An Intelligent Personalized e-Assessment Tool Developed and Implemented for a Greek Lyric Poetry Undergraduate Course

Maria Xanthou

Aristotle University of Thessaloniki, Faculty of History and Archaeology, Department of Ancient Greek, Roman, Medieval and Byzantine History, GR-54124 University Campus, Greece

mxanthou@lit.auth.gr

Abstract: The scope of this paper is to present the development of a web-based evaluation system that adjusts to the level of knowledge of each student. The underlying concept is to split the questions into pools (bins) of distinct difficulty at the same time assigning a different level of accuracy to each possible answer that the student may select. Given the weighted success score computed at each round, the system raises or lowers the level of the questions asked in the following round, thus adjusting the difficulty of the test according to the students' performance. Once the score of the student improves or declines, the system again adjusts the difficulty of the questions accordingly until convergence is achieved and the students' performance becomes stable. It is believed that the proposed evaluation system predicts the level of knowledge of the students in more correct ways compared with the traditional (non-adaptive) questionnaires, while giving the students the motivation to rerun the test until they become familiar with the course content. To illustrate the evaluation methodology proposed, an online system was developed and implemented for the evaluation of sixty-eight undergraduates attending a Greek Lyric Poetry course at the Department of Classics in Aristotle University of Thessaloniki, Greece. The results of the evaluation process, with the observations made, both about the operation of the system in real conditions and the performance of the students, are also discussed in the paper.

Keywords: e-learning; online assessment; adaptive assessment; learning management systems; web-based systems; ancient Greek literature; Greek lyric poetry

1. Introduction

The establishment of e-learning both as an academic educational discipline and as a learning medium through the vast expansion of the World Wide Web during the last three decades (Kahiigi et al., 2008) calls for the implementation of Learning Management Systems (LMS) as one of the prominent characteristics of developing e-tools in higher education worldwide (Duffy & Kirkley, 2004; Oncu & Cakir, 2011). Through the constant flow, the invention and storage of information on the web, e-learning tested the boundaries of time and space as strictly defined by the conservative mode of education (Graff, 2003; Collins & Halversont, 2010). Nowadays, e-learning tools have become more sophisticated (Oncu & Cakir, 2011).

1.1 Key features of advanced sophistication in effective e-learning

Adaptiveness is a key-notion to the effective sophistication of e-learning (Wang, 2010; Yalcinalp & Gulbahar, 2010), because the latter is inherently linked to the learners' prior knowledge (Mitchell et al., 2005; Smits et al., 2008). An LMS is called "adaptive", when, firstly, it includes a pre-planned reconstruction of the space in which the student operates (or should move); secondly, it directs the student toward the solution of a problem through rational decision making; thirdly, it provides personalized and ideal help at the right moment; fourthly and lastly, it repeats key-points of the course, when LMS "senses" that its user has deviated from its routine.

In general, there are two categories of adaptive, personalized, "smart" processes of learning incorporated into the modern composite distance learning environments. The first one is adaptive presentation: this smart application can adapt the content of hypermedia pages to the learner's objectives, knowledge and general profile. Therefore, these pages are not static but created or composed on a personal level. In this way, the well-read students receive more detailed and in-depth information, as opposed to the more unprepared to whom only explanations are mainly provided.

The second one is adaptive navigation, which supports the learner, while the learner navigates in a dynamically configured Internet environment (for example, through modulation of visible hyperlinks). The ways to customize the links have been studied and analyzed in the international secondary literature. Examples of this technology exist today in a number of sites (for example, www.cnn.com), which provide a personalized content (Kaplan et al., 1998).

It is noted that according to some researchers these two aforementioned categories, namely, adaptive presentation of teaching materials and adaptive navigation, form the category of Adaptive Educational Hypermedia (Triantafillou et al., 2003; Mampadi et al., 2011,). According to this distinction, through Adaptive Educational Hypermedia the user has sufficient freedom of choice during the user's web-navigation (Triantafillou et al., 2003; Mampadi et al., 2011), as opposed to Intelligent Tutoring Systems through which the system checks largely what is presented to the user.

1.2 Assessment and the effective use of online questionnaires

In this framework of constant e-evolution, assessment as an intrinsic corollary of the learning process has also been reinvented, either as its definitive conclusion (summative assessment), or as a helpful medium for weighing the "mental load" (formative assessment) provided to the students through the learning process (Graff, 2003; Hwang & Chang, 2011). Its role in the learning process is still highly valued (OECD 2010 PISA Executive Summary: Results), because its double purpose is to provide feedback not only to guide the learner throughout the learning process (Oncu & Cakir, 2011), but also to help the instructor reform the guidance offered to the students and the teaching activities (Wang, 2011). It is exactly this feedback that accounts for the positive effects of an assessment on learning (Wang, 2011) and, because of these effects, the idea of "assessment as teaching and learning strategy" has been proposed (Wang, 2010).

Today, the link of LMS with assessment remains a requirement (Oncu & Cakir, 2011), although the importance of CBE (Computer Based Assessment) has been documented (Terzis & Economides, 2011). So far, only Hyde et al., 2004 have comprehensively surveyed the role of assessment in LMS with exclusive emphasis on Vocational Education and Training (VET) programs. As already stressed, a similar survey on the role of assessment in LMS with exclusive emphasis on Academic Education (AE) is long overdue, although there is an explicitly expressed interest on this subject (Govindasamy, 2002; Byrnes & Ellis, 2006; Bang, 2006; Selwyn, 2007; Jones & Healing, 2010).

The use of online questionnaires for collecting data and providing feedback has been popularized by the increasing use of the Internet and the proliferation of Online Learning Environments (Ortigosa et al., 2010; Oncu & Cakir, 2011). Especially, in the case of Adaptive Hypermedia Systems (AHS), online questionnaires provide feedback and information about the user, and are stored and maintained to establish the user model (Kobsa, 2001). Adaptiveness as a key feature of AHS depends on this user model. Moreover, one of the students' features, usually detected through online questionnaires and often used for adaptiveness purposes, is their learning style (Ortigosa et al., 2010). In addition, the increasing number of students, enrolled in tertiary education (Costa et al. 2010), calls for alternates to on-campus teaching (Krause et al., 2005; Greyling et al., 2008). These alternates are provided using web-based delivery of course content and assessment (Anderson et al., 2002).

The proposed AH-questionnaire bears two major instructional characteristics of dynamic assessment. Firstly, it provides people with an opportunity to learn (Bransford et al., 1987; Graff, 2003; Wang, 2010). Secondly, both instruction and feedback could be built into the testing process (Elliott, 2003).

In this framework, it is possible to use the specific questionnaire in either the synchronous and asynchronous learning context, if adjusted appropriately (Offir et al., 2008). A major advantage of the proposed AH-questionnaire is that both formats of dynamic assessment can be used (Sternberg & Grigorenko, 2001), enabling its use for self-assessment and self-improvement, especially in e-learning contexts.

