ME – 56

Mathematical Algorithm on Conventional Computerized Adaptive Testing

Iwan Suhardi

Department of Electrical Engineering, Faculty of Engineering
State University of Makassar, Makassar, Indonesia
iwansuhardi@yahoo.com

Abstract - In education, it is important to determine the ability of the students on a subject. By knowing it, a teacher can take appropriate action to deliver good education. One method to determine the ability of the student is by performing adaptive testing. Adaptive testing is a testing method in which each examinee will be given a different set of questions based on the ability of each student. Thus, each examinee do not need to answer all questions. The items were selected with specific procedures based on the estimated level of ability of the students which reflected on their responses. Adaptive testing can be automated using a computer device, called a computerized adaptive testing (CAT). The CAT based on item response theory (IRT) with one parameter logistic model, two parameters, and 3 parameters. CAT has many advantages compared to other testing applications and has been shown to have high efficiency and reliability. CAT is also very capable to be developed further, both in terms of procedure and also the application. Eventhough there has a lot of references about CAT and its development, but in reality it is not easy to build a CAT application program as a whole. In this paper, the authors will share the experience in developing a conventional CAT application in a detailed mathematical algorithms and examples of results analysis obtained by the response answers from the students. This paper is expected to provide an initial basis for other developers and CAT researchers to build, develop and further improvement of the CAT program for large-scale applications.

Keywords: *education, adaptive, testing*

I. INTRODUCTION

In principle, the test was built to meet the needs of the groups for the test with a view to estimating the level of ability of the test participants. Giving the test is too easy for the person taking the test is a waste of time. Usually cause unwanted behavior such as fault for not careful or perhaps deluded by the answer of a trick question. Instead, the questions are too difficult, also produces test scores are not informative. Participants may stop with serious tests to try to answer the question, choosing to guess, or respond to other undesirable behavior. Adjust the test to bring the level of ability of each individual participant tests, a solution should be sought. How to test participants than if each person taking the test is given a different test?

Adaptive testing is also referred to as tailored test, which is a test that adjusts the ability of participants. Hambleton said that the definition of a computerized adaptive testing "would be to give every examinee a test that is 'tailored' or adapted, to the examinee's ability level" [1]. The use of computers to be used in the test is adaptive used to be called Computerized Adaptive Testing (CAT). Known as the implementation of computerized testing really was no longer using "paper and pencil". Adaptive, because the grains have been selected based on the results because the self-regulatory analysis and adapted to the needs or abilities of the examinees, works automatically through a computer software. According Wainer, adaptive testing is a test which was held for the participants of the test with a grain because determined by the answer (response) test participants [2].

II. THEORY OF CONVENTIONAL CAT

In general, conventional CAT system has components, namely (1) item banks, and working systematic CAT comprising components (1) item selection procedure, (2) ability estimation, and (3) stopping rule [2].

A. Items Bank

CAT taking items from a question bank that is based on Item Response Theory (IRT) using models 1, 2, or 3 parameter logistic (1 PL, 2 PL, or 3 PL) having the parameters of grains, namely b (difficulty), a (discrimination), and c (pseudo-guessing). Question bank for the purpose of CAT should have those items with a power level is high and the distribution is uniform (rectangularly) at every level of ability [3] and should contain those items with: different power (a) has a distribution that is uniform between 0, 4 to 2.0, the index of difficulty (b) be spread uniformly between -3.0 to 3.0, and the factor guess apparent (pseudo guessing) (c) be spread between 0 to 0.3 [4] [5]. More specifically, Urry [6] suggest a question bank that is ideal for both CAT must have: the power parameter is different items (a) above 0.8, difficulty index parameter (b) has a wide distribution, and pseudo guessing factor parameter (c) of less than 0.3.

B. CAT systematics

Diagram adaptive test algorithms can be seen in the following figure:

