

# Some Guidelines to Support the Selection of a Test Management Tool for Web-based Assessment

S. Valenti, A. Cucchiarelli, M. Panti

Computer Science Dept., University of Ancona - 60100 Ancona - Italy, {valenti, alex, panti}@inform.unian.it

## 1. INTRODUCTION

It is now a well-established and widely accepted concept that assessment plays a central role in the educational process.

The search for assessment methods able to reach objective judgement of student's knowledge is a crucial goal for both teachers and educational institutions. The teacher looks for homogeneous treatment of the students and for useful hints on her educational activity in terms of clarity, completeness and effectiveness, whereas the educational institution tries to log the teacher's activity and the quality of the service offered to the students. Moreover, the growing mobility of the manpower everywhere requires the educational institutions to comply with international standards of crediting courses, and the diffusion of computer based distance learning forces them to cope with problems posed by self-assessment procedures.

Dealing with large classes raises a number of problems both from the lecturer and the students point of view, and teaching large classes is often seen as a difficult and unwelcome assignment (Valenti, 2000). Furthermore, the lecturer is able to know only a limited number of students and since the lectures follow one another in a short interval of time, only a few of the students with questions about the material can be helped. The same considerations are equally true for the support that may be given in office hours. Furthermore, exams cannot be taken by all the students at the same time, due to lack of resources: this often reflects in exams being graded by different people with differences in grading styles, that may become relevant regardless to any "grading blending" policy. The freshmen, on their side, often have problems adjusting from teacher-led form of education used in high schools. Many of them need support as they make the transition to a learning style in which they have to take great responsibility for their own education. In particular they are often worried by the way exams are carried out, since the only possibility to verify the results of the learning process is just one final examination at the end of then course without the chance of any intermediate checkpoint. In particular they appreciate frequent feedback on their progress and reassurance that any misconceptions may be identified and fixed. However, the decrease in resources to be used for tutoring and the increase in class sizes often leads to poor feedback to students, which often reflects either in delaying their career or in poor grading.

In this scenario the Web-based Assessment (WBA) seems to have a number of applications as it is able to automatically capture all the information required by the actors of the educational process (teacher, student and institution) for large and/or distributed classrooms.

The interest in developing WBA tools grew up in recent years, thanks to the potential market of their applications. Many commercial products, as well as freeware and shareware tools, are the result of studies and researches in this field made by companies and public institutions. For an updated survey of course and test delivery/management systems for distance learning see Looms (1999). This site contains a description of more than 100 products, and is continuously updated with new items.

Now the point is: are there any criteria that may be useful to an educational team wishing to select the most appropriate assessment tool for their environment?

The answer to this simple question seems to be negative.

From a survey of all the material available on the net, starting from the YAHOO category on computers and education, and then going to a number of sites maintaining links related to educational resources as for instance the Educational Resources Information Center (ERIC® Clearinghouse on Assessment and Evaluation; ERIC, 1999) and TECFA (the academic unit active in the field of educational technology of the School of Psychology and Education of the University of Geneva, Switzerland;

TECFA, 1999) it appears that only two papers have been devoted to such an important topic (Freemont & Jones, 1994; Gibson et al., 1995).

Purpose of this paper is to present a proposal for a framework that may be useful to identify some guidelines for the selection of a Web based assessment tool.

More in detail, in section two we will start discussing the criteria identified by Gibson et al. that are by far more structured and general than those presented in the paper by Freemont and Jones. These criteria have been developed considering an assessment application as a single entity, and are focused both on assessment and on educational capabilities of such a tool.

First of all, it should be noted that we are not concerned with assessment systems that provide tracking capabilities with respect to the educational process, since our focus is on the selection of "stand-alone" assessment tools: i.e. assessment tools that are not integrated within course management systems as for instance Web-Ct or TopClass (Hazari, 1998). Furthermore it should be outlined that an assessment tool is composed by at least two sub-systems that may not necessarily belong to the same application: a Test Delivery System that is the module that interfaces with the students and allows to administer tests and collect results and a Test Management System that allows to create questions and tests and to collect and evaluate the results. In this paper we will provide only some criteria to select a Test Management System, while we are still working on the identification of a similar framework for analyzing existing delivery applications.

