International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Computer Based Adaptive Testing

S. Viswanandhne, G. S. NandaKumar
Kumaraguru College of Technology, TamilNadu, Coimbatore, India

ABSTRACT

Assessment is a fundamental part of the learning system. Now-a-days Computer based test is normal. Adaptive testing is a type of Computer based testing that depends on the student's ability. This adaptive testing is possible with two distinct theories. Those theories are Classical Test Theory (CTT) and Item Response Theory (IRT). CTT is moderately straightforward. This theory has simple theoretical model, which make CTT to apply in numerous testing circumstances. However, IRT has got exponential development in these late decades. IRT essentially concentrates on the item level information rather than the CTT's essential concentrate on test level information. The IRT framework incorporates a gathering of models, and the applicability of every model in a specific circumstance which relies on the test items and the viability of different theoretical assumptions about the test items. For test items that are dichotomously scored, there are three IRT models, known as three-parameter, two-parameter, and one-parameter IRT models. Finally Scoring is done, based on the parameters of the item. Students who react effectively for the most difficult question gets high score when contrasted with students who react accurately for the simplest question.

Keywords : Assessment, Adaptive Test, CTT , IRT , Parameters

I. INTRODUCTION

Assessment is the summons of documenting one‘s knowledge, science and their ability .Tests are conducted and information is obtained about the ability of the test taker.

Assessment are generally divided into 1. Initial 2. Formative 3.Summative

II. METHODS AND MATERIAL

A. Initial Assessment

All learners should undergo a period of initial assessment, the purpose of which is to identify their learning and support needs. The designation of a learner‘s acquisition and supporting needs is critical as it represents the first stage in the erudition wheel and on which all other stages depend. The learner lies at the center of the learning process and needs to be fully engaged in the initial assessment process and the exploitation of their learning plan. It is important that they feel that is done with them and not to them. It involves the assemblage of a wide range of information to form a coherent movie of the individual. This information should be used to place them on an appropriate pre-vocational or vocational learning program which couple their skills, knowledge and abilities, in an appropriate occupational area and to draw up a learning plan which addresses their individual needs.

B. Formative Assessment

The general destination of formative assessment is to collect detailed data that can be used to improve student acquisition while it‘s happening. What makes an assessment ―formative‖ is not the design of a test, or self-evaluation, but the way it is used.

For example, a formative-assessment technique could be as simple as an instructor asking students to raise their hands if they feel they have understood a newly introduced concept or it could be as sophisticated as

having students complete a self-assessment of their own writing that the teacher then reviews and comments on.

While formative assessment help instructor identify learning needs and problems, in many cases the assessments also help students develop a stronger reason of their own academic strength and weakness. When students know what they do well and what they need to work harder on, it can help them take greater responsibility over their own learning and academic progress.

C. Summative Assessment

Summative assessments are used to evaluate student, skill acquisition and academic achievement at the end of a project, whole course, semester, political program, or schooling year.

III. LITERATURE SURVEY

A. METHODS OF ASSESSMENT

PAPER & PENCIL METHOD

Paper and pencil exam is available for the traditional classroom situations. In this method, Testee read the question and responds in writing. Evaluators are to be appointed and correction process takes place separately. Hence Scoring and feedback are not immediate to the testee. The most of the college admission test such as SAT and ACT still rely on paper and pencil method because of the need to test very large number (in the millions) of testee. Here printing the test booklets and sending them to the exam center field involves a major expense. And also if any error is found then the booklets are to be reprinted and resend to all those exam centers. This becomes a great problem in this method.

Graduate Record Examinations (GRE) [1] uses Paper & Pen version for graduate school admissions purpose. This Test is administered five times a year at regular domestic test centers.

COMPUTER BASED TEST

Now-a-days Computer based test is common. CBT uses a computer to give the same mental test like a paper-and-pencil format. The Questions are same and are presented in the same order as the paper-and-pencil version of the test [4].

Large pools of questions are needed and testee gets unique randomized questions. Testee is asked to attend the test for the fixed duration. Thus the security system of the test is improved over paper and pencil test. Scores are generated at the end of the test and thus there is no need no any evaluator. Moreover scoring and feedback are instant to the testee.

With GRE, CAT each testee can get their preliminary view of score after their last question. As time is the important factor, all testee should be given with the same measure of time. Computers are good at keeping time. When we reply on human examiner who watches their clock, some testee may be given with more time[7]. ADAPTIVE TESTING Adaptive testing is a form of computer based testing that depends on the student‘s ability. It is a form of test where the student‘s next question is selected based on the performance of the response to the most recent question that is asked. It is also called tailored testing.