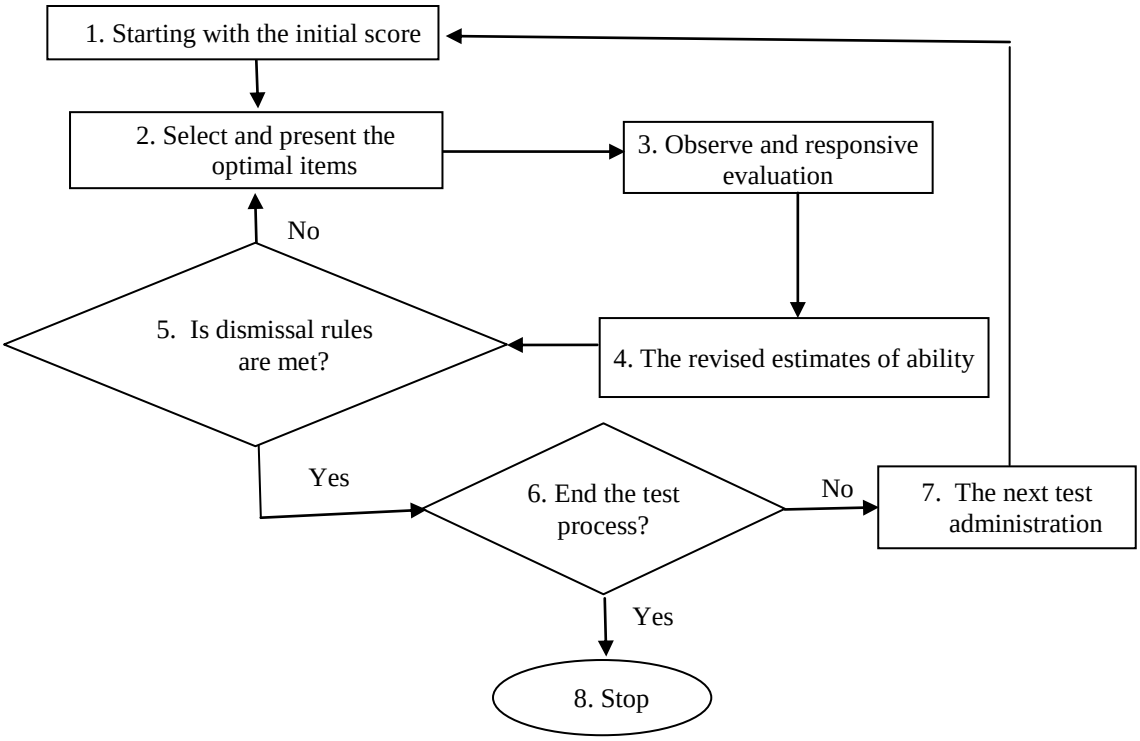


FIGURE 1. FLOW CHART ADAPTIVE TESTING

Based on the picture above, initially capabilities while participants estimated. Next awarded / presented items that optimally match the initial capability. Observe and evaluation of participants' responses. After the correct estimation of ability level of the participants. Then based on the rules of the dismissal of the test, to test whether the dismissal of the test criteria have been met or not. If you have met the test stops, otherwise if not met the participants are given optimal other items. This continues until the fulfillment of the criteria for dismissal of tests.

Systematics CAT contains the rules outlined in the steps that must be followed when participants carry out the test. The measures are commonly used to develop conventional CAT algorithm as follows [7]:

1. *How to Get Started* : The first items was given on the test taker?
2. *How to Continue*: After no response, the following items will be given to how the test taker?

3. *How to End*: When will the test be stopped?

Three main steps systematics CAT program, ie start, continue, and end, explained in more detail in the following sections:

1. CAT Starting Point

When CAT starts, no items were given to the participants of the test, there has been no response (response) given by the participants so that the test participants' ability levels can not be estimated. If no preliminary information about the ability of test takers, the CAT can begin by selecting the items beginning with a moderate level of difficulty [8] [9]. By selecting the items beginning with a medium level of difficulty, the participants answered any further tests that will be given items easily. Conversely, if answered correctly will be given items difficult.

Technically, it should be given a time limit for the test participants to respond to the answers. This is because the system will continue to wait for a response CAT test takers when there is no time restriction. Although given the limitations of time, Wise advise given sufficient time limit taking into account the factor of anxiety in the test participants take tests [10].

2. Continuing Process

After obtaining the response of the participants' answers to the test items given, CAT system gives a response assessment with a right or wrong answer. CAT system will decide whether or not to continue the test. There are two steps to continue the process of estimating the level of ability that CAT takers and how to choose the next items.

a. Methods of Estimating Capabilities

Having answered the item first given, the ability of the test taker is estimated based on the parameters of items, the estimated value of the initial capabilities, and answers to the items whether true or false. The general method used to estimate the ability of the test taker is Maximum Likeli-hood Estimation (MLE) [11] [12].