Therefore in section three we will present some remarks to the framework proposed by Gibson et al and then we will discuss a more detailed approach that we believe may be useful for the selection of Test Management System for web-based assessment. Some hints for further research and final remarks will be discussed in the last section of the paper.

## 2. AN EXISTING FRAMEWORK

Gibson et al., in a very well known paper (1995) describe an approach for evaluating a Web Based testing and evaluation system. The criteria identified are six: testing, tracking, grading, tutorial building, implementation and security issues. In this section we will shortly discuss these criteria, leaving the interested reader the burden of reading the paper in order to find the results of their comparison of four tools: Mkleesson, Eval, Tutorial Gateway and the package built on Tutorial Gateway by the Open Learning Agency of Australia.

### 2.1 Testing

In order to evaluate the testing capabilities of a Web/Computer based assessment system, the following sub-criteria have been adopted: the classes of questions allowed, the feedback provided, the existence of tools for providing help and hints to the student, the possibility of retrying more than once the answer to a question and use of multimedia as an integrating part of the testing system. In the following part of this section we will discuss briefly these points.

#### 2.1.1 Types of questions

Here the concern is focused on the classes of questions supported by the system under consideration. Some examples are multiple choice, true-false, simple numeric and simulations.

#### 2.1.2 Help and Hints

This item concerns the capability of the system to provide directions about the completion of the test and hints that usually are related to the

contents of the questions. In some way this item may be considered as a measure of the ease of use of the application from the student point of view.

### 2.1.3 Retries

This item is related to the ability to allow multiple attempts in answering a question. Obviously, this ability may be of great importance for self-assessment, since it may be useful to improve the knowledge of the student limiting the need of providing feedback and or tutoring.

### 2.1.4 Feedback

This criterion is related to ability to provide information to the student once the answer to a given question has been entered. Feedback is very important for self-assessment, since it may be used to correct misconceptions or to deliver additional material to deepen or broaden the coverage of the topic assessed by the question.

### 2.1.5 Multimedia

The use of questions incorporating multimedia, like for instance sound and video clips, or images, may improve the level of knowledge evaluation. This aspect may be of great importance for example in language assessment, where the comprehension of a talk or a movie can be assessed by recurring to multimedia only.

The use of multimedia can raise issues related to portability and interoperability since it may require special hardware and software, both for the server delivering the questions and for the client used by the students. Furthermore it may raise the costs for the adopted solution.

## 2.2 Tracking

Tracking is related to the ability of the system to remember where the student has traveled within a lesson and recording her performance on test questions and answers. This criteria is useful to allow the instructor to follow the specific pattern of progress and performance of each student, and to fine tune self-assessment by each student. On the other hand, Gibson et al. suggest on their paper that by tracking the student's progress, it is possible to provide dynamic guidance on how best to proceed through the lesson.

## 2.3 Grading

Obviously, any software for assessment should be able to compute student grades. Furthermore, grades must be delivered as feedback to the course coordinator, to the instructor and to the students. Each of these categories of users needs to obtain a different kind of feedback on the grades associated with a test. For instance, a student needs to know where she stands with respect to other students and to class average besides to her own individual and cumulative grades. This information need may raise obvious privacy concerns that may be faced through the security facilities provided with the assessment tool (see section 2.6).

## 2.4 Tutorial Building

This criterion is tied to the existence of some facility for automatic inclusion of tutorial in the testing system.

## 2.5 Implementation Issues

From the point of view of Implementation, Gibson et al consider only two main issues: Ease of Use and Platform.

### 2.5.1 Ease of Use

Ease of use focuses allows measuring how easy is for the author of the courseware and for the instructor to use the testing system to implement assessment. An important point to outline is that knowledge of HTML is assumed to be possessed by the lecturer using Web-based assessment tools.

