University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 4, Issue 4, December 2020
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/


# Efficient Creation of E-Assessment by Redefining Questions Items Using IRT

Nabila Jannat[1], Muhammad Yasir Imam[2], Adnan Naseem[3], Saman Fatima[4], Kashaf-ul-Khair[5]

Department of Computer Science and IT, Alhamd Islamic University, Islamabad, Pakistan

nabila.jannat, muhammad.yasir, adnan.naseem]@alhamd.pk
samanfatima.se@gmail.com, kashafulkhair30@gmail.com

***Abstract:*** Different student evaluation techniques are being used to evaluate students learning abilities. Most of learning techniques follow different evaluation methods but Item Response Theory (IRT) has strong impact among them. Mostly evaluation techniques lack some important evaluation measure such as difficulty level, time index and discrimination level. By redefining questions with the help of IRT, enhanced e-assessment method to be more accurate in terms of difficulty, time and discrimination level. This study will ease the examiners abilities to access the criteria of difficulty of exam. To check the level of difficulty and examiner's abilities accordingly, IRT will be suitable technique to follow. IRT can make a strong influence on aptitude tests like GAT, GRE etc. Meanwhile, many other job-related tests will be scheduled with the help of IRT. This method helps test makers to divide and allocate the time for every question to be asked in test. The IRT generates the test more effectively. Testing authorities and students can take equally benefit from this method.

**Keywords:** Assurance Technologies; Software Engineering; Software Evaluation, Item Response Theory, IRT;

## I. INTRODUCTION

Item response theory is basically set of hidden variables, a method that is intended to model interaction between subject's difficulty levels and item description level. IRT specially focuses on pattern answer rather than on composite answers. In IRT, the entry response is independent variable whereas examinees capability and entry characteristics are dependent variable. There is range of area of research in IRT and survey is largely used for assessment and evaluation of research to define association between examinees and test queries [1].

In recent years, IRT is dominant framework for assessment and development in education system on a large scale. Applications of Item Response Theory are not only available for large-scale assessments but also for small-scale assessment and development in education system such as sociological and psychological assessments. It is also applicable in scaling which includes tool progress, enhancement and equating which includes different methods for creating tests and computer adaptive testing [2].

Classical Test Theory (CTT) was principal method until 1953.While Classical test theory was lies between observed and true scores (including errors) based on linear relationship method. It also includes errors [2].

CTT (Classical Test Theory) has ruled the measurement of cognitive accomplishments in the educational system. Despite dominancy of classical test theory there are some limitation of section dependency for approaching of subject difficulty, and discrimination level.

Classical Test Theory simply assumes that worthy items distinguish by only conducting good tests in examinations. If test conducted very easy every examinee can solve that easily whether if the test is very complicated test cannot be solved by examines. Using Classical Test Theory (CTT) it is very difficult for educators to assume test items and measures examines ability and discrimination level because it is simply sample dependent.

### *A. Comparison of the CTT & the IRT*

In comparison to Item Response Theory and Classical Test Theory which is (CTT) IRT are used in education system for measurement of examinees ability (understand, guessing), discrimination level in educational system. Although research studies are ongoing of IRT in estimating assessment and evolutions of IRT application over last 50 years Classical Test Theory is still used many programs improvably and continuously researched.

However, many testing programs are still implement CTT tools for assessment and an improved testing result has some benefits over IRT. For CTT few advantages many researchers used e.g. CTT identifies relationship among true score and observed score on linear fashion to model for better understanding. The methodology of CTT has some limitations over IRT like CTT has small sample of data while IRT is nonlinear. Analyses in CTT are not to be too good as in IRT because the size of samples in CTT is very small. Minimum assumptions are required to model a problem while in IRT the assumptions are not straightforward [4].

Classical test theory has some disadvantages scores that define examinees expertise are not dependent on test difficulty. However, their scores may be greater in difficult

test and lesser in easier test so true score extraction is difficult. IRT can be used for better development and refinement in such kind of test [5].