Examinee is provided with easy level of question at the start of the test. When he/she answers it correctly then next question will be asked from the next level i.e., medium level. Now when this answer is also correct it further moves to the next hard level. Else it steps back to the easy level. Questions are selected and given until the test score value reaches certain constant value or for some test, test ends after specified number of questions [4].This is demonstrated in the Fig. 1.

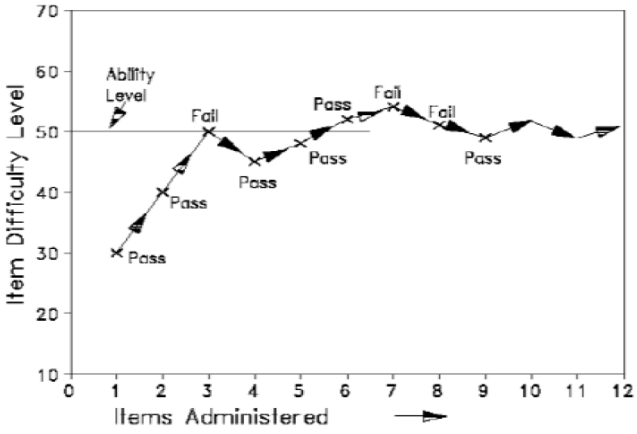

Figure 1. Adaptive Testing

The move to computerization with the GRE General Test has occurred in stages [1]. The primary stage has

included the organization of secure direct computer based GRE forms (i.e., paper-and-pencil forms managed by means of computer), to assess examinee acceptance also, comfort with the computer delivery system. This first stage is seen as temporary, with a move to the second stage, administration of adaptive tests that are comparable in content to the paper-and-pencil types of the test. During this stage, paper-and-pencil and adaptive testing will exist at the same time, with scores from both sorts of examinations used in the graduate school admissions-process. In the third and extreme stage, paper-and-pencil testing will be suspended and all testing will be done on the computer by means of the adaptive process.

CLASSICAL TEST THEORY-CTT

CTT is well suited for traditional testing environment. In this theory all the students who attend the test are administrated with the same or analog set of questions. Example of this type is conducting the test for college entrance test or job recruitment.

Classical test theory accept that each observed score (X) is a combination of a true score (T) on the concept of interest and random error (E). Classical test theory, otherwise called true-score theory, assumes that every individual has a true score, T,that would be obtained if there were no errors in estimation.

It is assumed that observed score (X) = True score (T) + some error (E).

True scores evaluate values on an attribute of interest, characterized here as the ability of the student (the "thing" expected to be measured). As estimations of the true score increases, responses to items about a similar idea ought to likewise increase (i.e., there ought to be a monotonically increasing relationship between true scores and item scores), assuming that item responses are coded so that higher responses reflect a greater amount of the idea[8].

It is also assumed that random found in observed scores are normally distributed .Therefore, the expected value of such random fluctuations is considered to be zero.

To estimate the student ‘s ability, their ability should match with the ―difficulty‖ of the question. When this matches, ability can be estimated. So in order to estimate student ‘s ability, we need large number questions to be administrated so that question‘s difficulty slowly matched the ability of the student[8].

ITEM RESPONSE THEORY –IRT

IRT is based on the idea that probability of correct response to the question is the mathematical function of student ‘ s ability and question parameters i.e., difficulty and discrimination of the question. It is also referred to as ―latent trait theory‖ .
In general it is referred superior to the Classical test theory. Examples of this type of test are Graduate Record Examination (GRE) and Graduate Management Admission Test (GMAT)[6].

IRT is more powerful than CTT [2].In UK, IRT has gained wide acceptance and Educational Research has published ― The Objective Interpretation of Test Performance‖ and it deals with Rasch model.

Every learner gets an alternate set of questions with different difficulty levels while taking the adaptive assessment (Weiss, 1983) [3]. IRT-based item selection strategies (Weiss, 1983) are maximum information item selection strategy, and Bayesian item selection strategy.

Kingbury and Weiss used maximum information item selection strategy [3], in which the item pool is searched for an item that can give greatest data about the examinee. Bayesian item selection strategy uses earlier data about the examinee than maximum information item selection strategy (Weiss, 1983).