### 2.5.2 Platform Issues

Among the points considered to belong to this criteria are server functionality, availability of viewers, ability of the hardware to support multimedia (like sound and video) and the requirements of the networking facilities.

## 2.6 Security

There is a wide range of security issues related to the use of both

Computer and Web based assessment system. Among these issues, it should be outlined that there are a lot of concerns on the security of the test material, of the HTML code that implements testing, of the identification of the user (both instructors and students), and so on.

## 3. EXTENDING THE SCOPE OF GIBSON FRAMEWORK

First of all it should be outlined that an assessment tool is composed by at least two sub-systems that may not necessarily belong to the same application: a Test Delivery System that is the module that interfaces with the students and allows to administer tests and collect results and a Test Management System that allows to create questions, tests and evaluate the results.

In this section we will identify some criteria that may be useful to select a Test Management System.

A Test Management System is a tool that should provide the Instructor an easy to use interface, the ability to create questions and to assemble them into tests, the possibility of grading the tests and to make some statistical evaluations of the results. Therefore, we have identified four main criteria for the evaluation of a Test Management System: Instructor Interface, Testing, Assessment and Implementation Issues. Each criteria is further divided in items so that, for instance, Testing is related to Question Items, Test Items and Tests Banks and sub-items so that Question Items are further put in relation with Type of Questions, Question Structure, Feedback, and Ease of Editing (as shown in Table 1).

Each of these criteria will be discussed in detail in the rest of this section.

As noted earlier in this paper, tracking allows the instructor to follow the specific pattern of progress and performance of each student, and to fine tune self-assessment by each student while allowing dynamic guidance on how best to proceed through the lesson. Therefore we believe that tracking is a criteria that may be used to select software systems falling in the wider range of products that go under the name of Computer based learning and teaching systems. For this reason, the "tracking" criterion discussed in the paper of Gibson's has not been taken into account in our framework.

Furthermore, we will not take into account the ability of the system to provide the possibility of multiple attempts to answer the same question (2.1.3), since we believe that this is one of the criteria that may be used to select a Test Delivery Application, whose scope is outside the purpose of this paper. For the same reason, we will not take in consideration the ability to provide directions about the completion of the test and hints that are related to the contents of the question (2.1.2).

Finally, we will not discuss the ability of the system to provide directions about the completion of the test and hints that are related to the contents of the question (2.1.2), since we believe that this represents a criterion for the selection of a Test Delivery Application, whose scope is outside the purpose of this paper. For the same reason, we will not take in consideration the ability to provide the possibility of multiple attempts to answer the same question (2.1.3).

### 3.1 Instructor Interface

It is an unstated axiom that Web-based tools automatically provide an

Table 1 - Criteria to evaluate test management applications

Criteria	Items	sub-items
Interface	Friendly GUI easy to learn and use	
Testing	Question items	Type of Questions Question Structure Feedback Ease of Editing
	Test Items	Test preparation Feedback Analysis of Tests
	Test banks	
Assessment	grading tools for test evaluation	
Implementation Issues	Server Platform Security Communication with other SW	

easy to use and learn environment. This is obviously false. There is a lot of work in the literature, on the criteria to be adopted in order to evaluate a Graphical User Interface (GUI) from the point of view of usability (see for instance Nielsen & Molich, 1990 and Gilham et al., 1995). In the following, we will list some well-known guidelines that may be used to evaluate a GUI. As Nielsen & Molich (1990) simply proposed, the interface must be:

- easy to learn
- efficient to use
- easy to remember
- error free
- subjectively pleasing.

The set of criteria that may be adopted to evaluate the usability of a GUI is summarized in the following list:

- use dialogues simple and natural
- speak the users' language
- be consistent
- provide feedback
- provide clearly marked exits
- provide shortcuts
- have good error messages
- prevent errors.

The capability of the system to provide directions on the construction of a question and of a test and hints for the analysis of the obtained results, will be discussed in the following sub-sections, even if it represents a measure the easiness of use of the interface.

