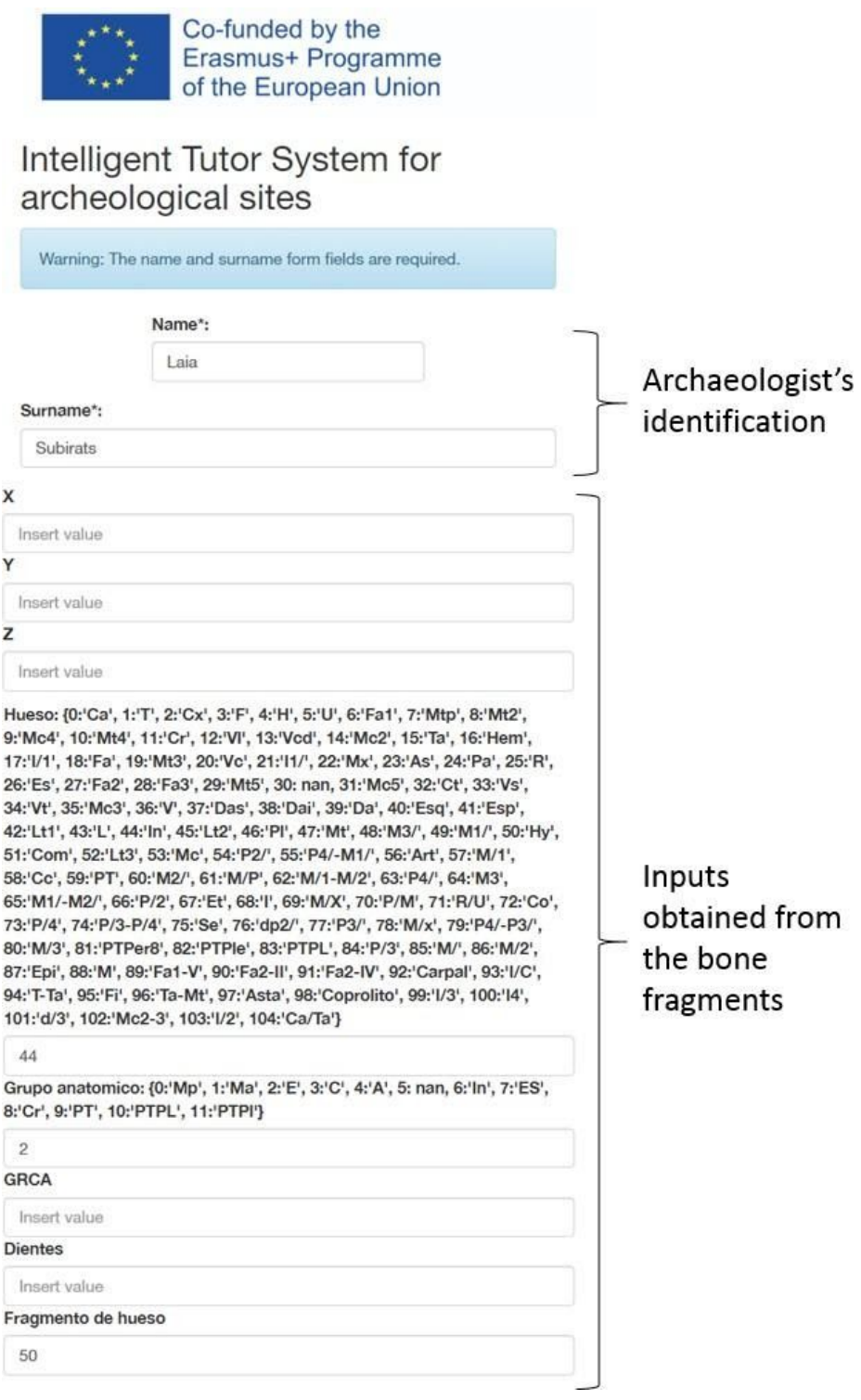
Figure 12 Confusion matrix of the fine granularity.

This first version of the ITS included the coarse granularity and was used in August 2019 by 3 students to classify 51 remains that were characterized following the format of the same database but belonged to a different archaeological site called "Abric del Pastor". Abric del Pastor belongs also to Middle Palaeolithic and is near "El Salt". Two of the three students also answered the questionnaire. The results of the tutor, compared to the predictions of the students frequently agreed even when the remains used for training corresponded to a different site ("El Salt"). Students, due to their limited knowledge and the difficulties of analyzing very fragmented bone remains, usually describe the remains based only on the size, which usually agree with the families predicted by the tutor. For example, a lot of bone remains predicted as "small size" are described by the tutor as Bovidae, which is a family included in the category of small animals because most bovids belong to the wild goat (*Capra pyrenaica*). This information can help students to give more accurate answers or verify their initial guess. On the other side, sometimes the tutor fails and not all the sizes and families are correctly classified. For example, some bones that are correctly described as "small size" by students are characterized as Leporidae by the tutor, which is a mistake. Other mistakes are related to the

Funded by the Erasmus+ Programme of the European Union

wrong characterization of birds as Leporidae.

Funded by the
Erasmus+ Programme
of the European Union



Funded by the Erasmus+ Programme of the European Union

Figure 13 Tutor interface to introduce the bone fragments.

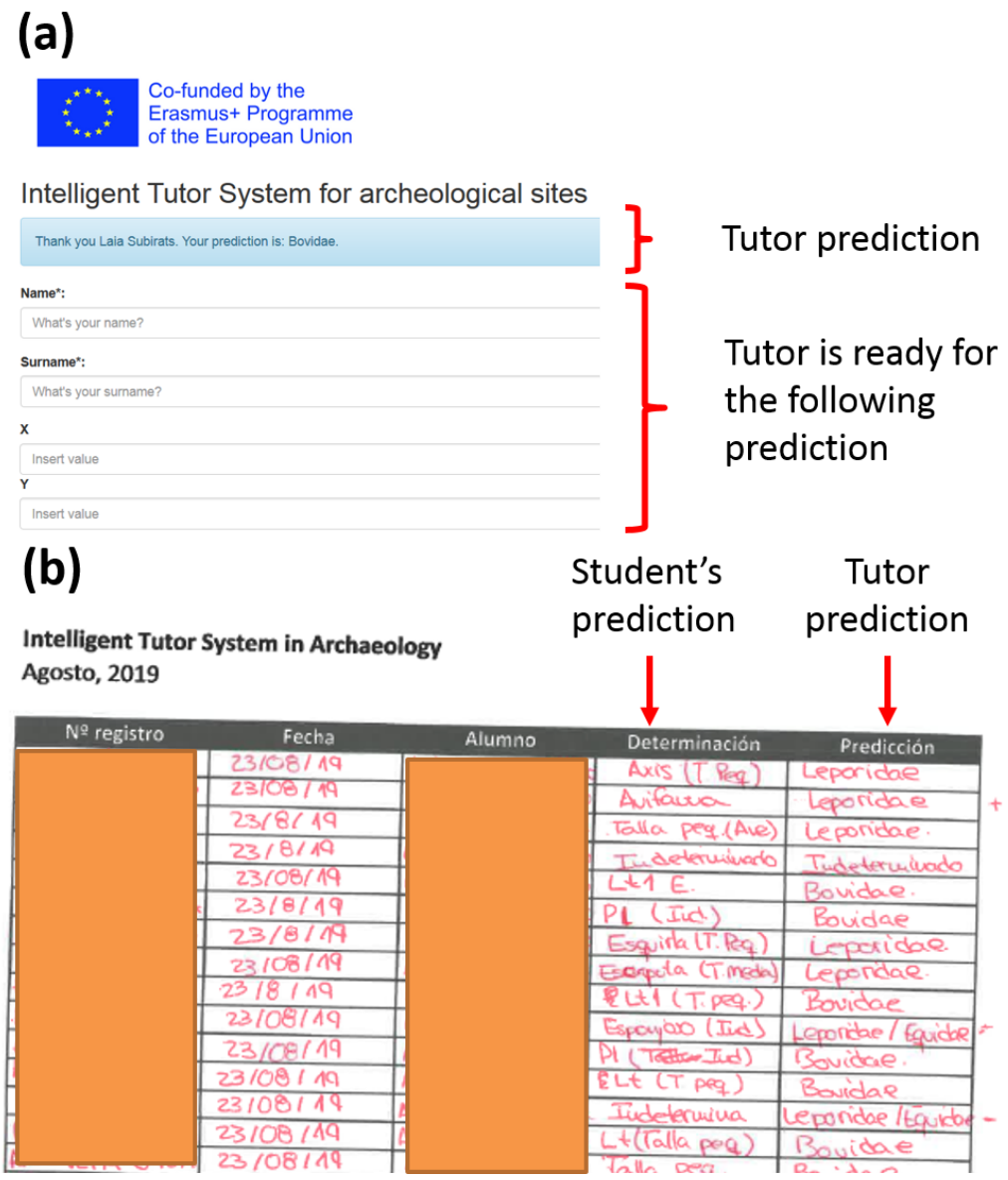


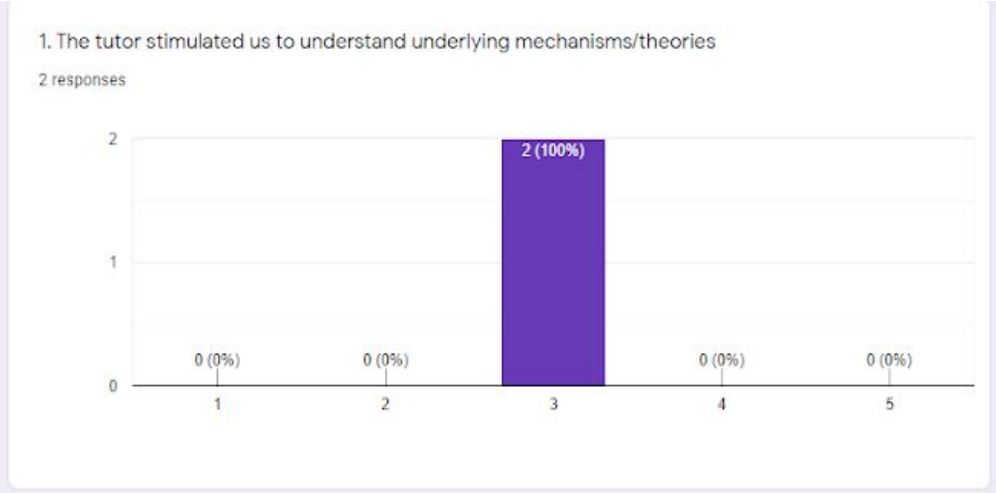
Figure 14 Tutor prediction (a) and student's/tutor prediction table (b).

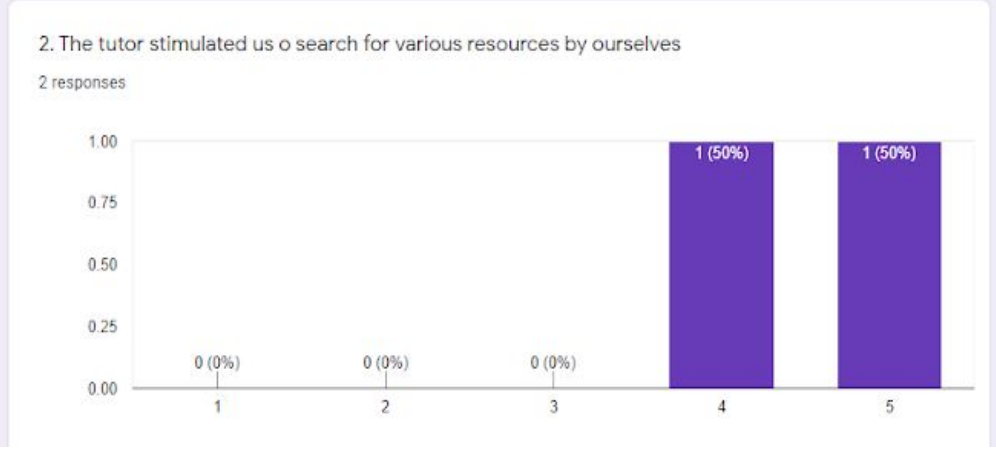
Funded by the Erasmus+ Programme of the European Union

Analyzing the results of the questionnaire, students appreciated the use of the tutor as a complementary tool in their activities. In the subjective open answer, they highlighted two concepts:

(1) The need of improving the interface and
(2) the need of increasing the accuracy of the predictions.

In the following Figures you can see the detail of their answers:





Funded by the Erasmus+ Programme of the European Union

Figure 15 Students' answers detail

4. Discussion

Random forests using SMOTE is the best machine learning method for this domain. In Table 3 the F1-score of the different animals is displayed. We can see that coarse and fine granularities have the best F1-scores (0.86 and 0.85 respectively). However, the fine granularity is more suitable because it has more granularity and it classifies better small animals.

Let us now analyze the feedback obtained from the archaeological students and teachers. From the pedagogical point of view, and analyzing their answers, we can see that the ITS moderately stimulates underlying mechanisms but, on the other hand, it strongly stimulates the search for resources by the users. There are a few different opinions about ITS's ability to stimulate the application of knowledge related to the discussed problem, but the overall performance of the ITS is

Funded by the Erasmus+ Programme of the European Union

quite high. Finally, the final tips for improvement were to improve the interface and to have more data in order to have a more accurate prediction.

Although predictions given by the ITS are usually accurate and useful, the presence of some mistakes must be analyzed. There are two ways of improving it. First, the excessive number of characterizations as Leporidae can be related to the granularity used in this version. An increment in the number of classes would increase the accuracy of the prediction by increasing the amount of information in the system. The second way of improving the accuracy of the tutor is taking into account the spatial distribution of the fragments in the sites. Since the X, Y, and Z variables are important in the characterization, using information from different archaeological sites in the training would help in the correct characterization of the tutor by avoiding an excessive importance of the bone distribution in a specific site.

Funded by the Erasmus+ Programme of the European Union

Table 3 F1-score of the random forest in the different granularities.

Animal	Coarse	Medium	Fine
Bovidae	0.57	0.59	0.56
Cervidae	0.69	0.69	0.69
Equidae	0.68	0.63	0.66
Leporidae	0.97	0.94	0.96
Unknown	0.50	0	-
Canidae	-	0	0
Corvidae	-	0	0
Felidae	-	0.22	0.5
Phasianidae	-	0	0
Rhinocerotidae	-	0	0
Suidae	-	1	0
Testudinidae	-	1	1
Anatidae	-	-	0
Bufonidae	-	-	0
Erinaceidae	-	-	0
Ursidae	-	-	0
Total weighted	0.86	0.83	0.85

5. Conclusions

In this work [49] we have presented a method to create intelligent tutoring systems in archaeology to help students in specialized tasks that require analysis of huge amounts of data. The method proposed here implies the application of classification algorithms that must be trained with a

Funded by the Erasmus+ Programme of the European Union

complete data set in order to give accurate results. We have tested our method by developing an intelligent tutoring system in the field of zooarchaeology. We have found that random forest is the method with the best performance in this domain, and the fine granularity approach is the more useful approach for archaeologists because the main doubts appear with small animals, which are not common. In the coarse granularity, random forest was the method which had better results between the tested ones, with an accuracy and f1-score of 0.86 for both. As questionnaires to classify the remains are quite long (63 attributes), the first 10 more relevant attributes for the classification have been shown, which are width, bone, thickness, length, bone fragment, anatomical group, long bone circumference, X, Y, and Z. In addition, a classification using a medium and fine granularity has also been performed obtaining an accuracy of 0.84 and 0.86 and a f1-score of 0.83 and 0.85 respectively.