In this Paper we will use Item Response Theory to evaluate examines ability of item difficulty, discrimination and time index. By taking some test of different subject we assess and analyze examines difficulty level, discrimination and time index on different nature of questions coming easier to more difficult questions.

## II. Literature Review

This article is about the diagnostic tests that play a main role in student learning level. Various techniques are used for this purpose. The main method is IRT. The IRT gives a variety of techniques for assessing students' abilities, such as charts and also math functions. Curves involves various learning features.

There are two methods which is utilized to understand the students' abilities: one to test score and the other to observe the student performance in the classroom.

The E-Assessment utilizing the IT for Object Learning process. The purpose of this article is to give an easy method to teach, test goals, and also evaluate goals. The Assessment allows students to make better their performance by utilizing an online source such as the WWW. Diagnostic testing is utilized in e-diagnosis. The Courses are organized online to enhance e-learning [7].

This article mainly describes the agenda for maintaining security and controlling a form of fraud in online diagnostics, and defines IRT, scale-related and score equivalents to generate appropriate scores. Will ISOM works for teachers, comparing test parameters under Point CTT and Metro to evaluate how we distinguish between the two techniques. CTT depending on the theory. It cannot effectively measure the level and degree of discrimination. The IRT analysis tool uses two dimensions to fit the model's dimensions. The teachers were positive about student's educational rating, and tried to involve in the process. Also, the allocation was more encouraging for administrative staff than teachers.

Cole available age, not of the test, the socio-economic status, sample size and its results is that the majority of students performed and preferential paper exams. Jackson works on students, teachers, managers, quality officers and its results & the censors are watching e-assessment constructive innovation. Teacher and pupil's training to gain the consistency and acceptance of technologies in e-assessment. Newhouse worked on students and its attainability for any of the games before exams. These two tasks were accomplished deprived of a lot of technical problems.

Eva-Heinrich worked on scholars and its results Moodle and light wave owner were fit for task organization and labelling, also teaching and learning process was maintained.

Adegbija et al. worked on students and it's concluded that all students do not prefer e-assessment. Female students required IT skills because they consider e-exam stressful. Detail table as below with complete authors names and their findings.

Table no.1

| Author | Findings |
|---|---|
| Chigona A. & Chigona, W. 2010 | - Factors mentioned impede ict use in Kenya |
| McDowell etal 2008 | - One tool cannot fit all<br>- Action research can be used to engage lecturers in investigating assessment practices. |
| Ang'ondi E.K. 2013 | - Teachers confirmed that learning of ict add more burden to teachers<br>- They belief, they are not competent enough to handle ICT task without support |
| Sato, M 2008 | - Improvement in practice attributed to standards and Assessment task. |
| Christian, B. 2014 | engaging in assessment task, pre-service teachers became more value literate<br>Ere was a shift in perception of the role teachers play in developing the values of literacy for students |
| Aojula, H. et al 2006 | % of students agreed that integrated is useful in helping them to learn<br>A should be applied to other courses |
| Coulby et al 2011 | - Students found completing of assessment using PDA straight forward |
| Jordan 2011 | - Evaluation leads to improvement in student performance |
| Sanna A. et al 2012 | - Assessment is objective and efficiently implemented |
| Doukas 2007 | - E-exam was preferred by students<br>- It makes exam processes easier |
| Bauer et al 2008 | - E-exams benefits are enormous<br>- Objectivity in marking is critical |