IRT OVER CTT

Availability of more advanced computational tools Questions and student‘s ability are formulated by distinctive parameter

Test Score Equating

In IRT student ability never change across different set of questions. In order to compare two students across different set of questions, the test score can be directly compared .In CTT, one test score should be converted into metric of other test.

Handling missing value

IRT is straight forward to analyze question that have various methods like likelihood. In CTT, those missing values are to be imputed.

ITEM RESPONSE THEORY

A. ITEM CHARACTERISTIC CURVES (ICC)

The ICC is the essential building piece of item response theory; the various constructs of the theory rely on this curve[4]. There are two specialized properties of an ICC that are utilized to depict it.

i. Difficulty of the item Under IRT, the difficulty of a item depicts where the item functions along the ability scale. For instance, a easy item functions among the low-ability test takers and a hard item functions among the high-ability test takers; subsequently, difficulty is a location index.

ii. Discrimination

This depicts how well a item can separate between examinees having abilities underneath the item location and those having abilities over the item location. This property basically mirrors the steepness of the ICC in its center segment.

The more extreme the bend, the better the item can. The flatter the bend, the less the item can discriminate since probability of correct response at low ability levels is nearly the same as it is at high ability levels. Utilizing these two descriptors, one can depict the general type of the ICC.

The probability of correct response is almost zero at the lowest levels of ability. It increments until at the highest levels of ability, the probability of correct response approaches 1. This S-formed bend depicts the relationship between the probability of correct response to an item and the ability scale. In IRT, it is known as the item characteristic curve (ICC). This is additionally called Item Response Function (IRF).

TABLE I. ITEM CHARACTERISTIC CURVE MODELS

Model	Name	Parameter
Single Parameter	Rasch Model	Item difficulty 'b'
Two Parameter	Birnbaum Model	Item difficulty 'b'
		Item discrimination 'a'
Three Parameter	Fred Lord Model	Item difficulty 'b'
		Item discrimination 'a'
		Item guessing 'c'

RASCH'S SINGLE PARAMETER MODEL

$$P\,i\,(\Theta) = e^{\,a(\Theta - b)} / 1 + e^{\,a(\Theta - b)}$$

Pi (θ) = Probability of getting correct answer to item I with ability θ
e=2.718, base of a natural log
a=scale constant determining the units of θ (For Rasch it is 1)
b=location parameter related to the difficulty of the item i (also referred to as item threshold)

The modified equation for Rasch model after rationalizing the numerator and the denominator by multiplying with e-a (θ-b)

$$P\,i\,(\Theta) = 1/1 + e^{\,-a\,(\theta - b)}$$

C. BIRNBAUM'S TWO PARAMETER MODEL

In the Two Parameter show, at the purpose of contra flexure (inflexion) corresponding to 0.5 probability value, a typical tangent is drawn and the slope of the tangent is assigned as "item discrimination ‖ and this value is shown by the letter "a". The "b" and "a" values are assessed for the given items.
The equation to the curve is:

$$P_i\,(\theta) = e^{\,a\,(\Theta - b)} / 1 + e^{\,a\,(\Theta - b)}$$

Where
Pi (θ) =Probability of getting correct answer to item i with ability θ
θ = Person abilitb = Item difficulty

a = Item Discrimination

The modified equation for Birnbaum model after rationalizing the numerator and the denominator by multiplying with e-a (θ-b)

$$P_i(\Theta) = 1/(1+e^{-a(\Theta-b)})$$

D. FRED LORD'S THREE PARAMETER MODEL

In the Three Parameter model, there is in addition, a third parameter called guessing for the item and is assigned by the letter "c". The guessing parameter is unique to the item and is autonomous of test taker ability. In this way the guessing parameter remains a consistent for all test takers of different abilities. The equation for the curve is:

$$\bar{P}_i(\theta) = c + (1-c)\,[e^{a(\Theta-b)} / 1+ e^{a(\Theta-b)}]$$

Pi (θ) = Probability of getting correct answer to item i with ability θ
θ = Person ability
b = Item difficulty
a = Item Discrimination
c = Guessing Parameter
The modified equation for Birnbaum model after rationalizing the numerator and the denominator by multiplying with e-a (θ-b)

$$P_i(\Theta) = c + (1-c)\,[1/(1+e^{-a(\Theta-b)})]$$

ITEM INFORMATION FUNCTION

Any item in a test gives some data about the ability of the examinee, however the measure of this data relies on upon how nearly the difficulty of the item matches the ability of the individual[4].

