МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2006
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2006
Proceedings of the Thirty Fifth Spring Conference of
the Union of Bulgarian Mathematicians
Borovets, April 5–8, 2006

A FRAMEWORK TO COMPARE THE TESTING CAPABILITIES IN e-LEARNING PLATFORMS*

Daniela I. Dureva-Tuparova, Georgi T. Tuparov

The aims of this paper are to develop a framework for analyzing of testing capabilities of e-learning platforms from didactical and technological point of view and to apply this framework in comparing of testing features of some popular open source e-learning platforms. The paper is organized as follows: In Section 1 the framework for comparing different sides of testing based on didactical and technological criteria is suggested. Section 2 deals with analysis of characteristics of some popular open source e-learning platforms according to the proposed framework. In Conclusions some ideas for improving of testing capabilities are suggested.

Introduction. Today "e-learning" becomes one of the most popular terms. Recently a lot of e-learning platforms – commercial or open source have been developed. Also in the web space found comparative studies for most popular platforms, based on different comparative frameworks. [1,2,4,10]. The main goals in most of comparative studies of e-learning platforms are directed to brief overview of their general features: used technologies, standard interoperability for presentation of learning materials, type of used communications among the teachers and students etc. Comparing of testing facilities is insufficiently presented. It is well-known that assessment and evaluation of student's achievements have a control function in educational process. Feedback takes an important role both in "face to face" learning and in e-learning. The contemporary e-learning environments of course offer the testing features. These features are performed in different ways.

The aims of our study are:

- to develop a framework for analysing of testing capabilities of e-learning platforms from didactical and technological point of view;
- to apply this framework in comparing of testing features of some popular open source e-learning platforms with interface localization in Bulgarian language and possibilities to do a real study of their capabilities.

In our study we do not consider and analyse the commercial e-learning platforms due to impossibility to do real experiments with them.

The paper is organised as follows:

In Section 1 a framework for comparing different aspects of testing based on didactical and technological criteria is suggested. Section 2 deals with analysis of characteristics of

*The study is partially supported by the Bulgarian National Science Foundation under contract No VU MI–111/2005

some popular open source e-learning platforms according to the proposed framework. In Conclusions some ideas for improving of testing capabilities of e-learning platforms are suggested.

1. A Framework to Compare the Testing Capabilities of e-Learning Environments. We propose a comparing framework based on two general issues of e-learning environments – didactical and technological. Didactical issues of the framework are grounded on pedagogical theories for test and test's items classification, construction and analysis. [3] The technological issues are directed to the concrete technological implementation of testing capabilities of e-learning platforms.

1.1. Didactical issues.

- Used type of test items according to the pedagogical classification's [3]
  - Short-answer item/ Completion item – Supply-type test items that can be answered by a word, phrase, number or symbol
  - True-False or Alternative Response Item – Consists of declarative statement that the pupil is asked to mark true or false, right or wrong, yes or no, agree or disagree, and the like.
  - Matching Exercises – the matching exercise consists of two parallel columns with each word, number, or symbol in one column being matched to a word, sentence, or phrase in the other column.
  - Multiple-Choice Item – consists of a problem and a list of suggested solutions. This type could be divided in two basic subtypes – with only one answer and with more than one answer.
  - Interpretative exercise – consists of a series of objective items based on a common set of data. The data may be in the form of written materials, tables, charts, graphs, maps, or pictures. The series of related test items may also take various forms but are commonly multiple choice or true-false items.
  - Essay – The student could explain her/his own opinion and decision.
    - Restricted response question – usually limits both the content and the response.
    - Extended response question – allows students to select any factual information that they think is pertinent, to organize their answer in accordance with their best judgment.
- Used properties of the test item such as:
  - Weight or marks or points – number that describes the "weight" of the item according to the measured content and level of leaning objectives. For example if you have to give a mark for the item that measures the description of concrete concept the suitable "weight" is 1. If the item measures the analysis of the procedure the weight could be set on 3 or more. If you use a matching type of item the weight will depend on the number of matching.
  - Knowledge domain – describes the affiliations of the item to the concrete knowledge domain;
  - Learning objectives – describes the learning objectives, which the item measures. This description could be done according to the suitable taxonomy- Bloom, Merril etc.
- Intent of the test – Self-assessment; Exam; Controlling of learning paths.

