Economides, A.A.: **Software for Test Development and Automated Grading.** *Conference on English for Specific Purposes and Information Technology*, University of Macedonia 2000.

SOFTWARE FOR TEST DEVELOPMENT & AUTOMATED GRADING

Anastasios A. Economides
economid@uom.gr

University of Macedonia
Thessaloniki 54006, Greece

1999

Summary: To evaluate the degree of success of teaching, it is necessary to measure the student's knowledge on the subject taught. Automated test assessment tools have been developed to facilitate the evaluation of the student's learning. Their basic components are the question bank, the test creation, the test distribution, the collection of answers, the automatic assessment, the analysis of results and the report. They support different types of questions, feedback for each question, help on each question, capability to retry, multimedia use, movement control (forward, backward, to main menu, at any point), pace control, exit capability at any point and capability to return to the point of break. In this paper we examine automated software tools for developing and automatic grading tests and exams.

1. INTRODUCTION

With the immense expansion of the Internet, it is evident the role that Internet can play in the training process. After the training material is taught, the appreciation of learning is very useful for both the student and the teacher. Hence, automated software tools have been developed, that assist teaching through the Internet as well as evaluating the student's degree of understanding the subject taught.

Most tools available at the market, offer a variety of question types. They offer questions that can accept as an answer:
a) text (definition, type, summary, comparison, suggestion, ideas…),
b) one correct answer among many,
c) several correct answers among many,
d) wrong/right, yes/no choice
e) gap filling with words, types, numbers…
f) formula, algebra expression,
g) putting in the right order,
h) matching members between different groups,
i) finding a mistake,

These software tools provide the capability to sort questions by their difficulty, their topic or whatever else the teacher decides. Questions may be chosen randomly based on desired probability distributions from the question bank. Moreover, each time the tool runs the question parameters may change so that each student sees different questions. Concerning grading, these tools offer the ability to weight each question with a different weight in the assessment, and to group questions in categories (different grading, different statistical analysis).

Assessment results are useful both to the student and to the teacher. The student can immediately learn about his grade in each question, his final grade, the average grade and the grades diagrams of all students at each question, at the whole test, at the whole lesson. In the same way, the teacher has information about each student and each question. Moreover, a very important capability offered by the tools is tracking the student's progress, keeping a record of what he has studied, his participation to conversations, his performance in self-evaluation exercises, in examinations and papers.

A basic element of the training process is the active participation of the student, his permanent, frequent and direct supervision and guidance. Hence, interaction and feedback are very important. Multimedia (sound messages, moving graphics, interactive animation, instantaneous curve design etc.) preserve the student's interest to continue the training process. Consequently, multimedia is a very fundamental component in automated assessment tools.

Another important feature of these tools is the ability to impose time and access restrictions on the test. More specifically: a) ability of defining the time period during which the test will be accessible, b) ability of limiting the time available for the student to answer the test, c) ability of limiting the number of trials that the student might attempt, d) ability of limiting the access to the test from certain IP addresses or from certain sub-networks and e) ability of controlling student's access to the test.

The easy learning and use of those tools by the student, the examiner and the author of the questions is very important. Functionality, ergonomy, coherence in the way of using, confirmation/undoing of actions, storage, printing of screens/ text/ exercises/ graphics/ path, search of keywords/ training units/ terms/ students, grouping of commands, uniform use of symbols/ icons/ tables/ designs are elements of the successful acceptance of these tools. The reliability of their use as well as the absence of operational and management problems are also basic elements for the successful use of such tools. The security of the question bank, the tracking of student's progress, the HTML code and the student's identity constitute main pre-assumptions for their introduction in the training process.

Easy registration of the student to take the test, access safety, navigation speed, appearance speed of multimedia and of the answers' corrections, dealing with possible breakdowns, capability to recover after a breakdown and back up storage contribute to their effective operation. These tools keep data for each student in a database in order to control access to the test and to different areas of the training material. They record the results of the self-evaluation exercises and results, the grades of the student and the particular questions he failed. Statistical processing of these data may result into useful conclusions about each exercise, each student, each group of students, each teacher, each group of teachers, etc. The results of the statistical processing of those data may be a histogram presenting students and grades, the grades of all students at the test (average grade, mean, variability), the grades of all students in a particular part of the test (average grade, mean, variability) etc. Finally, questions with a low success rate may be detected.

