# WHAT SHOULD THE PERFECT ONLINE ASSESSMENT SYSTEM LOOK LIKE?

MOSTERT EL-MARIE
University of Pretoria, Department for Education Innovation, South Africa
el-marie.mostert@up.ac.za (Corresponding)

DE BRUYN ERIKA
University of Pretoria, Department for Education Innovation, South Africa
erika.debruyn@up.ac.za

PRETORIUS GABY
University of Pretoria, Department for Education Innovation, South Africa
gaby.pretorius@up.ac.za



## ABSTRACT

The focus of this paper is on the characteristics of an effective online assessment tool or system for use in higher education institutions. This refers to a system that can be used to deliver objective assessment items (highly structured tasks and limited responses). This type of assessment is also referred to as computer-based testing (CBT) and our institution has implemented it as part of the assessment strategy as well as a valuable learning tool in various faculties since 1991.

Due to changes in technology or insufficient functionalities, systems have had to be replaced a few times over the years. As part of the support team for academic staff using these systems for more than two decades, the authors have identified and recorded the functionalities needed to enable educators to utilize it to its fullest capacity (not only for assessment but also for teaching and learning). A list of criteria against which a CBT system can be evaluated has been developed and used to identify suitable systems. These criteria specifications encompass the full range of criteria necessary for a "perfect" CBT system from the perspective of the authors. The development of these criteria is based on research, needs expressed by academic staff, personal experience with different systems and educational best practice principles.

It was found that, although not one system adhered to **all** the requirements on this list, many useful systems are available although none of them yet "perfect" according to this list of criteria. This does not mean that these systems could not be used to achieve positive outcomes with regard to good assessment practice, but with some innovative workarounds it can be used effectively. It is hoped that the criteria that will be presented will be useful to other institutions when deciding on an appropriate CBT system.

**Key words**: Computer-based Testing; Online assessment; Computer Assisted Assessment; Objective Assessment

## INTRODUCTION

The use of E-learning, sometimes also referred to as online learning (Moore, Dickson-Deane & Galyen, 2011) in higher education has grown exponentially and more than 80% of higher education institutions present online courses in some or other format (Bichsel, 2013). According to Dabbagh (2005) using technology in education also "redefined the boundaries and pedagogies" resulting in

possibilities to enhance teaching and learning, not readily available previously. An important component of the teaching and learning process is assessment and although it plays a vital role in education, it is often not given the careful attention it deserves. It is inevitable that changes in the delivery mode (using technology) will bring about changes in the way assessment is done as well. Various forms of assessment in face-to-face teaching models are used ranging from class tests, assignments, portfolios, oral and written examinations. In the online environment, some of these methods are still used, but other possibilities are also available.

The ideal strategy is to integrate different forms of assessment to test various knowledge levels in order to obtain valid student marks, although the focus for this paper is on the use of objective assessment. Objective assessment has been used for a long time in a paper-based format and is one form of assessment that is used extensively at the University of Pretoria (UP) – both paper-based and online. Objective assessment is often referred to as multiple choice questions or MCQs and it consists of test questions with highly structured tasks and limited responses to choose from (Linn & Gronlund, 1995 in: Le Roux, 1999). Although in the paper-based format, MCQs are mostly one answer questions, when using technology many other variations are available. Some examples of these variations include multiple response questions (more than one answer from a list of options can be correct), matching columns, drag and drop and hot spot questions, to only mention the most regularly used items. These different question types can be implemented effectively using technology and can be referred to as one of the methods of online assessment or, in the case of our institution, Computer-based Testing (CBT). Technology provides lecturers with a powerful tool to manage and implement this component of their assessment strategy.

Computer-based Testing at any tertiary institution will only be successful if it is based on a sound, reliable and comprehensive system that satisfies the needs of its users. Although every Learning management system (LMS) has some or other form of a quiz or test tool – it was found that these were not suitable for all the users in our institution and therefore an investigation into other available systems was conducted. This paper will present a list of criteria that, according to this investigation, consulting with our users and the experience of the instructional designers, are the functionalities that the authors believe should be available in a “perfect” CBT system.

