

# An Evaluation and Assessment System for Online MCQ's Exams

Asem Omari

College of Computer Science and Engineering, Hail University, Hail, Kingdom of Saudi Arabia

Email: a.omari@uoh.edu.sa

**Abstract**—Examination is one of the common methods to assess the level of knowledge of the students. In order to improve the quality of teaching, it is believed that teacher must be able to set good or proper exam questions. A good and reasonable exam consists of questions that are able to find out students learning levels. In this paper, we propose a computer system designed to evaluate the quality of online exam questions. We took the online exam of computer skills course introduced to undergraduate students at Jarash University in Jordan as a case study. The system takes online exam questions that are stored in a question bank as input, and based on some exam evaluation criterion, the system gives a report of difficult, easy, and medium level, and excellent questions. Based on the report, the teacher can delete bad questions, improve weak question, and sustain good questions. Furthermore, the system saves a lot of time and efforts needed to evaluate exam questions in the traditional methods.

**Index Terms**—exam evaluation system, multiple choice questions (MCQs), evaluation methodology

## I. INTRODUCTION

Educational measurement at the university level has been moved in the last years from the paper-and-pencil testing towards the use of computer and/or Internet-based testing. Computer-based testing refers to performing examinations via stand alone or network computers [1]. Computer-based tests can be found at all educational levels and in many universities all over the world [2]. Many researchers compared the equivalence of computer-based and paper-and-pen tests and most of them conclude that computer may be used in many traditional multiple-choice test settings without any significance on student performance [3], [4], [5].

Based on [6], the anatomy of multiple-choice test contains two basic parts: a statement or a situation, a problem (question) and a list of suggested solutions (alternatives or options). The question may be constructed in the form of a question or of an incomplete statement and the list of options must always contain at least one correct or best alternative one and a number of incorrect options (distracters).

The multiple choice questions cover the following classes:

- A. Single correct answer: all except one of the options are incorrect; the remaining option is the correct answer;
- B. Best-answer: the alternatives differ in their degree of correctness. Some options may be completely incorrect and some incompletely incorrect, but at least one option is completely correct;
- C. Multiple responses: two or more of the options but never all four options are keyed as correct answer.

In all levels of education, students have to take tests and assessments to demonstrate their ability, for example, to show whether they have fulfilled the course objectives or to guide them in their further learning. The tests are one of the assessment methods varied, and is an important way to measure the level of student achievement, and to identify the extent to which the curriculum goals set for him, and reveal the strengths and weaknesses in it.

Normally, the exam lifecycle consisting of four parts which are:

- Designing the exams.
- Conducting the exams.
- Checking the exams.
- Assessment of exam.

An important part of exam's lifecycle is exam assessment. Certainly test assessment by a group of different people in the absence of a common evaluation and assessment tool will lead to divergent results of the evaluation. Therefore, the existence of such a tool is essential for those who want to issue decisions with respect to exam and exam questions based on a solid infrastructure. Exam assessment is the discovery of the advantages that can be highlighted and/or disadvantages that can be avoided in exam questions based on clear scientific standards. The assessment results need to be valid and reliable. This can only occur when the assessment instruments that are used to assess the exams are of good quality.

The purpose of this paper is to introduce the outline of a new evaluation system for online exam questions. This paper is constructed as follows: Related work is provided in section 2; we will go through some available evaluation systems. Then in section 3, the principles and standards on which we built our evaluation system are

specified and the design of the system is discussed. Experimental work is provided in section 4. And finally, in section 5, we summarize and derive the future work.

## II. RELATED WORK

When the area of subjects in a course is large, a practical idea is to create a multiple choice examination system. The practice is prevalent because multiple choice examinations provide a relative easy way to test students on a large number of topics. More, for large number of students, a classical evaluation system (written evaluation) consumes a considerable amount of time. The characteristics of a good multiple choice exam are introduced in [7] and [8].

One of the reasons to evaluate test quality is that it is necessary to decide whether the use of a certain test for an intended decision is justified. We would like to know whether a test is good enough for the stated purpose [10]. An answer to a question may not be an evidence of whether the question is good or not. To evaluate test quality, several evaluation systems and standards are available [11]. The currently available evaluation systems, however, tend to focus around one specific type of test or test use [9]. Standards are often more broadly defined, but are aimed at guiding test developers during the development process and are not suited for an external evaluation of quality. In the next section, we propose our online exam assessment system and discuss our proposed assessment and evaluation criterion.