Despite the above advantages in the use of online questionnaires, major difficulties are often present as in all problems solved by automated processes. These problems are usually associated with traditional assessment methods (either electronic or manual) that ask a specific, predefined set of questions, uniformly to all students while disregarding issues raised by the following disclaimers: some questions are answered at random and their potential accurate guess cannot be trapped during evaluation; the inherent rigid nature of the non-adaptive questionnaires does not promote the use of the assessment as part of the overall learning process; even if the answers to the questions posed are indeed provided, this is often made after the examination, therefore, having limited impact to the learning outcome; it is rare that a student goes through the same questionnaire more than once, without being compelled to; students often act in an unpredictable way (Graff, 2003), not only with their answers but also in actions (for example, logging out of the system without a reason,

unexpected selection of non-relevant navigation buttons, abrupt termination of the evaluation and such like); any questionnaire, however detailed, is inherently a different process compared with a personalized one-to-one oral examination. It is typical, during an oral assessment where the student provides inconsistent answers, for the tutor to ask a larger number of questions of various difficulties, until the tutor comes to a conclusion on the student's actual level of knowledge.

In this context, the purposes of this research is, firstly, to develop a methodology for adaptive e-assessment and e-learning system that can quickly evaluate the current level of students' knowledge and automatically build a student-dependent scenario by adjusting the difficulty of the questions; secondly, to implement this methodology into an online system that can integrate the e-assessment and e-learning process; thirdly, to ensure that the developed system can operate in academic conditions and keep track of the history of student's behavior so it can automatically compute the best successful rate out of all student efforts; fourthly, to apply the tool in an actual class, with emphasis on a course of Ancient Greek Literature where, to the best of the authors' knowledge, it has never been used, with the only exception of Camastra et al., 2005, focusing exclusively on Latin literature; fifthly, to study the relevance of the system in an academic environment and at the same time the behavior of the students while using the adaptive tool; and lastly, to investigate the contribution of the tool in the improvement of the students' performance.

It is foreseen that the methodology and system developed simultaneously act as both as e-assessment and e-learning tool. There follows a discussion of the structure of the system and its application in class.

2. Methodology

2.1 The concept and principles of the proposed adaptive system

As already mentioned, the purpose of this study is to propose the development and the use of an up-to-date learning management tool to enable the intelligent adaptation of e-learning teaching and assessment depending on the level of knowledge and behavior of its end-user. This entails that the help and knowledge, which the end-user acquires through the system is tailored to the knowledge level that the user has acquired at the time of the user's participation in the electronic environment. This goal is pursued through an electronic assessment of adaptive difficulty, decided during run-time based on the system, by, firstly, distinguishing the questions according to their varying difficulty and allocating them in various pools, and, secondly, assigning an "accuracy" index to each possible answer provided to the end-user (in other words, the student's response is not treated on a "right / wrong" basis, but also on the additional "how right or how wrong the reply was" information, both defined *a priori* by the instructor.

Based on the student's (user's) answers, the algorithm computes a temporary (namely, partial) score and draws the following questions from a specific pool (bin) of the corresponding difficulty. In each way, each test is eventually a different experience of a different evaluation scenario. The process finishes when the student, while shifting from one knowledge (difficulty) level to the other, exhausts all twenty-five questions corresponding to the relevant level.

Based on this weighted structure between questions and students' responses, the system computes the weighted score and defines the difficulty of the questions of the next evaluation round. In this way, each student has a unique and different experience every time the student runs the evaluation test, ensuring a different path and, thus, a different "learning scenario". Through the student's continuous effort to achieve higher scores and discover different evaluation paths, the e-evaluation environment is turned into a complementary e-learning scheme.

In terms of the overall concept, the above process is a form of adaptive evaluation, the natural continuation of different levels of intelligence (curriculum sequence, problem solving, adaptive presentation and adaptive navigation).

2.2 System structure

When looking at the system architecture, the e-evaluation system structure can be broken into various steps as follows:

2.2.1 Step 1: The classification of questions into categories based on their difficulty

The instructor has to create a large pool of various questions, each corresponding to several levels of difficulty, the latter defined in the scale 0 to 100 (with 100 as the most difficult question). It is necessary to group the questions into sub-pools of similar difficulty, each (pool) having the same number of eligible questions. For the demonstration and the implementation presented here, one-hundred and twenty-five questions are used, divided into five categories (sub-pools) of differentiated difficulty. The coefficient of difficulty of questions in each category is given in Table 1:

Table 1: Question difficult levels and corresponding difficulty coefficient ?

Difficulty Level	Difficulty level coefficient ? (in the scale 0 to 100%)	Description
1	0% to 20%	Trivial
2	20% to 40%	Quite simple
3	40% to 60%	Medium difficulty
4	60% to 80%	Quite demanding

2.2.2 Step 2: Classification of possible answers based on their level of accuracy

Each question can be answered using multiple choices, which in the present demonstration is decided as equal to 5. It is the instructor's responsibility to assign a weighted score to each answer based on the students' different success rate, which can be equal to 100%, 75%, 50%, 25% or 0% respectively (100% corresponding to a correct answer).

2.2.3 Step 3: Evaluation initialization (1st round)

Once the user (student) starts the evaluation test, the user has to answer the first bulk of five questions, each drawn from a different category of difficulty. Depending on the question's difficulty level (where i=1 to 5 different level of question difficulty), and the user's response success rate (where j=1 to 5 different answer "correctness"), a preliminary (first round) partial evaluation score $PS_{k=1}$ (k corresponding to evaluation round 1), is computed using a simple weighted average formula:

$$PS_1 = \frac{\sum_{i=1}^{5} a_i \beta_i}{\sum_{i=1}^{5} a_i} = \frac{a_1\beta_1 + a_2\beta_2 + a_3\beta_3 + a_4\beta_4 + a_5\beta_5}{a_1 + a_2 + a_3 + a_4 + a_5}$$

A typical example of this first round partial score, where the students answers correctly (100%) only the simplest question a_1 (with difficulty 15%), while also answers a question of difficulty level 2 (with a_2=37%) partially correct (at a success rate 75%), a question of moderate difficulty (with a_3=56%) with moderate success (50%), a demanding question (a_4=73%) in a wrong way (success rate 25%) and completely fails to answer the most difficult question a_5=92%) is given below:

$$PS_1 = \frac{\sum_{i=1}^{5} a_i \beta_i}{\sum_{i=1}^{5} a_i} = \frac{15x100 + 37x75 + 56x50 + 73x25 + 92x0}{15 + 37 + 56 + 73 + 92} = 32.60\%$$

2.2.4 Step 4: Decision on the difficulty level of each following round of questions

At this stage the system classifies the user in a preliminary way, and, thus, makes a first assessment of the level of questions that correspond to the user's knowledge level (as defined in Table 1). For instance, in the example presented above, a first partial score of 32.6% would lead the system to draw five new questions, all from the pool of difficulty 2. In case the user achieves a good score, the system poses the difficult questions, which offer better difficulty rates so as the user keep the score high. On the contrary, in the event the user achieves a moderate or average score, the system adapts to a corresponding difficulty level and offers the user a new set of questions that better match the level of knowledge so the user can achieve a better score. If the user's

(student's) score drops further, the system offers even easier questions so the user can recover and improve his total score, then more difficult questions are gradually posed.