**BACKGROUND AND LITERATURE REVIEW**

The Department for Education Innovation (EI) is responsible for all e-learning activities at UP including web-supported learning, multimedia development, mobile learning and CBT. This institution has used CBT since 1991 (Le Roux, 1999). The number of tests administered has grown from 37 000 in 1992 to 149 989 in 2014. CBT forms only one facet of the assessment strategy of the academic departments of the University of Pretoria and are used for formative and summative assessment. Due to changes in technology and delivery platforms, systems need to be replaced regularly and the selection of alternative systems proves to be a very difficult task. To assist in the selection process an investigation of the features of different systems was done while also taking the unique needs of the users (lecturers) from different disciplines into consideration. It was soon evident that no one system contains all the required functionalities. This was confirmed by a study in the United Kingdom conducted by Mackenzie (2003:185) who stated that “No assessment system currently available delivers all the features that are likely to be required by tutors wishing to deliver assessments online”. The authors believe that this is still the case in 2014. Improvements to CBT

systems (especially within LMSs) are on-going in an attempt to provide users with the best possible assessment tools (Hunt, 2014).

Various terms are used when referring to CBT, e.g. Online assessment, Computer-aided Assessment (CAA), Computer Assisted Assessment (CAA), Computerised Assessment (CA) and Computer-based Assessment (CBA) (Bull & McKenna, 2003) but in this paper CBT will be used.

In the literature, even today, very little information on all the features necessary for an effective CBT systems are available as were confirmed by Valenti, Cucchiarelli & Panti (2001) and Sclater & Howie (2003). They indicated that minimal research has been carried out regarding requirements for CBT systems. Sclater & Howie (2003) emphasised the user requirements based on twenty-one user roles they identified for the 'ultimate' CBT system and categorized these roles as follows:

- System, user and group administration
- Questions
- Tests
- Sales
- Instances of a test
- Responses and results

Although some of these roles fit into the structure of our institution, not all are applicable. At our institution, the roles can be divided into the following categories:

- Technical (IT related)
- Question Management (functional)
- Test Management (functional)
- Implementation (IT related)

Both the IT related and functional roles are important and dependant on each other for successful implementation and use of a CBT system.

Sclater & Howie (2003) identified a lack in research with regard to requirements for different question types that can be used in testing students' application of concepts and deep learning. At UP the aim of objective assessment is not merely to test knowledge, but also address all levels of Bloom's Taxonomy, namely, knowledge, understanding, application, analysis, synthesis and evaluation (Bloom, 1956) and therefore some research in this regard was done at UP. De Bruyn, Mostert & Van Schoor (2011) indicated that the different question types available in CBT systems and the inclusion of multimedia elements enable lecturers to use objective assessment optimally to assess high-order thinking skills.

The study of Valenti, *et al* (2001) named 'A Framework for the Evaluation of Test Management Systems', indicated that a lack of criteria for educational teams that wish to purchase an appropriate CBT system to their environment exists. They attempted to provide some guidelines for evaluating a Test Management System (TMS) and compiled the framework as presented in Table 1.

*Table 1: Valenti et al (2001) evaluation criteria*

| Metric | Criteria |
|---|---|
| Interface | - Friendly graphical user interface<br>- Editing questions and tests should be easy |
| Question Management | - Different types of questions can be used<br>- Question structure should comply with the assessment strategy |
| Test Management | - Questions for tests and examinations can be prepared<br>- Test evaluation and analysis can be done<br>- Analysis of responses is available<br>- Test banks of questions are available |
| Implementation Issues | - The test environment must be secure<br>- Compatibility with other software and the ability to import and export data |

Valenti, *et al* (2001) highlighted two studies undertaken in this field, namely Fremont & Jones (1994) and Gibson, Brewer, Dholakia, Vouk & Bitzer (1995). Valenti, *et al* (2001) stated that when evaluating a CAA system, one should look at both the Test Management System (TMS) as well as the Test Delivery System (TDS). A TMS allows the instructor to create questions and tests and to evaluate the tests, while a TDS enables the administration of examinations and the collection of tests by interfacing with students. For the purpose of this paper no distinction between a TMS and TDS is made.