Concerning the expandability of the tools and the easy updating of the question banks, most tools are usually upgraded while preserving their compatibility with former editions. A basic element is the capability of many students and teachers using the tool simultaneously by different places. Of course the existing tools still have some limitations compared to other examination methods (e.g. capability of a group of students to solve one exercise), but they are constantly being improved.

In the next section, we describe various automated assessment tools and suggest the best ones for pilot implementation by educational institutions.

2. TOOLS

CAPA (Computer-Assisted Personalized Approach) is based on the Web and provides capability to use pictures, graphics, tables etc. with variables that can be chosen randomly for each student. The system records the participation and performance of each student at the test. The student may correct his own answers before the deadline. A special characteristic of the tool is that the student must choose the N answers among the M given for each question that are correct. The number N is unknown to the student and may differ from student to student. The M answers' order is random and different for each student.

Contest uses linear programming to find the optimum test, retrieving questions from the teacher's question bank.

Course Test Manager provides a ready to use question bank from Course Technology's books as well as the capability to create new questions. It can automatically create up to 9 versions of the same test.

C-Quest, which is based on the Web, creates tests choosing questions randomly among different categories according to the examiner, the desired subject and level of difficulty. It can produce the same test with the questions in different order. Questions may be: Multiple choice, True/False, Selection of all true answers, Text. It supports graphics (.bmp, .wmf, .jpg, .gif), sound (.wav, .mid) and video (.avi). It offers protection with password and capability to print the test, the answers and the grades, as well as capability to store, sort and search. The test is distributed via three ways:

1) C-Quest Test: via a local area network, capable of limiting time, of presenting help comments and correct answers after a certain time or number of tries,
2) C-Quest Echo: dispatched by diskette or e-mail,
3) C-Quest Web: on the Web, with the capability to disclose correct answers, given answers by the student and its grade. Automatic assessment can be brief or analytical. In the analytical report, statistics are provided for all students, for all tests, for a certain student, for a certain test or any combination.

Cyber Exam is based on the Web and supports multimedia (picture, sound, video) and access by password. It supports the following question types: multiple choice, true/false, brief answer, filling in gaps, text match and free text. It supports time-programming of activities. The automatic assessment is followed by the production of statistical reports about the highest/ lowest grade, the average and deviation of all grades.

Exam Bank creates, stores and prints questions that accept one of many answers or arithmetic result or text or matching, or commands. It provides the capability to use pictures and statistics.

Exam Mail offers security via password access through the Web. The student chooses the correct answer from an unlimited number of answers to each question. It supports multimedia and help comments for any questions. After the automatic assessment the student receives his grade and success percentage in all questions and for all students. He also gets the correct answers (wherever he did not answer correctly) and a link to a page where all answers are explained and more information is available.

EXAMaker creates tests by shorting out the questions in up to 20 categories. The correct answer may be one or more choices, a choice of picture, or a correct gap filling.

EXAMiner is capable of using graphics, sound, video and CD-ROM. The examiner may define weight factors to each question, to create an exercise map and organize them into groups. He may easily add, change and reuse the exercises. The Examiner provides capability of random choice of exercises from the exercises bank according to the time and subject of the test or the capability of specific selection of exercises and formats.
The tool may be used autonomously or be integrated into another educational environment. It provides capability of encoding the exercises and specifying multiple levels of safe access by password. The exercises are automatically put in a random order and a personalized test for each student may be randomly created.
After the automatic assessment the final result or the analytical performance or nothing is returned to the student. The report includes the average grade, mean, deviation, standard error, KP21 reliability, Z grade etc.
The tool creates an MS Access database for all students, including each one's results as well as the total result for all students and in all tests. Searching or shorting is possible in this database.
Finally the statistical analysis of the results and other useful information (the number of times that an exercise and each one of its answers have been presented, if an exercise is too difficult or too easy) are saved for future use.

Eval is based on filling in forms on the web. It supports multiple choice questions, true/false, explanatory answers. After the automatic assessment, it provides statistics for the grades (average, maximum, minimum, percentage of correct answers, percentage diagram). It provides safety for the question bank and control over the student's identity.

FastTEST, FastTEST Pro enables the use of multimedia. Graphics, sound and video can be part of both the exercise and the answer (up to 20 points). A question may refer to the matching of members (up to 50 members) of two groups, where each member may be text, graphic, sound file, video or report in text or graphic. Moreover, the screen may be divided into two or three parts.

