Jurnal CoreIT, Vol.9, No.2, December 2023
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Science Interest Detection Using Computerized Adaptive Testing Based on Fuzzy Item Response Theory

Fitri Wulandari[1], Haryanto[2], Samsul Hadi[2], Elin Haerani[1], Fadhilla Syafria[1]

[1]Faculty Sains and Technology/ State Islamic University of Sultan Syarif Kasim, Riau
[2]Yogyakarta State University (UNY)
*Corresponding author e-mail: fitriw1974@gmail.com[1]

Abstract

Choosing a major or interest at the beginning of High School is a very important process for the future development of students. A test may be performed to determine the learner's ability in a particular field in this research, an interest test was developed to determine the students' ability in science. Students will be measured for their cognitive ability in Mathematics and Science subjects for junior high school level. The research was developed using an adaptive test system called Computerized Adaptive Testing (CAT). CAT is an adaptive media-based model, test participants will receive the test according to their ability. The test item selection procedure uses the fuzzy algorithm using item difficulty parameters, item strengths and participants' response data as input data. While the rule or procedure for terminating the test is done with the maximum likelihood estimation method, MLE. Based on the test results, each student received different test items according to their ability level and the difficulty indexs that received by the students according to the characteristics of the item information. Therefore, the CAT program with the fuzzy item response theory can be used as a support for measuring the students' ability and interest in a major.

Keywords— Computerized adaptive test, maximum likelihood estimation, fuzzy item response theory

Received July 2023/ **Revised** October 2023 / **Accepted** 25 December 2023

1. Introduction

As the world and information technology evolve, there has been a change in human behavior over time. It has also changed the development of the education system in the world and particularly in Indonesia. These changes can be seen from the changes in the education system, which include learning, teaching, curriculum, learning methods, learning tools, facilities and infrastructure, as well as graduate competence from time to time. The current curriculum in the education system in Indonesia not only emphasizes on the achievement of quantitative goals in the form of test scores of a number of academic subjects, but also emphasizes on process-based assessments and student achievement [1]. Learners are given more opportunities to select subjects that they are interested in, to study and develop their potential more flexibly based on their general basic skills (intelligence), talent, interests and personality characteristics.

The function of education is not only in the teaching learning process, but it also includes guidance / counseling, selection and placement of students according to their individual capacities. The selection of interest at the beginning of the Secondary School or High School enrollment is essential process for the future development of the students. This is because majoring can be a reference for each student for further study. In addition, students who study a subject that they are passionate about will enjoy it, so that students will be able to develop their potential and perform well.

Test is set of planned questions to obtain information on trait (educational or psychological attributes) in which each of the questions has a certain answer regarded as correct [2]. According [3] test may be performed to determine the learning achievement or competencies that the learners have achieved in a particular field. Test results are information about the characteristics of a person or group of people in terms of their cognitive abilities or skills. Reference [4] shows that cognitive domain learning is oriented to thinking ability, including simpler skills and problem-solving skills. This testing activity is one of ways of predicting the learner's ability level indirectly, in response to a number of stimuli or questions. Test

DOI: 10.24014/COREIT.V9I2.27258

results are expected to produce data with as little error as possible. Therefore, to get accurate data requires a valid and reliable test.

The use of computer technology to improve the quality of test results has been extensively done. In the advanced technology and information era, it is very feasible to conduct a computer-based test [5]. According [6] [7] and [8] states that the test using a computer not only able to generate test that quickly and accurately, but also can improve the effectiveness and efficiency of the implementation and maintenance of the test. Currently test administration is moving from paper and pencil test to computerized.

In this form, the computer is used to display similar test items to those on the test paper. This kind of test model still doesn't feel proper, because the order of the test items appearing over time does not change, making it easy to memorize. The mechanism of retrieval of test items from a bank database is either sequential or random, according to the theory of measurement that reduces the validity of the test results. All test participants will receive the same set of test items or equivalents so that there is still a mismatch between the difficulty level and the ability of the test participants. Such measurements become inefficient because each item is presented in large quantities and cannot provide enough information to distinguish the test participants' ability to scale. High-ability test participants get a few simple questions so they have a small chance of answering the wrong thing. Similarly, low-test participants will find some difficult questions that they have little chance of answering correctly. Items like these do not provide enough information about the level of ability of the test participants.