Suppose a test participants with ability θ answered tests containing n item multiple-choice items with unknown parameters (previously estimated). Joint opportunities of test participants can be written as $P(U_1,U_2,....,U_n \mid \theta)$. In practice, $U_1,U_2,....,U_n$ replaced with a score of items to participants who actually written as $u_1,u_2,....,u_n$ (u_i = 0 If the answer on items to i wrong, and u_i = 1 If the answer on items to i correct). Furthermore, if the assumption of local independence is applied then the likelihood function; $L(\theta)$, written as follows :

$$L(\theta) = P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n \mid \theta) = \prod_{i=n}^{n} P_i(\theta)^{u_i}\, Q_i(\theta)^{1-u_i}, \quad (1)$$

with $i = 1, 2, \dots, n, -\infty < \theta < \infty$.

The objective of MEE is finding value maximization $L(\theta)$. The parameter values that maximize the likelihood function capability, *L*, referred to *the maximum likelihood estimate of ability*. Mathematically, it is the same as to find a value that maximizes the value of the natural logarithm, *ln* $L(\theta)$. The core value can be obtained by making the first derivative of *ln* $L(\theta)$ toward θ equal to zero.

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} [u_i - P_i(\theta)] \frac{P_i'(Q)}{P_i(\theta)\, Q_i(\theta)} = 0 \quad (2)$$

In practice, to solve systems of equations above is done by using the Newton-Raphson iterative procedure. Score θ in iteration ($m + 1$) can be expressed using recurrent relations. The iteration process stops when $error < \varepsilon$, with ε very small numbers. In this study used value ε = 0.0001.

One problem with the application of MLE method on adaptive testing is the inability of the likelihood function to find a solution when there is a maximum of examinees who earn a score of 0 (answered wrong on all items) or a perfect score (answered correctly on all items), except examinees who earn scores were excluded from the estimation process. Examinees who obtain a score of 0 would acquire $\theta = -\infty$ (due to $u_i = 0$ only be met by $\theta = -\infty$), and a perfect score would be obtained $\theta = +\infty$ (due to $u_i = 0$ only be met by $\theta = +\infty$). Both of these scores are difficult to interpret in the application.

To overcome the problem of the inability of MLE method in estimating the level of response capability when the participants have not figured test participants can use the method *step size* [13] [14]. Based on the method step size, ability level test participants increased or decreased by a certain number of participants during the test have not been patterned response. Suppose CAT using a step size of 0.5 and a degree of prior knowledge of participants test setup value 0. This means that when one participant answered correctly all the tests on the first three items given, the estimated level of ability of (0 + 0.5 + 0.5 + 0.5) = 1.5. Conversely, if the participant answered incorrectly are all on the first three items, the estimation of his ability level (0-0.5-0.5-0.5) = -1.5.

b. Selection of next items

Once the ability of participants is estimated, the next computer select the next items. Lord suggests using items maximum information procedure to select the next items to be given to the participants of the test [15]. Based on this procedure, item that has a value function greatest information on the ability of certain participants have to be given to the test taker. This ensures that the value of the function test information for each person taking the test is maximum, meaning that the standard error of measurement (SEM) minimum because no other test information function is the inverse of the variance of the measurement error. In other words, this method guarantees will yield prediction skill level of participants with high accuracy [16].

Value item information function (IF) illustrates how accurate some items can estimate the ability of the test taker. Using the information function, the accuracy of measurement in estimating the ability of test takers can be calculated at every level of ability. Function Birnbaun information item to be stated by the following equation [1]:

$$I(\theta) = \frac{2.89\, a_i^2\, (1-c_i)}{\left[\left(c_i + exp\left(1{,}7\, a_i(\theta - b_i)\right)\right)\right]\left[1 + exp\left(-1{,}7\, a_i(\theta - b_i)\right)\right]^2} \quad (3)$$

The above equation shows that the information is only dependent on grain parameter (eg a, b, and c for the model 3P) and the level of ability (θ). Thus for every level of ability (θ), the contribution of the function information for each item in the question bank can be calculated.

Function test information is the number of item information function test developers such [17]. Information function test device is mathematically written as follows:

$$TIF = \sum_{i=1}^{n} I_i \quad (4)$$

As a function information item, the information function tests illustrate how accurate the estimate test different ability levels. The greater level of information on given ability, the more accurate the estimated ability of the test device.

