

Computerized Adaptive Test based on Item Response Theory in E-Learning System

Yeni Kustiyahningsih

Informatics Management Department, Universitas
Trunojoyo Madura, Indonesia

Andharini Dwi Cahyani

Informatics Engineering Department, Universitas
Trunojoyo Madura, Indonesia

ABSTRACT

Computerized Adaptive Test (CAT) is a computer-based test framework which has ability to customize questions items given to the learner based on their estimated ability. In this research, the CAT system is build using Item Response Theory (IRT) techniques to develop an adaptive system based on question item's difficulty level and students' ability level. Moreover, to figure out the effectiveness of this CAT system, we do some experiments by comparing the average post-test score of students in CAT system and conventional system. The experiments result reveals that the average post-test score of students in the CAT system is much higher than the average post-test score of students in traditional test system.

General Term : e-learning, Adaptive Test

Keywords

Computerized Adaptive Test, Item Response Theory, students ability level, maximum likelihood estimation

1. INTRODUCTION

1.1 Computer Adaptive Test

Nowadays, many universities, corporations, and educational organization develop and deliver online course materials for distance learning programs (Georgieva, Todorov, & Smrikarov, 2003). E-learning, also known as distance learning or web-based learning, is self-learning using electronic appliances through the Intranet or Internet (Barker, 2002). The reason for the rapid growth of e-learning is its convenience and efficiency so that learning process can take place at any pace, anywhere and anytime (Kabassi & Virvou, 2004). The other perspectives of using e-learning is to solve the limitations of conventional learning, such as flexible schedule, higher interactivity, and better quality and a variety of lecture materials (Jayasimman & George, 2013).

In order to help the learner learn more efficiently, then the web-based learning system need to personalized (Xu & Wang, 2006; Brusilovsky, 1999; Andharini, 2012). Some of the personalization systems consider the learner's experiences, preferences, goals and existing knowledge (Huang, Huang, & Chen, 2007). The term personalized e-learning has become popular in learning context over this decade. By using the personalization in e-learning system, then the learner can be uniquely identified and their abilities can be individually monitored and assessed (Alavi & Leidner, 2001). One of the personalization system in e-learning is computer testing system that encourage the learner to improve their abilities (Davidovic, Warren, & Trichina, 2003).

To date, the educational assessment models applied in Indonesia are still employing the same test questions for all learners. These practices are based on the assumption that learners with the same age or level of education have the same capabilities. Meanwhile, the learners' ability is different even

though they are in the same age or class. In the same class there are not only intelligent and bright learners, but also unintelligent and slow learners. The adaptive test scoring model can be used to overcome these disadvantages. Moreover, in this model, the test questions given to each learner is different and based on learner ability.

In the beginning, the CAT system calibrates the difficulty level of each questions test using Item Response Theory in non-adaptive manner (Van der Linden & Hambleton, 1997; Wiliam, 2011). Then, the system tries to recognize learners' ability by giving test question with the middle difficulty level. If learners answer is correct, then the CAT system will give learners a question with a higher difficulty level. Otherwise, the CAT system will give learners a question with lower difficulty level. By using this system, therefore, can make intelligent learners feel more challenged and in slower learners will be encouraged at once.

There are various IRT models, with different complexity level. The simplest IRT model is the Rasch model that reveal a learner's response to a question item depends on the level of learner's ability and question item's difficulty (Van der Linden & Hambleton, 1997; Kim, 2006). More complex IRT models include additional parameters, such as an item discrimination parameter, a pseudo-guessing parameter or the effects of person or item characteristics (Ozaki & Toyoda, 2006; Kim, 2006; Wauters, Desmet, & Van den Noortgate, 2010).

1.2 ITEM RESPONSE THEORY (IRT)

Item response theory (IRT) is a psychometric model which is based on the idea that the probability of a correct response to an item is a mathematical function of person and item parameters (Kim, 2006). IRT provides a basis for estimating the parameters, determine how well the data conformed to the model and investigate how to measure the psychometric properties. Figure 1 shows the diagram of the steps in IRT algorithm.

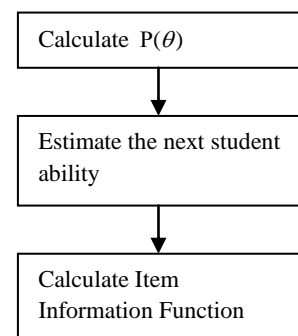


Figure 1. Steps in IRT algorithm

Here is the explanation of each step.