### 3.2 Testing

Testing represents the vital part of any assessment tool. We suggest adopting three main categories to evaluate the characteristics of an evaluation system: Question items, Test Items and Test banks.

#### 3.2.1 Question Items

Among the issues that may be used to qualify an assessment tool with respect to this co-ordinate, we suggest to adopt Types of Questions, Question Structure, Feedback and Ease of editing. All of these items will be discussed in some detail in the next paragraphs.

##### 3.2.1.1 Types of Questions

A list of the most common types of questions, along with a simple definition for each class is summarized in table 2 (Valenti, 2000).

An important point is tied, in our opinion, to the learning objective that must be assessed through the questions. Each of the classes listed in table 2 may be used to evaluate different types of knowledge. Therefore the choice of a system may be driven by the ability that needs to be verified.

It is worth while to outline that many universities are adopting the same tool in all courses in order to reduce costs, and to allow students to interact in the same way in each phase of their evaluation process. This obviously imposes the requirement of selecting a tool that provides the wider range of questions available, since within different courses, different learning outcomes may be assessed.

##### 3.2.1.2 Question structure

Each class of questions has a given structure and contains a number of

Table 2 - Summary of question types

Question Type	Meaning
multiple choice	questions where the user is asked to choose the correct answer from a list of alternatives
multiple response	questions where the user is asked to select a number of correct answers from a list
true/false	questions where the user is asked to evaluate the truthfulness of a statement
selection/association	questions where the user answers by matching items from two related lists
short answer	questions that may be answered by entering a word, a short phrase or number
visual identification/hot spot	questions that may be answered by moving a marker onto a part of the screen to identify a hot spot on an image.

fields. Multiple Choice Questions (MCQs), for instance are organized into three parts:

- 1) the stem
- 2) the key
- 3) some distractors

The problem to which the student should give an answer is known as stem. The stem should be stated either as a direct question or as an incomplete statement.

The list of suggested solutions may include words, numbers, symbols or phrases and are called alternatives, choices or options. The user is asked to read the stem and to select the alternative that is believed to be correct. The correct alternative, which must be one, and only one, is simply called the key, whilst the remaining choices are called distractors, since their intended function is to distract students from the correct one.

According to Bloom's Taxonomy (1956) six cognitive levels may be defined, ranging from Knowledge (simplest) to Evaluation (most difficult). Therefore, for each question, the type of cognitive level that is aimed to be assessed should be identified.

Furthermore, if the test should be used to evaluate the instructional process, fields to store the source of each question, the chapter to which it is related, the topic covered along with the author of the question itself should be provided.

Finally a number of additional numeric fields for storing the statistical performance of each question should be at disposal of the instructor.

#### 3.2.1.3 Feedback

A question may provide feedback that contains the mark to the given response along with optional comments reflecting the users' performance. The feedback could be presented either after any single question (this solution being preferable for self-evaluation tests) or at the end of the test and may be based on the overall performance.

#### 3.2.1.4 Ease of editing

This item is related to the easiness with which the system allows to construct questions and tests. This ability can be enhanced through the existence of a GUI that provides standard features as a "wyswyg" editor, a clipboard and cut-and-paste and undo operations.

Furthermore the existence of spelling and grammar checking may greatly improve the usability of the tool by helping the instructor to build up well-formed questions. Gronlund, for instance, in (1985) suggests that in order to create good MCQs:

- all of the alternatives should be grammatically consistent with the stem of the item.
- verbal association between the stem and the correct answer should be avoided.

These two requirements may be easily satisfied through a good spelling and grammar-checking tool. The existence of ad-hoc dictionaries tailored on the domain, to which questions are related, may represent a plus to improve the ease of editing.

Finally, the possibility to include text, graphic images for diagrams and properly display mathematical, chemical or other symbols may be of great importance for the instructor.