In every evaluation round, the partial evaluation score PS_k (k corresponding to the round number), can be derived as follows:

$$PS_k = \frac{\sum_i^{5k} a_i \beta_i}{\sum_{i=1}^{5k} a_i}$$

In other words, at the end of each round k, the partial evaluation score PS_k is derived as the ratio of the sum of the products of the difficulty weight (coefficient) by its corresponding success rate $\alpha_i \beta_i$ of all questions posed to the student (whose number is equal to the number of rounds k multiplied by five [5] questions per round) divided by the sum of the individual difficulty weights α_i of each question.

2.2.5 Step 5: Next rounds and evaluation termination

Every time a question is asked it is marked as used and cannot be posed twice in the future (namely, in the same test session). The adaptation of the question's difficulty based on the partial score of the user is continued until the questions of a specific difficulty level are exhausted and the system finally returns the final overall score. Clearly, the minimum number of rounds that a student has to complete is five, and the minimum number of questions is 5 x 5 = 25. But this situation corresponds to the rare case where the student's score constantly corresponds to a single difficulty level (from those defined in Table 1). For instance, the student starts with 32.6% but keeps in the range 20% to 40% (difficulty level 2) for five successive rounds.

However, in the typical case that the student's performance is not consistent, namely, the student starts with a good score, not kept high, or the studentscores low in the first rounds, but significantly improves in the following, the user gradually changes difficulty levels and, hence, the system is compelled to offer additional sets of questions until the user's score stabilizes and the student's evaluation can be deemed reliable. More details on the proposed concept are available elsewhere (Xanthou, 2006).

3. Architecture of the web-based system

3.1 Adaptation of the system for an actual Greek Lyric Poetry undergraduate course

Apart from introducing the above concept for e-evaluation and e-learning, and presenting the fundamental ideas behind it, it was deemed necessary to develop an online system so the merits and drawbacks of the algorithm could be better identified. Furthermore, it was aimed to use this system for a student evaluation at university level. The following section presents the implementation of the above algorithm through the analytical pool of one-hundred and twenty-five questions devised to support the teaching of a compulsory undergraduate course on Greek lyric poetry [GLP (=Greek Lyric Poetry), course ID=105], offered in the curriculum of Classics at the School of Philology of Aristotle University of Thessaloniki, in Greece. More specifically, the content of the course included a description of the life of the Greek lyric poet Pindar and a critical analysis of his literary works. Moreover, the course content was articulated in modules that were provided in two forms: (a) as seventy five RLOs (Reusable Learning Objects) and (b) as lectures delivered in an amphitheatre with students keeping notes and references made by the instructor to both primary sources (ancient Greek texts) and secondary bibliography (monographs and articles) for further reading. In addition, during lectures the instructor discussed important issues of critical analysis of Pindaric texts. As a result, the subsequent questions of the e-assessment tool focused on the mental load of information provided through RLOs and viva voce teaching. Based on the previously mentioned provision, the one-hundred and twenty-five questions have multiple answers (each characterized by a different success rate), and each question has its own difficulty level. It is noted that a similar pool of questions with suitable modifications can be used as a model for teaching the work of other Greek poets, both ancient (e.g., Aeschylus, Sophocles, Euripides) and modern (e.g. Solomos, Palamas, Sikelianos).

3.2 PHP module development and system substructuring

The implementation of the specific system was based on the use of the PHP: Hypertext Preprocessor (PHP) programming language, particularly efficient for developing web applications with dynamic content that enable user interaction. A PHP page is processed by a compatible web server (typically an Apache server with a My Structured Query Language [My SQL] database), to produce, in real-time, the final content sent to the users' browser as HyperText Markup Language (HTML) code. It is noted that the specific programming language was used not only because of being popular (today more than 16.000.000 web sites, at a rate of more than 35% of web pages, are using scripts written in PHP language), but primarily because PHP is essentially an open and free web development environment. A sample view of the developed MySQL content is shown in Figure 1:
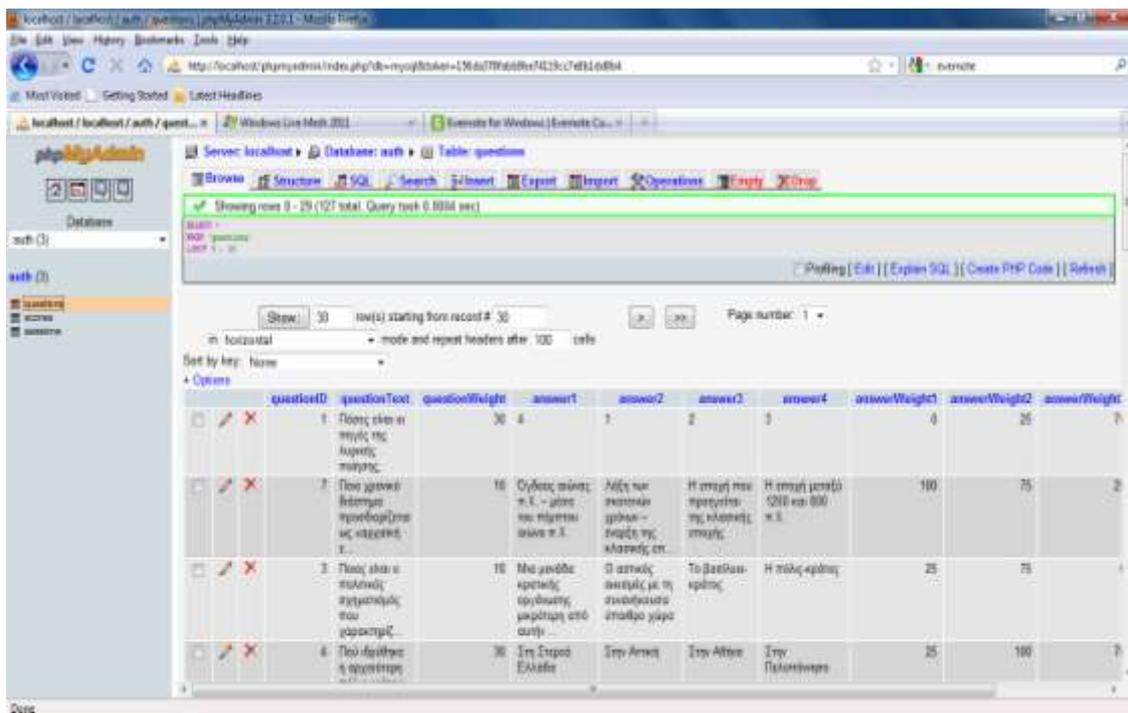
Question ID	Question Text	Question Weight	Answer 1	Answer2	Answer 3	Answer 4
1	How many are the cultural sources of lyric poetry?	30	4	1	2	3
7	What is the time span designated as "archaic age"?	10	Eight century B.C.E. – middle fifth century B.C.E.	The end of the Dark Ages and the beginning of the classical era	The period which precedes classical era	The period between 1200 and 800 B.C.E.
3	What is the preponderant political formation of archaic age in Greece?	10	A unit of state organization smaller than the city	The city along with the rural space	The kingdom-state	The city-state
4	Where was the oldest political formation of city-state founded?	30	In Continental Greece	In Attica	In Athens	In Peloponnese