Based on observations on programs used by agencies / institution based on computer-based testing, there are a few things that need to be studied and researched further. One of them is the random process of retrieving test items from the database of banks in computer-based testing. To solve this problem, in this study developed an adaptive test system, called the term Computerized Adaptive Testing (CAT). CAT is an adaptive-based medium where test participants receive test questions according to their ability, [9]. [10] further explains that CAT-based testing can improve efficiency and accuracy and practicality in its implementation. In this study, the interest test developed using CAT is to find out the students' ability in science. Students will measure their cognitive abilities in math and science subjects.

Some rules and procedures are set in the development of CAT, including test starting rules, test item selection procedures, and test delivery rules. The starting rule was by providing a pre-test to measure the initial ability of the test participant. Rules of selection of test items using fuzzy Algorithm with granularity parameters, different items strengths and participants' response data as their input. Fuzzy logic algorithms are used because they are suitable and appropriate for problem solving that map the qualitative values of one's cognitive ability (low, medium, and high) into quantitative values [11]. While the rule or procedure for ending the test and estimating the student's ability is done with the maximum likelihood estimation method, MLE.

2. Research Method

The development of computerized adaptive testing methods for measuring students' ability in this research consisted of three stages, namely (1) preparing a bank of test materials, (2) selecting test items using Fuzzy CAT, and (3) estimating student ability.

2.1. Preparing an Item Bank of Test Materials

Item bank (test materials) is a system that contains items, from a test that has a specific purpose, including its utilization system [12]. In developing a question bank, the details of the existing questionnaire were compiled into a test device and then tested. A test device is a good idea if it has good questionnaire characteristics, so it needs to first perform a characteristic analysis of the item. This is possible with both classical and modern approaches (item response theory) [[11], [13] , [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. But as science and technology advance, the use of item response theory approaches is increasingly popular. There are three assumptions underlying the theory of particle response, namely, undimensionality, local independence and parameter invariance [22]. In item response theory, the relationship between probability of responding correctly on a given ability scale is related to the relation to the item parameter used. Many of the parameters used are defined by the model equation.

The assumption of undimensionality can be proven by using factor analysis to see the eigenvalues of the inter-particle covariance matrix. Data analysis with factor analysis was preceded by sample adequacy analysis. This study would prove the unidimensional assumptions in the test participants' data on Mathematics and Science. Testing devices were tested using CBT for 770 high school students in Yogyakarta city. The test participants' response pattern data were calibrated with program R. The following

DOI: 10.24014/COREIT.V9I2.27258

questionnaires that met the ideal criteria were those with different power levels in the range of 0 to 2 and difficulty levels of -4.0 to +4.0 were selected as items - bank questions / test materials about CAT requirements.

2.2. Selecting Test Items using Fuzzy CAT

According to [24] and [25] adaptive testing is a test conducted for test participants whose questions /items are determined based on the participant's initial answers/ responses. Computerized adaptive testing not only efficiently shorten testing time and reduce the number of test items but can also accurately predict test participants' ability [26].

In CAT, the computer is set up to select and provide questionnaires, and the computer will then calculate and code the test participant's answers. The questionnaire given to the test participant is a questionnaire that is tailored to the test participant's response to the previous item. If the item is answered correctly, then the item is presented with a higher level of difficulty. If the questionnaire is answered incorrectly, it will be presented with the item with a lower level of difficulty [11]

By using fuzzy algorithms, the choice of questionnaire becomes quite different. The inputs for this algorithm are different strengths items, difficulty level items and test participant response. These parameters are processed through the membership function in a fuzzy set. The output obtained is the assurance of test items having different power and difficulty level up or down depending on the response of the test participant. The output is done by an inference mechanism based on the fuzzy algorithms in the form of the next test item to be given to the test participant.

The inference system also known as fuzzy control is a mechanism in fuzzy logic for decision-making. The inference model used in this study is Tsukamoto [27]. The fuzzy logic algorithm for generating output is done in four stages, namely:

a. Formation of fuzzy assemblies. The input variables and the output variables are divided into one or more fuzzy sets, which are based on the selected membership function.
b. Implications, namely, the formation of rules, based on knowledge. According to the Tsukmoto method, the implication function used is min (smallest value)
c. Inference, the affirmation of a decision based on the rule base is a set of rules used as the basis for making an inference.
d. Defuzzification is the affirmation of the inference result based on the weighted average value

Input from the defuzzification process is a fuzzy set derived from the mechanism of inference to the composition of fuzzy rules. The output generated from this defuzzification process is a number in the domain of the fuzzy set. If given a fuzzy set in a particular area, then a certain crispy value can be extracted as a result of the defuzzification process.