| | |
|---|---|
| Palak D & Walls R.T 2009 | achers attitudes towards technology are the most significant predictor for teacher /student use.<br>achers positive attitudes towards technology do not have the same influence on student technology use. |
| Hermans et al 2008 | - Teachers belief is significant in explaining teachers adoption of technology<br>- Experience, attitude, gender, constructivist belief showed positive effect<br>aditional belief showed negative effect on technology adoption |
| Bee et al 2008 | - They showed positive attitude towards technology<br>- Elderly teachers were eager to adopt ICT<br>- Experience users and on the job learners indicated high training & support needs. |
| Ilomaki 2008 | - Students are motivated users of new technology. Most teachers still find it difficult using ICT<br>- Male students show better skills in technical issues than female students.<br>- Current technologies used in schools are boring and does not provide competence |
| Peralta, H. 2007 | - Level of confidence varies from country to country depending on availability |
| Orlando 2014 | - Cultural, political beliefs and institutions policies affect technology practices and use |
| Buabeng- Andoh C. 2012 | Positive correlation between ICT use and teachers competence.<br>- Teacher's perception use of ICT was positive but not significant.<br>- Inverse correlation between ICT use, age and teaching experience |
| Teo, T. 2008 | No gender or age difference regarding attitude Significant difference for attitude by the subject area of studies<br>Association between years of experience and level of confidence and attitude. |
| Teo, T. Hwee, J. 2010 | - Study validated a 3 fact or scale for measuring teachers self efficacy construct, had High BCS, fairly high WBS and low MRS with regards to self efficacy |
| Oates, G. 2010 | - Aspects of assessment in maths need attention. |

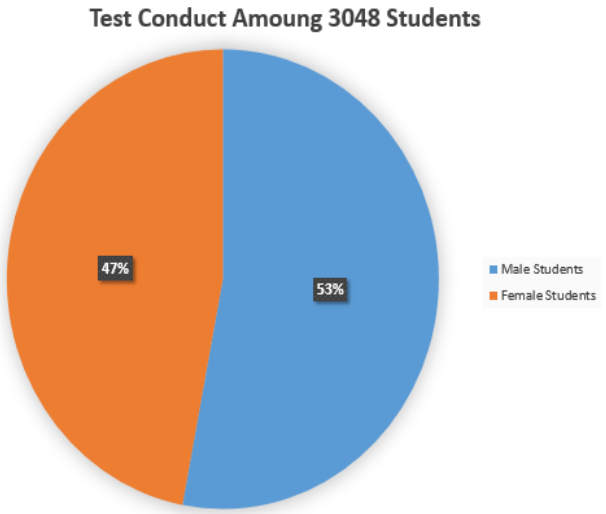


Figure 1 GAT test in Poonch University among 3048 students

## III. METHODOLOGY

The main objective of this paper is to evaluate examinees difficulty level, discrimination level and time index of test items by conducting test. We conduct a GAT test in Poonch University, AJ&K among 3048 students shown in Figure 1, in which 1510 were female and 1538 were male students. An entry test GAT paper of 100 question was consisting of 30 Mathematical, 30 English and 40 analytical questions. The main purpose to conduct GAT test among these students to make e-assessment more efficient and affective by using IRT (Item Response Theory).

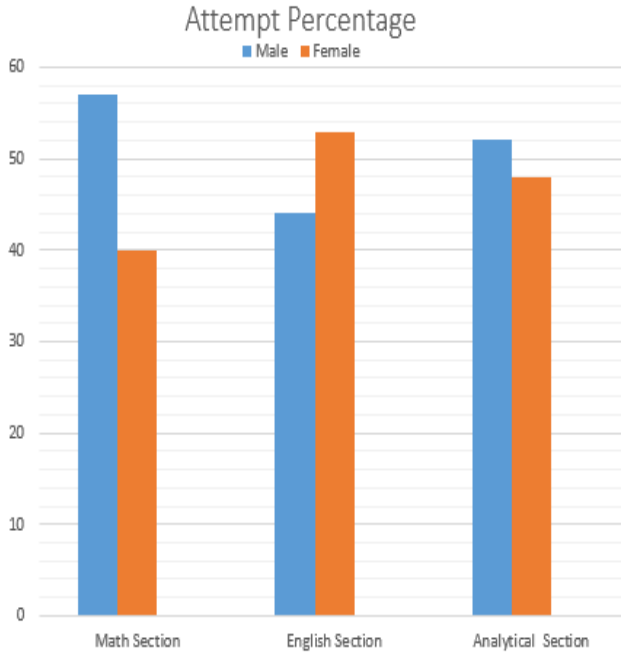


Figure 2 Analyze examinees difficulty level

After conducting the test, we analyzed examinees difficulty level (in English, Mathematics and analytical questions), discrimination level and time index for each question separately. Female student's responses were superior in English questions as compared with male students. Male student's responses in Mathematics questions was excellent as compared with female students as shown in the Figure 2.