Figure 1: Sample view of the MySQL content (in Greek, up), English index (bottom)

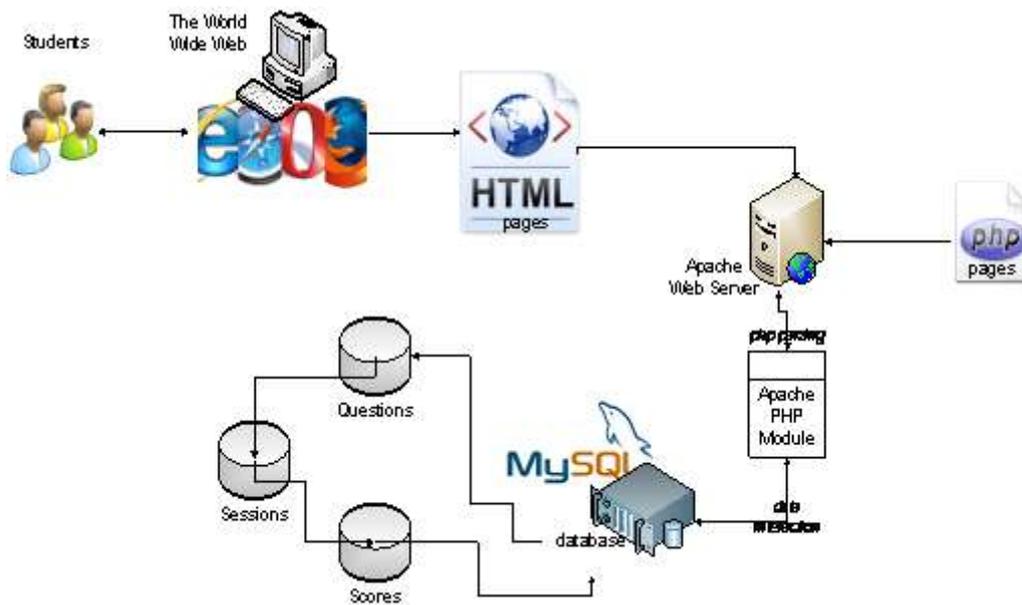While the complete structure of the system is illustrated in Figure 2:

Figure 2: Overview of the evaluation scheme developed

It is noted that the database consists of three relational tables, namely: firstly, sessions (i.e., a complete test comprising of a number of rounds independently taken by a student), storing data related to each test (session): time and date of the test, name of student, student identification, student e-mail, final score achieved; secondly, questions, storing data related to questions: question identification, question text, answers identification, question weight, individual answers success rate; thirdly, scores, storing data related to the history of partial scores in each test (session): partial scores and corresponding round identification, final session score.

Primary keys (such as session identifications) are used to make sure that each test will be treated and stored as an individual identity. All three database tables can be easily exported as spreadsheets for further post-processing. The outcome of the online test (questionnaire) in real-time is shown in Figure 3.

4. Implementation in class

4.1 Preparation of the evaluation process and the results management

Having established the concept of the adaptive e-evaluation and developed the web-based system for the case of the Greek lyric poetry course, it was decided to implement the system under real conditions, that is, for the actual evaluation of the students in class. For this purpose, all students were invited on a specific date and time to the Joint Computer Laboratory (Room 104) of the School of Philosophy at Aristotle University of Thessaloniki and were split into five groups. They were all notified beforehand that participation in the process is optional and their final scores will be used anonymously to extract results on their interaction with the e-tool. Upon their acceptance to enter the process they signed their name and identification number in a list. Responsible for organizing and implementing the evaluation process was the author (who developed the methodology and the software and is also familiar with the specific academic field (Xanthou, 2007; Xanthou, 2010) in the framework of her postdoctoral research funded by the Hellenic State Scholarships Foundation and supervised by Professor John N. Kazazis, the scientific and academic teacher responsible for the course. First, the students were given the web address (http://155.207.34.75/questions.php) to access the system online. Then, the students could finish the test in their own time at their own will or to repeat it as many times as they preferred. The motivation here was to improve their score, but, above all, to become familiar with the context of the course by exploring various paths and evaluation scenarios. All results were stored in real-time on the web server. After the test, the records related to the students' performance were exported into Excel files and processed separately in the form discussed below. Moreover, the students were encouraged to repeat the test at home, using their own computers and accessing the system with their own user name, password and

identification number. In addition, they were also encouraged to answer the questions posed by the adaptive system simultaneously using the reading material for the specific undergraduate course.
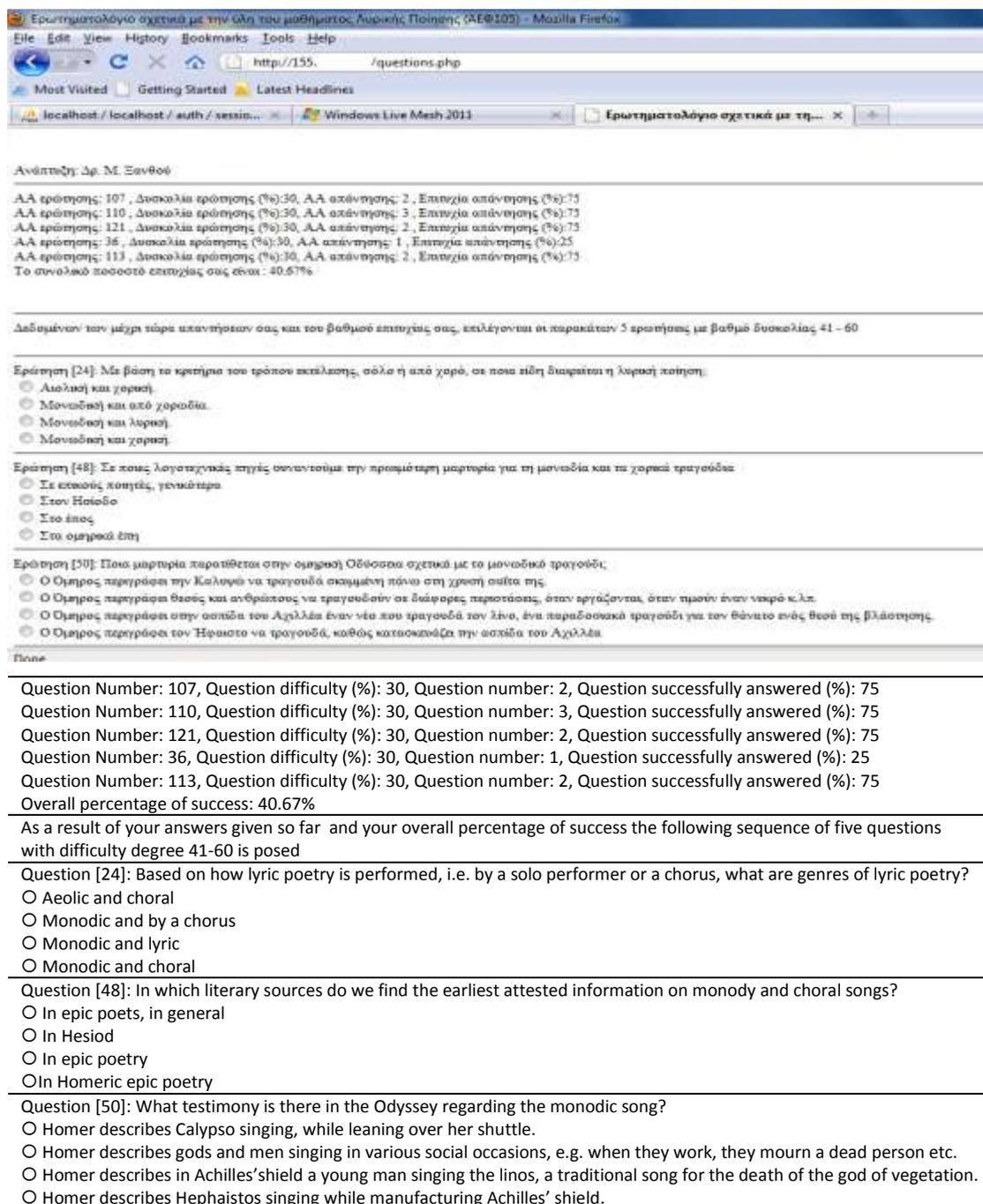