## III. EVALUATION SYSTEM DESIGN

Quality is defined as the degree to which something is useful for its intended purpose. In testing and assessment practice, the variety of intended purposes is very large and, furthermore, the solutions chosen to reach those purposes are endless. And, when quality is defined as being dependent on the purpose of a test, it seems hard, or even impossible, to develop an evaluation system with fixed criteria that are suitable for all possible tests and assessments [12]. Standards mention aspects of quality that you should comply with, in order to develop sound and reliable tests. Evaluation systems focus on evaluating a test, and decide what quality aspect must be met to ensure minimal quality.

Therefore, our evaluation system also includes other evaluation criteria that do lead to a result that states whether an exam question is good enough. These criteria are built into the system in such a way that, once the evaluation result is introduced an action of one of three actions is done. The exam administrator can delete, modify, and sustain questions.

Our evaluation system will be a computer application that consists of two modules. These modules are: evaluation, and reporting. The application is designed for use after the test conduction process, but can also be used for the evaluation and modifying of existing tests.

### A. Evaluation Criteria

The education scientists always introduce some guidelines to exams designers to be considered when developing good test questions [13] such as:

- The questions should be linked to the educational objectives to be achieved, which are represented in the learning outcomes.
- The questions should be formulated as precisely and clearly to enable the student to understand it easily.
- The number of questions should be suitable to the introduced exam time.
- Questions should vary to include easy and medium difficulty, and other difficult questions to verify the ability to distinguish between students.

Beside those guidelines, the exam questions have to be evaluated after exam conduction in order to guarantee for a high percentage the quality of the exam. In our system, we implemented different equations that measure the difficulty, easiness, and excellence of exam questions. Here we explain how to determine the coefficients of ease, difficulty, and excellence of exam questions and particularly in multiple-choice online exams:

- Difficulty coefficient

Difficulty Coefficient is defined as: the percentage of students who answered the question correctly. Difficulty coefficient is calculated as follows:

$$Dq = T/N \quad (1)$$

where:

$Dq$ : is the difficulty Coefficient.

$T$ : Number of students who answered the question correctly.

$N$ : The total number of students who answered the question.

For example; If we assume that (40) students from (100) answered the first question correctly, so the difficulty coefficient for this question is:  $40/100 = 0.4$ .

Since the difficulty coefficient is a ratio, so its value is between zero and one, and when the coefficient of difficulty is zero or close to zero it is a sign that the question is very difficult, and if its value is 1 or close then that means that the question is very easy. This means that the difficulty factor inversely associated with easiness of question in the sense that the high difficulty coefficient value of a question is an indication of ease of the question. So, from the same equation, we can calculate the easiness coefficient of a question. It is recommended that the difficulty values are between 0.50-0.75. The exam designers recommend putting some easy questions at the beginning of the exam to encourage students, but some hard questions that determine strong students are posted at the end of the exam.

- Excellence coefficient

A good test distinguishes between students who know the material and those who do not, and more than that distinguishes between those who know the material and understand more and those who understand less.

And the degree to which question distinguishes between students' knowledge and is able to see the contrast between students is called Excellence coefficient. To calculate Excellence coefficient, we take the top 25% of students to represent the upper group, and the lowest 25% to represent the lower group, then, we calculate correct answers to a question in both groups and then we calculate the excellence coefficient as follows:

$$\text{Excellence coefficient} = (X-Y)/0.25N \quad (2)$$

where:

X = Number of students who answered the question correctly from the upper class.

Y = number of students who answered the question correctly from the lower class.

N = sample size.

As a general rule, the question excellence coefficient of 0.2 or more is considered to be a good question.

### B. System Features

As shown in Fig. 1, the main page of the system has the following options:

- **Create Questions:** used to create or add new questions to the database of questions. In order to add a MCQ into database, the administrator must log in the system. When the administrator decides to add MCQs into database, he has to click on "Create Questions" button.

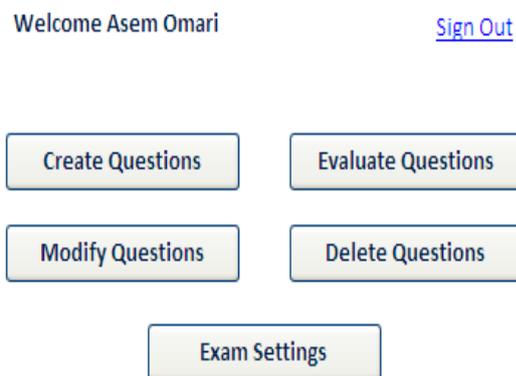


Figure 1. Main page of the exam administration and evaluation system

- **Modify Questions:** used to modify existing questions in the database of questions. In order to modify existing question, the administrator has to click on "Modify Questions" button and choose the question he wants to modify and then save changes by clicking on the save command button.
- **Delete Questions:** used to delete existing questions in the database of questions. In order to delete an existing question, the administrator has to click on "Delete Questions" button and choose the question he wants to delete.
- **Evaluate Question:** used to evaluate existing questions in the database of questions based on the answers of the students to that question which are stored in the database of the system. The result of the evaluation process gives a report of easy, difficult, and excellent

questions. This report can be used to delete, modify, and sustain questions.

