Scientific Journal Impact Factor (SJIF): 1.711

e-ISSN: 2349-9745
p-ISSN: 2393-8161

International Journal of Modern Trends in Engineering and Research
www.ijmter.com

An Adaptive Evaluation System to Test Student Caliber using Item Response Theory

Safa Hamdare
Computers Department, SFIT

Abstract: Computational creativity research has produced many computational systems that are described as creative [1]. A comprehensive literature survey reveals that although such systems are labelled as creative, there is a distinct lack of evaluation of the Creativity of creative systems [1]. Nowadays, a number of online testing websites exist but the drawback of these tests is that every student who gives a particular test will always be given the same set of questions irrespective of their caliber. Thus, a student with a very high Intelligence Quotient (IQ) may be forced to answer basic level questions and in the same way weaker students may be asked very challenging questions which they cannot response. This method of testing results into a wastage of time for the high IQ students and can be quite frustrating for the weaker students. This would never benefit a teacher to understand a particular student's caliber for the subject under Consideration. Each learner has different learning status and therefore different test items should be used in their evaluation. This paper proposes an Adaptive Evaluation System developed based on an Item Response Theory and would be created for mobile end user keeping in mind the flexibility of students to attempt the test from anywhere. This application would not only dynamically customize questions for students based on the previous question he/she has answered but also by adjusting the degree of difficulty for test questions depending on student ability, a teacher can acquire a valid & reliable measurement of student's competency.

Keywords: Intelligent Quotient (IQ), Adaptive Testing, Intelligent System, Artificial Intelligence, Item Response Theory (IRT), Tailored Testing, Computer Adaptive Theory (CAT).

I. INTRODUCTION

The traditional method of Testing using Pen and Paper, since ages around the world uses "One Size Fits All" approach. However there are many shortcomings that the traditional system faces like: the need to being tied to fixed class timings to appear for the test, the need to travel to distant places to reach the place where test is to be conducted, and the need to deal with different students in a different manner. Even if the traditional system of testing is web based, it also faces certain shortcomings like: No separate marking scheme for different level of questions and difficulty level of question doesn't change based on the previous response of the students. It means that still Computer tests are conducted but it is not adaptive to student's caliber, as the level of questions does not toggle between easy, medium and high based on response to previous question and same marking scheme is followed for a correct answer for each level. Despite its weaknesses, traditional test methods have ruled the amount of intellectual and thought processing capabilities in the educational system. In

spite of this dominance, traditional methods have the limitation of sample dependency for estimating the test items parameters namely the item difficulty and item refinement. Nowadays leading testing associations such as ETS (Educational Testing Service) have created adaptive tests using Item Response Theory that dynamically pose questions to the users based on how the user answers the previous questions. Based on this concept this paper proposes an Adaptive Evaluation System (AES) that adapts to the examinee's ability level. For this reason, it has also been called tailored testing. As this test dynamically customizes itself to an individual user thus can't be considered as a standard test having identical set of questions for all the students. As a result it provides very accurate results. An Adaptive system is a different and advanced means which will help in conducting test on computer as well as on mobile device allowing the students to access and attempt the test from anywhere and anytime around the world in order to overcome the limitations of traditional testing methods.