1. Calculate $P(\theta)$

In item response theory, there is an IRT main curve (Baker, 2001), that represents the characteristics of a problem that suggests the possibility of a learner with certain ability (θ) can answer the question correctly. In addition, this curve namely Item Response Function (IRF), is denoted by $P(\theta)$.

Furthermore, there are 3 item parameters of IRF curve. The item parameters simply determine the shape of the IRF and in some cases have a direct interpretation. Those 3 parameters are namely a (discriminant factor), b (questions' difficulty level) and c (pseudo-guessing parameter). In our study, we employ 2 parameter logistics (2PL). Equation (1) shows the formula of 2PL model:

$$P(\theta) = \frac{1}{1 + e^{-L}} = \frac{1}{1 + e^{-a(\theta - b)}} \quad (1)$$

where:

$P(\theta)$ = Probability that a learner answered a question correctly to parameter b

θ = Estimated ability of learners

a = Discriminant factor of each question

b = Difficulty level of each question

e = Exponential value (2,718)

2. Estimate the next learner's ability

To estimate the next learner's ability, we apply maximum likelihood estimation (MLE) theory. This process begins with estimated ability of learners and some parameters. In this process, we predict the learner's probability to answer correctly the next question item. This Maximum Likelihood Estimation (MLE) theory is quite efficient and able to distribute the error normally (Wang, 2006). Furthermore, to estimate the learner's ability is an iterative process that starts with an initial value for the learner's ability. In this research, we set the initial value = 0.5.

The learners' ability estimation formula is a modification of the Newton-Raphson iterative model. Equation (2) presents the formula of maximum likelihood estimation.

$$\theta_s + 1 = \theta_s + \frac{\sum_{i=1}^n -a_i [u_i - P(\theta_s)]}{\sum_{i=1}^n a_i^2 P_i(\theta_s) Q_i(\theta_s)} \quad (2)$$

where:

θ_s = Learner's ability estimation

a_i = Discriminant parameter of items i , with $i=1,2,\dots,N$

u_i = corresponding answers of items i , score=1 for a correct answer, otherwise score = 0

$P_i(\theta_s)$ = possible correct answer to the question i

$Q_i(\theta_s) = 1 - P_i(\theta_s)$

$Q_i(\theta_s)$ represent the possibility of learner can answer correctly the question i .

3. Calculate item information function (IIF) for each question

After we estimate the next learners' ability, the next steps is to calculate IIF for each item in question bank. Then, the system will select item question with the greatest IIF value. The item question with the greatest value IIF then will be given as the next test question.

To calculate IIF, we can use the equation (3):

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta) \quad (3)$$

After performing the IRT algorithm, then we need to set the criteria to stop the test. Usually, the stopping criteria are as follows.

a. Fixed length: after a certain number of questions, then the test will be terminated.

b. Time limit: after reaching a certain time limit, then the test will be terminated.

In this research, we use fixed length to terminate the test.

2. THE PROPOSED COMPUTERIZED ADAPTIVE TEST (CAT)

This section describes system requirements analysis, components and mechanism of CAT application.

2.1 System Requirement Analysis

This computerized adaptive system is developed and integrated within e-learning application. Therefore, learners are supposed to learn the material before they perform the test. Figure 2 below illustrate the responsibilities of each actor (teacher and student) towards e-learning application.

In this system there are three actors are admin / teachers, students / members and the public. Each actor has the right of access to the system, while the Admin permissions as a teacher is managing the material, question bank Manage, Manage syllabus, Manage term, Manage guest book, a management user guide, Manage polls, Manage settings, Manage forums, polls contents, guest book, help. While the access rights of students as members are working on the adaptive test, following the forum, setting profiles. The right of equal access to the public is between members. To register, searching materials, guest book, polls contents, help, view course materials, view help, and view the syllabus.

2.2 Flowchart and Mechanism of CAT application

Flowchart is a chart that shows the work flow or what is being done on the system as a whole and explain the sequence of procedures that exist in the system.

The following are description of each process in figure 3:

1. The first step begins with the input problem into the database, there are 180 items, where the problem is a combination of chapter 1 to chapter 9.

2. Further initialization capabilities provide students assumed 0.5.

3. Then the system displays the first question taken at random with a difficulty level of questions is assumed starting 5/10 or 0.5. Where in can of 5 answers correct in comparing the amount of matter that was done, which is 10.

4. Then the system will check the response of the students' answers. After getting the response from the students, the system will calculate the value of IRT, followed calculate MLE and the IIF last count.