Last, but not least, the inclusion of multimedia, like for instance sound and video clips, or animated images may improve the level of comprehension of a question. As stated in section 2.1.5 of this paper the use of multimedia may raise issues related with the portability and the interoperability of the application. These issues may not represent a problem whenever a Web-based assessment approach is selected, since the nature of the WWW is inherently multimedial. In this case, the choice of standard plug-ins for the available browsers may reduce risks of portability and of interoperability. Since most plug-ins used to grant access to multimedia sources are actually free of charge, their use may not interfere with cost problems.

Another criterion that is useful to evaluate the ease of editing is the programming abilities required to the instructor. The usability of the system may be dramatically reduced by the requirement of HTML, XML, Perl/CGI, Java or JavaScript knowledge.

#### 3.2.2 Test items

This criterion is concerned with the ability to build up a test from a set of questions and to deliver it. Among the issues that may be used to qualify an assessment tool with respect to this point of view, we suggest to adopt

Test preparation, Text Banks and Analysis of Tests. All of these items will be discussed in some detail in the next paragraphs.

### 3.2.2.1 Test preparation

Once questions have been defined, they should be selected and organized into a test. Test preparation is a non-trivial task, since it may require the ability to “manually” choose the questions from their base, or to construct automatically the exam through a random selection approach. This last point means that the tool should allow to compile tests by selecting questions with respect to educational objectives, keywords, contents, statistical value and so on. Furthermore the availability of facilities for building adaptive tests may be a plus for the selection of the tool. Adaptive testing is used to allow the student to move forward or backwards in a test depending on what has happened so far. This is a very powerful feature, since it allows creating material that reacts “intelligently” to what the student does.

Moreover, it should be possible to create multiple forms by rearranging questions, either by some instructor choice or automatically, in order to discourage cheating. Tools that provide the ability to randomize the order of answers for a question may further discourage cheating.

Finally, the possibility of correcting the grades to prevent guessing should be provided. See the discussion about grading (3.2.3.1) for some hints about this last point.

### 3.2.2.2 Test Banks

Questions can be assembled together directly in test or in a bank that is further referenced by the test. Test banks may be very useful in a number of ways, since organizing in a bank questions related to the same topic may simplify both random selection of questions and the evaluation of the understanding of the topic itself through statistical measures. Furthermore, it should be outlined that the same bank can be shared by different tests. This last point suggesting that it is possible to reuse the same material, saving time and effort. Obviously, different instructors may share the same questions thus obtaining synergies and homogenizing the way in which the same topic is assessed in different courses.

Furthermore, building well-formed questions is very hard and difficult task. The possibility of accessing question banks provided by well-known scientists or by professional organizations represents a great value for the educational community.

As an example, we can cite the effort made by a number of Student Chapters of the Association for Computing Machinery that are collecting test banks related to Computer Science (ACM-SC, 1999).

Therefore, a test management system should provide the possibility to create multiple banks with unlimited number of items in each bank, and the ability to import existing questions and corresponding data from existing banks.

### 3.2.2.3 Analysis of Test

Tests should be evaluated both before and after administration (Gronlund, 1985).

Evaluating a test before administration means analyzing it in terms of adequacy of test plan, text items and text format and directions. From the point of view of test plan, analyzing a test means finding an answer to the following questions among others:

- does the text plan adequately describe the instructional objectives, and the contents to be measured?
- does the text plan clearly indicate the relative emphasis to be given to each objective and each content area?

From the point of view of text item, analyzing a test means to evaluate each item in terms of appropriateness, relevance, conciseness, ideal difficulty, correctness, technical soundness, cultural fairness, independence and sample adequacy.

Finally, from the point of view of text format and directions, analyzing a test means, for example, finding an answer to the following questions among many others:

- Are the test items of the same type grouped together in the test or within sections of the test?
- Are the correct answers distributed in such a way that there is no detectable pattern?
- Are the test material well-spaced, legible, and free of typographical errors?

Evaluating a test after administration helps to verify whether it functioned as intended in order to adequately discriminate between low and high achievers; the test items were of appropriate difficulty and free of irrelevant clues and other defects (so, for instance, all distractors behaved effectively in MCQs);

The existence of training package that may be helpful to the instructor for setting good objective tests may be, in our opinion, a very important plus for the selection of a Computer Based Assessment tool.