In analytical questions, male and female students took nearly same time. Although, 120 minutes were allotted for 100 questions, students were facing trouble in solving complete question paper. Time management, to attempt such type of tests, was another issue which took our attention as shown in Figure 3. Student percentage between male and female is about to 47% female students and 53% of Male students are targeted in GAT test which is conducted in Poonch University, Rawalakot, AJ&K.
In Math section of GAT test the response rate of Female students is 40% and Male students 57%. In other hand, in English section Male students responses about to 43% and Female students 54%. Male students responses good at analytical section as compare to female students.

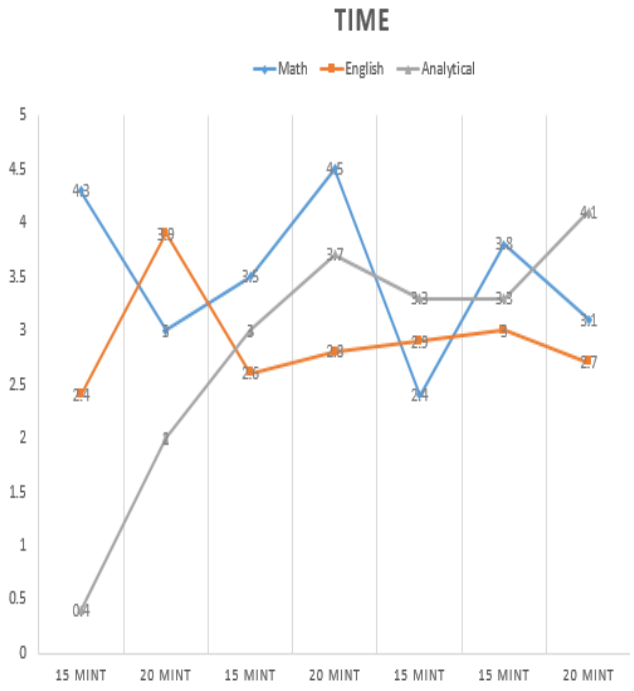


Figure 3 Time Management issue in attempting the test

By using IRT, one can assess examinees responses in solving a problem. Also, measure of item wise difficulty level, time index and discrimination level can be done. Examiners can estimate & manage time for conducting such type of tests. IRT is a very good practice for conducting state of the art GRE & GAT tests and also it is helpful in any type of class tests of Mathematics, Grammar and psychology as depicted in Figure 4.

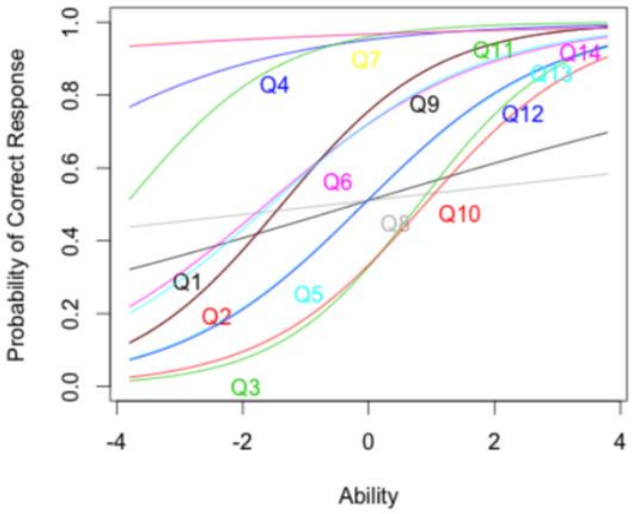


Figure 4 Measure the ability for correct response for questions [1]

## IV. Discussion

System basing on E-assessments have shifted the educational paradigm in the recent decades. Information Technology (IT) cannot be applied in full, because a lot of attention is not focused on the interplay between man and IT, and also the interactions leading to the achievement of the goals of any organization.