5. If the student answers incorrectly then the next question will be displayed <0.5 or about the difficulty level decreases, if the student answers correctly then the next question will be given > 0.5 or difficulty level will rise from the previous question.

6. Then the question is displayed according to the response of the students' answers.

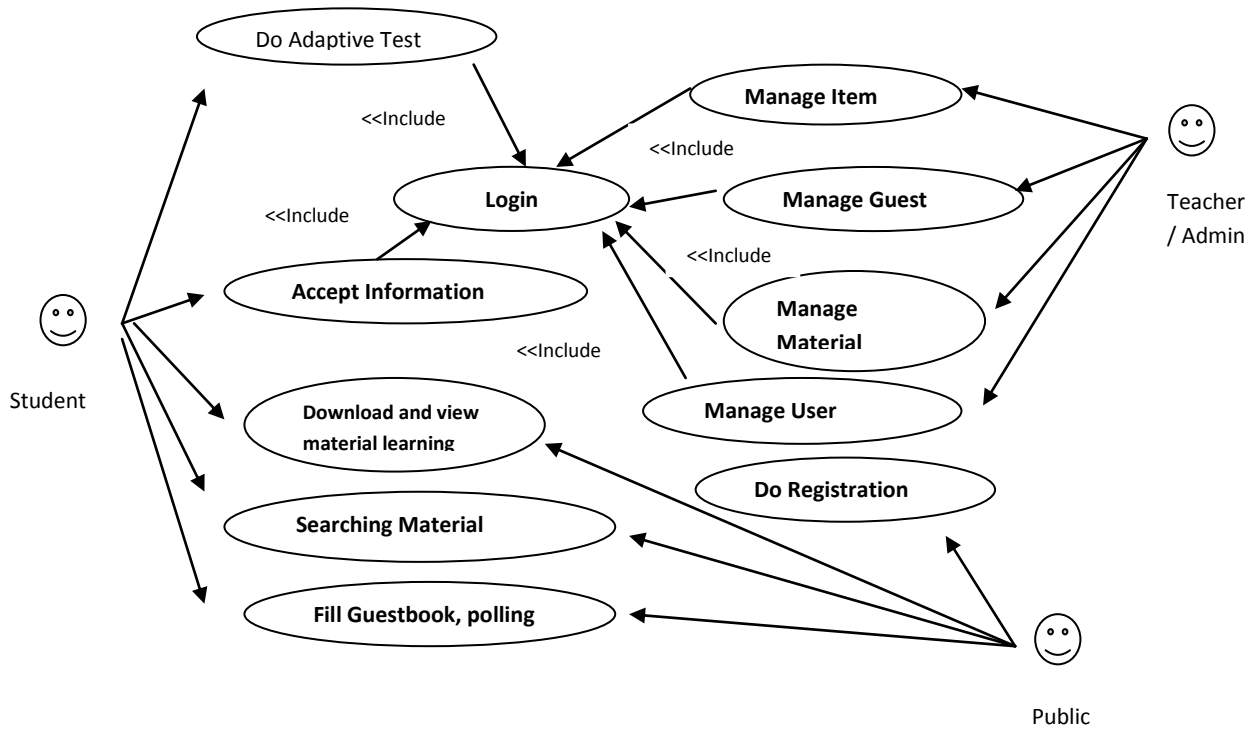


Figure 2. Use Case Diagram of the CAT system

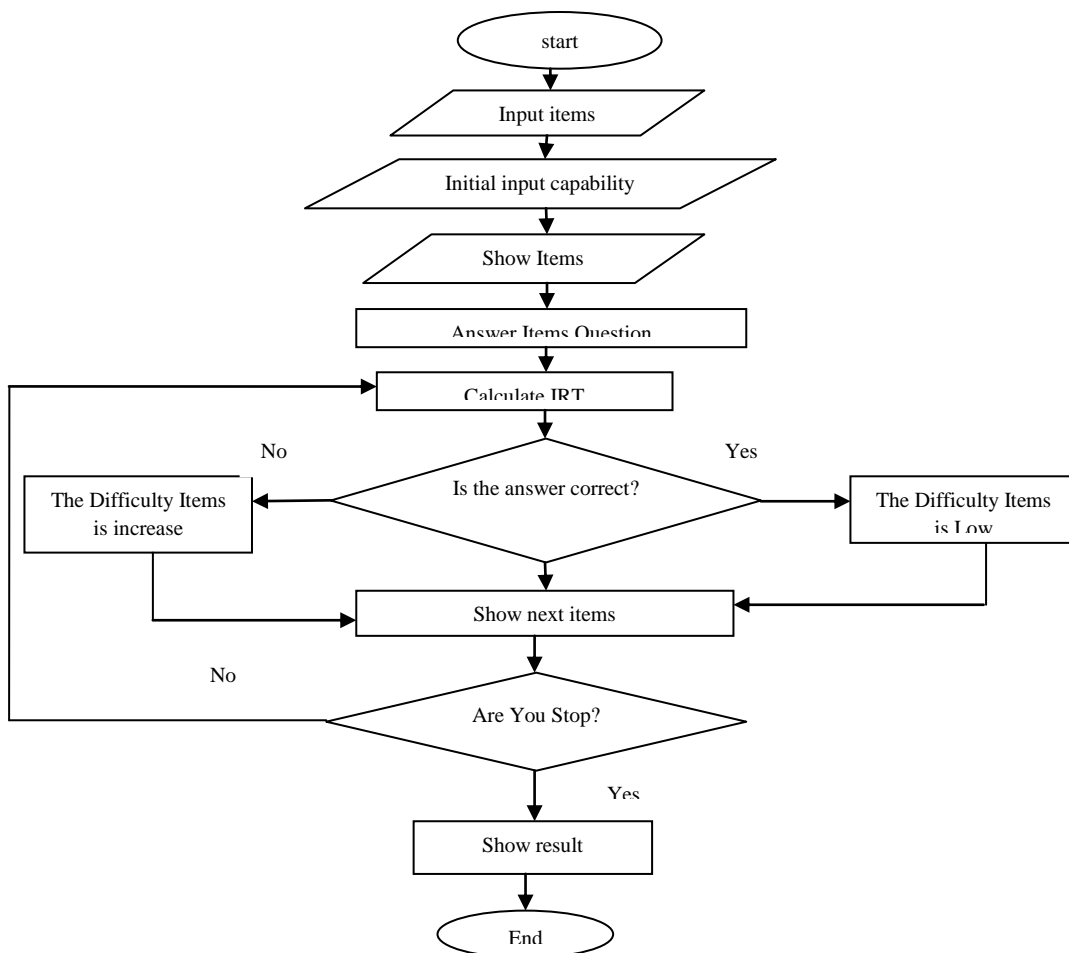


Figure 3. CAT flowchart system

The matter in which a given problem as much as 10 questions, if you do not meet the specified number of questions it will display the question and get back to number 4 to meet the problem stops. if it meets the number of questions it will display the results of the adaptive test work that has been done.

In general, the principle of CAT begins with initialization of the student's ability to assume the student has the ability to provide about the medium and medium difficulty levels anyway. The selection of the next question based on the answer to the question of examinees who was granted, if the answer is correct then the question will have a higher level of difficulty, but if the answer is wrong then the next question has a lower level of difficulty. There are three main steps in the CAT , the first, The whole matter of the question bank which has not been provided will be evaluated to select the best questions will be issued based on the estimated current level of ability. This process is also known as analysis item. The second, the best question was issued and students will answer that question. The third, the ability level newly calculated based on the answers of all questions given Steps 1 through 3 is repeated continuously until it reaches a limit with certain criteria. Suppose the problem has reached a certain amount, the ability of learners can be determined, has encompassed some particular topic, or based on certain indicators.