### 3.2.3 Assessment

Once questions have been designed and the test delivered, it is of fundamental importance to obtain an assessment of the student as a single and with respect to the class. This criterion will be examined in the following paragraphs, through the discussion of two main issues that may be used to qualify an assessment tool with respect to this point of view: Grading and Test Evaluation.

#### 3.2.3.1 Grading

It should be noted that each class of question provided by the assessment tool might show different scoring schemes.

As an example, we will discuss briefly two marking philosophies of MCQs.

The simplest way to assign a score to a MCQ is to mark 1 to the correct answer and 0 to the other options. This strategy allows students who make blind guesses or give random responses to all questions to obtain a score that may be evaluated as the number of questions divided by the number of distractors used: this means that a lucky student who is submitted to a test with 100 MCQs with 4 distractors may obtain a score up to 25.

Another approach called negative marking, assigns 1 for the correct response, 0 for no response and  $-1/(n-1)$  for an incorrect response. With this approach, a student who knows nothing, and therefore makes completely blind guesses may be marked with the plausible score of “about” zero.

The assessment tool should allow both of these marking schemes.

All the considerations done in section 2.3 about grading still hold in our extended framework. Therefore, the grader module of an assessment tool should be able to provide each class of actors involved in the educational process (students, instructors and course administrators) the corresponding feedback on the grades associated to a test or to a class of tests.

#### 3.2.3.2 Test evaluation

In section 3.2.2.3 we have discussed the importance of providing to the instructor some tools for the assessment of the evaluation process. To attain to such results, the assessment tool should provide at least the following information to the instructor:

- test performance report for each individual examinee, with percentage of correct answers and relative ranks;
- individual response summary by item, with an error report that lists wrong vs correct responses;
- class test performance with distributions, means and deviations;
- item statistics and analysis with indicators that may be useful to evaluate the questions in terms of reliability, discrimination, difficulty and so on.

Although the system may provide some numerical results to measure the test, it is actually completely left to the instructor the responsibility of evaluating them and to identify strategies and policies to modify the educational process in order to improve the understanding of mis-concepted topics.

### 3.2.4 Implementation Issues

Among the issues that may be taken into account to evaluate a Test Management Application from the point of view of Implementation, we have selected Security and Communication with other software, while the Ease of Use (2.5.1) is a criterion to evaluate the system interface.

#### 3.2.4.1 Security

As stated in section 2.6 of this paper, there is a wide range of security issues related to the use of Web based assessment system. Among these issues, it should be outlined that there are a lot of concerns on the security of the test material, of the HTML code that implements testing, of the



identification of the user (both instructors and students), and so on.

With respect to security concerns about the test material and the HTML code that implements test it must be outlined that, while commercial programs usually implement encrypting approaches, a lot of issues should be taken into account for freewares. In fact, most freeware applications rely either on Perl/CGI or on JavaScript. From the point of view of security, the use of CGI-based application may raise an important problem: since a CGI program is executable, it is basically the equivalent of letting the world run a program on the server side, which is not the safest thing to do. Therefore, there are some security precautions that need to be implemented when it comes to using CGI based applications. Probably the one that will affect the typical Web user is the fact that CGI programs need to reside in a special directory, so that the Web server knows to execute the program rather than just display it to the browser. This directory is usually under direct control of the webmaster, prohibiting the average user from creating CGI programs. On the other hand, since the JavaScript code runs on the client side of the application, the obvious drawback of this approach is that the assessment program cannot be completely hidden, and a "smart" student can access the source discovering the right answer associated to each question. In any case, some sophisticated techniques can be used to partially overcome the problem, which can be reduced to a minimum.

#### 3.2.4.2 Communication with other software

Communication with other existing software may be very useful both for exporting answers and for calling external applications.

Exporting answers is usually performed through test files and data conversion utilities. This may be useful to customize the reports generated by the application or whenever an analysis more detailed than that allowed by the assessment tool is needed to evaluate the results obtained.