ITEM INFORMATION FUNCTION OF SINGLE PARAMETER MODEL

On account of the Single Parameter model, the item information function relies on how nearly the difficulty of the item matches the ability of the individual, while in other models it consolidates with different variables. The item information function of this model is demonstrated as follows:

$$Ii(\theta, bi) = Pi(\theta, bi) * Qi(\theta, bi)$$

It is anything but difficult to see that the greatest estimation of the item information function is 0.25. It happens at the point where the probabilities of a right and of an off base reaction are both equivalent to 0.5. Any item in this model is most informative for examinees whose ability is equivalent to the difficulty of the item. As ability turns out to be either littler or more prominent than the item difficulty, the item information diminishes. This is appeared in Figure 2 [4] below:

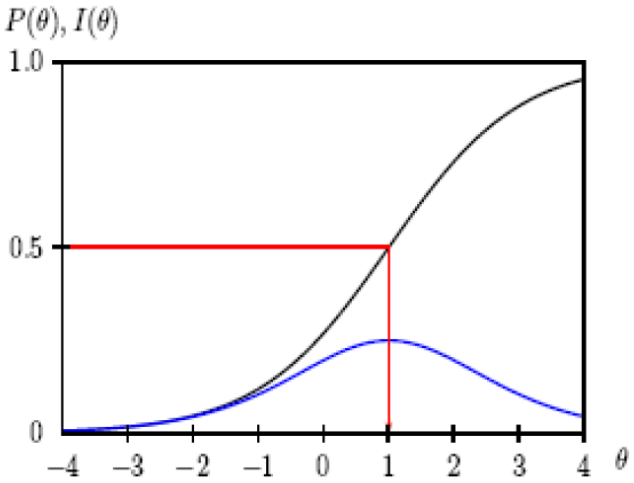

Figure 2. Item Information Function of single parameter model

ITEM INFORMATION FUNCTION OF TWO PARAMETER MODEL

The item information function for the Two Parameter model is as demonstrated as below:

$$Ii(\theta, bi, ai) = a2\ Pi(\theta, bi) * Qi(\theta, bi)$$

The discrimination parameter ai is the second parameter that has a significant impact since it shows up in the equation as a square[4]. This implies discrimination parameters below 1 can diminish the information function rather significantly, while discrimination parameters above one will build it generously.

ITEM INFORMATION FUNCTION OF THREE PARAMETER MODEL

The item information function of the Three Parameter model is more complicated as contrasted with the

Single Parameter or Two Parameter model. The item information functions of the three models are demonstrated as follows:

$$I(\theta, a, b, c) = a2\, Q(\theta)\, [P(\theta - c)2] / P(\theta)(1-c)$$

IV.CONCLUSION

The present article presents an overview of classical test theory and Item response theory. Classical test theory and item response theory give valuable techniques for content validity. Item response theory requires several items so that there is satisfactory chance to have a sufficient range for levels of item difficulty and person attribute. Single-item measures, or too few items, are not appropriate for IRT analysis. Depending on the particular type of measure and particular condition, either approach or both approaches may be useful to help maximize the content validity.

V. REFERENCES

[1] Daniel R. Eignor, Case studies in computer adaptive test design through simulation, in Educational Testing service Princeton, New Jersy, November 1993

[2] Brent Bridgeman, Experience from large-scale computer based testing in the U.S.A, Educational Testing Service, USA

[3] MansoorAl-A'ali, Implementation of an Improved Adaptive Testing Theory . Educational Technology & Society, University of Bahrain, 10 (4), 80-94

[4] Dr. V. Natarajan, Basic Principles of IRT And Application to Practical Testing & Assessment, in IACAT conferences at Amsterdam, Asilomar (US),2008 5Joseph C. Cappelleri, Overview of Classical Test Theory and Item Response Theory for Quantitative Assessment of Items in Developing Patient-Reported Outcome Measures, 5 May 2014.

[5] Mariana Lilley and Trevor Barker, An Evaluation of a Computer Adaptive test in a UK University context, Loughborough University,2003

[6] Richard M.Luecht and Stephen G.Sireci, A Review of Models for Computer – Based Testing, Research report of College board,2011-12

[7] Joseph C. Cappelleri, J. Jason Lundy and Ron D. Hays, Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures, in Elsevier HS Journals, 2014