The steps for selecting test item with fuzzy logic are as shown in figure 1 below:

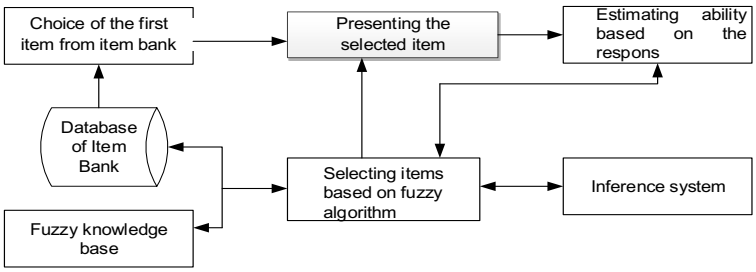

Figure. 1. The steps for selecting a test item with fuzzy logic

Based on Figure 1, the CAT item selection process with fuzzy logic algorithm starts with selecting the first questionnaire from the question bank. Once the questionnaire is selected, the questionnaire is then given to the test participant. Participants respond (right or wrong) to the questionnaire, and then the level of ability of the participant is re-estimated. Based on the details of the problem details, the different strengths of the test items and the response of the test participants, the parameters are processed through the fuzzy logic function. The output is done by the mechanism of the inference system based on the fuzzy algorithm in the form of the next test item that will be passed to the test server. This process is continuous and discontinued once as many specification items have been provided or after the precision of the ability level or the standard error of desired measurement has been achieved.

DOI: 10.24014/COREIT.V9I2.27258

2.3. Students' Ability Estimation

Item response theory (IRT) is a psychometric theory that provides the basis for measuring the scale of test participants and questionnaires based on the responses given to them. Modern test models with IRT are distinguished by the number of parameters. namely the model one logistic parameter (1 PL or Rasch model), two logistic parameters (2 PL), and three logistic parameters (3 PL) [22]. [28] mentioned that the parameters are difficult items, different strength items, and guesses. This study uses the IRT model for dichotomous problems of two logic parameters (2PL) namely difficult items and different strength items according to [27], [28], [29] is mathematically formulated as follows:

$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad ; \qquad i = 1, 2, 3, \ldots, n \tag{1}$$

Description:
$P_i(\theta)$: Probability of test participants having θ ability to answer item i correctly
θ : Subject ability level (as a free variable)
a_i : discriminant parameter item i
b_i : Index of difficulty of item i
e : natural number is close to 2,718
n : number of items in a test

The b_i parameter is a index of difficulty of item i is point on ability so that the probability of answering correctly an item is 50% [30]. The higher b_i indicates more difficult an item and needed greater ability to answer this item. The ai parameter is an item characteristic related to the item's ability to emphasize differences between participants who can answer correctly and answer incorrectly.

The three IRT concepts used in CAT development are (1) information function item (FIB), (2) standart error measurement (SEM), and (3) ability level estimation. The information function item is expressed as $I_i(\theta)$, which is a function that provides information by item i on θ. [20] state that each item has information on how well the item can differentiate between test takers with similar capabilities at different levels of ability. Mathematically, the item information function satisfies the following equation

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \; ; \qquad i = 1,2,3,\ldots n \tag{2}$$

Description:
$I_i(\theta)$: information function on item i
$P_i(\theta)$: chance participants with ability to θ correctly answer the item i
$P_i'(\theta)$: derivative of function $P_i(\theta)$ toward θ
$Q_i(\theta)$: chance participants with ability θ wrongly answer the item i

Equations (1) and (2) show that the value of information depends only on the parameters of items (a and b) and their capabilities. Thus, for each level of ability (θ), the information contribution to each bank item can be calculated. The test information function is the sum of the functions of the test item's granular information [23] which describes how accurate the test device is to estimate the level different capabilities. The more information at a given ability levels the more accurate that ability is estimated from the test device.