Regarding the pilot with the tutor, it correctly classified the families in most of the cases. However, some mistakes were also found. Mistakes can be partially corrected by changing the granularity since some of the mistakes are related to the reduced amount of classes in the granularity implemented in the pilot experience. The satisfaction questionnaire showed that the tutor was helpful. Students suggested an improvement in the interface and the accuracy of the predictions.

In future work, more data from different archaeological sites can be included in the training phase in order to increase accuracy and avoid effects related to specific characteristics of a single excavation. Also, natural language processing (NLP) and gamification techniques could be added to the tutor in order to enhance the student's learning process.

Glossary

ADASYN	Adaptive Synthetic	AI Artificial intelligence
ITS	Intelligent Tutoring System	
KNN	k-nearest neighbors	
NLP	Natural language processing	

Funded by the Erasmus+ Programme of the European Union

SMOTE Synthetic Minority Over-sampling Technique SVM Support vector machine

References

1. Maldonado-Mahauad, J.; Perez-Sanagustin, M.; Kizilcec, R.; Morales, N.; Munoz-Gama, J. Mining theory-based patterns from Big data: Identifying self regulated learning strategies in Massive Open Online Courses. *Comput. Hum. Behav.* **2018**, *80*, 179–196.
2. González-Sacristán, C.; Molins-Ruano, P.; Díez, F.; Rodríguez, P.; Sacha, G.M. Computer-assisted Assessment with Item Classification for Programming Skills. In Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM '13), Salamanca, Spain, 14–15 November 2013; ACM: New York, NY, USA, 2013; pp. 111–117.
3. Molins-Ruano, P.; González-Sacristán, C.; Díez, F.; Rodriguez, P.; Sacha, G.M. An Adaptive Model for Computer-Assisted Assessment in Programming Skills. *Int. J. Eng. Educ.* **2015**, *31*, 764.
4. Bravo, C.; van Joolingen, W.; de Jong, T. Using co-lab to build system dynamics models: Students' actions and on-line tutorial advice. *Comput. Educ.* **2009**, *53*, 243–251.
5. Sabitha, A.; Mehrotra, D.; Bansal, A. An ensemble approach in converging contents of LMS and KMS.

   Educ. Inf. Technol. **2017**, *22*, 1673–1694.
6. Graesser, A.C.; Chipman, P.; Haynes, B.C.; Olney, A. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Trans. Educ.* **2005**, *48*, 612–618.
7. Baker, R.S.J.; Corbett, A.; Koedinger, K.; Evenson, S.; Roll, I.; Wagner, A.Z.; Raspat, M.N.; Baker, D.J.; Beck, J.E. Adapting to When Students Game an Intelligent Tutoring System. In *Intelligent Tutoring Systems*; Ikeda, M., Ashley, K.D., Chan, T.W., Eds.; ITS 2006; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4053.
8. Long, Y.; Aleven, V. Educational Game and Intelligent Tutoring System: A Classroom Study and Comparative Design Analysis. *ACM Trans. Comput.-Hum. Interact.* **2017**, *24*, 20.
9. D'Mello, S.; Craig, S.; Gholson, B.; Franklin, S.; Picard, R.; Graesser, A. Integrating Affect Sensors in an Intelligent Tutoring System. In Proceedings of the Computer in the Affective Loop Workshop at 2005 International Conference Intelligent User Interfaces, San Diego, CA, USA, 10–13 January 2005; pp. 7–13.
10. Verdú, E.; Regueras, L.; Gal, E. Integration of an intelligent tutoring system in a course of computer network design. *Educ. Technol. Res. Dev.* **2017**, *65*, 653–677.
11. Baxter, M. *Statistics in Archaeology*; Arnold Publ.: London, UK, 2003.
12. Baxter, M. A review of supervised and unsupervised pattern recognition in archaeometry. *Archaeometry*

Funded by the Erasmus+ Programme of the European Union

2006, *48*, 671–694.

13. Barcelo, J.A. *Computational Intelligence in Archaeology*; Information Science Reference: Hershey, PA, USA, 2009.
14. Barcelo, J.A.; Bogdanovic, I. *Mathematics in Archaeology*; CRC Press: Boca Raton, FL, USA, 2015.
15. Barceló, J. Visual Analysis in Archaeology: An Artificial Intelligence Approach. In *Morphometrics for Nonmorphometricians*; Elewa, A.M.T., Ed.; Springer: Berlin, Germany, 2010; pp. 93–156.
16. Subirats, L.; Fort, S.; Hernández, C.; Pérez, L.; Vesisenaho, M.; Nousiainen, T.; Peltonen, M.; Miakush, I.; Sacha, G.M. Intelligent Tutoring System in Archaeology. In Proceedings of the EdMedia + Innovate Learning, Amsterdam, The Netherlands, 24 June 2019; pp. 745–750.
17. Bell, S.; Croson, C. Artificial neural networks as a tool for archaeological data analysis. *Archeometry* **1998**, *40*, 139–151.
18. Ardito, C.; Buono, P.; Costabile, M.; Lanzilotti, R.; Pederson, T. Re-experiencing History in Archaeological Parks by Playing a Mobile Augmented Reality Game. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. OTM 2007. Lecture Notes in Computer Science*; Meersman, R., Tari, Z., Herrero, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4805.
19. Vidal-Matutano, P.; Pérez-Jordà, G.; Hernández, C.M.; Galván, B. Macrobotanical evidence (wood charcoal and seeds) from the Middle Palaeolithic site of El Salt, Eastern Iberia: Palaeoenvironmental data and plant resources catchment areas. *J. Archaeol. Sci. Rep.* **2018**, *19*, 454–464.
20. Atienza-Fuente, J. The monumental thermal complex of the Roman city of Valeria: Proposal for virtual reconstruction of its cold rooms based on the archaeological findings. *Virtual Archaeol. Rev.* **2019**, *10*, 116–131.
21. Ferrari, I.; Quarta, A. The Roman pier of San Cataldo: from archaeological data to 3D reconstruction. *Virtual Archaeol. Rev.* **2019**, *10*, 28–39.
22. Molinero-Polo, M.A.; Pérez-Ruiz, S.; Acebo, A.; Sacha, G.M. Análisis del potencial motivador y docente del uso de las nuevas tecnologías en el aula. Interacción entre el aumento de la motivación y el aprendizaje; Artículo invitado en el monográfico "Cambios metodológicos: Nuevas formas en la enseñanza" de la revista OGE (Organización y Gestión Educativa) del fórum europeo de administradores de la educación. 2016. Volume 3, p. 26. Available online: https://dialnet.unirioja.es/servlet/articulo?codigo=5502970 (accessed on 29 September 2019).
23. Blanquez-Perez, J.; Roldan-Gomez, L. The Urban Reform of the Colonia Libertinorum Carteia

Funded by the Erasmus+ Programme of the European Union

during the Augustan Era. The Basilica Building. *Gerion-Rev. Hist. Antigua* **2017**, *35*, 443–468.
24. Garcia-Vegas, G.; Castelo-Ruano, R.; Lopez-Perez, A. Virtual Reconstructions of Archaeological Heritage. The Convivial Space of the Roman Villa of El Saucedo, Talavera la Nueva, Toledo. *Arqueología Iberoamericana* **2017**, *35*, 3–9.
25. Machado, J.; Pérez, L. Temporal frameworks to approach human behavior concealed in Middle Palaeolithic palimpsests: A high-resolution example from El Salt Stratigraphic Unit X (Alicante, Spain). *Quat. Int.* **2016**, 66–81.
26. Pérez, L.J.; Sanchis, A.; Hernández, C.M.; Galván, B. Paleoecología de macromamíferos aplicada a los conjuntos zooarqueológicos de El Salt y el Abric del Pastor (Alcoy, Alicante). In *Interaccions entre felins i humans. III Jornades d'arqueozoologia*; Museu de Prehistòria de València: Valencia, Spain, 2017; pp. 327–354.
27. Pérez, L. Estrategias de subsistencia y dinámicas de asentamiento en los Valles de Alcoy durante el Paleolítico medio. Análisis zooarqueológico, tafonómico y paleoecológico de la secuencia arqueológica de El Salt (Alcoy, Alicante). Ph.D. Thesis, University Rovira i Virgili, Tarragona, Spain, 2019.
28. Cady, F. *The Data Science Handbook*; Wiley: Hoboken, NJ, USA, 2017.
29. Lyman, R. *Vertebrate Taphonomy, Cambridge Manuals in Archaeology*; Cambridge University Press: Cambridge, UK, 1994.
30. Reitz, E.J.; Wing, E.S. *Zooarchaeology*; Cambridge University Press: Cambridge, UK, 2008.
31. Uerpmann, H.P. Animal bone finds and economic archaeology: A critical study of Osteoarchaeological method. *World Archaeol.* **1973**, *4*, 307–322.
32. Bunn, H.T. Patterns of skeletal representation and hominid subsistence activities at Olduvai gorge, Tanzania, and Koobi fora, Kenya. *J. Hum. Evol.* **1986**, *15*, 673–690.
33. Palomo, L.J.; Gisbert, J.; Blanco, J.C. *Atlas y Libro Rojo de los Mamíferos Terrestres de España*; Dirección General para la Biodiversidad-SECEM-SECEMU: Madrid, Spain, 2007.
34. Villa, P.; Mahieu, E. Breakage patterns of human long bones. *J. Hum. Evol.* **1991**, *21*, 27–48.
35. Real, C. Aproximación metodológica y nuevos datos sobre los conjuntos arqueozoológicos del Magdaleniense superior de la Cova de les Cendres. *Archivo De Prehistoria Levantina* **2012**, *29*, 99–120.
36. Real, C. Estudio arqueozoológico y tafonómico del Magdaleniense de la Cova de les Cendres (Teulada-Moraira, Alicante). Doctoral Dissertation, Universitat de València, Valencia, Spain, 2016.
37. Binford, L.R. *Bones: Ancient Men and Modern Myths, Studies in Archaeology*; Academic Press: New York, NY, USA, 1981.
38. Shipman, P. *Life History of a Fossil. An Introduction to Taphonomy and Paleoecology*; Harvard University Press: Boston, MA, USA, 1981.
39. Shipman, P.; Rose, J. Early hominid hunting, butchering, and carcass-processing behaviors:

Funded by the Erasmus+ Programme of the European Union

Approaches to the fossil record. *J. Anthropol. Archaeol.* **1983**, *2*, 57–98.
40. Blasco Sancho, M.F. *Tafonomía y Prehistoria. Métodos y procedimientos de investigación*; Departamento de Ciencias de la Antigüedad—Universidad de Zaragoza; Departamento de Cultura y Educación-Gobierno de Aragón: Zaragoza, Spain, 1992.
41. Yravedra, J. *Tafonomía aplicada a Zooarqueología*; UNED: Madrid, Spain, 2006.
42. Domínguez Rodrigo, M.; Yravedra, J. Why are cut mark frequencies in archaeofaunal assemblages so variable? A multivariate analysis. *J. Archaeol. Sci.* **2009**, *36*, 884–894.
43. Denys, C.; Patou-Mathis, M. *Manuel De Taphonomie*; Archéologiques; Errance: Arles, France, 2014.
44. Fernandez-Jalvo, Y.; Andrews, P. *Atlas of Taphonomic Identifications. 1001+ Images of Fossil and Recent Mammal Bone Modification*; Springer: Chame, The Netherlands, 2016.
45. Binford, L.R. *Nunamiut Ethnoarchaeology*; Studies in Archaeology Series; Academic Press: New York, NY, USA, 1978.
46. Shipman, P.; Rose, J.T. Cutmarks mimics on modern and fossil bovid bones. *Curr. Anthropol.* **1984**, *25*, 116–117.
47. Capaldo, S.D.; Blumenschine, R.J. A quantitative diagnosis of notches made by hammerstone percussion and carnivore gnawing on bovid long bones. *Am. Antiquity* **1994**, *59*, 724–748.
48. Dolmans, D.H.J.M.; Ginns, P. A short questionnaire to evaluate the effectiveness of tutors in PBL: Validity and reliability. *Med. Teach.* **2005**, *27*, 534–538.
49. Subirats, L., Pérez, L., Hernández, C., Fort, S., & Sacha, G. M. (**2019**). A Granularity-Based Intelligent Tutoring System for Zooarchaeology. *Applied Sciences*, 9(22), 4960.

Funded by the Erasmus+ Programme of the European Union

A31: create contents and guidelines for the students in order to facilitate their activities related to the competencies of information, communication, content-creation, safety and problem-solving.

Author: UAM

SUMMARY

- This task has been studied in different pilot experiences because, as mentioned before in other transversal tasks, conclusions related to contents and guidelines can be obtained from many pilot experiences.
- There is, however, one experience that is related to a subject which includes all the competencies described in the task. For that reason, we have selected it to be included here.
- In this subject, students should work in groups and individually for making activities such as problem-solving, safety and content creation.
- We have described the best scenarios for optimizing their learning in all these competencies.
- To also include information and communication, we have extracted and included here part of the results in computer skills, which is another important task in this IO.

1. INTRODUCTION

The development of adequate work group activities in engineering is a tough task, and their efficiency can be highly influenced by students' attitude. In this document, an evaluation of teamwork (chemical and computers laboratories) related to Chemical Engineering subjects is presented, as well as the conditions for an effectual development of work groups and students' attitude for guaranteeing an efficient learning. By using adaptive tests, the most effective self-regulated learning strategies and their relationship with work groups is defined. By doing so, it is demonstrated that teamwork can be helpful for students, but it is not risk free if students do not focus on the tasks. In this sense, results show that students with a passive attitude in the group reach minimal scores, i.e. do not learn concepts or, even do not pass the final examinations.

The use of technology opens new frontiers in learning and improves data mining from students, allowing us to develop and analyze learning strategies [1]. A paradigmatic example of the potential of

Funded by the Erasmus+ Programme of the European Union

communication technologies in learning is the Massive Open Online Courses (MOOC) [2,3], where a large number of students can be involved and makes manual monitoring of learning unfeasible. Information technologies convert this kind of initiative viable since data processing can be performed without real-time supervision of teachers. Another resource easily applied by computers is the use of numerical calculations to simulate realistic scenarios that are not easily accessible[4,5].

Moreover, one of the biggest challenges of including communication technologies in learning is the substitution of interaction between teachers and students by automatic procedures [6,7]. A concept that aroused great interest is tutoring and the possibility of turning an automatic system to an effective instrument for counseling and controlling students[8,9]. Yet, designing an effective automatic tutoring system requires also being familiar with the most effective learning strategies.

Engineering degrees include experimental disciplines in most of the courses that comprise the academic background of the future engineers. Engineering learning has specific attributes, such as analysis, constraints, modeling or optimization, and requires certain engineering mindsets such as accepting multiple possible solutions and the utility of productive failure [10]. In the specific case of the chemical engineers, its training is focused on bringing capability of conceiving, designing and operating chemical-industrial facilities. Experiential learning and teaching can also help the progress of viable engineering for empowering a global sustainable development [11] .

Most of the experimental activities carried out by the chemical engineering students are programmed to be executed in groups. Working in groups can be related to the limited resources in laboratories. Moreover, experimental activities are complex and the cooperation among schoolmates allow them to share the tasks. We should take into account also that engineers must acquire competences related to the ability of working effectively, both individually or as a team and working in multidisciplinary environments such as establishing the requirements for the verification of the official university degrees that qualify for the exercise of the profession of Engineer[12]. These competences are currently included in teaching guides; thus, they must be assessed properly.