Gibson *et al* (1995) identified six evaluation criteria that a CBT system should adhere to as summarised and described in Table 2.

*Table 2: Gibson et al's (1995) six evaluation criteria*

| Criterion | Description |
|---|---|
| Testing | The following criteria were considered:<br>- types of questions<br>- help & hints<br>- number of retries<br>- feedback<br>- multimedia |
| Tracking | A summary of the student's progress and performance |
| Grading | Grades are available for lecturer and students |
| Tutorial Building | The inclusion of tutorials |

| Criterion | Description |
|---|---|
| Implementation | This refers to ease of use and platform issues |
| Security Issues | Security for:<br>- test material<br>- student tracking information<br>- HTML source code<br>- test management |

Fremont & Jones (1994) embarked on a project to implement a computerised test question database to allow instructors to construct paper-based examinations based on a number of criteria. The general requirements were to provide a way to easily create and maintain a database of questions, and create tests from these questions. The ten specific requirements they compiled were:

- The ability to store a variety of question types in the database, including multiple choice, fill in the blanks, short answer, true/false, and long answer.
- The ability to include text, graphic images for diagrams, and properly display mathematical symbols for all questions.
- The ability to link each question to additional fields, which can be used to store information such as question type, learning objective, cognitive level, level of difficulty, actual usage, statistics, etc.
- The ability to communicate with other software products through text files and conversion utilities.
- The availability of the system on more than one computer platform.
- The availability of testing software to be used by various departments to ensure that development, maintenance and expertise are shared.
- The system should be easy to learn and use.
- The preparation of examinations should allow lecturers to select specific questions or to randomly select questions based on various fields.
- The flexibility to format paper-based examinations to meet the requirements of lecturers and departments.
- The availability of item analyses as well as how students performed in a group and individually.

Sun, Cheng and Finger (2009) emphasised that any system used as part of an e-learning strategy should take the functionalities needed by the users of that system into consideration, and therefore the input from users (in this case the academics) is an essential part of creating a criteria list for an effective CBT system.

One way to determine the most suitable system for an institution would be to use the criteria identified for such a system and select a number of off-the-shelf products that adhere to most of these criteria. The products must then be arranged from "best" (adhering to most criteria) to the "worst" (adhering to least of the criteria). The product that best meet the requirements will then be selected and implemented.

The authors recommend this process to be followed to identify a suitable system for an institution, while the evaluation of the systems are based on the detailed criteria presented in Tables 3 - 6.

**RESEARCH METHODOLOGY**

An action research perspective is taken in this study - research is all about generating new knowledge while action research creates this knowledge based on practical experience in a specific context (Koshy, Koshy & Waterman, 2011). Koshy *et al* (2011) emphasise the importance of sharing this knowledge as a basic principle of action research. In this study the identification of the different criteria happened over a long period of time, generating new knowledge as described below.

After studying the available literature on the subject it was clear that a need to specify criteria for an ideal or "perfect" CBT system for our institution existed. The objective of this research is to identify the criteria and functionalities for such a perfect system for our institution based on the following:

- A literature review
- A study of national and international systems
- The unique needs of academics from different disciplines in our institution
- Personal experience with different systems as encountered by instructional designers while consulting lecturers on the effective use of objective assessment
- Educational best practice principles.

Input from academic staff was obtained through discussions with experienced CBT users. There were also representatives from the different faculties serving on a steering committee whose main task it was to define the needs of lecturers with regard to an effective online assessment tool. Other members of the steering committee included the Deputy Director, the Head e-learning and the CBT project manager of the department for Education Innovation as well as representatives from Information Technology Services (ITS).