Hence technicians must create users-oriented systems while meeting the needs of population including their prime work on structuring efficient systems. Such efficient systems should be used after their successful creation. Malicious TTF technologies will result in exploitation or shutdown of systems. According to [49], to have positive effect on performance, technology should not only be used but also fit the needs of end user.

It is significant to consider that merely 2.4% articles discussed the accuracy of e-assessment tools [18]. So, there is less work in the articles which investigates user evaluations about the suitability or the adaptation of the various e-assessment tools provided by different options. Hence, the impression of IT on task under consideration cannot be reasonably described in details without considering features related to the stage of difficulty or fitness for work that is to be executed.

After investigating a number of research papers on the topic, 3 out of all stated slightly higher than 7% e-diagnosis technology articles a theory. The significance of establishing and showing data by using only relevant theoretical framework cannot be highlighted more than usual.

There is no combined paper on performance using technology or content and/or both presented for performance.

## V. CONCULSION

The main purpose of this study was to show the importance of IRT over CTT. This paper assessed the comparison of each item of test for difficulty level, discrimination and time indexing. Probability of answering a question with respect to an individual's ability and each item's characteristics was measured using Item Response Theory.

IRT predicted correct answer of each question of a particular difficulty. The examiner noted time to answer a question by reusing same difficulty level test. Study estimated the time required for each question by conducting a test on average student to learn their discrimination level, difficulty level and time index.

Considering the same test for future, one can determine the level of difficulty required for each question and the average duration for conducting a test according to an average student. This technique helps test makers to distribute time span for each question to be asked in a test. IRT models makes tests more efficient and reliable. The students can be benefited more by attending same nature of test in future.

## REFERENCES


[1] George, A. O. & Bola, O.O., Adegbija M.V., The Investigating of Students' Views on the Electron- Assessment: Detail Study of University of Ilorin, Nigeria. Global Media Journal: Pakistan Edition; Spring2013.

[2] Ajadi, T.O. Adewale, O.A., & Inegbedion, J.O. The Perception of Learners on the E-Examination in the Open and the Distance Learning centers: Detail Study of National Open University of Nigeria. US-China Education Review.

[3] Ang'ondi, E.K. (2013). The Teachers Attitudes and the Perceptions on the utilizing of ICT of the Teaching and also Learning as seen by ICT Champions. The X World Conference on Computers in Education July 2-5; Toruń, Poland.

[4] Cullen R., Aojula, H., Barber1 J., & Andrews J The Computer-based, on-line summative of the assessment in the undergraduate pharmacy teaching center: The Manchester experience. Pharmacy Education.

[5] Ayo, C. K. Akinyemi, I .O. Adebiyi, A. A. prospects of electronic-examination of the implementation in Nigeria'. Turkish Online Journal of Distance Education. 8 (4): 125- 134.

[6] Barrett, H. C. The Researching Electronic Portfolios and Learner Engagement. JAAL, (2007). 50(6), 436-449.

[7] Gilch, H., Kleimann, B., Bauer, Y., Degenhardt, L., & Wannemacher, K. The Online Exams as Part of the Information Technology-supported Examination Process Chain. 2008.

[8] Becta. . Becta's Review: The Electronic assessment and the e-portfolios. Coventry: Becta ICT Research.

[9] Chia Hua, Sim and Bee Theng, Lau. The Exploring the extent of" ICT" adoption among Secondary school teachers In the Malaysia. International Journal of Computing and ICT Research..http://www.ijcir.org/volume2-number2/article 3.pdf(2008).

[10] Buabeng-Andoh, C., (2012). An Exploration of Teachers' Skills, Perceptions and Practices of ICT in Teaching and Learning in the Ghanaian Second- Cycle Schools. Contemporary Educational Technology, 3(1), 36-49.