3. APPLICATION OF COMPUTERIZED ADAPTIVE TEST (CAT)

System CAT / Computer adaptive test is implemented on the server localhost, using the programming language HTML, PHP, CSS and Javascript. Software is displayed in a web page. The menu is available in this e-learning are as follows.

3.1 Adaptive Test Menu

Adaptive test menu is a special menu that can be accessed by students. This menu contains a collection of Indonesian as a matter of learning evaluation. Adaptive tests are provided in the form of questions with multiple choice models, many questions are worked out by the students depending on the admin (teacher). The menu is created dynamically adaptive test, meaning that at any time the admin (teachers) can make changes to the questions, changes made include, changes in matter. The workings of this test is adaptive menus student working on five questions, then students check the answers by pressing the answer button, the system will automatically check and display the next question based on the answers of the student if the student answers correctly then the next question that will appear about the difficulty level will higher, if the student answers incorrectly then the next question that will appear about the difficulty level lower. Page results of adaptive test students containing student grades, and a difficulty about the value and ability of the students (in the form of figures and graphs). The system also provides feedback in the form of comments or feedback to students in the form of a compliment if students have performed well exercises and warning if the student is still not good in doing exercise. On the menu adaptive test allows students perform exercises back next time, so that the system provides the results of students' value development exercise.