Furthermore, many available tools provide the ability of calling a program as a block within a question. The called program returns a score in points that may be added to the score of the test. This may be useful for assessing abilities that cannot be evaluated through the basic question-answer paradigm of many assessment tools.

Some tools allow activating a call to an external application at the very end of the test phase. In the user manual of Question Mark (Qmark, 1999), a very well known application to perform computer based assessment, the following reasons are listed to explain the utility of such a feature: "this may be useful for

- a) printing certificates for all users who pass the test;
- b) electronically submitting the answer file to a central location for analysis and evaluation;
- c) storing the results in a file to be accessed by a user program".

Finally, communication with other software is required in order to allow the integration with test delivery programs written by different providers.

#### 3.2.5 Other Issues

Web based assessment tools are available both as commercial and as freeware applications.

Commercial programs may be divided into two main classes: publisher and off-the-shelf systems (Freemont 1994). Publisher systems are licensed only for use with particular adopted textbooks and therefore usually are proprietary, thus limiting the sharing among different courses/departments.

On the other hand off-the-shelf programs are available with faculty wide site licenses that allow adopting the same system for all courses. This is an important aspect, since development, maintenance and expertise can be shared among instructors. At the same time students need to know only one interface, thus reducing their effort in learning something that is not directly related to their educational process.

Furthermore, commercial programs match most of the criteria discussed in this section, and therefore are preferable both to publisher systems and to freewares.

The main advantage of free application is tied to cost considerations, and on the availability of source code that allows tailoring of the web assessment to special needs that may not be fulfilled by existing tools.

In fact, although objective testing can be used to assess a wide range of learning outcomes a number of authors (Gronlund, 1985; Crabbe, Grainger & Steward, 1997; Ebel, 1979; Gagné & Briggs, 1979) agree on the fact that more complex patterns of achievements are very difficult to be evaluated through this approach.

For instance, the ability to state and to recognize inferences, the ability to recognize the limitations of data is very difficult to evaluate through question/answer mechanisms.

In our opinion, these tools may be very useful to improve the analytical abilities of the students.

Therefore whenever a model for the evaluation of complex patterns of interaction is needed it may be useful to start from a freeware program that may be enhanced through the integration with ad hoc facilities.

## 4. FINAL REMARKS

In this paper we have discussed a framework that may be useful to assist an educational team in the selection of a Test Management System. The framework has been obtained by modifying and extending existing work on the field (Freemont and Jones, 1994; Gibson et al. 1995).

Actually, we are trying to extend this framework in order to identify a set of criteria that may be used to support the selection of a Test Delivery Application, in order to provide an integrated approach to evaluate the usability of a Web-based Assessment System.

At the same time, our effort is aimed to review the commercial and freeware applications referenced in Looms (1999) using the criteria discussed.