Standard Error of Measurement (SEM) is a standard error of measurement that has a close relationship with the information function. The test information functions compared to the quadratic inversion with SEM, so the more the test information functions then the SEM is smaller or vice versa. The second relationship is stayed by [22]:

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta}} \tag{3}$$

DOI: 10.24014/COREIT.V9I2.27258

The assessment of students' abilities, first, is done by calculating the values of pi (θ) and qi (θ) of each test item. The ability value (θ) is taken in the range -3 to +3 with step 0.2. Furthermore, it is known that the values of pi (θ) and qi (θ) and θ can be calculated as the L values (L | U | θ) with the following equation:

$$L(U|\theta) = \prod_{i=1}^{n} p_i^{n} q_i^{1-n} \quad (4)$$

Based on the results L (U | θ) for the values of θ from −3 to +3 then the test participants' estimated ability is θ from the highest (L |Uθ) results. So the character estimation of test participants' capabilities is determined by the formula:

Estimation θ = Maximum [L(U| θ)]

The estimation states that the test participant's ability to characterize (θ) answers to the max L (U | θ) x 100% questionnaire correctly. On the contrary, the test participants' ability to (θ) answer the [1-max L (U | θ)] x 100% questionnaire was incorrect.

Determination of the degree of interest in exact (IPA) major based on the results obtained in mathematics and IPA tests. The fuzzy logic algorithm determines the test participant's ability (KMP), which is divided into 3 groups namely high ability (H) if KMP ≥ 90, (2) medium ability (M) if 75≤ KMP <90, and (3) low ability (L)) if the KMP is <75. Students join the science interest group if they meet the requirements as shown in the table 2

Table 2. Rule of Fuzzy

No	Ability		Interest Science
	Mathematics	Science	
1	Low	Low	No
2	Medium	Low	No
3	High	Low	No
4	Low	Medium	No
5	Medium	Medium	Yes
6	High	Medium	Yes
7	Low	High	No
8	Medium	High	Yes
9	High	High	Yes

3. Result

3.1. Item Bank (Test Instrument)

Calibration using the R program resulted in 160 math questions and 151 exact (IPA) questions. A summary of the statistical parameters bank items of the Empirical questionnaire is presented in Table 3 and Table 4 below

Table 3. Statistic of Mathematics Item Bank

Parameter	Mean	Standard Deviation	Minimum	Maximum
Discriminant Index (a)	0,9350	0,3667	0,17	2,47
Difficulty Index (b)	-0,3696	0,8256	-2,19	4,01

Table 4. Statistic of Science Item Bank

Parameter	Mean	Standard Deviation	Minimum	Maximum
Discriminant Index (a)	0.8063	0.38703	0,14	2,22
Difficulty Index (b)	-0.318	1.16581	-3.88	3,86

DOI: 10.24014/COREIT.V9I2.27258

An item is said to be good if the level of difficulty (b) from -2 up to 2 [23]it can be concluded every item has a normal difficulty level because it ranges from -2,19 up to 4,01. Likewise, the value of the discriminant index (a), produced a good item from 0.7 up to 2.47. As reference, a good item has discriminant index from 0 up to 2 [23].

Based on the analysis of sample adequacy in the math test the Khi-square value of Bartlet test was 29278,627 with degrees of freedom of 12720 and p-value less than 0.01, whereas for exact (IPA) subjects the Khi-square value of the Bartlet test 26801,505 with degrees of freedom 12720 and p-values less than 0.01. This result indicated that the sample size used in this study was sufficient and the variables can be further analyzed [31], [32]

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.895
Bartlett's Test of Sphericity	Approx. Chi-Square	29278.627
	df	12720
	Sig.	.000

Figure 2. KMO and Bartlett's test of Mathematics Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.828
Bartlett's Test of Sphericity	Approx. Chi-Square	26801.505
	df	12720
	Sig.	.000

Figure 3. KMO and Bartlett's test of Science Test

Further, the number of questionnaires in the bank is classified into three groups based on the level of item difficulty, which are difficult, moderate and easy level group. The results of the questionnaire classification are presented in table 5 and 6 below

Table 5. Range of Mathematics Test Items

Group	Frequency	percentage
• Ease	32	20%
• Moderate	95	59,4%
• Difficult	33	20,6%

Table 6. Range of Science Test Items

Group	Frequency	percentage
• Ease	41	27,1%
• Moderate	80	53 %
• Difficult	30	19,9%

The distribution of the groups is quite good, with the proportion of items on the medium difficulty level being higher than the items on the level of easy and difficult.