II. ITEM RESPONSE THEORY

Adaptive evaluation systems are computer-based systems for measuring the performance of individuals or of teams by tailoring the presentation order provided by the system based on ongoing user responses in order to match the assessment level to that of the learner or team [3]. Developing such Adaptive Evaluation System has become possible using the concept of Item Response Theory (IRT). Item response theory (IRT) was first proposed in the field of psychometrics for the purpose of ability assessment [2]. It is widely used in education to calibrate and evaluate items in tests and to score subjects on their abilities. During the last several decades, educational assessment has used more and more IRT-based techniques to develop tests [2]. Today, all major educational tests, such as the Scholastic Aptitude Test (SAT) and Graduate Record Examination (GRE), are developed by using item response theory, because the methodology can knowingly progress measurement correctness and reliability while providing potentially important reductions in assessment time and effort, especially via computerized adaptive testing. Lord (1980) invented IRT in the early 1950s, which utilized probability to explain the relationship between the examinee's ability and the question item response [5]. IRT is the probability of responding an item correctly or of accomplishing a particular response level. It is modeled as a function of an individual's ability and the characteristics of the item. An utmost goal of IRT is guessing the probability of an examinee's level of correct response to an item of a specific difficulty. Theoretical range of ability can be measured on a transformable scale from negative infinity to positive infinity; practical considerations usually limit the range of values from –3 to +3 [8]. IRT models are mathematical functions that specify the probability of discrete outcome, such as a correct response to an item, in terms of persons and item parameters. Specifically, a mathematical model, called Item Characteristic Function (ICF) that derives a continuous increasing curve for predicting the ability and test performance, and was developed to infer the examinee's ability or potential quality. There are traditionally three IRT mathematical equations termed, one parameter model (1PLM or Rasch model), two parameter model (2PLM), and three parameter model (3PLM) that are used to make predictions. All three models have an item difficulty parameter (b). The ICF model [6]

- One-parameter model equation is given as follows:

$$P_i\ (\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

- Two-parameter model equation is given as follows:

$$P_i\ (\theta) = \frac{e^{\alpha_i(\theta - b_i)}}{1 + e^{\alpha_i(\theta - b_i)}}$$

- Three-parameter model equation is given as follows:

$$P_i\ (\theta) = C_i + (1 - C_i).\frac{e^{\alpha_i(\theta - b_i)}}{1 + e^{\alpha_i(\theta - b_i)}}$$

In addition, the 2PL and 3PLmodels possess a discrimination parameter (α), which allows the items to discriminate differently among the examinees. The 3PL model contains a third parameter, referred to as the pseudo-chance parameter (c). The pseudo-chance parameter (c) corresponds to the lower asymptote of the item characteristic curve (ICC) which represents the probability that low ability test takers will answer the item correctly and provide an estimate of the pseudo-chance parameter [7]. All IRT-based models have some common features: They assume the existence of latent traits or aptitudes (in our particular case the trait is the examinee's knowledge level) that allow us to predict or explain examinee behavior; The relation between the trait and the answers that a person gives to test items can be described with an increasing monotonous function called the Item Characteristic Curve (ICC) [5].

Table 1: Advantage and Dis advantage of IRT

Advantage of IRT	Disadvantage of IRT
Item statistics are independent of the sample from which they were estimated.	The choice of a model may depend upon the sample available, particularly during the testing phase of an exam.
Examinee scores are independent of test difficulty	Assumptions made for using the IRT model are more rigorous and strict than those required of traditional test methods.
Test analysis doesn't require strict parallel tests for assessing reliability.	IRT models require large samples to obtain accurate and stable parameter estimates, although Rasch measurement models are useful with small to moderate samples
Item statistics and examinee ability are both reported on the same scale	The parameters of the ICC must be previously known for each item.

III. PROBLEM STATEMENT

In traditionally followed methods, system doesn't adapt to students ability level and based on the previous question answered by him/her the difficulty level of next question is not changed. Also students have to be present at the desired location on an exact time to appear for the test. This paper proposes an Adaptive Evaluation System that would evaluate students based on their performance in a test. For example, if a student performs well on medium level question, he will then be presented with a more difficult question of high level. Or, if he performed poorly, he would be presented with a simple question of easy level. Compared to static multiple choice tests where all the students are required to attempt a fixed set of questions, this system requires fewer test questions to be attempted by a student, to arrive at an equal accurate score. As student attempting many high level questions will attempt less questions as compared to student attempting only low level question. The main objective of this system is that it dynamically customizes the test question based on Individual user Caliber. It helps teacher to evaluate student and put student into different categories based on their caliber, this could further play a vital role for focusing more attention on weak students. Also, each

student is attempting different sets of question so there would not be any copying. This system also provides flexibility for the students to attempt the test using any internet enabled mobile handset making it easy for the examinee to appear for the test from anywhere.