If a group works well together, each classmate can achieve much more than working individually, accordingly it can be advantageous working together (a-c), although also some disadvantages can be detected (d-f), as the study guide from the University of Leicester collects [13].

a) Increasing productivity and performance: in fact, practical activities should allow students to share and discuss ideas.

b) Skills development: being part of a team can help schoolmates to achieve different roles such

Funded by the Erasmus+ Programme of the European Union

as leadership, and working with and motivating others, which can be useful not only at academic but at a professional level.

c) Strengths and weaknesses identification: working in groups can help students to identify the role into a group, for example, if they are better leaders than listeners, or better coming up with the 'big ideas' than developing them.

d) Unfair division among different group members': If someone feels they are doing all the hard work, this can lead to resentment.

e) Conflict between different group members: this might arise for many different reasons including leadership competition.

f) Tackling inappropriate tasks as a whole group: groups are not good environments for carrying out some activities such as writing first drafts, which are usually better developed individually.

In this document, we analyze data from different theoretical and practical tasks carried out individually or by groups, in order to measure the influence of teamwork in engineering learning. We use adaptive tests to measure parameters such as attention or motivation. For measuring experimental activities carried out by groups, we use traditional assessment tools (previous knowledge of the theoretical and experimental basis; autonomous management of the experimental installation; attitude, motivation and teamwork; critical analysis of the data obtained; content and format of the report; correction of the answered questions). These data are complemented with assessment by theoretical examination performed by the students before and after the experimental practices, where students worked in groups. All these data allow us to quantify the influence of the practical sessions on the students' training, determining the most adequate strategies to maximize the benefits of group assessment.

2. STATE OF THE ART

The inclusion of new technologies in the learning process gives teachers the opportunity of using new tools in the assessment of many different skills [14]. Work group skills are not an exception. A very interesting example is the use of wikis, that are websites that eases the collaborative development of interlinked web-pages. The massive collaboration process between students that can modify the contents from different places produces great amounts of data related to collaboration skills of students with the advantage of avoiding the problem of having them located in the same place [15]. However, traditional assessment methods do not scale well and new techniques that include self-assessment or involved students in different ways have been proposed. [16,17].

Funded by the Erasmus+ Programme of the European Union

Since the main objective of this document is to analyze and identify the most common and efficient learning strategies when working in groups, we must focus on the best way of measuring students' activities. In the past, self-reported surveys have been commonly used for data collection [18]. However, it contains errors since students are directly involved in the learning process and are subjective due to their own involvement in the learning activities. In this scenario, developing learning activities that can be easily monitored and give accurate information about students' performance is necessary. A great candidate for this task is Computer Adaptive Testing (CAT) since of their ability to adapt the content presented to learners as a function of their responses[19] . CAT has been commonly used in a wide variety of courses and activities: language [20], identification of learning styles [21] and programming [22] are a few significant examples. However, the advantages of using CATs in learning implies also a deep knowledge of the numerical models underneath to make a good calibration of the system [23].

Learning online increases the importance of Self-Regulated Learning (SRL) since many activities that previously were possible only in presence of teachers can be done now autonomously. In this scenario, one of the concepts that has been widely studied is the importance of attention when learning using online tools [24]. It has also been described the potential risks of digital devices when students use these technologies wrongly [25] and their use reduces students' focus on learning tasks. For these reasons, parameters related to attention are crucial when analyzing different students' attitudes when working in teams, which is the main topic of this document.

3. METHODOLOGY

3.1 Adaptive test application

The adaptive system proposed in this document has been developed with a view to providing a system of self-assessment that can also be used as a system for final student assessment. This system has the advantage of highly increasing student's motivation for using the system throughout their learning phase. In Figure 1 we show the teacher's interface of the adaptive tests application (e-valUAM). As we can observe, one of the options when creating questions is to include a Matlab file to calculate the correct answer. This is a very important fact since it allows the system to give different possible answers each time and students should make new calculations every time. By doing so, we are sure that they do not answer by memorizing the numerical answers.

Funded by the Erasmus+ Programme of the European Union

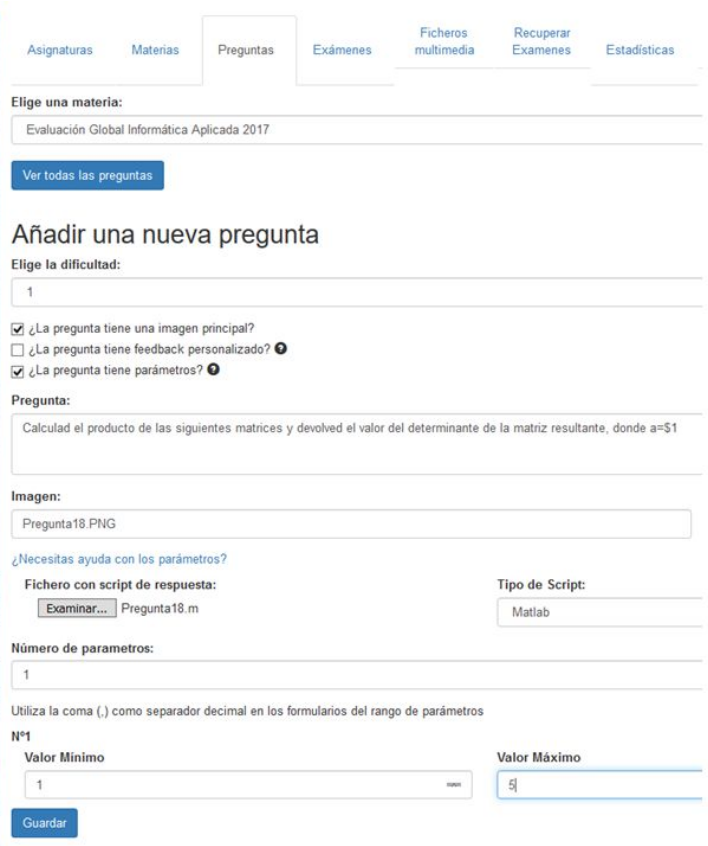


Fig. 1. Graphic e-valUAM interface for teachers in the "Add new question" section.

In Figure 2, we show the interface of e-valUAM for students. As we can see in the figure, students must answer the questions, after reading them without taking care of the configuration of the free parameters (see Figure 1) since this is an internal process. In the case of the question shown in Fig. 2, the free parameter is "a" and it takes a value of 4.447. This value will be changing randomly between two limits (set by the teacher).

Funded by the Erasmus+ Programme of the European Union

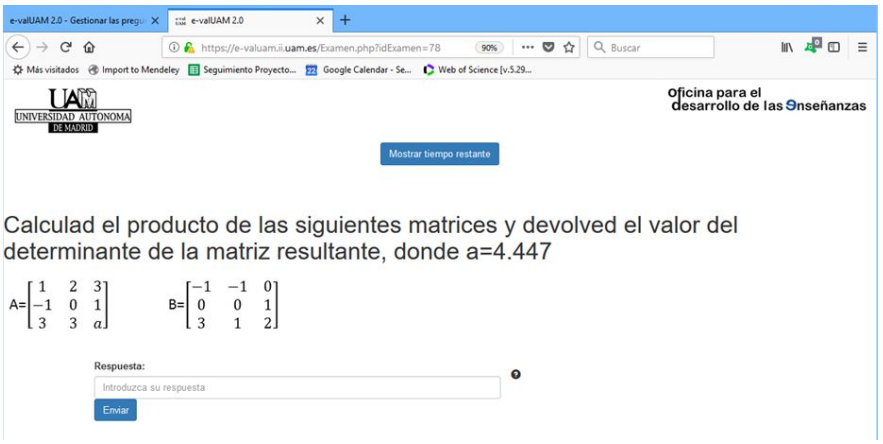


Fig. 2. Graphic interface shown to students during a self-assessment test.

This format is an adequate option for detecting students' attention. The results of a test include information about the time spent in every question, the answers and the final results of the test. In this way, if a test has not been completed, and contains only general answers, such as simple numbers (0, 1...), it would mean that the student only took a look at the questions without working seriously on them. On the other side, a test finished with not only simple numbers implies attention to the task by the student. Those conclusions are also supported by the analysis of the time spent to answer.

3.2 Subjects included in the experiment

The data used in this document are collected from two subjects that belong to the Chemical Engineering Degree of Universidad Autónoma de Madrid (Spain). The first dataset was acquired from students of the 2017/2018 academic year in the first-year subject "Applied Computer Science". 73 students were involved. This subject, which corresponds to 6 ECTS, was taught through theory lessons and practical classes. The practical classes were carried out in computer laboratories, individually and supported by an adaptive test developed in the e-valUAM platform, as pointed out above [26] [27]. In this subject, we have developed a test with 20 questions divided in four levels (the repository includes 50 different questions). Every time a student answers five questions correctly, the system starts selecting questions from the repository of the following level. Since the use of this practical resource is always individual, this first subject will serve essentially to evaluate the usefulness of computer-oriented practical activities.

To analyze the effect of working groups in practical activities, we have also obtained data from the subject "Design of Water Treatment Facilities" in the course 2017/2018. It is an optative 6 ECTS subject

Funded by the Erasmus+ Programme of the European Union

included in the 4th course of the degree. In this subject, we took data from 23 students.

3.3 Teaching methodology

The teaching methodology includes:

1. Theoretical and practical lessons in the classroom. The theory of the subject is divided in four sections covering the treatment of urban and industrial wastewater as well as water purification and regeneration. Practical exercises are proposed, most of them are solved at the classroom.

2. Chemical laboratory sessions. Experiments in the chemical laboratories are developed by groups conformed by three-four students, being them free to organize the groups. Outlines of the two experimental activities (including the aim of the session, a short introduction, the experimental set-up and procedure) are available for the students, before the laboratory sessions are carried out. The experimental activities are related with the theory lectured in the classroom. Specifically, they study:

    a. The operation of a sequencing batch reactor for the treatment of municipal wastewater, determining key parameters of the process.

    b. The treatment of a real industrial wastewater by coagulation-flocculation, optimizing the reagent doses.

3. Practical lessons with computers. Numerical simulations are developed in pairs. They are free to choose their team partners. The undergraduates use WEST, a modelling software for static and dynamic modelling and simulation of wastewater treatment plants.

4. A mandatory visit to a wastewater treatment plant. After theoretical classes, the students visit a wastewater treatment plant to identify, in situ, the different aspects studied during the course.

3.4 Assessment steps

The assessment includes the following elements:

1) Written examination, including the whole contents from the theoretical lessons. The tests

Funded by the Erasmus+ Programme of the European Union

include questions related to the theoretical contents and numerical problems. The students should give reasonable answers including information about the steps that must be done to solve the problems.

2) Reports of the experiments developed at the laboratories. These reports are guided by a set of short questions that students should fill with the results obtained in the experiments. Moreover, some discussion about the results obtained and how to improve them must be provided. The reports are presented by the whole group.

3) Reports of the computer laboratories. Such as for chemical laboratories, a document that works as a guide for the students should be filled with the results obtained in the numerical simulations. The reports are developed in pairs.

3.5 Tests adapted to the nature of the contents.

The autonomous learning was supported by different adaptive tests included in the e-valUAM platform. Due to the different nature of the subject contents', as it has been above commented (theoretical and numerical problems), four tests were developed in different formats to fit the intrinsic requirements of the contents. For the theoretical contents, multiple choice tests have been proposed. These tests include three possible answers. To solve the numerical problem, the adaptive test format implies open answers instead of multiple choice, following the explanation earlier mentioned. Two tests were developed for theoretical content and another two for numerical problems, including different parts of the subject, with direct correspondence with two stages of the subject that were independently evaluated with theoretical examination. At the end of the course, an additional written final exam was included in the assessment process. The adaptive tests were configured as follows:

1) Theoretical test for stage 1: A complete repository of 46 questions divided in four levels. Students must answer 16 questions (the adaptive test jumps to a new level every 4 correct answers).

2) Numerical problem test for stage 1: A complete repository of 12 questions divided in 3 levels. Students must answer 6 questions (the adaptive test jumps to a new level every 2 correct answers).

3) Theoretical test for stage 2: A complete repository of 44 questions divided in 3 levels. Students must answer 18 questions (the adaptive test jumps to a new level every 6 correct answers).

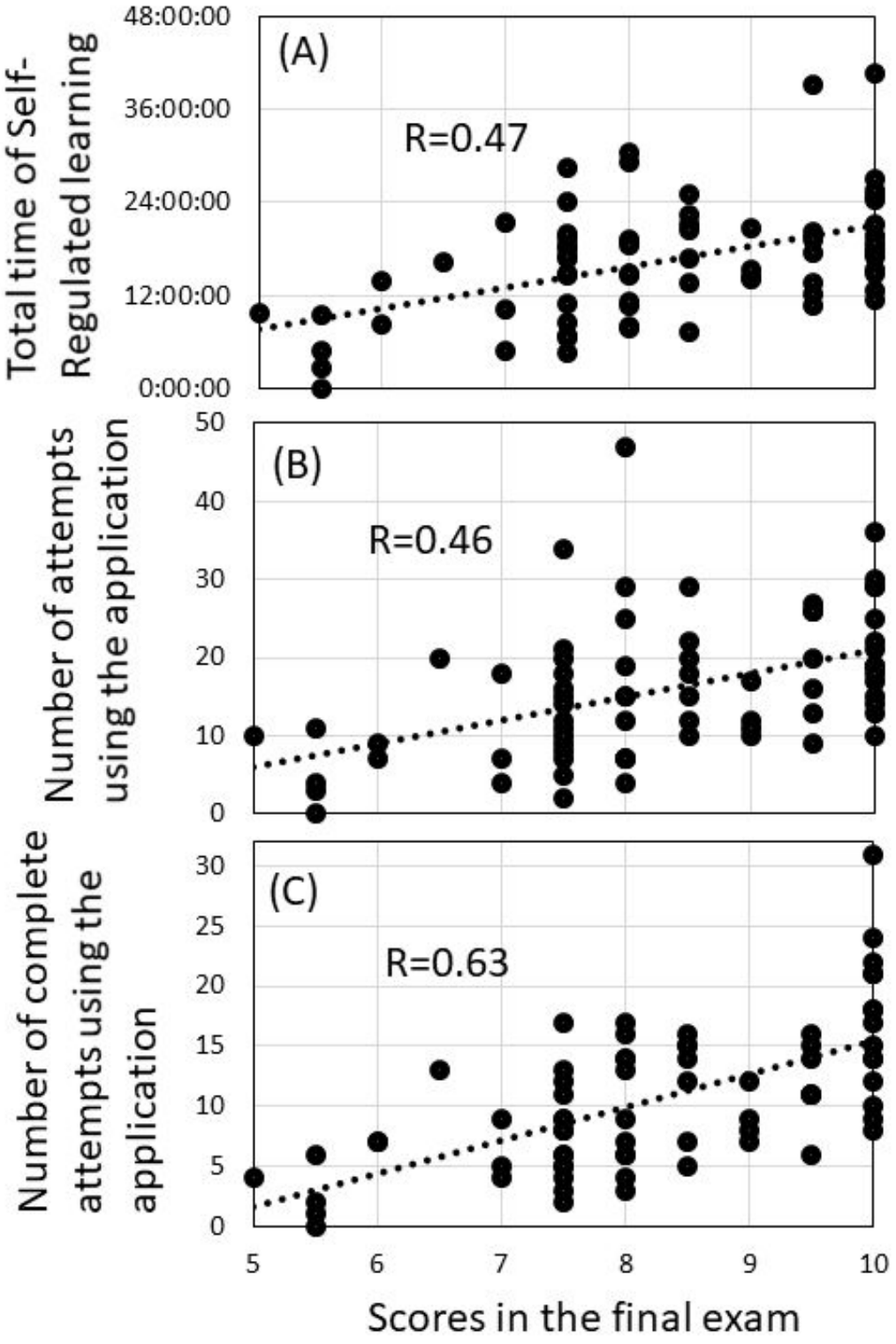
Funded by the Erasmus+ Programme of the European Union

4) Numerical problem test for stage 2: A complete repository of 18 questions divided in 3 levels. Students must answer 9 questions (the adaptive test jumps to a new level every 3 correct answers).