**RESEARCH FINDINGS**

Based on the five areas as described above the authors compiled the following criteria specifications – divided into the following four categories:

- Technical criteria (IT related)
- Question management criteria (Functional)
- Test management criteria (Functional)
- Implementation criteria (IT related)

Each of these criteria is described in more detail in Tables 3 - 6.

*Table 3: Technical criteria*

| Criterion | Description |
|---|---|
| Different platforms | The system runs on different platforms, namely Novell, NT, Linux or any platform as indicated by the general IT policy of our institution |
| Web-based system | The system has a web-based interface |
| Standalone version | A test can run on a standalone personal computer without using the web or network |
| Technical support | Support is reliable and fast at all times |

The technical criteria do not include all the requirements of the Information Technology (IT) department in an institution, but is only based on the requirements of the users (academics) and the instructional designers providing support for CBT users.

It is important that the system allows proper management of the question database and provide appropriate question types to enable lecturers to use it effectively. The question management criteria are presented in Table 4.

*Table 4: Question management criteria*

| Criterion | Description |
|---|---|
| **Question types** | |
| Information screen | - This screen provides information but does not contain a question<br>- There is no limit on the amount of text that can be added to this screen |
| Multiple Choice | - This question type allows the user to select only one correct answer<br>- Although only one answer can be selected, more than one answer can be indicated as correct<br>- Marks can be set per option |
| Multiple Response | - This question type allows the user to select more than one correct answer<br>- The number of responses a student can select can be limited<br>- Marks can be set per option |
| Fill-in the blanks | - This question type gives the user the opportunity to type short text answers<br>- Provision is made for more than one possible correct |

| Criterion | Description |
|---|---|
| | answer<br>- Boolean operators can be used for marking<br>- Case sensitivity can be set on or off |
| Hot-Spot(s) | - This question type gives the user the opportunity to identify an area on the screen (graphic/text) by clicking on it<br>- More than one hot spot area can be defined as correct and the student can select any of these areas<br>- More than one hot spot can be defined and must be selected by the student |
| Drag and Drop | - A marker or markers can be dragged to a specified area on an image<br>- There is no limitation on the number of markers / defined areas on an image<br>- Different markers can be selected<br>- Different markers can be added to the system |
| Matching | - This question type gives the user the opportunity to match data in columns.<br>- The options in the selection box can be matched to more than one option in the column<br>- There is no limit on the number of options that can be included in each column<br>- The options in both columns can be randomised<br>- 'Column B' can have more options than 'Column A'<br>- Marks can be set per match<br>- Marks can be set for "all or nothing" |
| Numeric | - This question type gives the user the opportunity to type in a numeric answer<br>- Only numeric answers are accepted<br>- Absolute or defined ranges can be set<br>- The number of decimal places can be limited |
| Calculated | This question type creates numeric questions that can be randomly created according to set parameters within specified ranges |

| Criterion | Description |
|---|---|
| Free format | - This question type gives the user the opportunity to type an open-ended answer with no limit to the amount of text a student can enter<br>- These answers can be exported to text file format to be marked by the lecturer<br>- The mark provided by the lecturer can be added into the system |
| **Creating questions** | |
| Importing questions | Questions can be imported from a text file in a specified format to indicate the stem, distractors and key |
| Question code | Each question has a unique question code that can be used to search for, and select questions for inclusion in a test |
| Randomising | The options in multiple choice, multiple response and matching questions can be randomised |
| Formatting | Text can be formatted easily within questions, e.g. bold, italics, underline, special characters, scientific notation, etc. |
| Multimedia / program call | - Multimedia can be embedded in any question (graphics, sound, video, animations)<br>- Links to multimedia calls to open video, sound, animation files and other programs can be included in questions |
| Graphics | One or more graphic can be imported into any question or option |
| Scoring | - The score can be set per option<br>- Marks can be indicated as fractions<br>- Negative marking is possible |
| Number of options | There is no limitations on the number of options that can be included in any question |
| Preview question | A preview option is available and it provides the exact view that a student will see when taking a test, including feedback that the student will receive |

At UP, managing tests is the responsibility of instructional designers and it is equally important that different possible scenarios in setting up tests are available, as described in Table 5.