3.2. Result of CAT Simulation

The process that takes place in the CAT program starts from the test participants logging into the system, performing math and Exact Science (IPA) tests, until they get the test results. The scenario is as follows:

1. Selection of first test items based on students' ability at pre-test.
2. Students will be given 3 introductory questions to initialize students' abilities with easy, medium and hard difficulty levels.
3. Student response to all three test items (Score = 1 if correct, and score = 0 if incorrect) can have one of the following patterns: 000,001,010,011,100,101,110, or 111. If the pattern is 011, 101, or

DOI: 10.24014/COREIT.V9I2.27258

111 then the ability (θ) = high, for the response pattern 110,010, or 001 then the ability (θ) = medium, and for the response pattern 100 or 000 the ability (θ) = low

4. The pre-test (θ) test result is estimated as the initial ability to provide the first test item. Test participants can choose math or exact (IPA) test as their first test.
5. Display the first item, take the response of the test participant (ABCD), then confirm the response with the answer key to get the score.
6. Secondary selection using fuzzy logic based on variable in terms of item difficulty level (b), power / strength different item (a) and answer response from previous item. If the bank in question does not find the item with the difficulty level in question, it will retrieve the item nearest to the group.
7. Display the third item based on the answer to the second item, and do the same as point 6. The probability that the score is 000.001, 100, and so on, or 111.
8. Calculate theta (three scores) with MLE formula, followed by SEM calculations.
9. Select the next item based on the third item's answer
10. Display the selected question, get feedback from the participants, and then confirm with the key to get the score
11. Calculate Theta values from the last 4 scores, and calculate SEM values as well
12. Loop (repeat):
    a. Repeat steps 8-11, and end the test if: (1) the number of problematic items is reached or (2) the SEM value is less than or equal to 0.01.
    b. Records the latest Theta and SEM values

3.3. Ability Estimation (Theta)

The method of estimating ability used in research is Maximum Likelihood Estimation (MLE). The CAT results of the test participants for the math test are shown in table 7, the students' ability graphs are shown in figure 4, the mathematical cognitive abilities in table 8, and the Function Graphs Likelihood Test Item Response Graphs in figure 4.

Table 7. The CAT Result of Mathematics

No	Item Code	Difficulty Parameters	Item Strength	Responce	SEM	FIB	Theta
1	MAT156	-2.148	0.596	True	13.7196	0.0053	3
2	MAT046	0.542	0.902	False	1.6501	0.3673	-1
3	MAT076	0.327	0.763	True	0.9680	1.0673	0.5
4	MAT091	0.533	0.872	True	1.0144	0.9719	1.5
5	MAT078	0.507	0.563	False	0.7863	1.6175	1
6	MAT017	0.326	0.792	True	0.7091	1.9889	1
7	MAT062	0.49	0.795	False	0.6031	2.7490	0.5
8	MAT137	-1.004	1.258	False	0.5984	2.7925	-0.5
9	MAT001	-1.041	0.897	True	0.5522	3.2801	0
10	MAT022	0.483	0.601	False	0.5326	3.5258	0
11	MAT045	0.297	0.87	False	0.5074	3.8836	-0.5
12	MAT113	-1.056	1.003	True	0.4731	4.4688	-0.5
13	MAT060	0.478	0.858	False	0.4564	4.8010	-0.5
14	MAT125	-1.09	1.533	True	0.4374	5.2270	0
15	MAT008	0.473	0.649	True	0.4260	5.5115	0
16	MAT028	0.463	0.809	False	0.4103	5.9395	0
	MAT002	-1.091	0.886	True	0.4001	6.2475	0
18	MAT048	0.442	0.408	False	0.3964	6.3650	0

Table 7 above shows the history of student test results using CAT consisting of question number, question code, difficulty level, different power, answer response, SEM, FIB, and Theta. In the first item, the MAT156 question code has a difficulty level of -2.148, a power difference of 0.596, and the answer is correct. The fuzzy algorithm was used to determine the next test item, and the second test item was chosen as MAT046 with a difficulty level of 0.542. The selection of the second test item is in accordance with the rule used, that is, if the answer response is correct then the level of item difficulty is increased. In the second test item the student answer was incorrect and the difficulty level for the third test was lowered, and the third test item was selected with the MAT076 questionnaire and item level of 0.327. The accuracy of item

DOI: 10.24014/COREIT.V9I2.27258

selection in this test is in accordance with [33] stating that CAT will adjust the presenting items to the ability level of the test takers (theta).