Standard error of measurement (SEM) is expressed by the following equation [1]:

$$SEM = {}^{1}/_{\sqrt{TIF}} \quad (5)$$

c. Stopping Rule

Two main methods are used to stopping CAT, equal measurement precision and fixed number item. Both of these methods produce different measurement error variance. Is the purpose of the method equal measurement precision is generating test scores with the same error rate measurements for each test taker's ability. But the predicted length of the test varies from one participant to the other test participants. Standard error of measurement equivalent capped at 0.03 with a reliability of 91% on conventional tests [18]. But in practice it is used also use criteria fixed number of items the dismissal rules CAT, eg using criteria fixed starting rule as much as 20 grains to avoid the process of tests that may not converge.

Selection criteria assumptions on the components of the CAT will have different consequences. With reference to some research results CAT was developed, then the assumption that the criterion selected in the CAT algorithm is as follows:

1) Selection of initial grain based on the level of difficulty was.
2) Estimation of the level of capability by using Maximum Likelihood Estimation (MLE).
3) Selection of the next item using the procedure maximum value of the function information item.
4) Rules dismissal of tests using *equal measurement precision* and *fixed number of item*.

III. CAT PROGRAM ALGORITHM

Conventional CAT algorithm is expressed as follows :

A. Starting CAT
- *Starting with the login process CAT program, including entering username, password, name, identity number, and others*

B. Selection of the first item that appears.
- *Select the first item with the medium level of difficulty by randomly ie -0.5 ≤ b ≤ 0.5*

C. Selection of the second item that appears
- *Because there is no pattern then it using step-size (assumed to be step-size using a value of 0.5). If the response is correct answer, select the item with a value of θ = 0.5 and if the response is incorrect answers, select the item with a value of θ = -0.5*
- *Calculate the information function I (θ) to (3), namely:*

$$I(\theta) = \frac{2.89\, a_i^2\,(1-c_i)}{\left[\left(c_i + exp(1{,}7\, a_i(\theta - b_i))\right)\right]\left[1 + exp(-1{,}7\, a_i(\theta - b_i))\right]^2}$$

with value :
b = difficulty parameter
a = discrimination parameter
c = pseudo-guessing parameter

- *Calculate and find value I (θ) the maximum on all items, display items that have a value I (θ) the maximum.*

D. Selection of third item and the rest items
- *If the response to the second answer and so have not been patterned (always right or always wrong) then using the step-size with a value of θ added if the response answers +0.5 and -0.5 if the response is correct wrong answers.*
- *If the response answers the third and so have figured it using MLE*
  - *Calculate the estimated ability (ability level) test participants (θ) with the Newton-Raphson iterative procedure.*

$$\theta_{duga} = \theta_{duga\,0} + error \quad (6)$$

where

$$error = \frac{\sum 1.7\, a\,(u-P)\,(P-c)/\left(P\,(1-c)\right)}{\sum\left[-1.7^2\, a^2\,((1-P)/P)\right]\left[(P-c)/(1-c)\right]^2} \quad (7)$$

with value :
u = 1 if the student answers correctly
u = 0 if the student answers wrong
P = participants the opportunity to answer item correctly by the formula

$$P = c + \frac{(1-c)}{\left(1 + exp\left(-1.7\, a\left(\theta_{duga\,0} - b\right)\right)\right)} \quad (8)$$

  - *Iterating until got error ≤ 0.0001, then θ_duga = θ, iteration will take place at convergent and fast, usually iterations already completed less than 10 cycles.*
  - *Calculate the information function I (θ) with the above formula, find the value of I (θ) maximum on all items, display items that have a value I (θ) the maximum.*
  - *Calculate Test Information Function with (4), namely :*

$$TIF = \sum_{i=1}^{n} I_i.$$

by Ii = Item Information Function (IF)