4. RESULTS

In Figure 3 we compare the final assessment (i.e. scores in the final examination) of the whole group in the subject "Applied Computer Science" with three parameters related to the learning stage. Firstly (Fig. 3A), we compare final scores with the total amount of time invested using the application. Secondly (Fig. 3B), we compare final scores with the number of times students used the application. In this case, we have included all the attempts, without excluding anything. In other words, we have counted all the times a student starts a test, even if they do not finish it. Finally (Fig. 3C), we have counted the number of real attempts, which means attempts that were finished. These attempts are of huge relevance to this kind of system, because answering questions of our test implies a mental effort and some mathematical calculations, i.e., completing tests implies a considerable effort for students and are an adequate measurement of the focusing of students in the task.

Funded by the Erasmus+ Programme of the European Union



Funded by the Erasmus+ Programme of the European Union

Figure 3. Comparison of different parameters related to Self-Regulated learning with the scores in the final examination from the subject Applied Computer Science.

As **Figure 3** shows, the three measured magnitudes indeed correlate (R>0.45) with the final scores. However, the correlation was only medium-strong (0.63 of correlation coefficient) in the case of completed attempts, which involves active participation of students in programming and mathematical calculations. Ergo, we conclude that this kind of activity is only useful with an active participation of students.

In **Figure 4**, we show the time spent for a student of Applied Computer Science to finish the test since they started to study the subject until the final examination (marked point), when they reached the highest grade. The figure shows a jump at the point corresponding to the first attempts of the student that involves complex programming questions. Before this point, the student had not yet received enough information in the theoretical classes to face these questions. A clear decrease in the invested time occurring after this point, indicating that the student is learning those difficult contents. The increase in the scores also supports this fact. It is worth pointing out that these conclusions can be only obtained by using adaptive tests since the order of the questions presented in the test gives us an accurate description of what is happening at every point of the learning phase.

At this point we can conclude that using supporting tools, like computers, in learning is useful only if students pay enough attention to the task. In other words, students need to be active and focused on the practices. After that, we are going to check the validity of this hypothesis for working groups.

Funded by the Erasmus+ Programme of the European Union

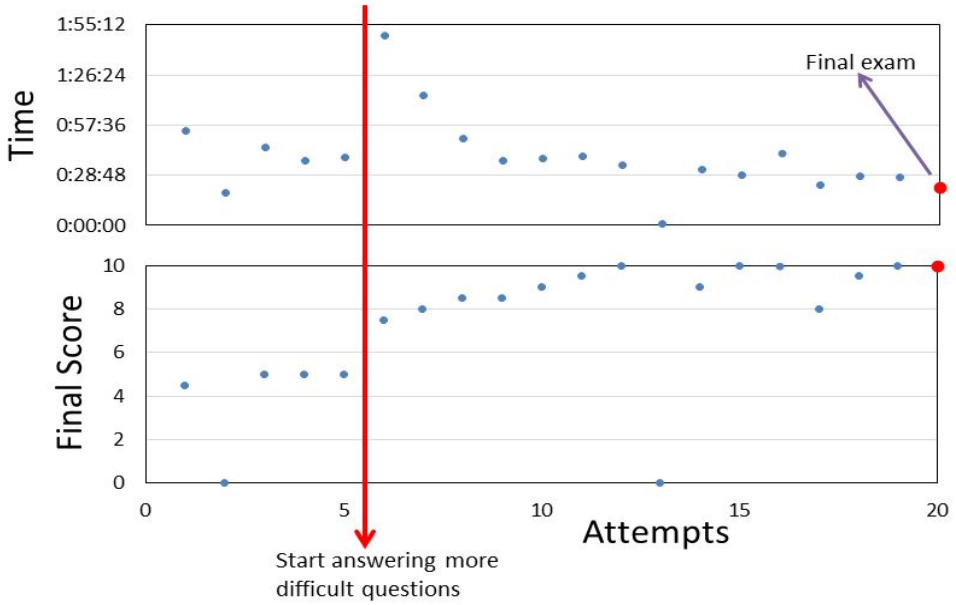


Fig. 4. Time invested using the online application as a function of the attempted number.

In the subject "Design of Water Treatment Facilities", the laboratory computer sessions with the software WEST, the students were divided into 10 pairs. Taking into account the scores of the final examination, those pairs have been divided into two clusters. In other words, we have ordered the students inside each group by their scores in the final examination, obtaining a group of "clever" students (group 1) and a group of "lazy" students (group 2). By doing so, remarkable differences between them in 8/10 groups were observed. In these 8 groups, we found an average difference of 3.36 points (in a 0-10 scale) in the final scores between teammates of the pair (group 1 compared with group 2). The average scores obtained for group 1 were 6.52 while only 3.16 was reached for group 2. By following those results, it seems that students (which were free to choose their pair) were arranged in groups with a teammate being much better than the other. However, this remarkable score difference is not detected in the individual written test performed before the teamwork practical sessions, as can be seen in Figure 5 (down), where the scores of the two students of each group show an average difference of 1.01. In Figure 5 (up), the scores obtained for both groups (black for initial written examination before practical experiment, and grey for final examination) are shown. It is clear that the differences are much higher in the final exam (i.e. after the practical session). It is particularly relevant that the huge differences in the final exam are not related to students that were different at the beginning, since we can find many different scenarios before the practical sessions. For example, students from working groups 4, 5 and 7 even changed their tendency. In those cases, one student from the working group was better at the beginning but got much worse scores in the final

Funded by the
Erasmus+ Programme
of the European Union

examination than their partner. This heterogeneous initial scenario demonstrates that the significant differences found after the practical sessions should be influenced by something related to the teamwork in this activity (computer assisted practical sessions). However, it is worth noting that this effect is not found in the chemical laboratory teams, formed by three teammates. In this case, the heterogeneous profiles found at the beginning are also found after the experimental sessions.

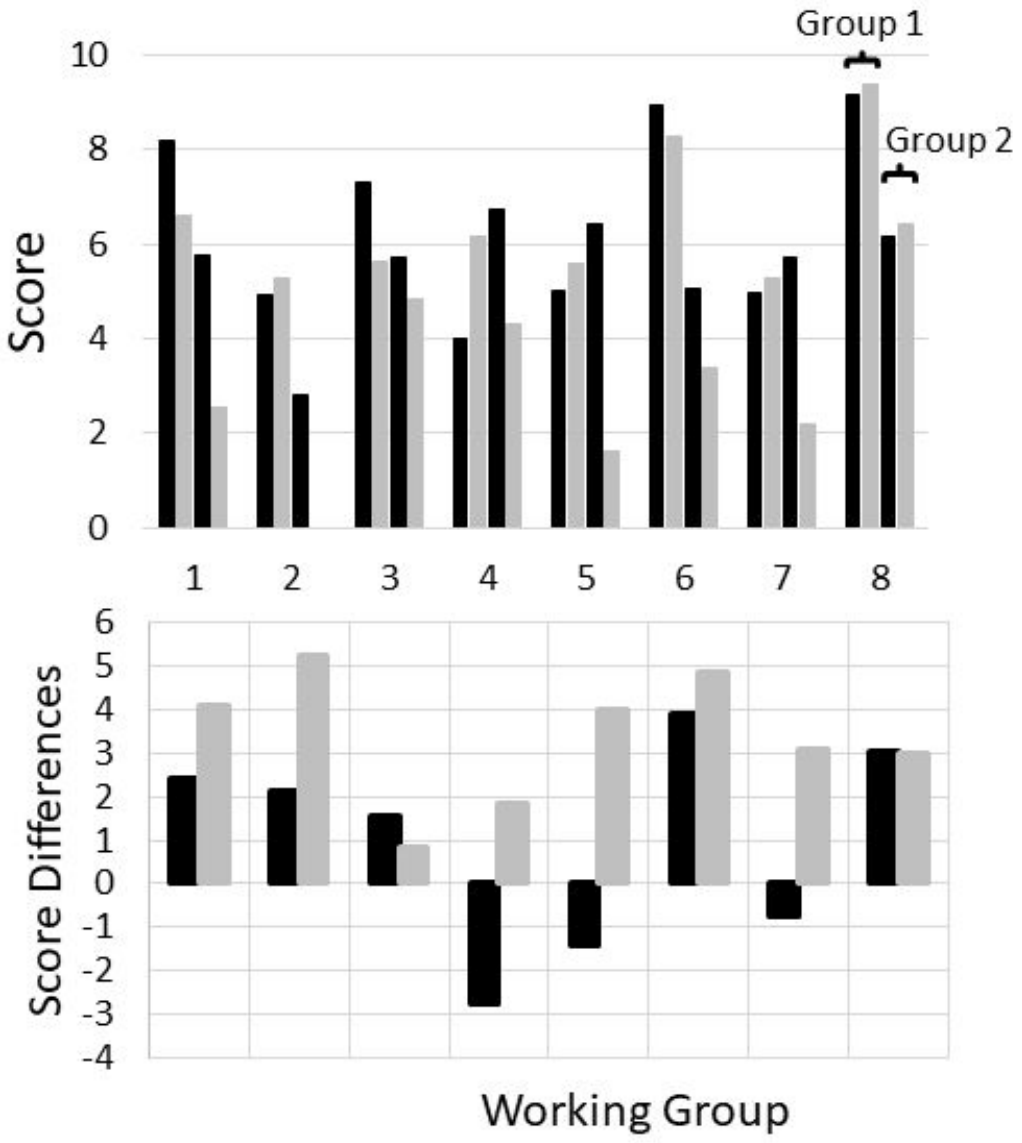


Fig. 5 Scores (up) and score differences (down) of all the students in the 8 groups where a significant effect of the working groups has been detected. Black: Scores from the test done before the practical sessions.

Funded by the Erasmus+ Programme of the European Union

Grey: Scores of the final exam.

5. Discussion

It is well-known that many parameters such as motivation, enjoyment or perceived usefulness are factors that strongly influence students' performance in many different tasks in their learning process. Indeed, this is also happening when analyzing teamwork. An active behavior is also considered an appropriate strategy for improving knowledge acquisition and even teamwork skills. Assuming those premises, we are analyzing the effect of team working and how this kind of activity can change or modify some students' attitudes and how these changes can have a significant influence in their learning efficiency.

It becomes clear that teamwork does not have the same influence on the different teammates. In the case of practical sessions with computers, in 8 of the 10 groups formed, one teammate is positively influenced and the other one is negatively affected. This can be explained bearing in mind the results obtained for individual work (Applied Computer Science), where we observed a better students' performance when they focused on the tasks. In other words, they should be active when using computers or developing their algorithms. In our opinion, the problem detected in the work groups of "Design of Water Treatment Facilities" is that the pair tends to be led by one teammate and the other one only "pay attention" to the work done by their partner. This passive attitude has been found to be almost useless in our first analysis (Fig. 3B compared to Fig. 3C). This hypothesis is also supported by the not detected negative effect of work teams in the groups that performed the laboratory activities, where the complexity of the practical work does not allow students to assume a passive attitude, and working as a real team.

In order to maximize the work groups' effectiveness, we suggest being aware in designing team activities in engineering, especially when the activities allow students to work individually when they should do it as a team. Wrong strategies can induce huge differences between the involvements of the students conforming to a group, due to some teammates tending to lead (obtaining better results) and others simply watching or listening to (being negatively affected in their learning effectiveness).

6. Conclusions

In this document, we have analyzed the influence of experimental activities in work groups on the final assessment of Chemical Engineering subjects. We have concluded that some tasks are inappropriate to be carried out by a group. We must be aware of group activities' limitations and proposing some experimental activities such as individual tasks. We have realized that, in pair groups, the group's

Funded by the Erasmus+ Programme of the European Union

leader-student role benefits the students who assume it, while the listener-student tends to worsen its academic performance. These results are also supported by the analysis of practical activities developed individually using adaptive tests, where student's scores are highly correlated with the tasks that students finished focusing on the activity. We can conclude that teamwork can be of great interest, but potential troubles must be considered before planning task developments in groups, to avoid the inadequate academic performance of some students due to a wrong planning. An alternative would be including new subjects focusing on undertaking group projects as part of their course.

The experiments developed in this document are related to subjects from the Chemical Engineering grade. However, the kind of activities proposed to students in the laboratories are closely related to many others from different engineering grades. Even in the cases where specific contents are strictly related to Chemical Engineering, the procedure of the experiments and the documents that must be completed usually follows general patterns. For this reason, the results of this document can be also extrapolated to other engineering disciplines.

Although this work can be extended to other disciplines with similar format, it is also limited by the methodology followed in different subjects. It is reasonable to extend our results to subjects where numerical problems are solved. However, in disciplines where the teachers do a subjective assessment, we cannot be sure about the validity of our conclusions. In those cases, different approximations such as the use of rubrics have been proposed to evaluate teamwork activities.

Future work should involve the extension of our analysis to other engineering and natural sciences disciplines. In those studies, it will be convenient to evaluate the creation of groups following the findings of this document, as well as comparing students' performance when working in groups or individually.

REFERENCES

1. J. Maldonado-Mahauad, M. Perez-Sanagustin, R.F. Kizilcec, N. Morales and J. Munoz-Gama, Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses, Computers in Human Behavior, 80, 179-196, 2018.
2. Y. Lee, Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses, Computers in Human Behavior, 86, 174-180, 2018.
3. A. McAuley, B. Stewart, D. Cormier and G. Siemens, In the Open: The MOOC model for digital practice. SSHRC Application, Knowledge Synthesis for the Digital Economy, 2010.

Funded by the Erasmus+ Programme of the European Union

4. P. Molins-Ruano, C. Sevilla, S. Santini, P.A. Haya, P. Rodríguez and G.M. Sacha. Designing video games to improve students' motivation. Computers in Human Behavior, 31, 571, 2014.
5. G.M. Sacha and P. Varona. Artificial Intelligence in Nanotechnology, Nanotechnology, 24, 452002, 2013.
6. A. Kumar, Using online tutors for learning - what do students think? In IEEE Proceedings of Frontiers in Education Conference (FIE), 524–528, 2004.
7. S. Hrastinski, M. Cleveland-Innes and S. Stenbom, Tutoring online tutors: Using digital badges to encourage the development of online tutoring skills. British Journal of Educational Technology, 49, 127-136, 2018.
8. C. Bravo, W.R. van Joolingen and T. de Jong, Using co-lab to build system dynamics models: students' actions and on-line tutorial advice, Computers & Education, 53, 243–251, 2009.
9. A.S. Sabitha, D. Mehrotra and A. Bansal, An ensemble approach in converging contents of LMS and KMS, Education and Information Technologies, 22, 1673-1694, 2017
10. K. Shirey. Breaking the Silos of Discipline for Integrated Student Learning: A Global STEM Course's Curriculum Development, Engineering 4, 170–174, 2018.
11. B. Gladysz, M. Urgo, L. Gaspari, G. Pozzan, T. Stock, C. Haskins, E. Jarzebowska and H. Kohl, Sustainable Innovation in a Multi-University Master Course, Procedia Manufacturing, 21, 18-25, 2018.
12. Spanish Ministry Order CIN/351/2009, of February 9, which establishes the requirements for the verification of the official university degrees that qualify for the exercise of the profession of Industrial Technical Engineer.
13. https://www.le.ac.uk/oerresources/ssds/studyskills/ (16 September 2019).
14. A. Balderas, J. Caballero-Hernández, J. Dodero, M. Palomo-Duarte and I. Ruiz-Rube, Assessment of Generic Skills through an Organizational Learning Process Model, Proceedings of the 14th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST, 293-300, 2018.
15. M. Palomo-Duarte, J. M. Dodero, I. Medina-Bulo, E. J. Rodríguez-Posada and I. Ruiz-Rube, Assessment of collaborative learning experiences by graphical analysis of wiki contributions, Interactive Learning Environments, 22, 444-466, 2014.
16. M. Cubric, Using wikis for summative and formative assessment. In Assessment design for learner responsibility. Re-engineering assessment practices (REAP) international online conference, Glasgow: University of Strathclyde, 1–8, 2007.
17. E. Panadero, J. Broadbent, D. Boud and J.M. Lodge, Using formative assessment to influence self- and co-regulated learning: the role of evaluative judgement, European Journal of Psychology of Education 34, 535-557, 2019.
18. P. Wagner, B. Schober and C. Spiel, Time students spend working at home for school, Learning and Instruction 18, 309-320, 2008.

