

A survey of AulaWeb question types for continuous evaluation in a mandatory course

M.J. Villamide^{*}, Y. Alegre, M.R. Alvir, R. Carabaño, N. Nicodemus and D. Menoyo

Departamento de Producción Animal, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, España

The use of an on-line resource that self-evaluates the knowledge of students in a fast and dynamic way is especially useful when dealing with numerous groups of students where the continuous evaluation becomes a difficult and tedious task. AulaWeb is an interactive e-learning system which includes a self-assessment module with multimedia questions and several types of answers. The present study evaluates question accuracy and difficulty index assessment in a mandatory course at the U.P.M. To this end, a survey of the question type included in the AulaWeb self-assessment module was carried out. It is concluded that multiple choice and randomly generated questions give reliable results to evaluate the knowledge of a subject. The single choice ones are very affected by the random answers, and therefore their difficulty index is less stable. Difficulty indices computed by AulaWeb is a useful tool to check the standards of the questions included in the database towards a continuous evaluation in a mandatory course.

Keywords B-learning; question types, difficulty index

(Spanish Abstract) El uso de un entorno virtual que autoevalúe el aprendizaje de los alumnos de forma rápida y dinámica es especialmente interesante al trabajar con grupos numerosos donde la evaluación continua por los medios tradicionales es lenta y tediosa. AulaWeb es una plataforma virtual educativa interactiva que incluye un módulo de autoevaluación con distintos formatos de preguntas. Este estudio evalúa el tipo de preguntas y su índice de dificultad en una asignatura troncal en la UPM a través de la corrección automática del Aula Web. Se concluye que las preguntas de selección múltiple y de enunciado variable son las más adecuadas para evaluar los conocimientos del alumno. Las preguntas de selección simple se ven afectadas por el efecto del azar y por tanto su índice de dificultad es menos estable. Además, el nivel de dificultad calculado automáticamente por el entorno virtual Aula Web es una herramienta útil para determinar la calidad de la pregunta y así gestionar la base de datos de preguntas a fin de llevar a cabo la evaluación continua de la asignatura.

Palabras clave B-learning; tipo de pregunta, índice de dificultad

1. Introduction

The European Credit Transfer System (ECTS) is based on the student workload required to achieve the objectives of a programme, objectives preferably specified in terms of the learning outcomes and competences to be acquired [1]. Therefore, learning outcomes should give insight to the degree to which the student is actively engaged in learning, including demonstrating critical thinking, problem solving, knowledge and application. In other words, expressing what the student will know, understand or be able to do after completion of a process of learning, long or short. To properly assess the learning outcomes achieved by the students there is a need to use innovative teaching techniques and effective instructional materials allowing a higher interaction and communication between learners and educators in a more flexible and customized learning environment. This is possible through the implementation of blended learning referred as a combination of face-to-face teaching with on-line activities [2,3]. According to Loidl [2] an accurate application of the e-learning tools should help learners to manage their own learning since educators may be able to select the learning opportunities that best match their abilities, aspira-

^{*} Autor de contacto: e-mail: mariajesus.villamide@upm.es, Teléfono: +34 914524859

tions, needs and learning styles. Blended learning is also referred as continuous process of knowledge gathering and guidance [3]. This is especially useful when dealing with numerous groups of students where the continuous evaluation becomes a difficult and tedious task. Therefore, the use of an on-line resource that self-evaluates the knowledge of students in a fast and dynamic way is worth necessary.

AulaWeb is an interactive e-learning system which has been developed by the División de Informática Industrial of the Universidad Politécnica of Madrid (UPM) which is the largest technological university in Spain [4]. This e-learning system was installed in the academic year 2002-03 at the Escuela Técnica Superior de Ingenieros Agrónomos, being at this moment implemented in Zootecnia II, a mandatory course for students in Agricultural Engineering. The contents of Zootecnia II is designed to give students an idea of the fundamentals in animal science and the management of the main productive species (swine, poultry and sheep). The system includes a self-assessment module with multimedia questions of several types of answers: one-word, single or multiple choice, numerical, questions with randomly generated values [5]. This kind of module enables the students to check how far they have got and which parts of the subject matter they need to go over in more depth [2] but also provides the educators a useful tool to create and shape exercises to continuously gather the student learning progress [5]. However, once we decide to evaluate the knowledge of students based on this on-line tool, there is a key point that has to be raised. What kind of questions are the most accurate to be included in the self-assessment module?. The present study tries to answer this question in the particular case of Zootecnia II through a survey of the question type included in the AulaWeb self-assessment module.

2. The survey

The self-assessment module is based on a questions database created by the educators which is composed by a set of 513 questions structured in 18 didactic units. An important part of these questions come from the exams database of the last years adapted to the on-line format. Currently the database is composed mainly by single choice (54%) that included not only the true/false ones but also one among several choices, and multiple choice (29%) questions. Also there are numeric questions (8%), randomly



Fig. 1 Self-assessment test template in the student interface
generated questions (5%), that each one can generate an average of 200 different parameters, and one word answer (4%). See Figure 1 for a self-assessment question template in the student interface. The way

of scoring each question is correct or incorrect, with no intermediate scoring. This is important to assess the difficulty level ((number of answered questions – number of correct answers)/number of answered) of the questions as the module provides some statistical tools to grade the questions and to compare its theoretical and experimental difficulty level.