The exercises are encoded. The teacher may ask the programme about statistical results and the usage of the exercises. He may create exercises according to his specifications e.g. rare usage, success percentage, and subject. There is also the capability of interrelation among exercises. It may store and print the test and the results.
It also offers time-programming constraints on the test. The questions are set in the database of the examinant as soon as the start time of the test comes. The student may sort questions in any order.

GAME (Groningen Advanced Measurement Environment) offers multiple choice, checking, number filling, and betting odds exercises. Parts of the exercise may be used again in other exercises. It offers safety and statistical analysis of the results (variability, correlation, reliability).

Hot potatoes uses JavaScript. It supports any language in Latin characters. The questions may be multiple choice, gap filling, text answers, word crossing.

HTML to Quiz is based on the transformation of multiple-choice tests into HTML forms that are either sent by electronic mail or assessed electronically and returned to the Web Server. The answers that are given by the student in HTML forms are encoded with RSA in chipper block chaining mode. The answers that are sent by e-mail include a time stamped MD5 hash encoded with RSA. A secure server like the Netsite must be used. There does not seam to be a solution to the problem of answers existing in the Web Server's cache by a former student.

JBC is based on Java and creates multiple-choice tests. The student answers through the Web and is being assessed automatically.

IntelliTest supports multimedia questions of the following types: true/false, short text, multiple choice. It offers time-programming of the test and report after the automatic assessment.

IntelliExam supports questions with answers: text, multiple choice, true/false, and gap filling. The questions in the test may be in random order and have different weights of difficulty/grading. The questions may be sorted by subject, weight, and keywords. Statistical analysis of the results is conducted and the success rate per question is given.

LXR*TEST is based on the Web and supports multiple choices, true/false, matching, brief answer, numerical answer, and text. Multiple correct answers may exist and different weight for different answers may be given. It supports colored graphics and QuickTime movies. The teacher can short the questions in 8 categories. It also supports optical readers and links to relative material. The creation of the test may be based on multiple criteria (goals, keywords, content, statistical prices etc).
Questions may be chosen randomly from 10 different question banks. Automatic reformatting of the questions and automatic creation of multiple versions of the same test are possible.
The automatic assessment is followed by statistical analysis of the questions/answers, and reports are produced for each student, group of students, educational goal, as well

as frequency diagrams, group comparisons etc. Finally, groups may be chosen and sorted in order to compare them with each other.

Markin32 provides a button system for easy text correction, comments, assessment, statistics, links to URLs etc. The student sends his text by e-mail or diskette or ftp. The examiner uses the programme to state his remarks. He may then send it back as an .rtf or .htm file or put it on the Web Server.

MicroCat offers the facility to create text exercises, graphics, sound (.WAV files), digital pictures, and print the test. It provides flexibility concerning the time of the test. Its particular characteristic is the Rasch data analysis, 2 & 3 parameter IRT calibration and the examination of the test's validity by examining regressions and descriptive statistics. The measurement of information and of the reliability of the test and the detection of test errors are very important too.

Mklesson creates lessons consisting of multiple chapters on the Web. Each chapter consists of units. Each unit includes text and a question. Each question has multiple answers one of which is correct.

OLLA (Open Learning Agency of Australia) is built upon the Tutorial Gateway; hence it supports the same as the last one, with access security to the HTML code of the answers.

PHATE (Parametric Homework, Administration, Testing and Examination) is based on the Web. It uses Homework Mark Up Language (which is an extension of HTML) in order to create and manage tests.

PREP Online supports multiple choices, performance simulations, gap filling, true/false. After the automatic assessment, it sends comments to the student and creates reports. It offers the capability to create a chat room, distance view of the student's screen, past discussions' bank, Frequently Asked Questions (FAQ) and distribution of announcements at a scheduled time.

Question Master is written in Java and developed as CGI script. It supports multiple choice questions (in random order), true/false, numerical answer, and text. The student has access to the test by password.

QuestionMark Perception offers capability to create a question bank, choose questions in the test, test delivery via the Web, or an Intranet using a Windows NT Server. It supports creation of random tests from multiple question banks, explanatory pages, multiple choice questions, multiple correct answers, numerical answers (precise number or number within an error margin), text answers, choice answers (matching/sorting), maps, graphics and diagrams. The examiner can set the weight to each grade for each answer to the question. It can organize questions (according to the subject, the level of difficulty, etc.) in units, sub-units, as well as perform reformatting of the questions and the answers to each question each time the test is given out. Finally it can adjust the test to the student's level.