Funded by the Erasmus+ Programme of the European Union

19. S. Narciss, S. Sosnovsky, L. Schnaubert, E. Andres, A. Eichelmann, G. Goguadze and E. Melis, Exploring feedback and student characteristics relevant for personalizing feedback strategies, Computers & Education 71, 56-76, 2014.
20. C.A. Chapelle and D. Douglas, Assessing Language Through Computer Technology. Cambridge University Press. 2006;
21. A. Ortigosa, P. Paredes and P. Rodriguez, AH-questionnaire: An adaptive hierarchical questionnaire for learning styles. Computers & Education. 54, 999-1005. 2010.
22. P. Molins-Ruano, S. Atrio, P. Rodríguez and G.M. Sacha, Modelling experts' behavior with e-valUAM to measure computer science skills. Computers in Human Behavior, 61, 378-385, 2016.
23. A.J. Martin and G. Lazendic, Computer-Adaptive Testing: Implications for Students' Achievement, Motivation, Engagement, and Subjective Test Experience. Journal of Education Psychology, 110, 27-45, 2018.
24. J.Y. Wu and T.Y. Cheng, Who is better adapted in learning online within the personal learning environment? Relating gender differences in cognitive attention networks to digital distraction, Computers & Education 128, 312-329, 2019.
25. J. Aagaard, Drawn to distraction: A qualitative study of off-task use of educational technology, Computers & Education 87, 90-97, 2015.
26. P. Molins-Ruano, C. González-Sacristán, F. Díez, P. Rodriguez and G.M. Sacha, An Adaptive Model for Computer-Assisted Assessment in Programming Skills, International Journal of Engineering Education 31, 764, 2015.
27. E.A. Gomez, D.Z. Wu and K. Passerini, Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes, Computers & Education, 55, 378-390, 2010.
28. F.A. Ganotice Jr. and L.K. Chan, How can students succeed in computer-supported interprofessional team-based learning? Understanding the underlying psychological pathways using Biggs' 3P model. Computers in Human Behavior, 91, 211-219, 2019.
29. P. Sancho-Thomas, R. Fuentes-Fernandez and B. Fernandez-Manjon, Learning teamwork skills in university programming courses, Computers & Education, 53, 517-531, 2009.
30. M. Comas-Lopez, K. P. Hincz, N. de Haro, M. Mazalu and G. M. Sacha, Soft Skills Assessment in Art and Globalization, Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, ACM International Conference Proceeding Series, 199-204, 2018.
31. P. Molins-Ruano, C. González-Sacristán, F. Díez, P. Rodriguez and G.M. Sacha, An Adaptive Model for Computer-Assisted Assessment in Programming Skills, International Journal of Engineering Education 31, 764, 2015.

Funded by the Erasmus+ Programme of the European Union

32. E.A. Gomez, D.Z. Wu and K. Passerini, Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes, Computers & Education, 55, 378-390, 2010.
33. F.A. Ganotice Jr. and L.K. Chan, How can students succeed in computer-supported interprofessional team-based learning? Understanding the underlying psychological pathways using Biggs' 3P model. Computers in Human Behavior, 91, 211-219, 2019.
34. P. Sancho-Thomas, R. Fuentes-Fernandez and B. Fernandez-Manjon, Learning teamwork skills in university programming courses, Computers & Education, 53, 517-531, 2009.
35. M. Comas-Lopez, K. P. Hincz, N. de Haro, M. Mazalu and G. M. Sacha, Soft Skills Assessment in Art and Globalization, Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, ACM International Conference Proceeding Series, 199-204, 2018.

Funded by the Erasmus+ Programme of the European Union

A26: Determination of the best methodology to analyse formal and informal computer skills.

Author: UAM and EURECAT

SUMMARY

- In this deliverable, we present two numerical methods for the prediction and correction of students' bad performance and wrong learning strategies.
- To test these methods, we have applied our results to two subjects that are representative of two different learning and teaching strategies.
- The first one includes a high percentage of subjective contents. The second one is easily evaluated by objective numerical scales. To evaluate the first subject, we propose the analysis of students' peer reviews.
- The second one is analyzing by Artificial Intelligence algorithms. In both cases, teachers have proposed and executed preventive measures to the students that were in higher risk of failing.

1. INTRODUCTION

The use of computers in education is a hot topic nowadays [1]. The automatization that computers can add to the learning process is a concept that must be taken into account when facing courses or subjects that include a huge number of students [2] or a scenario where students are not available for face-to-face sessions [3]. Thanks to the use of computers, teaching in these complex circumstances is achievable [4]. However, the automatization of learning implies that teachers will not have a close interaction with students, which is a disadvantage that must be compensated by well-structured learning environments [5]. The correct configuration of these environments strongly depends on data analysis, which is the origin of Learning Analytics [6].

Learning Analytics is the discipline that deals with the treatment and analysis of data in learning environments. Indeed, it is strongly supported by computers and closely related to the wider concept of e-Learning. Many different approximations of data analysis have been done within the Learning Analytics discipline [7,8,9]. In this document, we will focus on the prediction of students' performance. The research objective is the development and analysis of new effective techniques for the prediction of students' effectiveness in their learning process. These methods will be useful in two different ways. From a theoretical point of view, they will help in the knowledge of the parameters and

Funded by the Erasmus+ Programme of the European Union

conditions that are more important for an effective learning activity. From a practical point of view, they will help teachers to find students with problems with enough time to prepare activities to impulse their learning.

We propose two different approximations that strongly depend on the nature of the contents of every subject. As representative examples, we will focus on two very different subjects. The first one is called Art and Globalization and all its content is closely related to the interpretation of students when analyzing an art piece. Indeed, this subject includes a huge percentage of subjective content, which is very difficult to measure numerically. The second subject is Applied Computing, which is related to numerical algorithms. In this case, it is easy to evaluate by an objective numerical scale.

We propose two different techniques for the prediction of students' performance that will fit the different requirements of both subjects. In both cases, we have used data that has been collected in previous academic years due to the necessity of acquiring data related to the final scores in the subjects. With this information, we have predicted the potential risks of students in the 2018/2019 course and we have applied some preventive measures to the students that were identified as the ones with the higher risk of failure. We finally show the result of those actions.

In section 2 we present the characteristics of the two subjects included in the study. In section 3 we show the methods employed for the prediction. In section 4 we present the results of the preventive measures and in section 5 we show the conclusions.

2. DESCRIPTION OF THE SUBJECTS INCLUDED IN THE STUDY

In this section the two subjects included in the study will be described, which are art and globalization and applied computing.

2.1 Art and Globalization

The experiment with students at Universidad Autónoma de Madrid was carried out in the subject Art and Globalization during the 2017/18 and 2018/19 academic years. In both cases, the total of students taking this subject was of 51. Art and Globalization is a compulsory module in the 2nd year of the International Studies Degree at the Philosophy and Letters Faculty in Universidad Autónoma de Madrid. It corresponds to 6 ECTS. It is taught through theoretical lessons.

The aim of this subject is that students question and know some of the most relevant alternatives to Euro-centered perspectives regarding the understanding and interpretation of artistic and cultural

Funded by the Erasmus+ Programme of the European Union

productions. It is expected that the students in this course demonstrate proactive and critical attitudes towards knowledge. They should also be able to engage into productive conversations and debates about relevant issues concerning the uses and interpretations of art and culture. Therefore, in this subject, the development of communication skills and critical thinking is as important as the development of hard skills. This means that, in this subject, a good assessment of soft skills is essential for a comprehensive evaluation of students.

2.2 Applied Computing

The experience with real students took place during the 2016/2017, 2017/2018 and 2018/2019 academic years. 97, 73 and 79 students were involved respectively in the first year elective course "Applied Computer Science". This subject was taught through theory classes and practical classes in the laboratory. This course corresponds to 6 ECTS and belongs to the Chemical Engineering Degree in the Faculty of Sciences of the Autonomous University of Madrid (UAM). The method applied to self-regulated learning in this subject has been described elsewhere [10] and only a brief description is given here. The most important fact is that students use the e-valUAM platform for both autonomous learning and final exams, which is a high motivation for them in order to use the platform for almost their whole learning process. This fact is extremely useful when measuring autonomous learning because it gives us an accurate environment with many parameters to analyze.

3 Prediction methods

In this section the methodology applied in the two subjects included in the study will be described, which are a rule-based system and machine learning techniques.

3.1 Art and globalization

Each year the group of students was divided by the lecturer into working groups. In 2017/18 there were 10 groups of 5-6 students each and in 2018/19 there were 12 groups of 4-5 students each. Every group worked on a topic of their choosing among those in a list prepared by the lecturer. Groups had to give two presentations on their chosen subject: the first one after the first month of the course and the second during the last weeks of the semester.

Therefore, the group's work on the topic and the presentation preparation period spanned several weeks over which the team had to divide responsibilities amongst all members. When half of those weeks had passed, each team member was asked to fill in an evaluation form to assess the performance of the group members in their own team, including themselves. The prescribed peer

Funded by the Erasmus+ Programme of the European Union

evaluation form included the following 6 statements: (1) attends group meetings regularly, (2) contributes meaningfully to group discussions, (3) completes group assignments on time or makes alternative arrangements, (4) prepares work in a quality manner, (5) demonstrates a cooperative and supportive attitude, (6) contributes significantly to the success of the project. It also included open questions which have not been considered for our analysis in this study. Each person had to indicate their agreement with the statement ranging from 1 to 4 for all their peers. In the course 2017/18 there were 50 students who carried out this task. In the course 2018/19 there were 42 students who completed it.

Taking this peer evaluation into account, an average student grade was calculated on a scale 1-4. The resulting grades for each student were compared to the final grade on the whole course given by the teacher on a scale 0-10. To make this analysis, both scales were normalized to 0-10.

In figure 1, we show both peer assessments and final scores of all students. Students are ordered by their final grades. Our objective here is the establishment of a criteria for selecting students that are below the fail limit and increase their efficiency when learning by some additional monitored activities. We propose to select the students by the scores obtained in the peer assessment process. In the figure, we show 4 potential criteria based on different limits in the peer assessment.

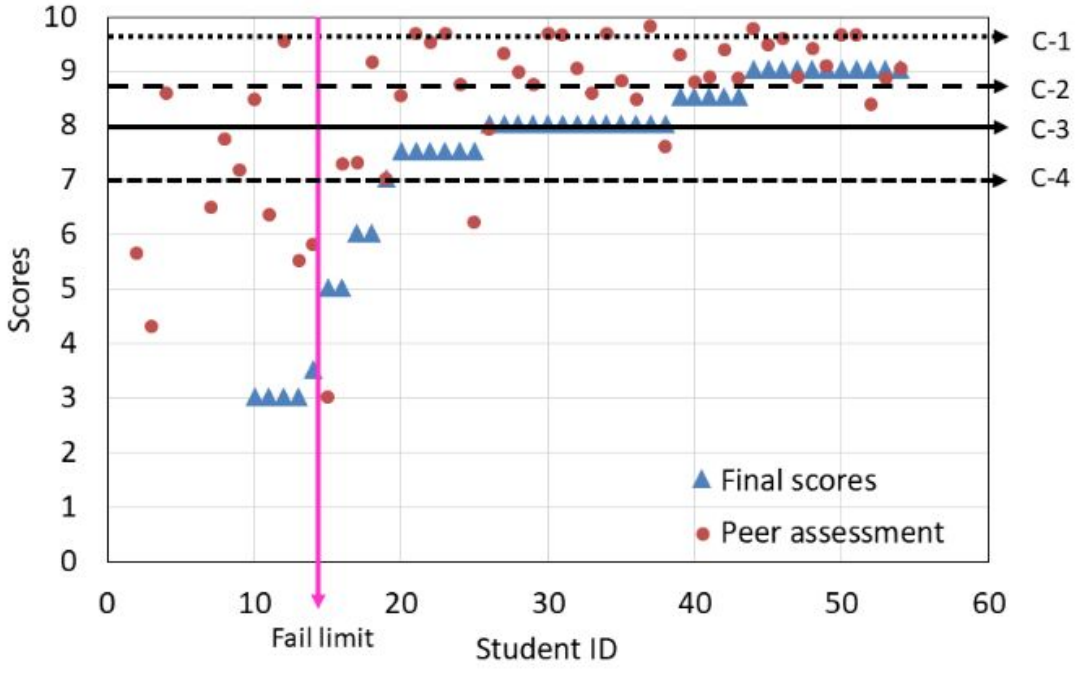


Figure 1: Students' peer assessment average and final scores in the subject Art and Globalization. Course

Funded by the Erasmus+ Programme of the European Union

The first criterion selected (C-1) includes all students that are below the fail limit (11). However, by applying this criterion, we must also call 95% of students who actually passed the exam. By applying C-2, we lose 1/11 students who failed (9%), but we reduce the percentage of students wrongly selected to 32%. C-3 loses 27% of students who failed but reduces to 17% the number of students selected from the group that passed the exam. The last criterion (C-4) only selects 5% of students who passed the exam but loses 45% of students who failed. The selection of a criterion depends on this balance and it is a subjective decision. By one side, we should take into account the number of students from the fail group that we are not calling, but we must also take into account the additional effort for both teachers and students that implies the call of a huge number of students who do not need any additional task. In our experiment we have considered C-3 as the most rational criterion.

It is worth noting that selecting a criterion is a subjective process that depends on the conditions on the subject, teachers' availability, etc. Selecting one criterion implies the call of a certain percentage of students. This fact, however, can change in different subjects since students' answers to the rubrics could be quantitatively different. For this reason, it is important to obtain as much information as possible before starting using this methodology. It is recommended to have at least one previous course with the rubrics being applied in order to obtain an accurate correlation between students' answers and final scores. Having results from more than one year would be also convenient in order to avoid effects of small groups or the influence of the intrinsic characteristics of a single generation of students. For example, we have found that in some courses there is a huge resistance against judging partners, which has been strongly influenced by some students that joined their doubts with their friends, spreading their feelings along the whole group. This fact affected a lot of the final quantitative results.

3.2 Applied Computing

In order to predict the final marks of the subject, the method Random Forest [11] is used to predict the binary class (alert or OK). Regarding the dataset 4 parameters were considered when predicting the final marks of the subject: the number of marks, the mean, standard deviation and median of e-valUAM.

Random forests are chosen because they provide good predictive performance, low overfitting, and easy interpretability (they provide the importance of each attribute). According to [12], Random

Funded by the Erasmus+ Programme of the European Union

forests are a set of decision trees, where the overall result is the average result of all the decision trees. It is important to point out that each decision tree is trained with a random subset of training data and attributes.

66.67% of the data was used for the training set and 33.33% for the test set, and 10-cross fold validation was used. For the random forest, the following parameters where used classifier__n_estimators: [100,500], classifier__max_features: ['auto', 'sqrt', 'log2'], classifier__class_weight: ['balanced',None] and 'classifier__min_samples_leaf': [1,10,50].