  - *Calculate SEM (5) ie :*

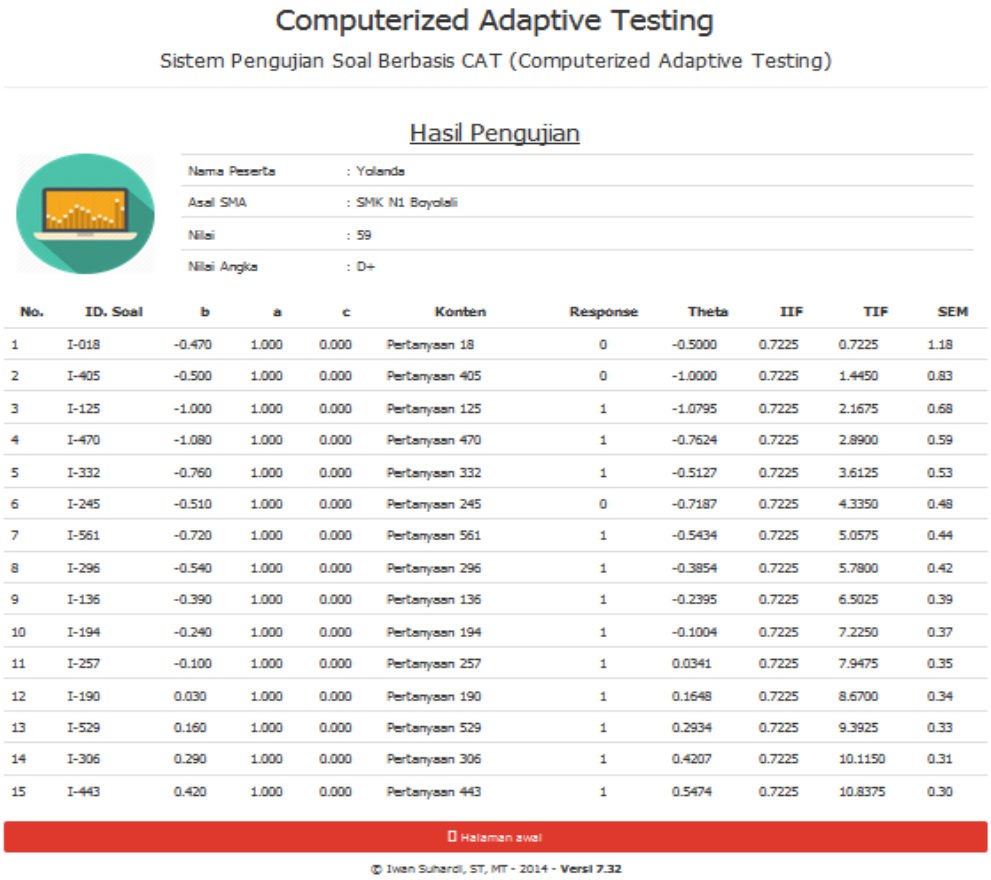
$$SEM = {}^{1}/_{\sqrt{TIF}}$$

E. Rules dismissal of tests using equal precision measurement and a fixed number of items..
- Rules discharge test with equal precision measurement
  - *The test will be stopped when the SEM ≤ 0.3, get the latest θ*
- Rules discharge test with a fixed number of items.
  - *The test will be stopped when the item appears achieve maximum number (eg 20 items), get last θ*

- *If the item is not patterned, always answer true or always answered incorrectly, the test will stop when $\theta \geq 3$ or $\theta \leq -3$*
- *Give limitation rule with :*
  When obtained $\theta \geq 3$, then $\theta = 3$, and if obtained $\theta \leq -3$, then $\theta = -3$

- Conversion value obtained by the formula

$$Score = 50 + \left(\frac{50}{3}\ \theta\ \right) \quad (9)$$



Computerized Adaptive Testing

Sistem Pengujian Soal Berbasis CAT (Computerized Adaptive Testing)

Hasil Pengujian

Nama Peserta : Yolanda
Asal SMA : SMK N1 Boyolali
Nilai : 59
Nilai Angka : D+

No.	ID. Soal	b	a	c	Konten	Response	Theta	IIF	TIF	SEM
1	I-018	-0.470	1.000	0.000	Pertanyaan 18	0	-0.5000	0.7225	0.7225	1.18
2	I-405	-0.500	1.000	0.000	Pertanyaan 405	0	-1.0000	0.7225	1.4450	0.83
3	I-125	-1.000	1.000	0.000	Pertanyaan 125	1	-1.0795	0.7225	2.1675	0.68
4	I-470	-1.080	1.000	0.000	Pertanyaan 470	1	-0.7624	0.7225	2.8900	0.59
5	I-332	-0.760	1.000	0.000	Pertanyaan 332	1	-0.5127	0.7225	3.6125	0.53
6	I-245	-0.510	1.000	0.000	Pertanyaan 245	0	-0.7187	0.7225	4.3350	0.48
7	I-561	-0.720	1.000	0.000	Pertanyaan 561	1	-0.5434	0.7225	5.0575	0.44
8	I-296	-0.540	1.000	0.000	Pertanyaan 296	1	-0.3854	0.7225	5.7800	0.42
9	I-136	-0.390	1.000	0.000	Pertanyaan 136	1	-0.2395	0.7225	6.5025	0.39
10	I-194	-0.240	1.000	0.000	Pertanyaan 194	1	-0.1004	0.7225	7.2250	0.37
11	I-257	-0.100	1.000	0.000	Pertanyaan 257	1	0.0341	0.7225	7.9475	0.35
12	I-190	0.030	1.000	0.000	Pertanyaan 190	1	0.1648	0.7225	8.6700	0.34
13	I-529	0.160	1.000	0.000	Pertanyaan 529	1	0.2934	0.7225	9.3925	0.33
14	I-306	0.290	1.000	0.000	Pertanyaan 306	1	0.4207	0.7225	10.1150	0.31
15	I-443	0.420	1.000	0.000	Pertanyaan 443	1	0.5474	0.7225	10.8375	0.30