The student has access to the test for a scheduled duration time and for a specified number of efforts. After the automatic assessment, the student receives his grade, information about the answers, links to relevant educational material, the percentage of students who choose each answer, the number of students that answered correctly etc. Finally an analysis of the results is done and reports on each class, each question and each student are created.

QuestionWriter helps at creating and automatic assessment of tests through the Web. It accepts multiple-choice questions, true/false, filling in gaps and text answer. Access is given by passwords. There can be a time limit (start time and deadline) of the activities, supervision of deadlines and reminding through e-mail. The student has the capability to repeat the test choosing the way of re-assessment (average grade of the repeats, last grade, best grade etc.). It offers links under condition (time and grade results), which means that the student proceeds according to the number of correct answers he gives as well as the time of his answers. Comments are returned after each answer (correct or wrong) to each question. The student has the capability to stop temporarily and continue the test later.
Discussion Groups can also be created with the discussions being recorded so that the teacher can grade them on a later time. A special function of QuestionWriter is the management of a group of students. It detects when all group members are ready, and notifies them to proceed to the next step of the activity. It also identifies students that have problems and notifies the teacher.
The grade is calculated in different ways: i) in its initial form (raw), ii) normalized in relation to the best grade, iii) weighted. Finally it offers reports on each activity, each repeat of the test, each student and on all activities and all students.

QuickQuiz is based on the Web offering multiple choices and question grouping. It offers each student's grade, his correct answers' percentage, the final score and the first N best grades. It also creates statistics on each question, the number of correct and incorrect answers and of the unanswered questions. It can accept assessment with weight per question.

QuizMaker is written in Perl. It supports security with password. It offers time-limited accessibility of the test. After the assessment, it can return comments on each answer.

SAMaker (Self-Assessment Maker) is written in Perl. The questions accept multiple-choice answers, true/false, with comments to each answer.

QuizPlease is based on the Web and is written in Perl. It supports multimedia. The questions are in an MS Excel file and accept multiple choice answers or text. It offers repeat capability and help to each question in text, Web link, picture, sound, video.
Questions are sorted and selected randomly from the question bank. After the automatic assessment, each question can be followed by comments.

QuizSite offers a question bank with categories. Questions accept multiple choice answers, matching, numerical result, gap filling, and text. It supports multimedia (pictures, photographs, diagrams, charts), random choice of questions from the bank and time-programming of the test. It is accessed by password.

After the automatic assessment, it can return comments on the answers. It offers statistic analysis of the grades, categorization and analysis of the questions' level of difficulty.

Test Maestro offers creation of multiple tests with graphics and OLE objects. Access to the database is protected by password. It uses MS Access files. It provides a grade book with multiple units, capability of reports on different periods, each student's position in the teaching room, each student's attendance, and the meetings with the student, the student's background, and statistical analysis of the assessment. The student can see his grades, his attendance, and his papers. It offers graphics design (square, full-log, semi-log, polar, isometric-orthographic), best-fit curve, formula illustration, equation solution, and calculator. The results of the grading, the student's papers, reports with average prices per category, absences, paper deadlines, examination dates, meeting dates can appear on the Web.

Test Pilot is based on the Web and supports text, sound and picture (.au, .mid, .mov, .gif, .jpg files). Questions can be put in different random order each time the test is given. They can also be grouped and a specific number of questions from each group can be chosen randomly.
Answers can be of the following types: Yes/No, True/False, multiple choice (up to 9 choices), text, 5 levels of choice from Like (Agree, Always, Important) to Dislike (Disagree, Never, Indifferent), declaration of all valid (up to 9 cases), formulas with random variables. Its basic advantage is the capability of questions/answers based on mathematical formulas. The examiner gives the possible values of the variables (e.g. a=10 to 20, step 2), and the mathematical formula which calculates the result. Test Pilot randomly chooses (for each student) values for the variables and asks him to calculate the result. It checks if the result number is correct. It accepts a number as correct if it is the same as the one calculated by the program within a margin.
It also offers capability to impose time and access restrictions to the test. More specifically, it offers capability of: a) defining the time during which the test will be accessible, b) limiting the time available for the student to answer the test, c) limiting the number of efforts the student can take for the test, d) limiting access to the test from certain IP addresses or certain sub-networks, e) controlling students that have access to the test.
Results appear in a spreadsheet for statistical analysis and are stored. The student can see or not see his grade and the correct answers. The examiner can see the test and the answers given as a total and by each student anytime.