Given that in a confusion matrix we have true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), we define in the usual way the following measures:

$$
\begin{aligned}
precision &= \frac{TP}{TP+FP} \\
recall &= \frac{TP}{TP+FN} \\
F1 &= \frac{2 \times precision \times recall}{precision + recall} \\
accuracy &= \frac{TP+TN}{TP+FN+TN+FP}
\end{aligned}
$$

where the confusion matrix is

		True condition	
		Condition positive	Condition negative
Predicted condition	Predicted condition positive	True positive	False positive
	Predicted condition	False negative	True negative

Funded by the Erasmus+ Programme of the European Union

	negative		

Confusion matrix

In Table 1 there are the details of the prediction of the classifier detailed in the methods section computing the accuracy, precision, recall and f1-score, and the confusion matrix of the prediction is shown in Table 2.

Grade	Precision	Recall	F1-score	Support
Alert [0-7)	0.26	0.78	0.39	9
OK [7-10]	0.94	0.60	0.73	50
Total	0.83	0.63	0.68	59
Total accuracy	0.63			

Table 1. Detailed performance of the random forest considering years 2017 and 2018 of e-valUAM

Funded by the Erasmus+ Programme of the European Union

	Actual class [0-7)	Actual class [7-10]
Predicted class [0-7)	7	20
Predicted class [7-10]	2	30

Table 2. Confusion matrix

If we compare results with the state-of-the-art [13], the precision is close to research studies (0.83) while the global recall and accuracy are a bit lower (0.63). This is because some final marks close to the threshold (7) are not well predicted. As a consequence, in the validation there are 20 students which are false positives. This can be interpreted that there are students who are alert [0,7), but at the last moment they worked hard and finished the course successfully. This fact is something that we cannot control with this method because we need to analyze data from the beginning of the course to a date not so close to the final exams in order to have enough time to prepare additional sessions for those students. This fact reduces the performance of the method. However, the very small value in the false negatives cell implies that a small number of students who are in danger are not called, which is adequate for the objectives of the method.

4. Results

In this section the results in the two subjects included in the study will be described.

4.1 Art and Globalization

In figure 2 we show the scores obtained by the students in the course 2018/2019. We show 2 data series. The first one is the average of the peer assessment made by their partners (and themselves). The second one is the final score in the subject, that includes several evaluations such as the two oral

Funded by the Erasmus+ Programme of the European Union

presentations, a written essay and the final exam. Data are ordered by the final scores. Students that appear with an average score of 0 are those who had not finished their studies in this specific subject at the time the data used in this analysis were gathered. In the figure, we also show the threshold that has been used to distinguish students that are performing well in their learning process and those who can be in trouble (C-3 in previous section). As we can see, only 9 students are placed below this limit.

All of these nine students were called by the lecturer in order to schedule an individual meeting. There were 5 students who responded to the call and attended this meeting. There, the lecturer talked with them about their experience with the subject in general and with the members of their own group. They were encouraged to ask any questions they might have about the subject, all of them asked about how the exam would be like and wanted to check what kind of materials they should use to prepare it. In addition to answering their questions, the lecturer confirmed a specific session would be devoted to discussing the exam and to working with exams that had been used in previous years. In some cases, the conversation showed that there had been internal conflicts in some groups which might have been the main cause for the poor grade of some team members.

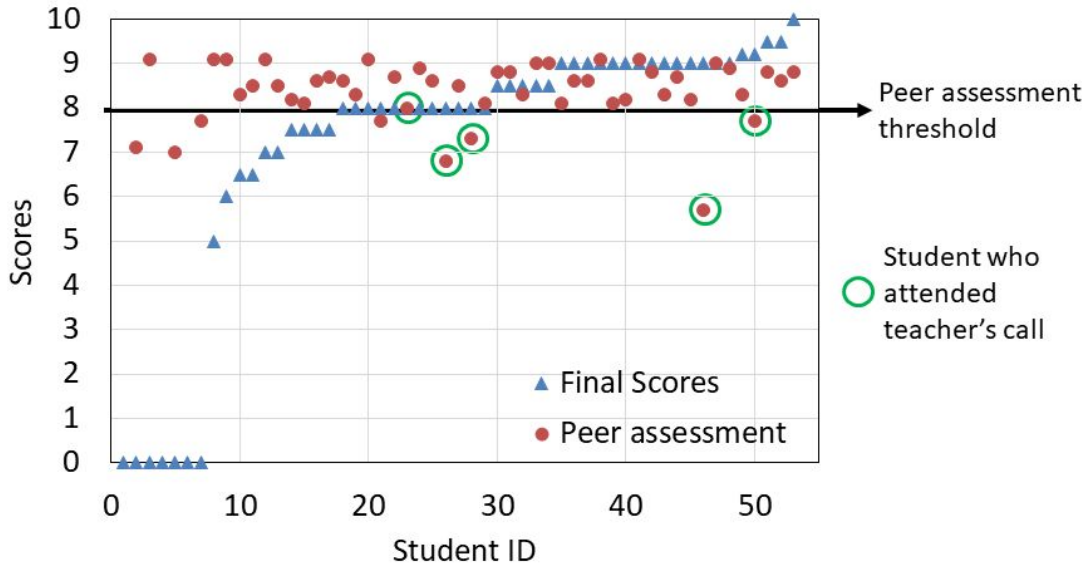


Figure 2: Students' peer assessment average and final scores in the subject Art and Globalization. Course 2018/2019.

As we can see in the figure, students who attended this session obtained higher scores than those who did not answer the lecture's call.

Funded by the Erasmus+ Programme of the European Union

4.2 Applied Computing

Based on the results, we have found that the attributes with higher importance for the prediction are the number of times using the application, the mean of the scores obtained in the autonomous study, the standard deviation and the median in this order.

Academic performance of students in the year 2019 is predicted obtaining 25 alerts out of 72 students. All of these 25 students were called by the lecturer in order to schedule a group meeting. There were 8 students who responded to the call and attended this meeting. 2 of them were not actually called. They attended the meeting only because they wanted to increase their efficiency, but they were not considered in the warning group by the algorithm. In other words, only 6/25 students answered the call since 2 were not actually called. There, the lecturer talked with them about their skills and some guided activities were suggested. Only 3 of them finished those tasks. Since these numbers are too small, we cannot make any quantitative analysis from this experience.

5.CONCLUSIONS

In this work we have proposed two different methods for the prediction and correction of students' wrong learning strategies. In both cases, the prediction must be done by the data analysis of complete courses in order to establish a correlation between data collected along the learning process and final scores in the subjects. Without the final scores, this analysis is not possible.

The reason why we propose two methods is related to the different nature of subjects taught at different grades and universities. Sometimes, numerical quantitative data can be acquired because the subjects are related to numerical contents. However, there are many cases where subjective interpretation or soft skills are involved, and this numerical analysis cannot be done easily. For this reason, we have developed a numerical analysis for the cases where quantitative data can be acquired, and an alternative method based on student's peer review for the cases where it is not possible.

In both cases we have been able to distinguish students who are in potential danger of failing. By applying additional sessions with those students, we have seen that students increase their performance when attending lecturer's calls. However, we have also found a very low rate of students who answer this call. This is one of the most important facts that must be corrected when applying our method since it becomes useless. To correct this problem, the only possible solution would be starting the preventive measures earlier, but the algorithms performance would be smaller. A second option

Funded by the Erasmus+ Programme of the European Union

would be applying some techniques to increase students' motivation for attending lecturers' calls.

REFERENCES

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

[2] Lee, Y. (2018). Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses. Computers in Human Behavior, 86, 174-180.

[3] McAuley, A., Stewart, B., Cormier, D., & Siemens, G. (2010). In the Open: The MOOC model for digital practice. SSHRC Application, Knowledge Synthesis for the Digital Economy.

[4] Bravo, C., van Joolingen, W.R., & de Jong, T. (2009). Using co-lab to build system dynamics models: students' actions and on-line tutorial advice. Computers & Education, 243–251.

[5] Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. The International Review of Research in Open and Distributed Learning, 16(3). http://doi.org/10.19173/ irrodl.v16i3.2112

[6] Siemens, George (20 de agosto de 2013). «Learning Analytics: The Emergence of a Discipline». SAGE Publications. doi:10.1177/0002764213498851

[7] Louw, J., Muller, J., & Tredoux, C. (2008). Time-on-task, technology and mathematics achievement. Evaluation and Program Planning, 31(1), 41-50. http://doi.org/ 10.1016/j.evalprogplan.2007.11.001.

[8] Eriksson, T., Adawi, T., & Stohr, C. (2016). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. Journal of Computing in Higher Education, 29(1), 133-146. http://doi.org/10.1007/s12528-016-9127-8.

[9] Sujatha, R. & Kavitha, D. (2018). Learner retention in MOOC environment: Analyzing the role of motivation, self-efficacy and perceived effectiveness. International Journal of Education and Development using Information and Communication Technology (IJEDICT), 14, 62-74.

[10] Molins-Ruano, P., González-Sacristán, C., Díez, F., Rodriguez, P. & G. M. Sacha. (2015). An

Funded by the Erasmus+ Programme of the European Union

Adaptive Model for Computer-Assisted Assessment in Programming Skills. International Journal of Engineering Education. 31, 764.

[11] Ho TK. Random Decision Forests. In: Proceedings of the Third International Conference on Document Analysis and Recognition. vol. 1 of ICDAR'95. Washington, DC, USA: IEEE Computer Society; 1995. p. 278–282.

[12] Cady, F. (2017) The Data Science Handbook. Wiley.

[13] Rovira, S., Puertas, E., Igual, L. (2017) Data-driven system to predict academic grades and dropout. PLOS ONE 12(2): e0171207.

Funded by the Erasmus+ Programme of the European Union

A32, A34, A36: Piloting experiences on self-evaluation

Author: UAM

SUMMARY

- This deliverable *(Piloting experiences on self-evaluation)* presents the results of the piloting experiences for the three years of the project in subjects that included material related to students' self-evaluation.
- In the document, we will give information about the subjects, the number of students and the material used in the experiences.
- Main conclusions of the experiences are also shown.
- However, the creation of new materials due to the conclusions obtained here are not shown in this document since this information belongs to tasks 3.3, 3.5 and 3.7.

Starting point

The purpose of this deliverable was to give the objective data related to the pilot experiences and their main conclusions in order to facilitate the creation of useful material for the following years and, in general, for being used for a wider community.

Next steps

After every year pilot experiences, new e-Learning material has been created to solve the difficulties found or to improve the advantages or suggestions detected in previous courses. After the third year, final versions of all the e-Learning material developed in previous courses will be produced.

1. METHODOLOGY

Methodology 1: Multiple Answer vs Open Answer tests. Short term vs Long term self-evaluation.

This first methodology implies the analysis of two dimensions. The first one is related to the format of the tests used in the self-evaluation process. The second one is related to the amount of time used by the student in their self-evaluation process. The analysis of these two dimensions have been developed in a single experiment where all the possible combinations have been tested.

Funded by the Erasmus+ Programme of the European Union

In this section we describe the proposed method and the study that has been carried out to test its effectiveness. Since the dimension related to the time requires at least two stages, all the material developed has been duplicated. Although details on the nature of these stages will be given later, the main difference is the length of time the self-assessment material is presented to the students. The first was one day and the second extended to three weeks.

First, we will show the method and support material. Two types of adaptive tests were used: Multiple Answer Test (MAT) and Open Answer Test (OA-T). These kinds of tests are used in both stages. Therefore, an additional effort has been made to keep the format of the questions as similar as possible in terms of difficulty due to issues of consistency in the subsequent study.

In the first stage, the material was presented to the students at a very early stage of the evaluation test, so all the contents of the tests had already been presented in class. In the second one, it was possible to carry out tests with contents not yet exposed in the classroom since the material was available from the beginning. To compensate for this effect, the contents of each level of the adaptive test were made to coincide with the contents explained each week in class, so that the contents of the first week coincide with the questions of the first level of the test and so on.

As for the support material, the adaptive tests used were designed using the Unidirectional Levels Method (ULM), in the sense that once you have reached a level you cannot return to a previous one. When the student answers a certain number of questions correctly, they will automatically change the questions and their difficulty, thus advancing to the next level. In the case of the first stage, advancing to a more advanced level implies facing questions of greater difficulty or accuracy.

In the second stage, moving up leads to questions related to content explained later, which can lead to the student not having the necessary knowledge because those contents have not yet been exposed in the classroom. The most appropriate level format will be determined after analysing the results of the study. To manage the theory contents, a MA-T has been created, where the students have 3 possible answers to each question and only one is correct. 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 questions to 18 minutes.

On the other hand, 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 in each execution of the application. This kind of questions requires the following elements to be created:

Funded by the Erasmus+ Programme of the European Union

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

-Minimum and maximum values of each modifiable parameter.

-Programming code (Matlab in our case) that calculates the solution to the problem.

The OA-T tests had 6 and 9 questions in the first and second stage respectively.

Both figures were represented with the interface that is shown to the students when using the platform. While the first question has a limited number of possible answers, the second has a space for entering the numerical answer. In this figure we also show the feedback of the application when a correct answer has been introduced.

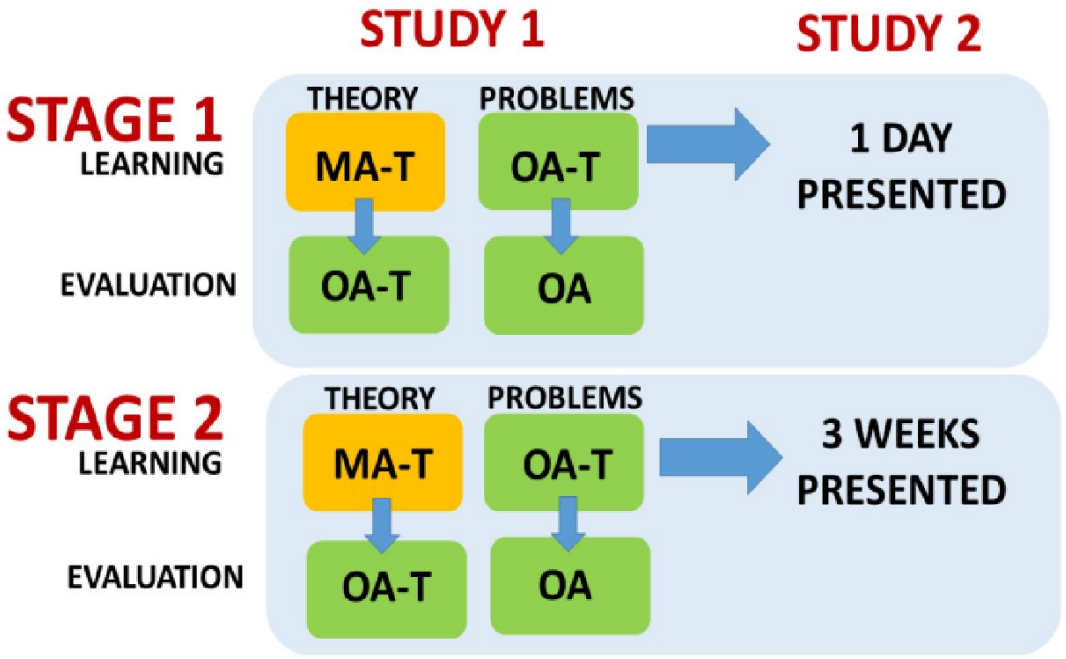


Figure 1: Protocol used in methodology 1.

The protocol of the study is shown in Fig.1. The proposed study is divided in two sub-studies. Both sub-studies have been divided in two stages. The first stage covers the first month of teaching and the second the following one. The first sub-study, called Recognition / Remembrance Study, seeks to evaluate the effect of studying with a system in which recognition memory is essentially used and, subsequently, an evaluation test is performed in which remembrance memory is required. The second

Funded by the Erasmus+ Programme of the European Union

sub-study, called Punctual/Continuous Study, analyses the differences between having access to the support material in a short period of time against having access throughout the whole learning process. Although the first study can be carried out in a single stage, the second one requires both stages in order to present the material to the students in the two formats.