Question Number: 107, Question difficulty (%): 30, Question number: 2, Question successfully answered (%): 75
Question Number: 110, Question difficulty (%): 30, Question number: 3, Question successfully answered (%): 75
Question Number: 121, Question difficulty (%): 30, Question number: 2, Question successfully answered (%): 75
Question Number: 36, Question difficulty (%): 30, Question number: 1, Question successfully answered (%): 25
Question Number: 113, Question difficulty (%): 30, Question number: 2, Question successfully answered (%): 75
Overall percentage of success: 40.67%
As a result of your answers given so far and your overall percentage of success the following sequence of five questions with difficulty degree 41-60 is posed
Question [24]: Based on how lyric poetry is performed, i.e. by a solo performer or a chorus, what are genres of lyric poetry?
O Aeolic and choral
O Monodic and by a chorus
O Monodic and lyric
O Monodic and choral
Question [48]: In which literary sources do we find the earliest attested information on monody and choral songs?
O In epic poets, in general
O In Hesiod
O In epic poetry
O In Homeric epic poetry
Question [50]: What testimony is there in the Odyssey regarding the monodic song?
O Homer describes Calypso singing, while leaning over her shuttle.
O Homer describes gods and men singing in various social occasions, e.g. when they work, they mourn a dead person etc.
O Homer describes in Achilles'shield a young man singing the linos, a traditional song for the death of the god of vegetation.
O Homer describes Hephaistos singing while manufacturing Achilles' shield.

Figure 3: System run-time (in Greek, up), English index (bottom)

4.2 Implementation results and observations

4.2.1 *Sample size and scores achieved*

The first statistics that the instructor had to compute and evaluate were related to the scores achieved by the students as a whole (independent of how many times each student took the test). As seen in Figure 4, the distribution of the final score is reasonable, in the sense that from a qualitative point of view, it resembles a normal distribution around a mean value equal to 59.5%. This observation implies that the test was neither too easy nor too difficult, a fact that is paramount given the inherent adaptive nature of the algorithm used. It is also noted that the number of tests (one-hundred and one) was 60% higher than the number of students that

took the test (sixty-eight), and clearly, almost half the students repeated the test to improve their score. It is also noted that in thirteen cases, the students restarted the test at their own will, because their score in the first round was equal to zero. Therefore, it can be considered that the overall sample size was equal to one hundred and fourteen tests.
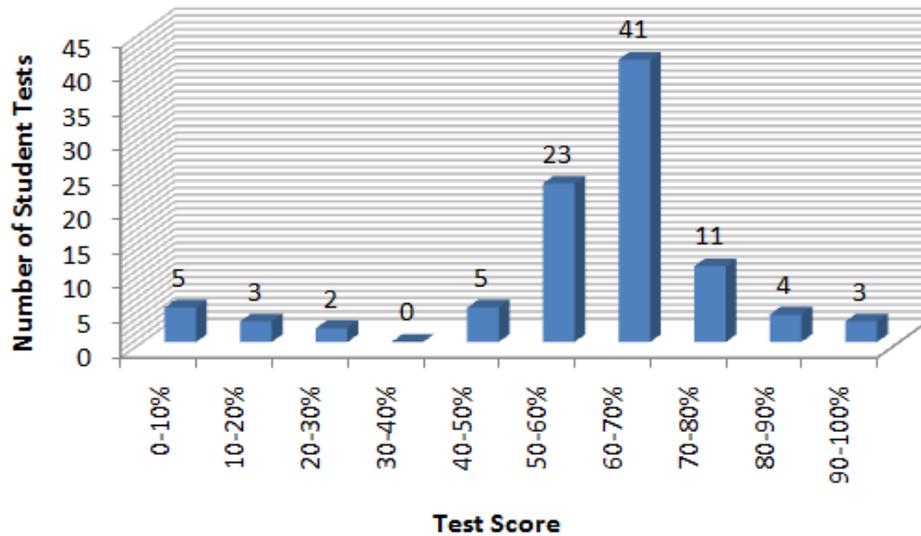

Figure 4: Distribution of final (total) scores achieved by the students during the examination in class (sample size: one-hundred and one sessions completed by sixty-eight students)

4.2.2 Rounds needed until a final score was assigned

As already described above in Section 4, the number of rounds required until the system can assign a final score to the student is not fixed. On the contrary, it depends on the consistency of the students' response. In other words, if the student's performance abruptly drops or improves from one round to another, the system imposes additional questions until the performance of the student stabilizes.
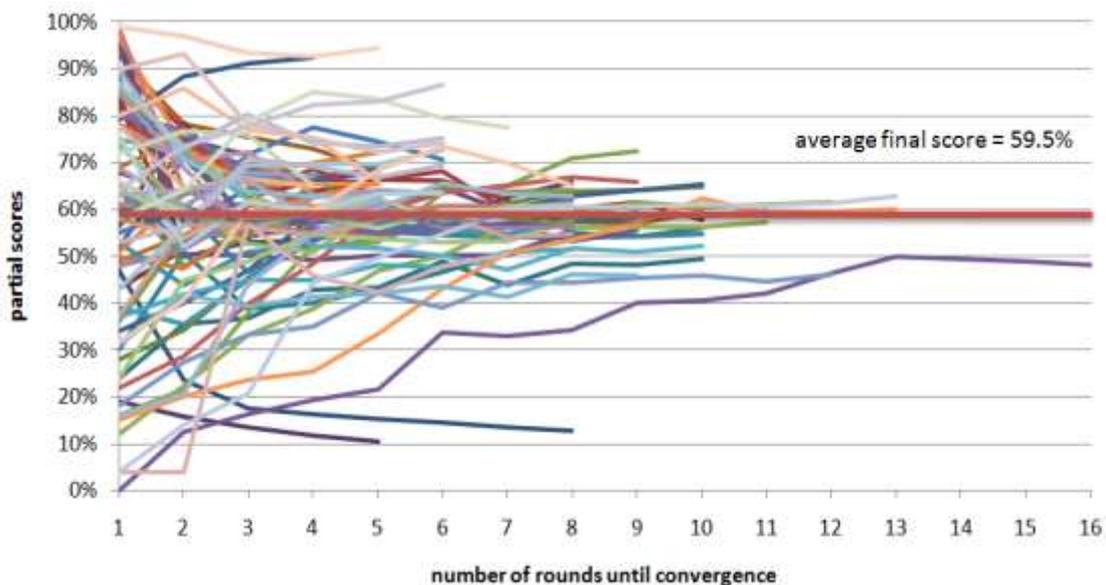
Figure 5: Overview of the complete cloud, i.e. trend, of partial scores achieved in the same session against the number of rounds needed until the system converged and assigned a final score