Tutorial Gateway supports multimedia, multiple choice answers, true/false, number, and algebra expression. Each question can be followed by a help text and each answer can return comments. It provides capability for the student to retry the test.

TRIADS (TRipartite Interactive Assessment Delivery System) supports multimedia. Each question accepts as an answer: multiple choice, true/false, multiple correct answers, selection of points on diagrams, shapes and text lines, object movement to the appropriate position, line or group, text and numbers insertion. It provides the capability to create random numbers for each question. It accepts a numerical result within an error margin. It supports the design of points, lines, arrows, and circles. Movement of an object can result in movement of another object.

A significant capability of TRIADS is the creation of questions of multiple levels, where every different answer leads to other questions. The examiner can group the questions and TRIADS can choose randomly some questions from each group.
Each question can have weight in the assessment with possibility of negative assessment. Each answer can return comments. The results can be stored in MS Excel files. It offers different ways of examination: a) serially (the student cannot return to any previous question, b) randomly (the student can go back and forward and re-answer a question, c) in cycle (the student can retry the questions he failed to answer). The test can be time-limited concerning its duration as well as the time period that it can be taken. The programme returns the following information to the student and the teacher: name of student, grade, date of test, duration of test, duration of dealing with each question, correct answers, given answers, comments on each question, assessment weight of each question, question group, author of each question.
For the evaluation of each programme, the users fill in forms before the examination, after they answer the test's questions (but before they see their assessment) and finally after they see their assessment.

WebWork is written in HTML and Perl. Each student sees a different version of the same question (e.g. different arithmetic values). After the automatic assessment, the student can retry all the questions he did not answer correctly, getting part of the grade he would get if he had answered correctly the first time. Grades are processed with MS Excel.

WEBTEST is written in Perl. The questions take answers: a) multiple choice through list boxes (the student chooses a line from the list), or radio buttons (the student chooses a diagram pushing the button next to it, b) arithmetic answers with an optional error margin (e.g. because of rounding), c) text with optional tolerance of caps/small letters. It offers capability of multiple correct answers, use of variable arithmetic data, presentation of questions and possible answers in random order. Time limitations can be set concerning the time that the test can be taken. A soon as the student starts the test, an electronic message is sent to the administrator so that time starts running. This electronic message integrates the secret test certification code, so that the student cannot copy an electronic message with correct answers and sent it as his own.
It offers self-evaluation capability with feedback at any answer given by the student to each question. Moreover, the transition to a former page can be deactivated.

Web Worksheet supports multimedia and security by password. The questions take multiple choice answers, true/false, gap filling, and text. Each question can have a different weight. The student can try each question more than once. The following test types are offered: a) practice without an assessment report to the teacher, b) homework exercises with an assessment report to the teacher, c) assessed exams. As soon as the student sends the answer to a question, this answer is recorded and a different answer cannot be written over the previous one. The student can pass a question without answering it, and come back to answer it later on. The student can see the total number of questions of the test, the maximum grade for each question, the number of question he tries to answer at any moment, the grade he took in the previous question, the total grade until that moment, how many times he can try each question, the number of his present attempts. After the automatic assessment, a direct feedback is given to the student with the questions, his answers, the correct answers and explanations. The teacher takes a partial report on each student and a total report for all students with a definition of the question where many students have failed.

Web@ssessor written in Java. It supports multimedia (video, sound, moving pictures, graphics). The questions can be multiple choice, True/False, Gap filling, Text Answers.

WebQuiz offers random choice of questions from a question database distributed to chapters, semesters, etc. according to the desired distribution. It has the capability to choose questions added to the database after of before a certain date. After the automatic correction, explanations are sent to the student.

WWWAssign is written in Perl. It supports picture and sound. Questions can be multiple choice, true/false, gap filling, arithmetic answers, and text answers. It supports random arithmetic data and small error margin to the answer (e.g. because of rounding). After the automatic assessment it offers statistic processing of the results.

3. CONLUSIONS

Realizing the education opportunities offered by the Internet, many software companies have released tools for automatic test assessment. Those tools support the creation, distribution, automatic assessment and statistical analysis of the grades. We evaluate many of these tools and find that they offer many capabilities, although they still have some unsolved problems, for example the safety issue. At present, Cyber Exam and QuestionMark Perception take some special attention. However, not only the existing tools are continuously being improved, but also new ones are continuously being created, the best tool is yet to come.