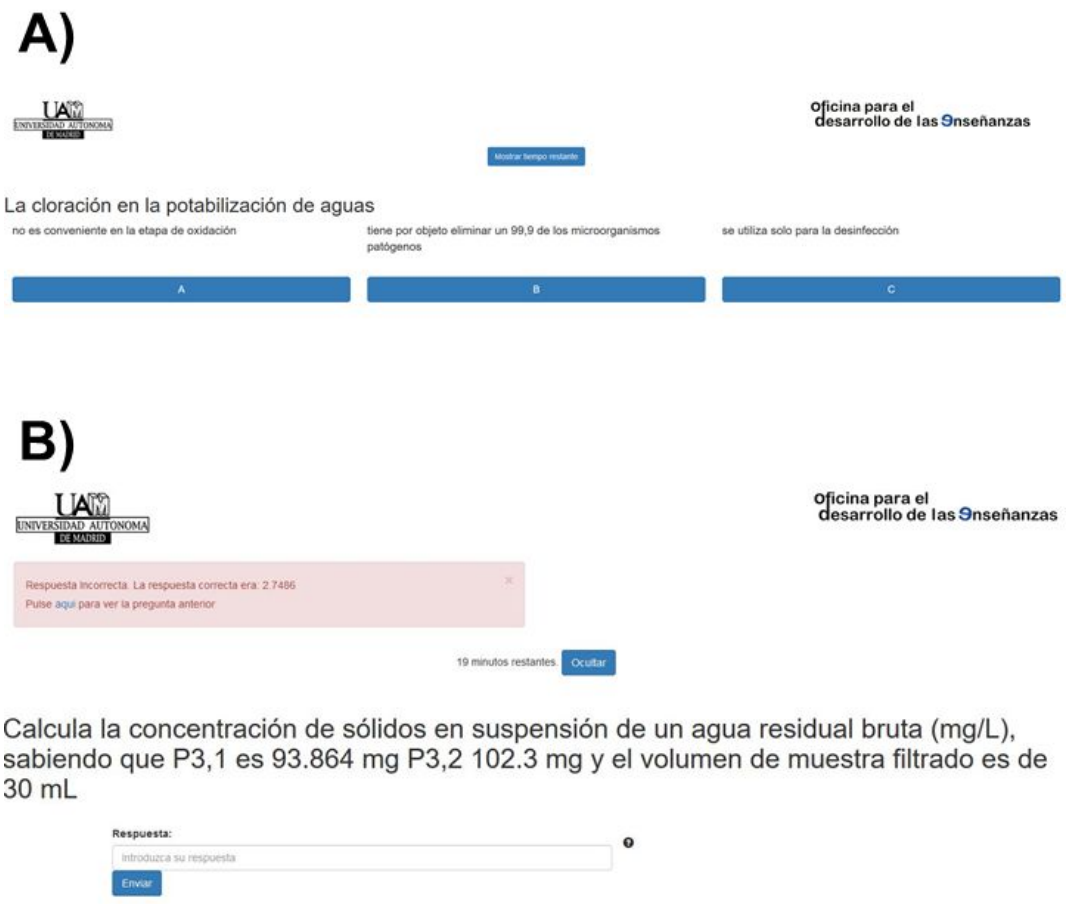


Fig.2. Questions in Stage 1. A) question to be answered from the Multiple Answer Test (MA-T) of the Learning process. B) question from the Open-Answer Test (OA-T) of the Evaluation phase. The notification in red indicates that the previous answer is wrong.

Funded by the Erasmus+ Programme of the European Union

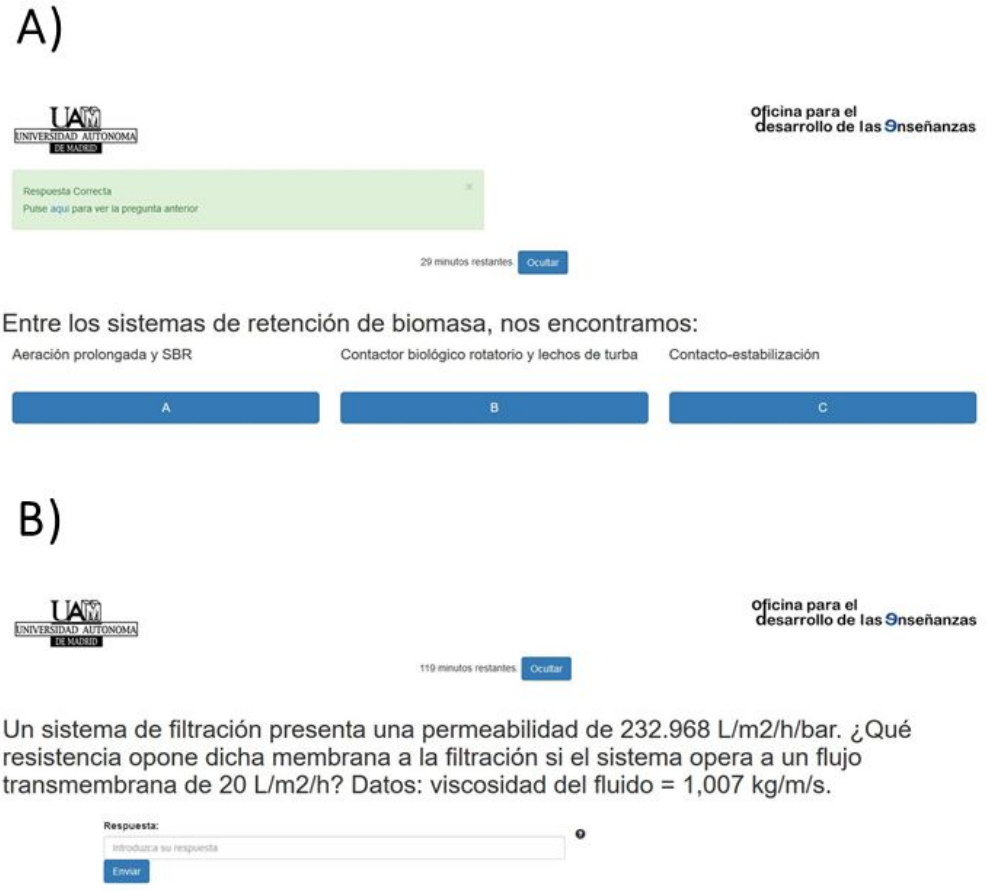


Fig. 3. Questions in Stage 2. A) question to be answered from the Multiple Answer Test (MA-T) of the Learning process. The notification in green indicates that the previous answer is right. B) question from the Open-Answer Test (OA-T) of the Evaluation phase.

Methodology 2: Avoiding wrong self-regulated learning strategies.

The system proposed in this case has been developed with a view to providing a system of self-assessment that can also be used as a system for final student assessment. This approach is very dangerous in the sense that students could focus their study strategy on learning only how to pass the

Funded by the Erasmus+ Programme of the European Union

assessment test instead of learning the contents correctly. To avoid this effect, the system will have the following characteristics:

1) A large number of questions. The more questions the students have to answer, the less likely they are to focus on learning the answers by heart. In addition, a sufficiently large number of questions would cover the entire agenda, so it would be irrelevant for the student to learn them by heart. In other words, if the student memorized how to answer all the questions, they would have learned the whole content of the subject, so that the objectives would be covered.

2) Even with a large number of questions covering the content of the course, there is the additional problem that students may learn the answers by heart rather than by procedure. To this end, we have designed a dynamic system in which the questions focus on numerical problems in which the parameters of the statement vary in each execution of the problem and, therefore, the answers also vary.

This system has the following additional advantages:

1) Questions are shown to students randomly (except for the restriction of the level of each question outlined below). This fact, together with the different results due to the dynamic parameters of the statement, makes it virtually impossible for students to copy from each other. In order to be able to copy, they would initially have to be lucky that two students located nearby were working on the same question. But, even then, the parameters of each one would be different, so they would have to copy the entire procedure, as well as being able to detect the points where the parameters have influence and modify them accordingly.

2) Being able to use literally the same tool during the learning/self-assessment phase and the formal assessment phase has the great advantage of providing very accurate and continuous feedback to learners about their learning process.

3) Another advantage of using the same tool in both phases is that the students have an exact knowledge of what evaluation system is going to be used, not being able to argue that the evaluation test surprised them, was very difficult or inadequate. These are the typical arguments that are used in exam reviews and they are eliminated with the use of our system.

4) The tool on which the assessment tests were developed generates a great deal of information on the attempts made by students throughout their learning process, which is very useful for teachers when it comes to detecting possible learner deficiencies and being able to prevent them with

Funded by the Erasmus+ Programme of the European Union

sufficient time.

In this methodology, the questions will have the following elements:

Statement: This is the text field in which the question is asked. As a special feature the numbered variable parameters need to be entered after the $ sign. The following text would correspond to a statement that includes two variable parameters.

"Write the result of the area of a rectangle of height $1 and width $2."

In this case, both the width and the height of the rectangle will be values that will change between different executions of the application, generating different results.

Maximum and minimum values of the variable parameters: Once the application is told how many variable parameters exist in the statement, limit values must be given for each parameter. The system will generate a random actual number between the two values.

Image: The system additionally accepts the inclusion of images in the statement. This content is optional.

Question level: The proposed method is included in evalUAM [23], which is an adaptive test application. The model used requires that the questions be divided into levels, where the lowest levels are initially shown to the students and only when a minimum number of correct answers are given, are they moved on to the next levels.

Source code of the solution: Since the numerical solution of the exercises varies in each execution due to the variable parameters, a code must be included to solve the problem and generate the solution. The application has been developed to support mathematical functions programmed in Matlab.

The graphical interface of the application at the time of adding a question is shown in Fig. 2. Format in which all the required parameters should be added can be seen.

Funded by the Erasmus+ Programme of the European Union

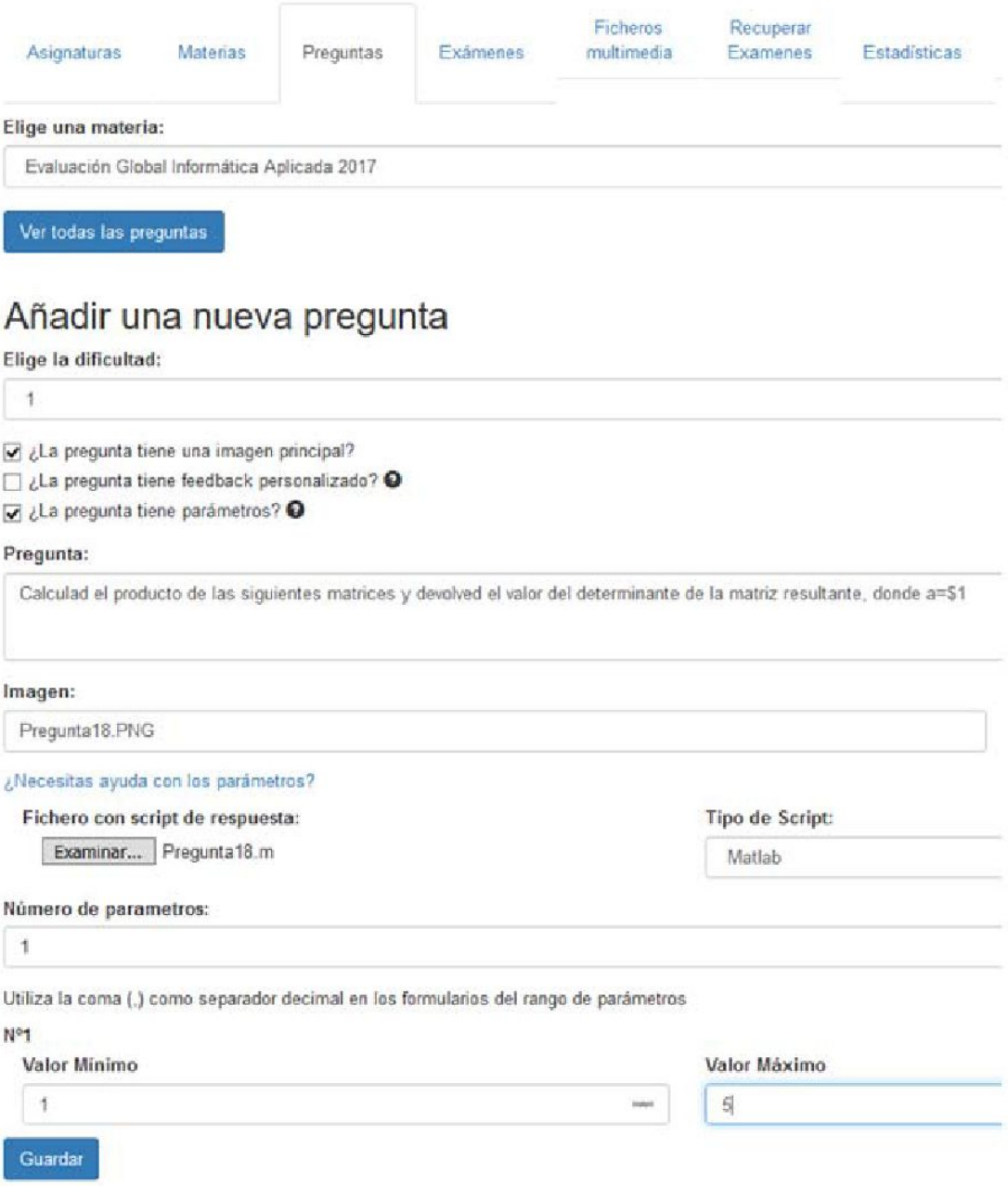


Figure 2: Graphical interface of the e-valuam application for teachers

Funded by the Erasmus+ Programme of the European Union

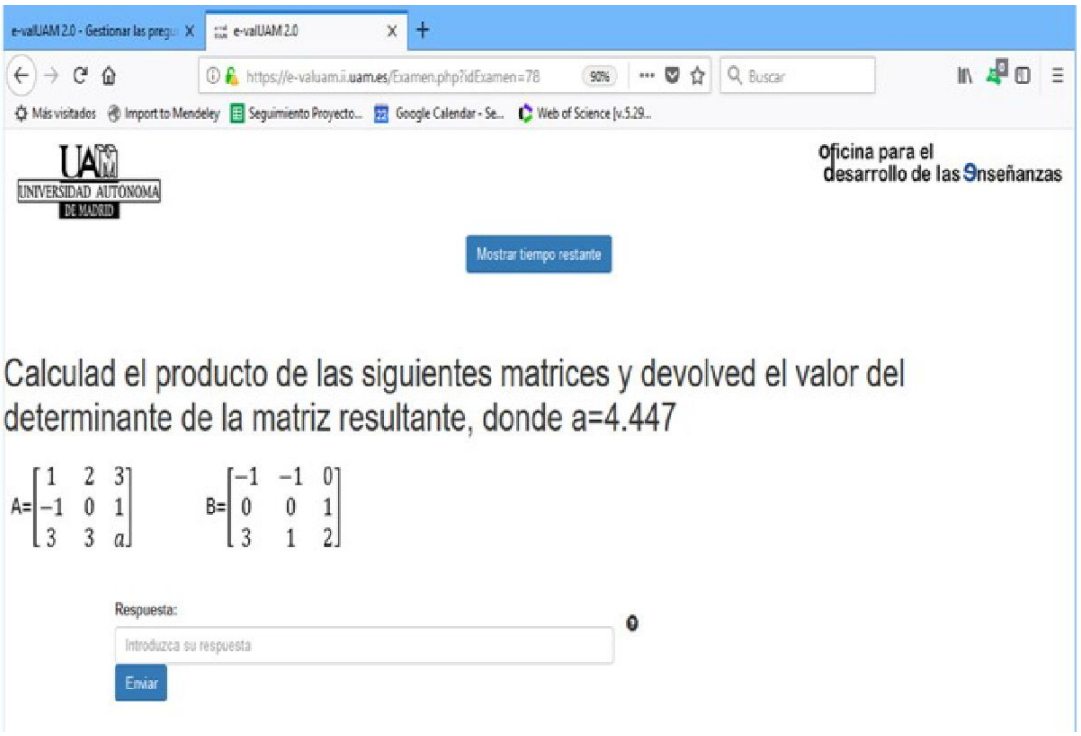


Figure 3: Graphical interface of the e-valuam application for students.