*Table 5: Test management criteria*

| Criterion | Description |
|---|---|
| Categorisation | Questions can be categorised according to topic, types, difficulty value etc. |
| Compilation of test | - A test can be compiled with questions from different topics<br>- Tests can be created on a random basis per student<br>- Tests can be created on a random basis per group of students<br>- Questions can be flagged to be included/excluded in a test |
| User rights and access | - Different levels of administrator rights can be assigned<br>- Lecturers have access to their own data in the system<br>- A password is needed to log into the system<br>- Students have to enter a unique password to access a specific test<br>- Only students linked to a specific test can take the test |
| Surveys | The system is suitable for surveys |
| Printed test | A printed test and a memorandum can be compiled |
| Saving a test | A test can be saved in text format |
| Control mechanisms | - A time limit can be set for each test<br>- The sequence of questions can be randomised<br>- The distractors / options per question can be randomised<br>- The system limits the number of times a student can write a test<br>- Navigation buttons can be selected per test, to be included / excluded, e.g. backward / forward buttons<br>- Students can be “forced” to go through all the questions at least once before exiting the test<br>- Students can be allowed to exit the test before completing all the questions<br>- After exiting a test, students can continue the test from |

| Criterion | Description |
|---|---|
| | the last question they answered<br>- Navigation in a test can be allowed or disallowed<br>- Jumps to specific questions can be added to a question based on the selection made by the student<br>- A specific date for a test to be active can be set<br>- A specific time for a test to be active can be set<br>- Students can access tests they have completed for revision purposes<br>- One test can branch to another<br>- The following information is displayed on the screen while a student is taking a test: test name, student number, time available, current question number, total number of questions in the test |
| Feedback | - Feedback can be set on / off<br>- The event of feedback can be set, e.g. after all the questions / after each question / not at all<br>- The format of the feedback can be determined, e.g. graphical / text / own format<br>- The feedback indicates what the student answered as well as the correct answer<br>- Feedback of final marks can be set on / off<br>- Extra time can be set for students to work through the feedback after test completion<br>- Feedback time can be limited<br>- Students can provide comments per question while viewing the feedback per question |
| Reporting | - The following results on student performance can be obtained:<br>- Student number, name and mark in Excel, Word and txt format<br>- Student number and mark per question in Excel<br>- Results per topic per student<br>- Average per group<br>- Time taken for individual students to complete |

| Criterion | Description |
|---|---|
| | the test<br>- Average time of the group<br>- Date and time of test taken per student<br>- A full report per question is available, and includes:<br>- Difficulty value of questions<br>- Discrimination index<br>- Standard deviation<br>- A graphical presentation of results<br>- The number of times each option has been selected<br>- Student results can be remarked after editing a test, e.g. if a question is deleted or an answer edited<br>- Questions can be sorted in a report according to the original sequence after a test has been taken |

When implementing a system, stability is one of the key aspects that should be taken into consideration. In Table 6 all the criteria that should be taken into consideration during implementation are presented.

*Table 6: Implementation criteria*

| Criterion | Description |
|---|---|
| Stability and speed | - The system is stable at all times<br>- The system is able to handle large groups of students (more than 200) per session<br>- Answers are saved in real time (if a power failure occurs the answers must be saved up to that point)<br>- The delivery speed of the test from the server to the workstation is acceptable |
| Security | - Only registered students are able to access a test<br>- The test can be made available on specific dates and times<br>- The number of times a student can access a test can be set<br>- The login and logoff time of a student is available |

| Criterion | Description |
|---|---|
| | - The system can limit logins to a specific subnet<br>- The IP address of the computer the test is taken on, is captured |
| Support and training | - Local technical support is available<br>- There is quick response time in the event of technical problems<br>- Extensive training is available to enable clients to use the system to its full capacity<br>- There is a service level agreement between the user and the support team |

Commercial, off-the-shelf products can be rated against these criteria. It is the authors view that if **all** the criteria are adhered to, an institution would have a "perfect" system.