3.2 Adaptive Test Form

In this form students work on adaptive items that have been provided. The Items created by multiple-choice models. After replying to the question about the level of difficulty that will be displayed in accordance with the truth or falsity of the students answered that question. If you want repeat adaptive test with different questions students return to the adaptive test. Can be seen in figure 4.

Hallo Feni

e-Adaptive

2J. Mengantuk, gangguan pencernaan, mulut kering, retensi urine dalam teks petunjuk penggunaan obat kalimat di atas di sebut dengan

A. Komposisi

B. Cara Kerja obat

C. Indikasi

D. Efek Samping

Rese

Jawab

Figure 4. Adaptive Test Form

3.3 Graph Form

After work on items, the students can see difficulty level of questions and ability of students and also score from the adaptive answering questions. This can be seen in Figure 5. and Figure 6.

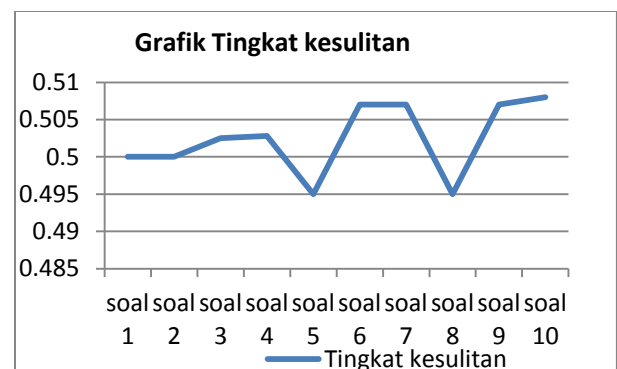


Figure 5. Graph Form Difficulty items



Figure 6. Graph Form ability student

3.4 Scores Menu

The scores menu is menu for students that contains historical scores adaptive test that has been done. scores displayed is based on an existing session id. Each scores of each session id adaptive display graphs that have followed, the graph shows level of difficulty about value and ability of the students. When results of adaptive pressed, the system will display value of adaptive and graphs on the session id. Can be seen in figure 6.

3.4.1 Scores Form each id Session

scores form students presented in the form of graphs adaptive development based on the existing session id. This page also lists historical results displayed Adaptive ever done.

e-Adaptive Test SMP Negeri 1 Sumenep







Detail Nilai adaptive					
NO	IDSESSION	TANGGAL	NILAI	OPERASI	
1	3167672473	2013-06-04	70		
2	4165820989	2013-06-03	100		
3	2472198751	2013-06-03	60		
4	4423153936	2013-06-03	60		
5	1519529189	2013-06-02	50		

Figure 7. Scores form each id Session

4. EXPERIMENTAL RESULT

4.1 Experimental Design

4.1.1 Data is used

The data used in this research is data class seven at SMPN1 Sumenep. The amount of items that is used as much as 180 items. This test is divided into 2 groups, each group consisting of 88 students. The first group uses adaptive test, while the second group did not use the adaptive test or just conventional tests. The first scenario each group pretest conducted in order to determine students' initial ability. The average pretest score for group 1 was 72.46 while the group 2 was 72.67. Next, the first group do adaptive testing while the second group do non-adaptive testing. Adaptive test was conducted to determine the ability and the difficulty level of each participant, as well as the average value of which is derived. Where as non-adaptive or conventional tests done to determine the average student value, the results can be seen in Table 2. After each group performs adaptive and non-adaptive test, then two groups at post-test done in order to determine how much of an increase scores of pre-test to post-test of both groups.

4.2 Experimental Analysis

In the adaptive test trials following examples are take from the results of one students. Where it is clear that the level of difficulty of questions given to students in accordance with the response of the students' answers indicated in the table. 1 named adinda on-1 skills test to early 0.5 and the first question is given by the level of difficulty of 0.5 and then when answering the question the answer is 0 it indicates that the answer is wrong, so the ability of the students fell into it for about no.2 -0.833333 given the level difficulties under 0.5 is 0.488889, while answering the student's response is correct then the ability of students rose to -0.0749264, then no.3 given a 0.5 degree of difficulty, and answers students' incorrect responses then down another student's ability to -2.3208. Until such no.10 continued ability of the students will go up or down depending on the response of students and answer questions that will be given will always be in accordance with the response of the students' answers.