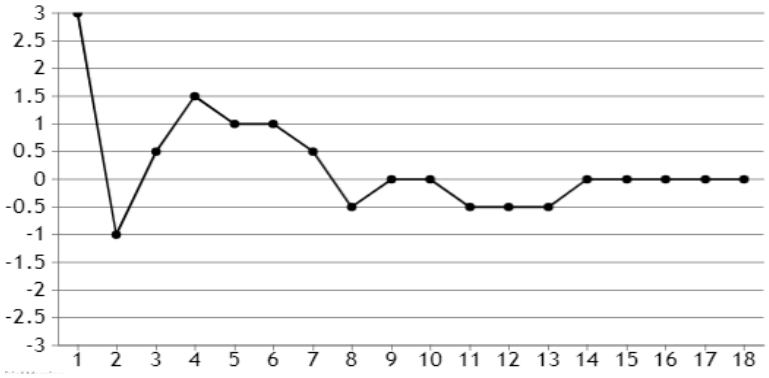

Figure. 4. The Graph of Student's Ability

Table 8. Student's Cognitive Ability of Mathematics Test

Ability (θ)	**L(U | θ)**	**L(θ | U)**
-3	0.000000	0.0000
-2.5	0.000000	0.0002
-2	0.000000	0.0026
-1.5	0.000002	0.0244
-1	0.000009	0.1197
-0.5	0.000020	0.2840
0	0.000023	0.3215
0.5	0.000013	0.1815
1	0.000004	0.0551
1.5	0.000001	0.0098
2	0.000000	0.0011
2.5	0.000000	0.0001
3	0.000000	0.0000
Amount	**0.000072**	**1.0000**

The students' ability estimation in using CAT depends on the difficulty level of the item that is answered correctly and is not only determined by the number of correct answers. Figure 4 is a graph of the test results of table 7, showing how students' ability estimated. Beside to being influenced by the answer, ability estimation is also influenced by the difficulty level of the test items and the different strengths of the test items. From item 14 to item 18 students' ability to converge, and to stop at item 18 because the difference between SEMs in items 17 and 18 is ≤ 0.01. This result is supported by [34] where CAT generally requires fewer items than long-form instruments and can achieve the same precision. CAT also optimizes items are managed and can produce the most significant information in a measurement of the ability of test-takers.

Based on Table 8 it can be seen that the maximum value of L (θ | U) is 0.3215 with an ability position (θ) of 0. If table 5 is graphed the Likelihood function will look like Figure5. The likelihood results for values θ from -3 to +3 (figure 5) represent the estimated ability of test takers, namely θ from the highest likelihood results (maximum). The likelihood value of 0.000023 indicates that the maximum Likelihood of 0.3215 and ability (θ) is 0. This implies that the probability of students with the ability [θ] = 0 to answer

DOI: 10.24014/COREIT.V9I2.27258

the test item correctly is 32%. Furthermore, Theta value (θ) is converted in the form of numbers from 0 to 100. The conversion result for theta 0 is 50.

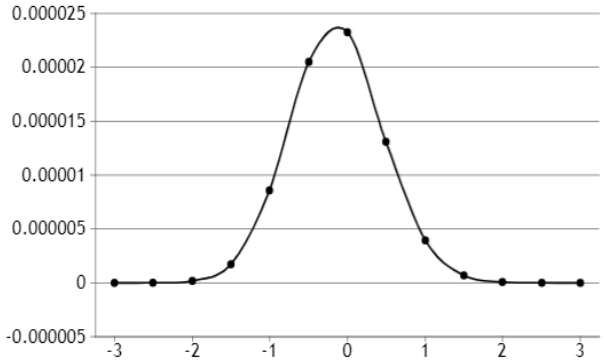

Figure 5. The Graph of Test Item Likelihood Function of Mathematics Test

The same thing was done in the Science (IPA) test, the results obtained as in table 9. In item 14, the SEM was 0.4893. The difference of SEM value between item 14 and item 13 is ≤ 001, so the IPA test is stopped. Based on 14 items which were done by the test takers, there were 7 items which were considered as correct items and 7 items were incorrect. The estimated ability value (θ) is 1, and then converted to a value of 66.67.