- **Exam Settings:** used to set exam duration time, number of questions, and number of easy, difficult and medium level questions. The default settings are to select 10 questions from each level.

### IV. EXPERIMENTAL WORK

The implementation of online testing and evaluation system was performed by the use of a relational database that stores multiple choice questions, student's information and information regarding the evaluations. The questions database contains more than 700 questions. Each exam contains 30 problems with four possible options.

The exam is time limited. The moment when the test is generated and the moment when the test solutions are sent to database are recorded. It is not possible to give up the exam after it was generated. The examination system generates questions randomly from a pool of difficult, easy, and excellent level questions. As mentioned previously, the default settings are to select 10 questions from each pool. This can be adjusted through exam setting command button.

With 30 multiple choice examinations, the tests were applied on first year undergraduate students in Computer skills course from Jarash private University, Jordan. The multiple choice evaluation tests were applied on 10 groups of first year undergraduate students who took the same course in the university in the first semester of study. The evaluation system found a lot of very easy and very difficult questions. Some of them were deleted and the others were modified.

After each exam, a questionnaire was filled by the students in order to evaluate their opinions on the exam questions. The questionnaire results showed an increasing positive opinion about the exam question in parallel to the evaluation and assessment process progress. This indicates the efficiency of the evaluation and assessment process used.

### V. SUMMARY AND FUTURE WORK

This paper introduces a new evaluation system for the quality of online exam questions. This new evaluation system includes different quality criterion on which the evaluation is conducted according to them. As a future work, we plan to make the decisions with respect to good or bad questions automatic. This means that the system will automatically remove very easy and very difficult questions and notify the administrator of that change.

### REFERENCES

- [1] J. B. Olsen. (June 2012). Guidelines for computer-based testing. [Online]. Available: <http://www.isoc.org/oti/printversions/0500olsen.html>
- [2] G. Frosini, B. Lazzarini, and F. Marcelloni, "Performing automatic exams," *Computers & Education*, vol. 31, no. 3, pp. 281-300, 1998.
- [3] P. Peak. (July 2011). *Recent Trends in Comparability Studies*. [Online]. Available:

- www.pearsonedmeasurement.com/downloads/research/RR\_05\_05.pdf
- [4] M. Russell, A. Goldberg and K. O' Connor. (July 2013). Computer-based testing and validity: A look back and into the future. [Online]. Available: <http://www.bc.edu/research/intasc/PDF/ComputerBasedValidity.pdf>
- [5] R. MacCann, "The equivalence of online and traditional testing for different subpopulations and item types," *British Journal of Educational Technology*, vol. 37, no. 1, pp. 79-91, 2006.
- [6] L. JÄNTSCHI, "Auto-calibrated online evaluation: Database design and implementation," *Leonardo Electronic Journal of Practices and Technologies*, no. 9, pp. 179-192. 2006.
- [7] D. P. Wegener. *Text Construction*. [Online]. Available: <http://www.delweg.com/dpwessay/tests.htm>
- [8] K. Scouller, "The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay," *Higher Education*, vol. 35, no. 4, pp. 453-472, 1998.
- [9] S. Wools, *Towards a Comprehensive Evaluation System for the Quality of Tests and Assessments*, 2011.
- [10] M. Kane, "Certification testing as an illustration of argument-based validation," *Measurement*, vol. 2, pp. 135-170, 2004.
- [11] M. Kane, *Validation*. In R. Brennan (Ed.), *Educational Measurement*, 4th ed. Westport, CT: American Council on Education and Praeger Publishers, 2006, pp. 17-64.
- [12] "The Cambridge approach. principles for designing, administering and evaluating assessment," *Cambridge: Cambridge Assessment*, 2009.
- [13] D. Bartram, "The development of international guidelines on test use: The international test commission project," *International Journal of Testing*, vol. 1, no. 1, pp. 33-53, 2001.



**Asem Omari:** is an Assistant professor of computer science at Hail University, Kingdom of Saudi Arabia. Dr. Omari obtained his PhD in Computer Science from Heinrich Heine University, Dusseldorf, Germany in 2008. He obtained a Graduate Certificate in Computer and Information Sciences from the University of Michigan/Dearborn, USA in 2002, and a bachelor degree in Applied Mathematics, System Analysis and Programming orientation from Jordan University of Science and Technology in 1999. His research interests include Data Mining and knowledge discovery from databases, e-commerce, e-learning and e-government.