This adaptive behavior of the online evaluation system is clearly illustrated in Figure 5, where the cloud of all sessions recorded is shown as the partial score history against the round number. It is also observed that in general, most students' performance history tends to converge after five to eight rounds around the mean value of 59.5% described earlier.
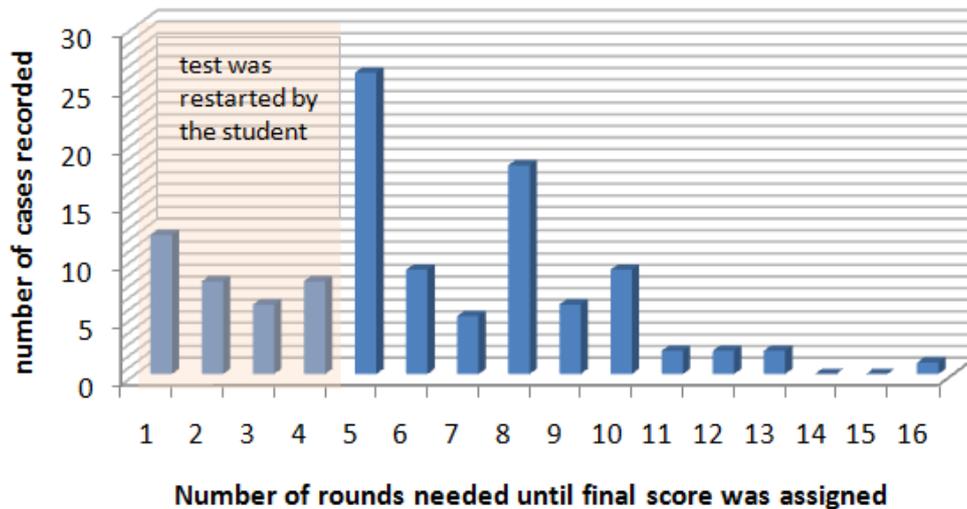
Figure 6: Overview of the distribution of the number of rounds needed for the student to be assigned a final score by the evaluation system (final score assigned once any bin of questions of a given difficulty is exhausted). It is noted that the scores achieved in only 1-4 rounds are temporary and correspond to cases where the students quit the test due to poor initial performance

This is further illustrated in Figure 6 where the distribution of the number of rounds needed for the student to be assigned a final score is plotted.

4.2.3 Discussion of the results

Since Figure 5 presents only a broad overview of the system performance and no detailed observations can be made to the many data plotted in the same chart, it is deemed preferable to present below few characteristic cases of partial scoring history. In particular, the progress of the partial score development in five characteristic sessions (tests) is plotted in Figure 7:
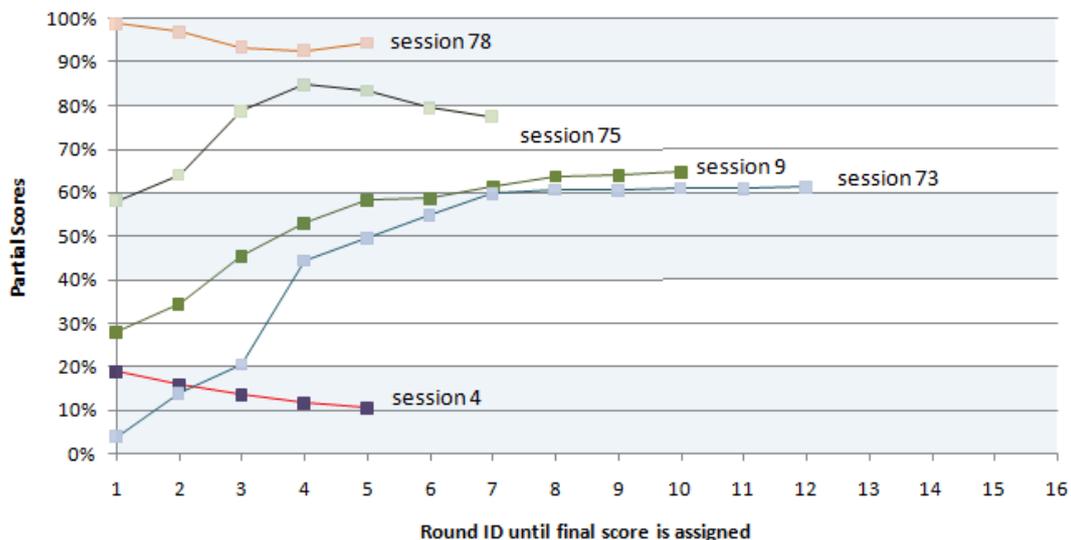
Figure 7: Characteristic sessions stored during the implementation of the system in class. Partial scores progress with round identification until final score is assigned

Clearly, sessions 78 and 4 finish at the fifth round, because the students that took the specific tests were consistent in their performance (namely, the student in session 78 never dropped below 80% and remained in difficulty level 5 during the whole test, while on the other hand, the student in session 4 could never make it higher than difficulty level 1 because the score remained below 20% in all rounds). Session 75 and 9 required a larger number of rounds until the final score was assigned, because during the tests the student's partial score

shifted between two and three difficulty levels respectively. As for session 73, the most interesting of all, it took the system thirteen rounds until it could assign a final score, because the student started by scoring low (4%), then significantly improved the student's performance by rising to just higher than 60% on the seventh round. From this stage onward, the student could not score higher and kept relatively close to the boundary between difficulty level 3 and 4. The system offered the student the opportunity to score higher, but this was not made feasible, given the level of knowledge, finally assessed at 63%, as shown in Table 2.

Table 2: Example of score build up for the case of session 73 (as it relates to Figure 7

Round Identification	Partial Score	Bin used for questions	times bin was used
0	-	All Bins (one question each)	0
1	4.00%	Bin 1 (0% to 20%)	1st
2	14.00%	Bin 1 (0% to 20%)	2nd
3	20.57%	Bin 2 (20% to 40%)	1st
4	44.40%	Bin 3 (40% to 60%)	1st
5	49.60%	Bin 3 (40% to 60%)	2nd
6	54.70%	Bin 3 (40% to 60%)	3rd
7	59.76%	Bin 3 (40% to 60%)	4th
8	60.63%	Bin 4 (60% to 80%)	1st
9	60.51%	Bin 4 (60% to 80%)	2nd
10	61.05%	Bin 4 (60% to 80%)	3rd
11	60.90%	Bin 4 (60% to 80%)	4th
12	61.28%	Bin 4 (60% to 80%)	5th (Bin exhausted – session terminated and final score assigned)

It is believed that the system interacts with the student. This means that its gradual adaptation to the student's performance provides an insight that would not be available using conventional questionnaires, which given the student's response during the first five rounds (namely, twenty-five questions), would have provided a significantly lower score (25%).