Halaman awal

© Iwan Suhardi, ST, MT - 2014 - **Versi 7.32**

FIGURE 2. SAMPLE RESULTS OBTAINED CAT

TABLE 1. DESCRIPTION OUTPUT CAT

No.	Item ID.	b	a	c	Content	Response	Theta	IIF	TIF	SEM
1	I-018	-0.470	1.000	0.000	Question 18	0	-0.5000	0.7225	0.7225	1.18
Information:		The first point is drawn at random by selecting the item with a difficulty level was $-0.5 \leq b \leq 0.5$. Student answered incorrectly, then the step-size models will appear $\theta = -0.5$. The IIF calculated values for $\theta = -0.5$ obtained in a matter of numbers with Item ID. I-405 as the second items. Because the student answers incorrectly, the following items appear easier.								
2	I-405	-0.500	1.000	0.000	Question 405	0	-1.0000	0.7225	1.4450	0.83
Information :		For items the student answers incorrectly, it still uses the step-size models because they still have not figured in order to extract the value of $\theta = (-0.5) + (-0.5) = -1$. The IIF calculated values for $\theta = -1$ obtained in a matter of numbers with Item ID. I-125 as the third items. Because the student answers incorrectly, the following items appear easier.								
3	I-125	-1.000	1.000	0.000	Question 125	1	-1.0795	0.7225	2.1675	0.68
Information :		For items three students answered correctly, the next is already used models MLE for the order item has been patterned, so that by way of iterating obtained value $\theta = -1.0795$. The IIF calculated values for $\theta = -1.0795$ obtained in a matter of numbers with Item ID. I-470 as a fourth items.								
4	I-470	-1.080	1.000	0.000	Question 470	1	-0.7624	0.7225	2.8900	0.59
Information		Although the student answers correctly, the fourth item that appears to have slightly lower levels of difficulty, than ever before. It is ideal not supposed to happen. Supposedly items appear to be more difficult than the previous items. This could happen because there is a question bank that has a difficulty level parameters uneven.								
5	I-332	-0.760	1.000	0.000	Question 332	1	-0.5127	0.7225	3.6125	0.53
Information		For items five students answered correctly, then the model is used MLE sought IIF maximum value to								

		get items that will appear. Because the students answered correctly, the next items that appears more difficult. And so on.								
6	I-245	-0.510	1.000	0.000	Question 245	0	-0.7187	0.7225	4.3350	0.48
7	I-561	-0.720	1.000	0.000	Question 561	1	-0.5434	0.7225	5.0575	0.44
8	I-296	-0.540	1.000	0.000	Question 296	1	-0.3854	0.7225	5.7800	0.42
9	I-136	-0.390	1.000	0.000	Question 136	1	-0.2395	0.7225	6.5025	0.39
10	I-194	-0.240	1.000	0.000	Question 194	1	-0.1004	0.7225	7.2250	0.37
11	I-257	-0.100	1.000	0.000	Question 257	1	0.0341	0.7225	7.9475	0.35
12	I-190	0.030	1.000	0.000	Question 190	1	0.1648	0.7225	8.6700	0.34
13	I-529	0.160	1.000	0.000	Question 529	1	0.2934	0.7225	9.3925	0.33
14	I-306	0.290	1.000	0.000	Question 306	1	0.4207	0.7225	10.1150	0.31
15	I-443	0.420	1.000	0.000	Question 443	1	0.5474	0.7225	10.8375	0.30
Information		CAT will stop because the value of SEM meets the criteria, namely the dismissal of ≤ 0.3. Last θ value obtained by the 0.5474 conversion into the final value, a score 59								