Table 1. Result of student 1

NO	Id item	A	b	Respon	Theta
1	153	1.5	0.5	0	0.5
2	14	1.5	0.488889	1	-0.833333
3	163	1.5	0.5	0	-0.0749264
4	50	1.5	0.464286	1	-2.3208
5	167	1.5	0.5	1	-1.64391
6	64	1.5	0.514286	0	-0.950496
7	38	1.5	0.457143	1	-3
8	117	1.5	0.521739	0	-2.3296
9	168	1.5	0.454545	0	-3
10	58	1.5	0.444444	1	-3

Further trials using adaptive and non-adaptive tests done in stages that students there are 3 session. From the results of trials that have been conducted so as to compared the average value of using adaptive tests with an average value of using non-adaptive tests, can be seen in Table 2.

Table 2. Comparison Between Adaptive Test and non-Adaptive Test

Session Test	Adaptive test average	Non adaptive test average
test 1	52,80193	55,34091
test 2	76,42512	62,84091
test 3	86,18357	72,15909

After the students through this stage then performed posttest to both groups, the results can be seen in Table 3 below.

Table 3. List of scores pretest posttest results

Session Test	Adaptive test average	Non adaptive test average
Pre Test	72.46	72.67
Post Test	78.57	76.65
the average increase in value	6.11	3.98

From the table above shows that scores average student after doing adaptive test higher than using conventional test. Where the average increase in the posttest score of 6.11, whereas the conventional tests or non-adaptive test is 3.98 so the use of the CAT method is better than conventional.

5. CONCLUSIONS

The conclusion that can be drawn from this research is that using adaptive testing or CAT (Computer Adaptive Test), the evaluation system is more accurate in measuring the ability of the user and can accommodate the diversity of user capabilities to provide learning materials for the system according to the level of proficiency tests students. While the average value when using the adaptive and non-adaptive tests in the test get the average increase between pre-test and post test at 6.11 for students who use adaptive and 3.98 for that use conventional or non-adaptive test, so from here scores average posttest students further increased by using a CAT or adaptive tests compared with conventional tests.

6. REFERENCES

- [1] Andharini, D.C., & Ari Basuki. (2012). Personalized Learning Path of a Web-based Learning System. *International Journal of Computer Applications* 53(7):17-22
- [2] Alavi, M., & Leidner, D. E. (2001). Research commentary: Technologymediated
- [3] Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. *Künstliche Intelligenz*, 13(4), 19–25.
- [4] Baker, F. (2001). The basics of item response theory. ERIC clearing house on assessment and evaluation.
- [5] Barker, K. (2002). E-learning in three step. In *School business affairs*. Available <http://www.asbointl.org>
- [6] Davidovic, A., Warren, J., & Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Transactions on Education*, 46(2), 241–251.
- [7] Georgieva, G., Todorov, G., & Smrikarov, A. (2003). A model of a Virtual University some problems during its development. In *Proceedings of the 4th international conference on Computer systems and technologies: e-Learning*. Bulgaria: ACM Press.
- [8] Hamdi, M. S. (2007). MASACAD: A multi-agent approach to information customization for the purpose of academic advising of students. *Applied Soft Computing*, 7, 746–771.
- [9] Huang, M. J., Huang, H. S., & Chen, M. Y. (2007). Constructing a personalized elearning system based on genetic algorithm and case-based reasoning approach. *Expert Systems with Applications*, 33, 551–564.
- [10] Kabassi, K., & Virvou, M. (2004). Personalised adult e-training on computer use based on multiple attribute decision making. *Interacting with Computers*, 16, 115–132.
- [11] Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of Educational Measurement*, 43(4), 355–381.
- [12] L Jayasimman & George Dharma Prakash E Raj. (2013). A Soft Computing Approach for User Preference in Web based Learning. *International Journal of Computer Applications* 61(21):25-29
- [13] Learning—A call for greater depth and breadth of research. *Information Systems Research*, 12(1), 1–10.
- [14] Ozaki, K., & Toyoda, H. (2006). Paired comparison IRT model by 3-value judgment: estimation of item parameters prior to the administration of the test. *Behaviormetrika*, 33(2), 131–147.
- [15] Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- [16] Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: possibilities and challenges. *Journal of Computer Assisted Learning*, 26(6), 549–562.
- [17] Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.
- [18] Wang, F.-H. (2006). Application of componential IRT model for diagnostic test in a standard conformant elearning system. In *Sixth international conference on advanced learning technologies (ICALT'06)*.
- [19] Xu, D., & Wang, H. (2006). Intelligent agent supported personalization for virtual learning environments. *Decision Support Systems*, 42, 825–843.