## REFERENCES

- ACM-SC (1999), [www.cs.uidaho.edu/~acm/test\\_bank.html](http://www.cs.uidaho.edu/~acm/test_bank.html), [www.utdallas.edu/orgs/acm/testbank.html](http://www.utdallas.edu/orgs/acm/testbank.html), [www.csc.lsu.edu/~acm/testbank/index.html](http://www.csc.lsu.edu/~acm/testbank/index.html)
- Bloom B. (1956), *Taxonomy of Educational Objectives, Handbook I, Cognitive Domain*, New York: David McKay Co. Inc.
- Crabbe J., Grainger J. & Steward R. (1997), *Quality assessment of Computer Based Learning*, Educational Computing, 8(3), 17-19.
- Freemont D.J., Jones B. (1994) *Testing Software: a review*, New Currents 1.1, <http://www.uccalcary.ca/Newsletters/Currents/Vol1.1/TestingSoftware.html>
- Ebel R.L. (1979), *Essentials of Educational Measurement*, Prentice Hall, Englewood Cliffs, NJ, 1979.
- ERIC® (1999) Clearinghouse on Assessment and Evaluation, <http://ericae.net/>.
- Gagné, R. M. & Briggs, L. J. (1979), *Principles of instructional design*, (2nd ed.), New York: Holt, Rinehart and Winston.
- Gibson E. J., Brewer P.W., Dholakia A., Vouk M.A., Bitzer D.L. (1995), *A comparative analysis of Web-based testing and evaluation systems*, Proceedings of the 4th WWW conference, Boston.
- Gillham, M., Kemp, B. & Buckner, K. (1995), *Evaluating Interactive Multimedia Products for the Home*, in the New Review of Hypermedia and Multimedia Vol.1, p. 199-212.
- Gronlund N.E. (1985), *Measurement and Evaluation in Teaching*, MacMillan Pub. Co, NY.
- Hazari, S. I. (1998), *Evaluation and selection of web course management tools*, Retrieved [10, 01, 1999] from the World Wide Web: <http://sunil.umd.edu/webet>
- Looms T. (1999), *Survey of Course and Test Delivery / Management Systems for Distance Learning*, <http://tangle.seas.gwu.edu/~tlooms/assess.html>.
- Nielsen, J. & Molich, R. (1990), *Heuristic evaluation of user interfaces*, Proceedings of CHI 90, 249-256. New York, NY: ACM.
- Question Mark Software (1999), <http://www.qmark.com>, Question Mark Computing Ltd.
- TECFA (1999), [http://agora.unige.ch/tecfa/edutech/welcome\\_frame.html](http://agora.unige.ch/tecfa/edutech/welcome_frame.html).
- Valenti S., Panti M. and Cucchiarelli A. (2000), Web-based assessment of Student Learning, to appear in *Web-Based Learning and Teaching Technologies: Opportunities and Challenges*, A.K. Aaggarwal ed., Idea Group.

0 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/proceeding-paper/some-guidelines-support-selection-test/31565](http://www.igi-global.com/proceeding-paper/some-guidelines-support-selection-test/31565)

## Related Content

---

### Interventions Strategies to Promote Adaptive Behaviors by Persons with Acquired Brain Injuries

Claudia De Pace and Fabrizio Stasolla (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 5564-5572).

[www.irma-international.org/chapter/interventions-strategies-to-promote-adaptive-behaviors-by-persons-with-acquired-brain-injuries/113010](http://www.irma-international.org/chapter/interventions-strategies-to-promote-adaptive-behaviors-by-persons-with-acquired-brain-injuries/113010)

### Knowledge at Work in Software Development: A Cognitive Approach for Sharing Knowledge and Creating Decision Support for Life-Cycle Selection

Luca Iandoli and Giuseppe Zollo (2005). *Causal Mapping for Research in Information Technology* (pp. 312-342).

[www.irma-international.org/chapter/knowledge-work-software-development/6524](http://www.irma-international.org/chapter/knowledge-work-software-development/6524)

### Taxonomy for "Homo Consumens" in a 3.0 Era

Carlos Ballesteros (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 1638-1645).

[www.irma-international.org/chapter/taxonomy-for-homo-consumens-in-a-30-era/183878](http://www.irma-international.org/chapter/taxonomy-for-homo-consumens-in-a-30-era/183878)

### I-Rough Topological Spaces

Boby P. Mathew and Sunil Jacob John (2016). *International Journal of Rough Sets and Data Analysis* (pp. 98-113).

[www.irma-international.org/article/i-rough-topological-spaces/144708](http://www.irma-international.org/article/i-rough-topological-spaces/144708)

### Supporting the Module Sequencing Decision in ITIL Solution Implementation: An Application of the Fuzzy TOPSIS Approach

Ahad Zare Ravasan, Taha Mansouri, Mohammad Mehrabioun Mohammadi and Saeed Rouhani (2014). *International Journal of Information Technologies and Systems Approach* (pp. 41-60).

[www.irma-international.org/article/supporting-the-module-sequencing-decision-in-til-solution-implementation/117867](http://www.irma-international.org/article/supporting-the-module-sequencing-decision-in-til-solution-implementation/117867)