IV. PROPOSED SYSTEM METHODOLOGY

The Proposed Adaptive Evaluation System will be implemented over web and would be accessed using a computer or a mobile application for wired and wireless connection respectively. So that test for students could be conducted in Laboratories using computer as well in Classroom using mobile device remotely. In order to serve this functionality, a responsive website would be created and then set of objective type questionnaire to be used would be formed. The system would then be tested to check whether web application is compatible and customizable to screen size of any mobile and also works properly on any browser of the computer. The System successively selects questions so as to maximize the precision of the exam based on what is known about the examinee from previous questions. From the examinee's perspective, the difficulty of the exam seems to tailor itself to his or her level of ability. We have devised a system wherein different levels of questions will have different marking schemes. Initially system poses a question from an easy level to all the examinee. If an examinee answers correctly then adaptive tests require fewer test items to arrive at equally accurate scores. For the test, the following sample marking scheme will be used, assuming a student attempting a test of 50 marks:

Table 2: Marking Scheme for Adaptive Evaluation System

Question Level	Marks	Questions Need to be Attempted
Easy	+2	For Example: First question attempted is of 2 marks, if answered correctly then examinee moves to attempt the next level question of 3 marks, if this also answered correctly then he needs to attempt 9 more questions of difficult level correctly to complete the score of 50. So a student with high IQ would attempt only 11 questions. Whereas if a student is unable to answer correctly the first question of easy level then to complete the test he needs to go through at least 25 questions.
Medium	+3	
Difficult	+5	

Student can login and appear for a test. Based on the student's answer for the given question, the subsequent questions will toggle between different levels. The IRT model is used for this purpose. Item response theory places examinees and items on the same metric. Therefore, if the system has an estimate of examinee ability, it is able to select an item that is most appropriate for that estimate. Technically, this is done by selecting the item with the greatest information at that point. Information is a function of the discrimination parameter of the item, as well as the conditional variance and pseudo guessing parameter.

V. CONCLUSION

The proposed research work is a new and innovative method of conducting tests more efficiently and would be of great help in teaching learning process. This method of testing will allow the students to be assessed accurately in a shorter period of time compared to the traditional testing methods. In traditional testing methods, all students are provided with same set of questions irrespective of their performance in the test. This leads to longer tests with possibly redundant questions. As Questions

are administered to the students dynamically based on their performance in the test; no two students will be given the same set of questions. Thus by making use of this system, a dynamic testing environment would be created which will provide accurate results to students, while at the same time reducing testing time for the teacher. The applications of this project are several. Some possible applications include: It can be used by students preparing for competitive exams such as GRE, CAT, etc.; it can be used by educational institutions for conducting online tests; it can be used by schools and colleges for conducting exams instead of the traditional testing approach which is currently used.

REFERENCES

[1] Anna Jordanous, "Evaluating Evaluation: Assessing Progress in Computational Creativity Research", 2nd International Conference on Computational Creativity, ICCC 2011.

[2] Xinming An and Yiu-Fai Yung, SAS Institute Inc., "Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It", Paper SAS364-2014.
Encyclopedia of the sciences of Learning, Springer Link, Prof. Roy B. Clariana, Simon Hooper.

[3] Kavitha Rajamani and Vijaya Kathiravan, "An Adaptive Assessment System to compose Serial Test Sheets using Item Response Theory", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME) February 21-22.

[4] Omobola O. Adedoyin and J. A. Adedoyin , "Assessing the comparability between classical test theory (CTT) and item response theory (IRT) models in estimating test item parameters", Herald Journal of Education and General Studies Vol. 2 (3), pp 107 - 114 August, 2013.

[5] Yu-Hui Tao, Yu-Lung Wu and Hsin-Yi Chang, "A Practical Computer Adaptive Testing Model for Small-Scale Scenarios", Educational Technology & Society, 11(3), 259–274.

[6] Eduardo Guzmán and Ricardo Conejo, "Self-Assessment in a Feasible, Adaptive Web-Based Testing System", IEEE TRANSACTIONS ON EDUCATION, VOL. 48, NO. 4, NOVEMBER 2005.

[7] Embretson SE, Reise SP, "Item Response Theory for Psychologists", New Jersey: Lawrence Erlbaum Associates, 2000.

[8] Hambleton RK, Swaminathan H, Roger HJ, "Fundamentals of item response theory", Newbury Park. C.A: Sage, 1991.