Table 9. The CAT Result of Science Subject

No	Item Code	Difficulty Parameters	Item Strength	Respons e	SEM	FIB	Theta
1	BIO2052	-2.101	0.885	True	308,52 9	0,0011	3
2	BIO2057	0.553	0.227	False	28,272	0,1251	0
3	FIS1012	0.339	0.82	False	13,315	0,5641	-1
4	BIO1018	-1.054	1.303	True	10,103	0,9797	0
5	FIS1007	0.619	0.807	True	0,9224	11,752	0.5
6	FIS2041	0.913	0.922	True	0,9264	11,651	1.5
7	FIS1011	1.018	0.772	False	0,7096	19,860	1
8	BIO1028	0.47	0.572	False	0,6523	23,501	0.5
9	FIS1004	0.266	0.636	False	0,6260	25,519	0
10	FIS2040	0.25	0.668	True	0,5819	29,537	0.5
11	FIS1008	0.694	0.908	True	0,5511	32,920	1
12	FIS2057	1.035	0.899	True	0,5080	38,756	1
13	FIS1001	1.202	0.573	False	0,4932	41,105	1
14	BIO1034	0.326	0.306	False	0,4893	41,761	1

Based on the results of Mathematics and Science tests, the test taker's ability of both subjects are in low level. The results then were analyzed by fuzzy algorithm (Table 2). Based on the analysis, it can be concluded that the student is not proper to be included in the science class specialization.

4. Conclusion

Based on the results of the CAT program testing in this study, it can be concluded that the ability of the CAT program with fuzzy response items theory, has been consistent with the item response theory, namely that each student will receive different test items according to their ability level. In addition, the degree of difficulty that the student receives depends on the characteristics of the item information. This is in line with the nature of the CAT theory that demands adaptability in tests. The nature of the adaptability

DOI: 10.24014/COREIT.V9I2.27258

is embedded in the fuzzy theory of inference system that can determine the decision that each student should receive the right number of items, and each student must receive the right test item according to their ability characteristics.

The CAT program with the fuzzy item response theory developed in this study has been able to work in accordance with these demands, so the CAT program with the fuzzy item response theory can be used as a support for the students' ability to measure and interest.

5. References

[1] Gultom, Syawal, Pedoman peminatan peserta didik Jakarta, Kementerian Pendidikan dan Kebudayaan, 2013.
[2] Zainul, A. and Nasution, N. (2001). Penilaian hasil belajar. PAU untuk peningkatan dan Pengembangan Aktivitas Instruksional Ditjen Dikti, DepDiknas, Jakarta,
[3] Mardapi, Djemari., Teknik penyusunan instrumen tes dan non-tes. Yogyakarta: Mitra Cendikia Press, 2008.
[4] Anderson, L. W., & Krathwohl, D. R. A taxonomy for learning, teaching, and assessing. New York: Addison Wesley Longman. Inc, 28-32, 2011.
[5] Khoshsima, H. and Toroujeni, S. M. H. (2017). Computer-Based Testing: Score Equivalence and Testing Administration Mode Preference in a Comparative Evaluation Study. International Journal of Emerging Technologies in Learning (iJET), 12(10): 35-55. https://doi.org/10.3991/ijet.v12i10.6875
[6] Georgiadou, E., Triantafillou, E., and Economides, A. A. (2006). Evaluation parameters for computer-adaptive testing. British Journal of Educational Technology, 37(2): 261-278. https://doi.org/10.1111/j.1467-8535.2005.00525.x
[7] Chee, T.S., & Wong, A.F.L. Teaching and learning with technology. Singapore: Prentice Hall, 96, 2003.
[8] Towndrow, P.A., & Vallence, M., Using IT in the language classroom: A guide for teachers and students in Asia (3rd ed.), Singapore: Longman Pearson Education South Asia Pte. Ltd, 2004.
[9] Thompson, A. N.,& Weiss, D. J, "A framework for the development of computerized Adaptive Tests," Practical Assesment, Research & Evaluation, vol 16, pp. 49-65, 2011.
[10] Trianfillou, E., Georgiadou,E., & Economides, A.A., "CAT-MD : computer adaptive test on mobile devides. Current Development in Technology-Assisted Education, No 2, pp. 773-777, 2006
[11] Haryanto, "Optimasi Pengukuran Hasil Belajar melalui computerized intelligent measurement model fuzzy", in Seminar Nasional Psikologi UMS, Solo, 2014, p 316-331.
[12] Retnawati, Teori Respon Butir dan Penerapannya, Yogyakarta, Parama Publishing, 2014.
[13] Awopeju, O.A.,& Afolabi, E.R.I, "Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination", European Scientific Journal, Vol. 12, No. 28 pp. 263-284, 2016. https://dx.doi.org/10.19044/esj.2016.v12n28p263
[14] Zoghi, M.,& Valipour, V, "A comparative study of classical test theory and item response theory in estimating test item parameters in a linguistics test", Indian Journal of Fundamental and Applied Life Science, Vol 4, pp. 424-435, 2014.
[15] Guler, N., Uyanik, G.K., & Teker, G.T., "Comparison of classical test theory and item response theory in terms of item parameters", European Journal of Research on Education, Vol 2, No 1, pp. 1-6, 2014
[16] Ratnaningsih, D.J. and Isfarudi, I., "Analisis butir tes objektif ujian akhir semester mahasiswa universitas terbuka berdasarkan teori tes modern", Jurnal Pendidikan Terbuka Dan Jarak Jauh. Vol 14, no 2 , pp. 98-109, Aug 2013
[17] Abedalaziz, N.,& Leng, C.H., "The relationship between CTT and IRT approaches in analyzing item characteristics", The Malaysian Online Journal of Education Science, Vol 1, no 1, pp. 64-70, 2013.
[18] Thorpe, G.L.,& Favia, A., "Data analysis using item response theory methodology: an instroduction to selected programs and applications", Psychology Faculty Scholarship, July 2012, pp. 1-33.
[19] DeMars, C., Item response theory: Understanding statistics measurement. New York: Oxford University Press, Inc, 2010.
[20] Adedoyin, O.O.,& Mokobi, T., "Using IRTpsychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items", International Journal of Asian Social Science, Vol *3,* no 4, pp. 992-1011, 2013.
[21] Crocker, L.,& Algina, J, Introduction to classical and modern test theory, New York: Holt, Reinhart, and Winston, Inc, 2008.
[22] Hambleton, R.K., Swaminathan, H., Rogers, H.J., Fundamental of Item Response Theory, Newbury Park, CA: Sage Publication Inc, 1991.
[23] Hambleton, R.K., & Swaminathan, H., Items response theory:principles and application, Boston: Kluwer-Nijjhoff Publish, 1985.