- Type of the assessment – Norm-referenced, Criterion-referenced. This classification reflects how the results are interpreted. The Norm-referenced assessment "describes the students' performance in terms of the relative position held in some known group" [3]. The Criterion-referenced assessment describes the "specific performance" that are demonstrated by the student. The specification of result's interpretation strongly influences to the used statistical analysis of test and tests' items.
- Scales for grading – passed-failed, fixed or determined by the teacher.
- Statistical analysis of the test's items characteristics.
  Basic statistical characteristics of the test item are:
  - Coefficient of difficulty of the item. This coefficient has a different interpretation in the Norm-referenced assessment and Criterion-referenced assessment.
  - Coefficient of discriminative power of the item. For calculating of this coefficient usually the students are grouped in two extreme groups. The first group consists the first 27% (25%) of the students with upper scores and the second group consists the last 27% (25%) of the students with low scores.
  - Analysis of distractors for the multiple choice questions – the percentages of the accepted distractors like right answers are calculated and interpreted.
- Statistical analysis of the test characteristics. The basic statistical test characteristics are:
  - Reliability – refers to the consistency of measurement. The different statistical coefficients are relevant to the Norm-referenced and Criterion-referenced tests
  - Validity – answers the questions: Does the test measure the planed for evaluation objectives and knowledge domain. For the Norm-referenced and Criterion-referenced tests are used different methods, based mainly on the expert estimations.
- Analysis of the student achievements – describes the student's individual achievements. This analysis answers the questions – what kind of concepts, facts, procedures etc. at what kind of level of instructional objectives are attained.

1.2. Technological issues

Used technological implementation of the different test's items.

With the rapid development of the IT the diversity of technological implementations of test items could be found in the contemporary e-learning platforms.

Test storage organisation – describes the access to the test's items and possibilities of reusing one item in different modules of the course or different courses.

Used multimedia elements in the test items;

Used technologies for delivering of testing materials – on-line, offline, printed, export in different formats.

Automatic generation of the test – random or according to the didactical characteristics of the test's items specified by the teacher.

2. Comparative analysis. At the web site of the project "Open Software for Education in Europe" [12] 13 open source learning management systems are evaluated. Also 28 e-learning platforms at the UNESCO web site [13] for open source resources are presented. From these platforms we have chosen Moodle [7], a Tutor [8], Ilias [11], Claroline [9] / Dokeos [5] because of their popularity and Bulgarian language support for educational content and user interface.

Table 1. Results from the Comparative Analysis

Characteristics	**aTutor 1.5**	**Claroline 1.7/ Dokeos 1.6.2**	**Ilias 3.5**	**Moodle v 1.5**
Didactical				
Used type of test items, according to the pedagogical theories	True/false; Multiple choice with one answer; Short-answer item; Likert scale – suitable for surveys.	True/false; Multiple choice with one or more answers; Short-answer item; Matching type. Interpretative exercise	True/false; Multiple choice with one or more answers; Short-answer item; Matching type; Essay – with limited number of words in answer; Ordering question; Interpretative exercise	True/false; Multiple choice with one ore more answers; Short-answer item; Matching type – only matching text with text; Interpretative exercise.
Weight of the test's item	Yes	Yes	Yes	Yes
Knowledge domain	Could be described in item categories	No	Yes, the test items are organised in the question pools.	Could be described in item categories
Learning objectives (LO)	Could be described in item categories	No	Yes, but the definitions of LO are not based on the common taxonomy. Initially the LO have to be described for the whole course. After that the LO could be assigned to ready test and test items.	Could be described like an item category or sub category The objectives described in this way are not related to the objectives of the course.
Intent of the test	Self-assessment	Formative and Self assessment	Self assessment; Exam; Controlling of the learning paths.	Self assessment; Exam; Controlling of the learning paths (separate items in the lessons)
Type of the assessment	Not Specified	Not specified	Norm-referenced	Not specified
Scores and Scales for grading	Not available	Fixed system defined scale	Pass-Fail; ECTS; Custom description of the grading scale.	Teacher defined grading and scaling. Negative marks for the wrong answers could be given.
Statistical analysis of the test Items	Percentages of the students who choose each of the answers in each item	Not available	Total of answers. Percentage of right answer of the item (maximum points)	Difficulties coefficient; Coefficient of discriminative power (DI), Facility Index (% Correct) Standard Deviation (SD); Discrimination Coefficient (DC)
Statistical analysis of the test	Overall test average.	Not available	Rank of median, rank of participators, median of test results in points etc. Missing reliability	Not available