[11] Christian, Beverly J. (2014) "Using Assessment Tasks to Develop a Greater Sense of Values Literacy in Pre-Service Teachers". Education Papers and Journal Articles. Paper 46. http://research.avondale.edu.au/edu_papers/46.

[12] Chung, R.R. (2008). Beyond Assessment: Performance assessment in teacher education. Teacher Education Quarterly. Winter 2008.

[13] Cole, M.S (2013). Students' Performance and Preference on Computer-based Tests VS Paper- based Test: A case study. Msc Thesis/Dissertation State University of New York at Fredona.

[14] Conole, G., & Warburton, B. (2005). A Review of Computer Assisted Assessment. ALT-J, Research in Learning Technology, 13(1),17-31.

[15] Coulby C., Hennessey S., Davies N., and Fuller R. (2011). The use of mobile technology for work- based assessment: the student experience British Journal of Educational Technology Vol. 42, No. 2, 251–265.

[16] Doukas N. And Andreatos A. (2007). Advancing electronic assessment. International Journal of computers, communications and control. Vol II, no.1. pp. 56-65.

[17] Ertmer P. A. (2010). Teacher Technology Change: How Knowledge, Confidence, Beliefs, and Culture Intersect. Journal of Research on Technology in Education, Vol.42 No.3.

[18] Eva-Heinrich, J.M and Granshaw B. (2012). Pathways for improving support for the electronic management and marking of assignments. Australasian Journal of Educational technology. Vol. 28 No. 2, 279-294.

[19] Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. Australasian Journal of Educational Technology, 25(4), 509-523.

[20] Georgiadou, E., Triantafillou, E and Economides A.A. (2006). Evaluation parameters for computer- adaptive testing. British Journal of Educational Technology Vol. 37 No. 2. 261 – 278 doi:10.1111/j.1467-8535.2005.00525.x.

[21] Goldhaber, D., Gabele, B. & Walch, J. (2012). Does the Model Matter? Exploring the Relationship Between Different Achievement-based Teacher Assessments. CEDR Working Paper 2012-6. University of Washington, Seattle, WA.

[22] Hermans, R., Tondeur, J. Braak, J. and Valcke, M.G. (2008). The impact of primary school teachers' educational beliefs on the classroom use of computers. Computers & Education 51, 1499– 1509.

[23] Ilomäki, L (2008). The effects of ICT on school: teachers' and students' perspectives. PhD dissertation, Department of Teacher Education, University of Turku, Finland. www.doria.fi/bitstream/handle/10024/42311/B314. pdf?sequence=3.

[24] Isom, Jamie S., (2012). "Perceptions of Nebraska Teachers Regarding the Transition from STARS to NeSA and its Perceived Influence on the Implementation of a Balanced Assessment System" Educational Administration: PhD dissertations, and Student Research. Paper 85. http://digitalcommons.unl.edu/cehsedaddiss/85.

[25] Jackson, T. (2011). Stakeholders' perception of the utilization of electronic assessment in an HE environment. A Thesis submitted for the Degree of M.Phil by Research in the University of Hull.

[26] Jamil J., Tariq R.H., Shami P.A (2012). Computer- Based Vs Paper-Based Examinations: Percept of University Teachers. TOJET: The Turkish Online Journal of Educational Technology, Vol.11 Issue 4.

[27] Jordan S., (2011) Using interactive computer- based assessment to support beginning distance learners of science Open Learning Vol. 26, No. 2, 147–164.

[28] Karadeniz, S. (2009). Computer-based Tests: the impact of test design and problem of equivalency, Computers in Human Behavior, 23, 32–51.

[29] Kitchenham, B. (2004). Procedures for performing systematic reviews, Technical report. TR/SE0401, Leele University. http://www.scientificcommons.org/42461878.

[30] Lingard M. (2005). Introducing Computer-assisted Assessment: considerations for the new practitioner. Investigations in university teaching and learning Vol. 2 (2) ISSN 1740-5106.

[31] K. Elissa, "Title of paper if known," unpublished.

[32] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[33] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[34] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.