Three additional points should be discussed here based on the server records (Figure 8):
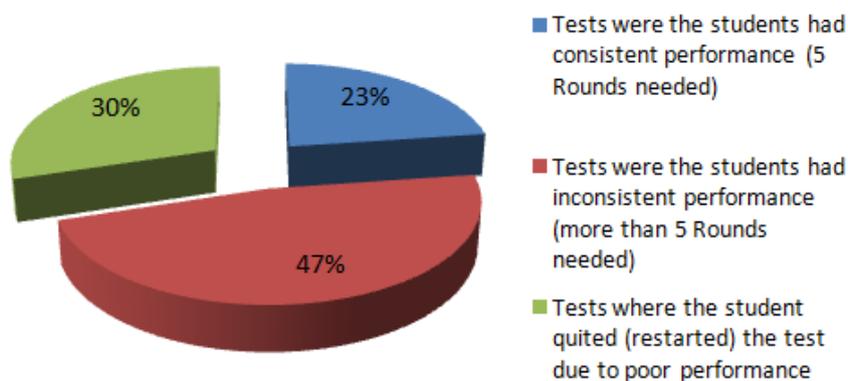

Figure 8: Overview of the students' performance and behavior

Firstly, the cases, where the student's performance was consistent (namely, no abrupt changes in their score was observed and as such all questions were drawn from a single pool of a given difficulty level), were only 23%; secondly, the cases, where more rounds were needed until the system could assign a reliable score, was approximately double (47%), a fact that demonstrates the necessity of adaptation in the evaluation process;

thirdly, in 30% of the cases, the students decided to restart the test before the assignment of a final score, that is, in fewer than five rounds, because they considered that the partial scores in the first rounds were low. It is believed that this indicates the motivation of the students to repeat the test by being more careful and better prepared. This motivation for rerunning the test is further demonstrated in Figure 9:
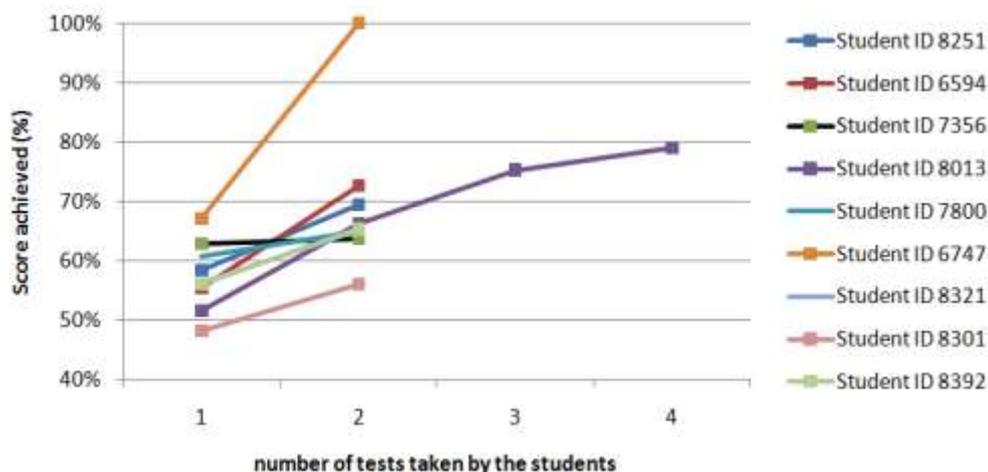

Figure 9: Score achieved and number of tests taken by the same student

where the final score achieved is plotted against the number of times that the same student repeated the test. Clearly, almost always, the more the students repeated the test, the higher was their performance, although the evaluation scenario was different by definition. It can, therefore, be claimed that the specific e-evaluation system is at the same time a useful e-learning tool that encourages students to be involved in the evaluation process, while managing to evaluate them in a justified and reliable way.

5. Key-issues and concepts related to the literature review and the research findings

The main goal of this paper was to propose an adaptive e-assessment tool for an academic course on Greek Lyric Poetry, devised exclusively for the students, who attended it. The author designed the proposed tool after taking into consideration the possible benefits, limitations and issues raised for the students who eventually used this tool. Its major long-term benefit that it proposed a personalized management of the "mental load" of a learning process. In that sense, it could be used either as a definitive conclusion (summative assessment), or as a helpful medium for weighing the "mental load" (formative assessment) of a learning process. Moreover, it could help a student restructure the way the student processes knowledge. This means that the student could use the adaptive e-assessment tool in combination with the hardcopy material provided by the instructor, or use it to check the credibility of the knowledge extracted from e-resources. Another advantage of this adaptive e-assessment tool is that it could be used simultaneously by more than students working as a group, thus promoting synergy and discussions between them on issues they believe are important for understanding major issues of the course regarding ancient Greek civilization and literature. This promotes the students' awareness as researchers and users of the internet, since they could look up their answer into the vast e-library of the internet and assess the credibility of the e-sources.

6. Conclusions and recommendations

The aim of this paper is to propose the concept and to develop the computational framework for an intelligent, e-evaluation and e-learning tool that adapts to the performance of the student during an online questionnaire with a special emphasis on teaching a course on Greek Lyric Poetry at academic level. As both the questions and the potential answers are weighted according to their relative difficulty and accuracy respectively, the system decides in successive rounds the new knowledge level that the student is to be tested against. After describing the idea behind this evaluation approach, the paper also describes in detail the development of the corresponding algorithm as well as the online system structure. Furthermore, the system was tailored to the needs of an actual undergraduate course on choral Lyric Poetry and was actually implemented in class. It is noted that, to the best of the knowledge of the author, this is the first time that such a tool has been developed and used for Ancient Greek literature.

The main conclusions drawn by the specific research effort can be summarized as follows: firstly, the adaptive evaluation framework proposed is a feasible and effective alternative, which provides a more realistic assessment of the student's level of knowledge, because it reveals cases where many questions have to be posed until a final evaluation can be made. Such cases in which the student had to pass through additional evaluation rounds, were found approximately double compared with the cases where the student's performance was consistent (or easily predictable); secondly, the adaptive nature of the online questionnaire, and the many evaluation scenarios and paths that the student may explore, provides the system with an attractive e-learning tool as well; thirdly, the performance of the students was usually found to be improved with the increasing number of tests taken, although the evaluation scenario was different in each case.

Based on the above information, it can be claimed that, while the physical presence and contact with the instructor is always the communicative medium par excellence, concerning the transmission of knowledge, similar adaptive educational tools can significantly contribute toward a more interactive, and hence, more efficient, meritocratic educational framework that cannot be reproduced using conventional or traditional means.

A major goal that could be met in the future is the adaptation of the e-assessment tool for other courses linked with teaching ancient Greek literature, which includes many different literary genres. Apart from that the designer could develop an added application for the students to write their comments, while they use the e-assessment, for the instructor to see and weigh further the reason for their answer. This could provide the instructor with further feedback on the students' choices and reach to a better assessment of their overall score.