			and validity.	
Analysis of the student's achievements	Show achieved marks for the whole test	Tracking of students attempts, average, maximum or minimum scores.	Student's results in points and in grade distance to median. Suggestion of solution in learning material in case the student has not given the right answer to the item.	Presents the obtained scores from the student for each question and final grade in points.
Technological				
Technological implementation of the different test's items	Radio buttons, fill in blanks of separate word, phrase, page	Radio buttons, drop down menu Check box, fill in blanks	Radio buttons, drop down menu, check box, image map, fill in blanks possibilities for matching text and graphics, drag and drop,	Radio buttons, drop down menu Check box, fill in blanks.
Test storage organisation	Use the categories of questions	Use the question pool, constructed on the questions, filtered according to the different exercises.	Uses the pools of questions. The pool could be described for the different knowledge domain.	The items could be stored in hierarchical categories that allow the items to be or not to be used in different courses on the current server. The categories are not relevant to the learning objectives.
Used multimedia elements in the test items	Yes in the	Yes in the body of the test's item. In Dokeos could be implemented Hot Potatoes test items	Only .jpg in description of the item and item's answers. Possibilities to implement of the Hot Potatoes test items.	Gif, jpeg, png in the description of the item. Possibilities to implement of the Hot Potatoes test items.
Used technologies for delivering of testing materials	On-line, offline.	On-line, export in IMS QTI	On-line, print version, SCORM compatible, export in .xml.	On-line, export in: GIFT, Moodle XML format, IMS QTI 2.0, XHTML
Automatic generation of the test	Random choice from the selected categories	No	Allows random choice and order of the items from different pools. The user set the pools and number of questions from each pool.	Allows manually or random choice and order of the items and item answers from manually selected categories and subcategories. The user could choice the category and number of questions from this category

For example, up to December 31, 2005 Moodle has 8204 worldwide registered installations (28 known in Bulgaria) [7], a Tutor – more than 100 official registrations [8], Claroline – more than 400 registered site's in 60 countries [9], and Ilias – more than 100 installations in 16 countries [11]. "Dokeos is a quite recent fork of Claroline. Both tools are similar, but Dokeos shows its own personality now" [5]

The results of the analysis are presented in Table 1.

3. Conclusions. The proposed framework is strongly based on the pedagogical theories for assessment and test's development. We could summarise that considered e-learning platforms have many features relevant to the didactical and technological issues of the discussed framework, but these features could be improved. The statistics capabilities in each of the 5 reviewed platforms should be enlarged too. In our opinion Ilias and Moodle cover the most of the proposed characteristics in the framework. Also Illias supports most suitable didactical model for testing.

Our study is still in progress. We will analyze another set of open source e-learning environments about different issues of their capabilities. Also we intend do enlarge the proposed framework for comparison of testing capabilities towards of their adaptive features.

REFERENCES

[1] S. Britain, O. Liber. A Framework for Pedagogical Evaluation of Virtual Learning Environments. http://www.jisc.ac.uk/uploaded_documents/jtap-041.doc.

[2] I. Clements. Virtual Learning Environment Comparison 2003. http://www.atutor.ca/atutor/files/VLE_comparison.pdf.

[3] N. Gronlund. Measurement and evaluation in teaching, Macmilan Publishing Company, New York, 1985.

[4] Vergleich von Blackboard und Learning Space. http://www.informatica-didactica.de/HyFISCH/Produzieren/MultimediaAG/OnlineLearning/vergleiche_online_systeme/Vergleich_von_Blackboard_und_Learningspace.doc.

[5] http://campus.dokeos.com/.

[6] http://hotpot.uvic.ca.

[7] http://moodle.org.

[8] http://www.atutor.ca.

[9] http://www.claroline.net.

[10] http://www.edutech.ch/lms/index.php.

[11] http://www.ilias.de.

[12] http://www.ossite.org.

[13] http://www.unesco.org/webworld/portal_freesoft/Software/Courseware_Tools.

Daniela Ivanova Dureva-Tuparova
Department of Informatics
South West University
2700 Blagoevgrad, Bulgaria
e-mail: ddureva@aix.swu.bg

Georgi Teoharov Tuparov
Dept. of Software Engineering
Institute of Mathematics and Informatics
Acad. G. Bontchev Str., bl. 8
1113 Sofia, Bulgaria
e-mail: georgett@avala.bg

МОДЕЛ ЗА СРАВНЯВАНЕ НА ВЪЗМОЖНОСТИТЕ ЗА ОЦЕНКА И ПРОВЕРКА НА ЗНАНИЯ ЧРЕЗ ТЕСТОВЕ В СРЕДИ ЗА ЕЛЕКТРОННО ОБУЧЕНИЕ

Даниела И. Дурева-Тупарова, Георги Т. Тупаров

В статията е представен модел за сравняване на функционалните характеристики на модулите за оценка и проверка на знания в системи за електронно обучение. Моделът е разработен с отчитане на педагогическите и технологичните аспекти на проверката и оценката на знания. Предложеният модел е използван за сравнение на 5 от най-популярните платформи за електронно обучение: Moodle, Ilias, Claroline/Dokeos и a Tutor.