IV. CONCLUSIONS AND RECOMMENDATIONS

A. *Conclusions*

In education, it is important to determine the ability of students in a subject. Adaptive testing is a testing method in which each examinee will be given a set of different questions adjusted according to the ability of each learner. Thus, each of the examinees do not need to answer all the problems that exist. CAT has many advantages compared to other testing applications such as paper and pencil test or CBT, and has been proven to have high efficiency and reliability. Participants ability test already can be estimated only by answering less than half of the items required in a paper and pencil test or CBT. In building a conventional CAT application basically covers started CAT (starting point), the process continues CAT (continue process) and CAT dismissal rules (stopping rule). CAT also be supported by IRT-based question bank and evenly distributed so that students can have the ability to estimate optimal accuracy.

B. *Recommendations*

Paper is expected to be able to provide a foundation early in the morning of developers and researchers to build and develop the CAT program for wider scale applications. Conventional CAT designs have the possibility of items given to participants of the test does not represent all of the modules / materials that exist. Development suggested among others by constraint content so that all modules / material can appear in a matter of CAT. Among other development control procedures and equity items appear (item exposure), and CAT integrate with other e-learning modules to determine which modules need to be studied further.

REFERENCES

[1] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, "Fundamentals of item response theory," Newbury Park, CA: Sage Publication, Inc., 1991.

[2] H. Wainer, "Computerized adaptive testing: A primer, " 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1990

[3] T. Wang and W.P. Vispoel, "Properties of ability estimation methods in computerized adaptive testing," Journal of Education Measurement, 2, 1998, pp.109-136.

[4] T.N. Ansley and R.A. Forsyth, "An examination of the characteristics of unidimensional IRT parameter estimates derived from two dimensional data," Applied Psychological Measurement, 1, pp. 37-48.

[5] V.G. Folk and B.F. Green, "Adaptive estimation when the unidimensionality assumption of IRT violated," Applied Psychological Measurement, 4, 1989, pp. 373-389.

[6] V.W. Urry, "Tailored testing: A successful application of latent trait theory," Journal of Educational Measurement, 2, 1977, pp.181-196.

[7] D. Thissen and R.J. Mislevy, "Testing algorithms", dalam H. Wainer (Ed.), Computerized Adaptive testing : A Primer, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates.Thissen & Mislevy, 1990, pp. 103-135.

[8] W.P. Vispoel, " Creating computerized adaptive test of music aptitude: Problem, solutions, and future directions," dalam F. Drasgow, & J.B. Olson-Buchanan (Eds.), Innovation in Computerized Assessment (pp. 151-176). Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 1999, pp. 156.

[9] C.N. Mills, "Development and introduction of acomputer adaptive graduate record examination general test," dalam F. Drasgow & J.B. Olson-Buchanan (Eds.), Innovation in Computerized assessment, Mahwah, NJ: Lawrencw Erlbaum Associates Publishers. Mills, 1999, pp. 123.

[10] S.L. Wise, “Examinee Issues in CAT”, The Annual Meeting of National Council on Measurement in Education, Chicago, March 25-27, 1997.

[11] A. Birnbaum, “Some latent trait models and their use in infering an examinee's ability,” MA: Addison-Wesley, 1986.

[12] F.B. Baker, “ Item response theory: Parameter estimation techniques,” New York: Marcel Dekker, Inc., 1992.

[13] B.G. Dodd, “The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model,” Applied Psychological Measurement, 4, 1990, pp. 355 – 366.

[14] D.J. Weiss, “Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education,” Measurement and Evaluation in Counseling and Development, 37, 2004, pp. 70-84.

[15] F.M. Lord, “A broad-range tailored test of verval ability,” Applied Psychological Measurement, 1, 1977, pp. 95-100.

[16] D.R. Eignor, M.L. Stocking and W.D. Way, “Case studies in computer adaptive test design through simulation (Research Report RR-93-66),” Princeton, NJ: Educational Testing Service, 1993.

[17] R.K. Hambleton and H. Swaminathan, “ Item response theory,” Boston, MA: Kluwer Inc., 1985, pp. 94.

[18] D. Thissen, “Reliability and measurement precision,” dalam H. Wainer (Ed.), Computerized Adaptive testing : A Primer (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates, 1990, pp. 103-135.