The self-assessment module allows two different kinds of exercises, the *customized* which is created by the student and the *programmed* configured by the educators to be saved in the student history [5]. The *programmed* one includes a set of questions (1..N) chosen by the educator at a random or in a predetermined way regarding the type of answer, the didactic unit, the difficulty level and the type of penalty for the wrong answers. Also, the timing and the deadline for the exercise have to be set [5].

3. Results

The difficulty index of the overall questions in the database, after being answered 360 times as average, varied in a wide range (from 5 to 94%), following a chi square distribution (Figure 2a) according to the way of scoring (correct vs incorrect). Therefore, the average difficulty was 31%, that is to say easy or medium easy, although 10% of the questions were difficult. The distribution of the difficulty index of each kind of question is shown in Figure 2b.

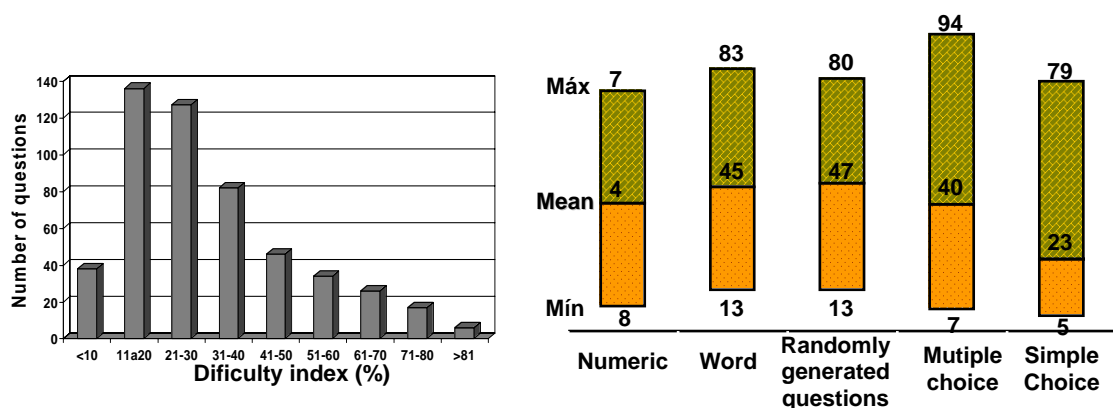


Figure 2. a) Distribution of difficulty index of overall questions, b) Distribution of difficulty index of questions according to their type

The average difficulty index is close related to the question type, being randomly generated questions the most difficult ones (47%), followed by one-word answer (45%) and numeric ones (40%). In these question types the range of difficulty varied from about 10 to 80%. The multiple choice type has the highest range of variation in the difficulty index, although the average difficulty is similar to above mentioned (40%). Finally, single choice questions were the easiest (23%), although the range of variation is also important. One of the reasons for the differences found in the difficulty index among questions is the positive and significant correlation (0.52, $P < 0.05$) between easiness and the probability to have a correct answer by random in multiple and single choice questions. When the easiness index is corrected by this random issue, multiple choice questions resulted to be the best to evaluate the knowledge of a subject (52%), whereas in single choice questions only the 30% of the easiness index is achieved after correction.

When question results were compared along the academic year (from 120 to 360 times answered as average) a decrease in the average difficulty index of the questions was observed, depending of the type of

question. Thus, the average difficulty index of one-word answer decreased 4 points, the numeric ones 1.4 points, randomly generated questions 1.3 points and multiple choice questions 0.8 points. This could indicate that the students search information about how to solve the most difficult questions, and therefore, this contributes to the learning process. On the other hand, the average difficulty index of single choice questions did not vary, although an important increase (maximum of 25 points) or decrease (maximum of 43 points) was obtained for the easiest and the most difficult questions, respectively. The most important decrease in the difficulty index of some questions was due to concepts not clearly treated in the lectures, or to concepts of general knowledge of other courses, and even to a non-common number of answers (i.e. all the choices in a multiple choice). On the contrary, the difficulty increases in the simplest questions due to confusions, no answers and random answers. A next step to consider is the reliability of those indices as a marker of difficulty through a survey of examinees perception.

4. Conclusion

It is concluded that among the questions examined those with multiple choice answer and randomly generated questions provides reliable results to evaluate the knowledge of a subject, because the single choice ones are very affected by the random answers, and therefore their difficulty index is less stable. Difficulty indices computed by AulaWeb is a useful tool to check the standards of the questions included in the database towards a continuous evaluation in a mandatory course.

Acknowledgements The authors acknowledge Raquel Martinez, Juan Antonio Criado and specially Angel García Beltrán their help for solving any doubt about AulaWeb platform.

References

- [1] European Credit Transfer System(ECTS)Key Features .http://ec.europa.eu/education/programmes/socrates/ects/index_en.html.
- [2] S. Loidl, Proceedings of the 3rd International Conference on Multimedia and Information & Communication Technologies in Education, Badajoz, Spain, 7-10 June 2005, pp. 84-88.
- [3] A. Jelfs, R. Nathan and C. Barrett, *Journal of Educational Media*, **29**, 2 (2004).
- [4] R. Martínez and A.García-Beltrán, Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, ED-MEDIA, Tampere, Finland, 2001 pp. 1239-1240.
- [5] A. García-Beltrán, R. Martínez. Proceedings of 2002 EDEN – *European Distance Education Network- Annual Conference*. Granada, Spain, 2002 pp. 302-307.