Figure 3 shows how students would view the question introduced in Figure 2. As can be seen, the variable parameter represented in the statement in Figure 2 as "$1" already appears with a value generated between the limits 1 and 5 that were specified when the question was created. For students, there is no way to know which parameter will change between one run and another.

Funded by the Erasmus+ Programme of the European Union

2. DATA

Methodology 1.

The experiences with real students took place during the 2017/2018, 2018/2019 and 2019/2020 academic years. Around 25 students were involved in the fourth year optative subject "Design of Water Treatment Facilities". This subject was taught through theory classes and practical classes in the laboratory. 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 Autonomous University of Madrid (UAM).

Fig. 5 shows the grades during the first stage of the study. In the case of MA-T, the grades have been ordered taking into account the number of times that the students used the application. In the case of the OA-T, the grades have been sorted according to the number of problems solved. In the data analysis, the results of students who tried answers from previous OA-T have been discarded. The problem with these students is that they do not try to solve the problems, as the parameters of the problems change with each execution and it is virtually impossible to get the same results in two consecutive executions. As this is an inappropriate use of the tool, this data has been omitted to avoid inadequate data in the analysis. Only 2 students have been omitted for this reason.

In both figures (5a and 5b) we can see a slight increase in the grades as attempts increase, which is in agreement with the initial premise that the use of this kind of tool is beneficial to students. An interesting fact extracted from the graphs is that students are more participatory in the MA-T. The number of students who have not used the application is 3 in MA-T compared to 8 in the case of OA-T.

Funded by the Erasmus+ Programme of the European Union

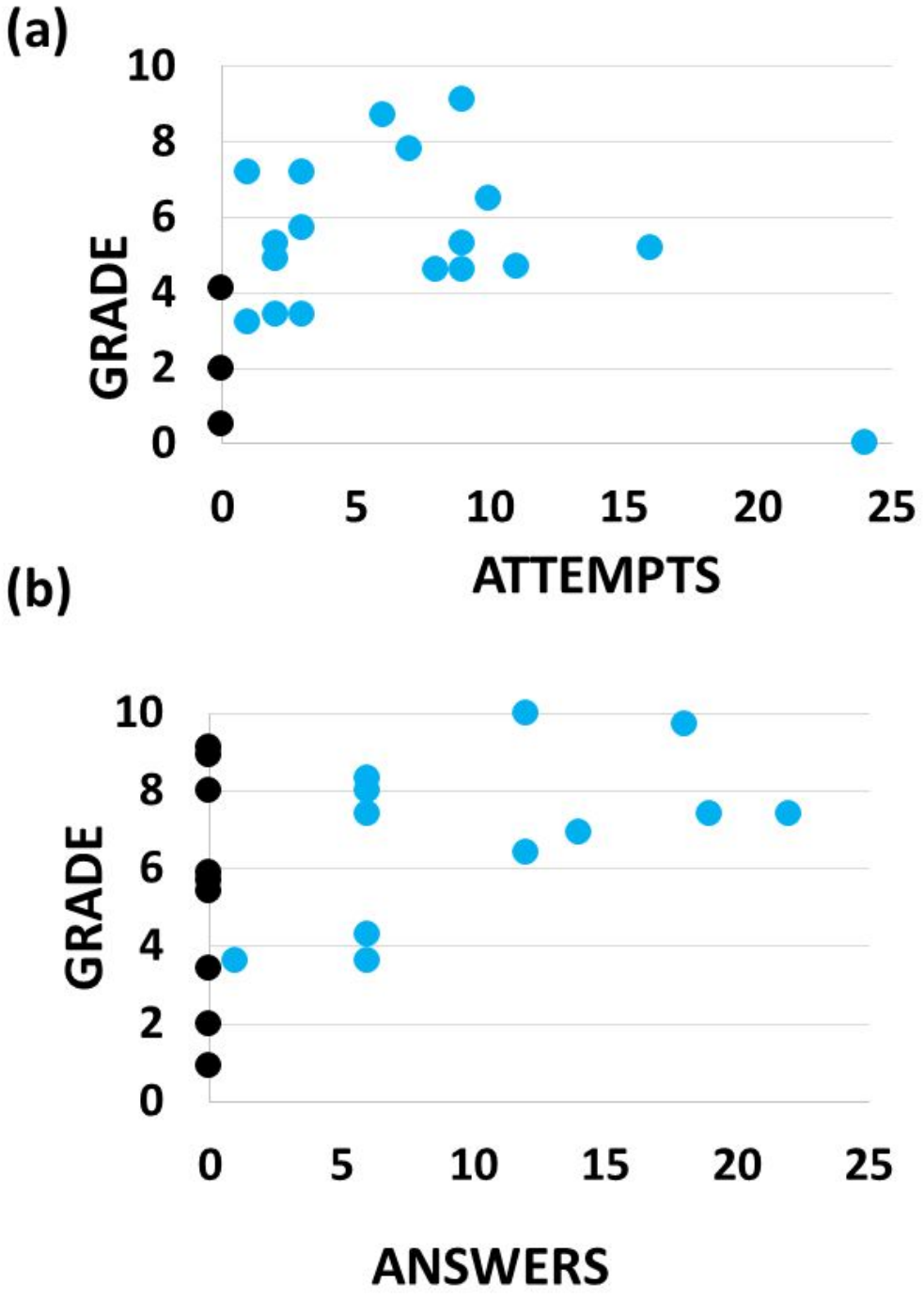


Fig.5. Student grades obtained in the first stage according to the number of attempts. In blue, students who used the tool 1 day and in black, students who didn't use it. (a) Grades on the MA-T. (b) Grades from OA-T.

Funded by the Erasmus+ Programme of the European Union

In the case of MA-T, there is a clear difference between the grades of students who have not used the application and those who did it, even when the number of attempts is very low. This can be explained by the similarity in content between the questions in this test and those in the assessment test. Additionally, since it is a multiple answer test, the application shows feedback on the correct answer, which is an additional support to the memorization of contents.

In the case of OA-T, grades from students who have not used the application show very varied results, which could imply that alternative study methods can be as effective as ours. However, observing the correlation between the grades and the number of attempts, we can conclude that the more persistent use of the tool is useful.

Fig. 6 shows the grades obtained in the second stage. In this case we have explicitly shown the results of the students who have used the tool for two or more days, which was not possible in the first stage. Surprisingly, in this second stage there has been less participation than in the first one. Also, there have been a greater number of students who have not used the tool.

Funded by the Erasmus+ Programme of the European Union

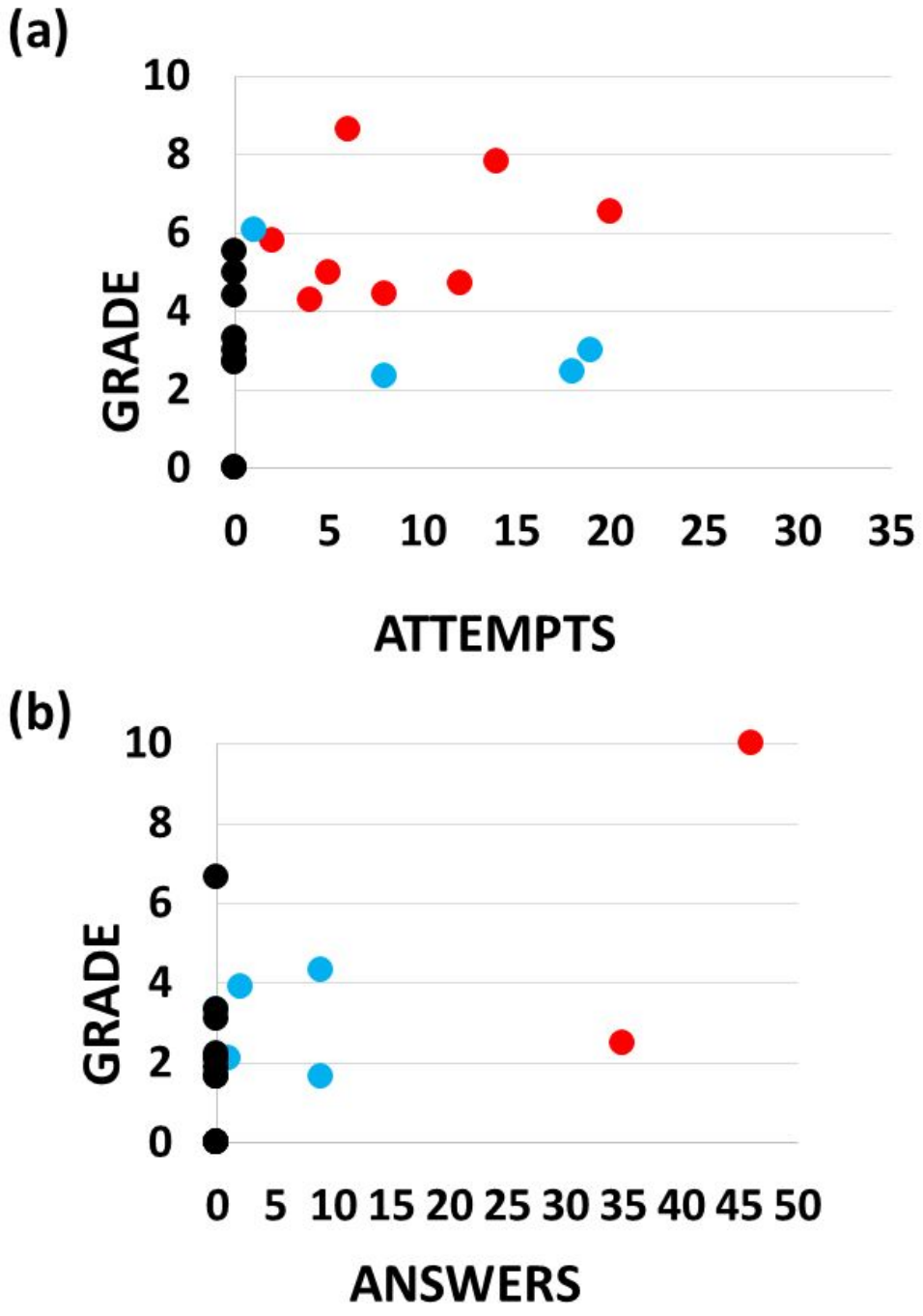


Fig. 6. Students' grades obtained in the second stage according to the number of attempts or questions answered, in ascending order. In blue, students who used the tool 1 day or less. In black, students who didn't use the tool and in red students who used it 2 days or more. (a) Grades on the MA-T. (b) Grades from OA-T.

Funded by the Erasmus+ Programme of the European Union

Focusing on the MA-T, 13/20 students have made more attempts in the first stage than in the second, which is shocking considering that in the second stage they have had the application available for longer. 3 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 have only 3 of them 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. Although only a few students worked continuously, Fig. 6a shows that the results of students working more continuously tend to be better, which is an expected result.

Methodology 2.

The experience with real students took place during the 2016/2017, 2017/2018, 2018/2019 and 2019/2020 academic years. 97 students were involved the first year. Around 85 students were involved in the following academic years in the first year elective course "Applied Computer Science". This subject was taught through theory classes and practical classes in the laboratory. This course corresponds to 6 ECTS and belongs to the Chemical Engineering Degree in the Faculty of Sciences of the Autonomous University of Madrid (UAM).

3. CONCLUSIONS

Methodology 1

As for the group of students who have used the application for more than one day, we have only 3 of them 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. Although only a few students worked continuously.

The proposed system makes use of a multiple-choice test for theoretical developments and an open question answering system for numerical problems. Both formats use adaptive tests in which the level is passed without the possibility of going back when a certain number of questions are adequately answered. For numerical problems, the test uses questions in which there is always at least one numerical parameter of the problem that changes in each execution. This way, students have the

Funded by the Erasmus+ Programme of the European Union

possibility to use the platform as many times as they want with the guarantee that the results will always be different.

In order to check the usefulness of the proposed system, a study has been carried out in which it has been demonstrated that the use of the application is beneficial for both objectives. It has also been found that students do not use the tool continuously, and are more likely to use it occasionally. Since it has been found that the continued use of the application can give beneficial results to students, the use of additional tools or motivation techniques is suggested to encourage the use of the platform more widely over time.

Methodology 2:

We have shown a method of developing adaptive tests that generate questions with numerical solutions in a dynamic way, not repeating the same values between different executions of the application. To ask these questions, the application has a specific format for writing the statements and a programming code that is responsible for generating the answers for each set of numerical values of the statement. Additionally, the questions are divided into levels that are introduced using an adaptive test model in which the questions are displayed from the lowest to the highest levels depending on the number of correct answers given by the student.

The system developed allows the same tool to be used during the learning phase and the final assessment phase. This greatly facilitates feedback between the students and the faculty, which is a requirement in current assessment models. A study has been carried out in which the test has been developed and used by real students from 2016/2017 to 2019/2020 academic years. This study demonstrates the effectiveness of the model and its potential to predict student workload together with the correct use of assessment tools dynamically throughout the course.

4. PUBLICATIONS

Results from these pilot experiences have been published in the following documents or conferences:

1) Adaptive Tests as a Tool for Evaluating Work Groups in Engineering

M.A. de la Rubia and G.M. Sacha

Funded by the Erasmus+ Programme of the European Union

International Journal of Engineering Education, 36, 411 (2020)

2) Self-assessment method based on questions with variable numerical values

Por: Comas-Lopez, M.; Sacha, G. M.

20th International Symposium on Computers in Education (SIIE) Ubicación: SPAIN SEP 19-21, 2018

Software Proc Improvement & Formal Methods Grp; Inst Univ Investigac Desarrollo Social Sostenible; ADIE; IEEE Sociedad Educac Capitulo Espanol; Univ Cadiz

2018 INTERNATIONAL SYMPOSIUM ON COMPUTERS IN EDUCATION (SIIE)

3) Sequential adaptive tests to improve the assimilation of contents during learning

Por: Comas-Lopez, M.; Molins-Ruano, P.; Atrio, S.; et ál..

20th International Symposium on Computers in Education (SIIE) SPAIN SEP 19-21, 2018

Software Proc Improvement & Formal Methods Grp; Inst Univ Investigac Desarrollo Social Sostenible; ADIE; IEEE Sociedad Educac Capitulo Espanol; Univ Cadiz

2018 INTERNATIONAL SYMPOSIUM ON COMPUTERS IN EDUCATION (SIIE)

4) Adaptive test system for subjects that simultaneously include theoretical content and numerical problem solving

Por: Comas-Lopez, M.; de la Rubia, M. A.; Sacha, G. M.

20th International Symposium on Computers in Education (SIIE) SPAIN SEP 19-21, 2018

Software Proc Improvement & Formal Methods Grp; Inst Univ Investigac Desarrollo Social Sostenible; ADIE; IEEE Sociedad Educac Capitulo Espanol; Univ Cadiz

2018 INTERNATIONAL SYMPOSIUM ON COMPUTERS IN EDUCATION (SIIE)

Funded by the Erasmus+ Programme of the European Union

Quality Evaluation:

Not included.

Funded by the
Erasmus+ Programme
of the European Union