DOI: 10.24014/COREIT.V9I2.27258

[24] Wainer, H., et al., Computerized adaptive testing: Aprimer, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1990.
[25] Wainer H., Bradlow, E.T., & Xiaohui Wang, Testlet response theory and its applications. Cambridge: Cambridge University Press, 2007.
[26] Weiss, D.J, "Computerized Adaptive testing for Effective and Efficient Measurement in Counseling and Education", Measurement and Evaluation in Conseling and Development, Department of Psichology, University of Minnesota, Minneapolis, Vol. 37, pp. 70-84, 2004.
[27] Yan J., Ryan M., & Power J., Using fuzzy logic, Englewood Cliffs: Prentice Hall Inc, 1994.
[28] van der Linden, W.J., & Hambleton, R.K., Handbook of modern item response theory. New York: Springer Verlag, 1997.
[29] Lord, F.M., Applications of item response theory to practical testing problems, Hillsdale, NJ : Lawrence Erlbaum Associates, Publishers, 1980.
[30] Afifi, A.A. And V. Clark, Computer-Aided Multivariate Analysis. 3 rd Edition. Chapman And Hall/CRC, New York. 1996.
[31] Dillon W.R dan Goldstein, Mathew, Multivariate Analysis Method and Application, John Wiley and Sons, New York, 1984.
[32] Olson, J.B., & Bunderson, C.V., "Validity and decision issues in selecting a CAT measurement model Paper presented at the CAT model and monitoring paper session", Juni 2007. http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/cat07 olsen.pdf
[33] Cella, D., Gershon, R., Lai, J. S., & Choi, S. (2007). The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*, *16*(SUPPL. 1), 133–141. https://doi.org/10.1007/s11136-007-9204-6
[34] Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., Kramer, J., Ni, P., … Ludlow, L. H. (2011). Accuracy and precision of the Pediatric Evaluation of Disability Inventory computer-adaptive tests (PEDI-CAT). *Developmental Medicine and Child Neurology*, *53*(12), 1100–1106. https://doi.org/10.1111/j.1469-8749.2011.04107.x

DOI: 10.24014/COREIT.V9I2.27258