Acknowledgements

References

Anderson, D., Johnson, R., & Saha, L. (2002). Changes in academic work: Implication for Universities of the changing age distribution and work roles of academic staff. http://www.dest.gov.au/sectors/higher_education/publications_resources/summaries_brochures/changes_in_academic_work.htm. Retrieved 06.03.2011.

Bang, J., (2006). "eLearning reconsidered. Have e-learning and virtual universities met the expectations?". http://www.elearningeuropa.info. Retrieved: 05/03/2011.

Bransford, J. C., Delclos, J. R., Vye, N. J., Burns, M. & Hasselbring, T. S. (1987). State of the art and future directions. In C. S. Lidz (Ed.), Dynamic assessment: An interactional approach to evaluating learning potential, Guilford Press, New York, pp. 479-496.

Byrnes, R. & Ellis, A. (2006). "The prevalence and characteristics of online assessment in Australian universities". Australasian Journal of Educational Technology, 22.1, 104-125. http://ascilite.org.au/ajet/ajet22/byrnes.html. Retrieved 07.03.2011.

Camastra, P., Fedeli, P., & Grattagliano, M. R. (2005). "The Ancient Latin Language e-learning Course at the University of Bari". Proceedings of the 1st WSEAS / IASME Int. Conf. on Educational Technologies, Tenerife, Canary Islands, Spain, December 16-18, 138-142.

Collins, A. & Halverson, R. (2010). "The second educational revolution: rethinking education in the age of technology". Journal of Computer Assisted Learning, 26, 18-27.

Costa, D. S. J., Mullan, B., Kothe, E. J., & Butow, P. (2010). "A web-based formative assessment tool for Masters students: A pilot study". Computers & Education, 54, 1248-1253.

Duffy, T. M., & Kirkley, J. R. (2004). Learner-centered theory and practice in distance education: Cases from higher education, Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Elliott, J. G. (2003). "Dynamic assessment in educational settings: Realizing potential". Educational Review, 55, 15-32.

Jones C. & G. Healing (2010). "Net generation students: agency and choice and the new technologies". Journal of Computer Assisted Learning, 26, 344-356.

Hwang, G.-J., Chang, H.-F., (2011). „A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students". Computers & Education, 56, 1023-1031.

Hyde, P., Clayton, B. & Booth, R. (2004). Exploring assessment in flexible delivery of vocational education and training programs. National Centre for Vocational Education and Training (NCVER) & Australian National Training Authority (ANTA), Adelaide. http://lofrances.edublogs.org/files/2009/07/hyde_13casestudies.pdf. Retrieved: 06.03.2011.

Govindasamy, T. (2002). "Successful Implementation of e-learning: Pedagogical considerations". Internet and Higher Education, 4, 287-299.

Graff, M., (2003). "Cognitive Style and Attitudes Towards Using Online Learning and Assessment Methods". The Electronic Journal of E-Learning, 1.1, 21-28.

Greyling, F., Kara, M., Makka, A., & van Niekerk, S. (2008). "IT Worked for Us: Online Strategies to Facilitate Learning in Large (Undergraduate) Classes". The Electronic Journal E-Learning, 6.3, 179-188.

Kahiigi, E. K., Ekenberg, L., Hansson, H., Tusubira, F.F. & Danielson M. (2008). "Exploring the e-Learning State of Art". The Electronic Journal of E-Learning, 6.2, 77-88.

Kaplan, G., Fenwick J., & Chen, J. (1998). "Adaptive Hypertext Navigation based on User Goals and Context". In Brusilovsky, P., Kobsa, A., & Vassileva, J. (edd.). Dordrecht, Academic Publishing, The Netherlands, pp. 45-70.

Krause, K.-L., Hartley, R., James, R., & McInnis, C. (2005). The first year experience in Australian Universities: Findings from over a decade of national studies. University of Melbourne.

Kobsa, A. (2001). "Generic user modeling systems user. Modeling and user-adapted interaction". The Journal of Personalization Research, 11, 49-63.

Mampadi, F., Chen, S. Y., Ghinea, G., & Chen, M.-P. (2011). "Design of adaptive hypermedia learning systems: A cognitive style approach". Computers & Education, 56, 1003-1011.

Mitchell, T. J. F., Chen, S. Y., & Macredie, R. D. (2005). "Hypermedia learning and prior knowledge: Domain expertise vs. system expertise". Journal of Computer Assisted Learning, 21, 53-64.

OECD (2010), PISA 2009 Results: Executive Summary. [online] (http://www.oecd.org/dataoecd/34/60/46619703.pdf) Retrieved: 05/03/2011.

Oncu, S., & Cakir, H. (2011). "Research in online learning environments: Priorities and methodologies". Computers & Education, 57, 1098-1108.

Ortigosa, A., Paredes, P., Rodriguez, P. (2010). "AH-questionnaire: An adaptive hierarchical questionnaire for learning styles". Computers & Education, 54, 999-1005.

Selwyn, N. (2007). "The use of computer technology in university teaching and learning: a critical perspective". Journal of Computer Assisted Learning, 23, 83-94.

Smits, M. H., Boon, J., Sluijsmans, D. M., & van Gog, T. (2008). "Content and timing feedback in a web-based learning environment: Effects on learning as a function of prior knowledge". Interactive Learning Environments, 16, 183-193.

Sternberg, R. J., & Grigorenko, E. L. (2001). "All testing is dynamic testing". Issues in Education, 7, 137-170.

Terzis V., & Economides, A. A. (2011). "The acceptance and use of computer based assessment". Computers & Education, 56, 1032-1044.

Triantafillou, E., Pomportsis, A., Demetriadis, S. (2003). "The design and the formative evaluation of an adaptive system based on cognitive styles". Computers & Education, 41, 87-103.

Xanthou, M. G. (2006). New Technologies and Classics: e-guidance in teaching ancient Greek texts through online learning management systems [=Νέες τεχνολογίες και κλασικές σπουδές: η ηλεκτρονική καθοδήγηση της διδασκαλίας κλασικών κειμένων μέσω τεχνολογικών περιβαλλόντων διαχείρισης μάθησης (LMS)], Unpublished MSc Thesis, Department of French Literature and Language & Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki.

Xanthou, M. G. (2007). Asyndeton in the Epinician Odes of Pindar and Bacchylides: exploring its function as a figura rhetorica and narratological mode. Unpublished PhD Thesis, School of Philology, Faculty of Classics, Aristotle University of Thessaloniki.

Xanthou, M. G. (2010). "Contextualizing Dikaiopolis' Persona: Urban Life, Rural Space, and Rural Perceptions of Urbanity in Aristophanes' Acharnians". Hellenika, 60.2, 297-314.

Wang, T.-H. (2010). "Web-based dynamic assessment: Taking assessment as teaching and learning strategy for improving students' e-learning effectiveness". Computers & Education, 54, 1157-1166.

Wang, T.-H. (2011). "Implementation of Web-based dynamic assessment in facilitating junior high school students to learn mathematics". Computers & Education, 56, 1062-1071.

Yalcinalp, S. & Gulbahar Y. (2010). "Ontology and taxonomy design and development for personalized web-based learning systems", British Journal of Educational Technology, 41.6, 883-896.