## CONCLUSION

When selecting any teaching and learning systems, the needs of the specific institution should play an important role, it is however also important to investigate international trends and best practice principles with regard to teaching and learning. Bichsel (2013) confirms that e-learning provides the opportunity to increase enrolment, while an effective CBT system provides the opportunity to assess the larger number of students regularly and enhance deep learning, without too much more effort from the lecturer.

UP used these criteria to select a CBT system and although it may not adhere to all the criteria yet, these criteria lists assisted the task team to make an informed decision. The decision is therefore not based on financial or IT related issues alone, but takes into consideration the needs of academic staff as well as educationally sound principles.

It is hoped that the criteria presented in this paper will be useful to other higher education institutions when deciding on an appropriate and effective CBT system.

## REFERENCES

Bichsel, J., (2013), The State of E-learning in Higher Education: An Eye toward Growth and Increased Access (Research Report), Louisville, CO:EDUCAUSE Center for Analysis and Research, available online: http://www.educause.edu/ecar.

Bloom, B.S., (1956), Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain. New York: David McKay Co Inc.

Bull, J., and McKenna, C., (2003), Blueprint for Computer-Assisted Assessment. UK: Routledge.

Dabbagh, N., (2005), Pedagogical models for E-Learning: A theory-based design framework. International Journal of Technology in Teaching and Learning 1(1), 25-44.

De Bruyn, E., Mostert, E., and van Schoor, A., (2011), Computer-based testing - the ideal tool to assess on the different levels of Bloom's taxonomy. Interactive Collaborative Learning (ICL), 14th International Conference. Available online: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6059623.

Fremont, D.J., and Jones, B., (1994), Testing Software: a review. New Currents 1.1. Available online: http://www.ucalgary.ca/pubs/Newsletters/Currents/Vol1.1/TestingSoftware.html.

Gibson, E.J., Brewer, P.W., Dholakia, A., Vouk, M.A., and Bitzer, D.L., (1995), A Comparative Analysis of Web-Based Testing and Evaluation Systems. Available online: http://renoir.csc.ncsu.edu/MRA/Reports/WebBasedTesting.html.

Hunt, T., (2014), Goals of an online assessment system. International Journal of Technology in Teaching and Learning, 1(1), 25-44. Available online: https://docs.moodle.org/dev/Goals_of_an_online_assessment_system.

Koshy, C., Koshy, V., and Waterman, (2011), Action research in Healthcare. SAGE publication. Available online: http://www.sagepub.com/upm-data/36584_01_Koshy_et_al_Ch_01.pdf.

Le Roux, I., (1999), Computer-based Testing in Higher Education. Unpublished Master's dissertation. University of Pretoria, Pretoria.

Mackenzie, D., (2003), Assessment for E-Learning: What are the Features of an Ideal E-Assessment System? Computer Assisted Assessment (CAA) Conference: Proceedings for the 7th CAA Conference. Available online: http://www.caaconference.com

Moore, J.L., Dickson-Deane, C., and Galyen, K., (2011), e-Learning, online learning, and distance learning environments: Are they the same? Internet and Higher Education (14), 129–135.

Sclater, N., and Howie, K., (2003), User requirements of the "ultimate" online assessment engine. Computers & Education, (40), 285-306.

Sun, P., Cheng, H.K., and Finger, G., (2009), Critical functionalities of a successful e-learning system – An analysis from instructors' cognitive structure toward system usage. Decision Support Systems (48), 293-302.

Valenti, S., Cucchiarelli, A., and Panti, M., (2001), A Framework for the Evaluation of Test Management Systems. Current Issues in Education, 4(6). Available online: http://cie.ed.asu.edu/volume4/number6/