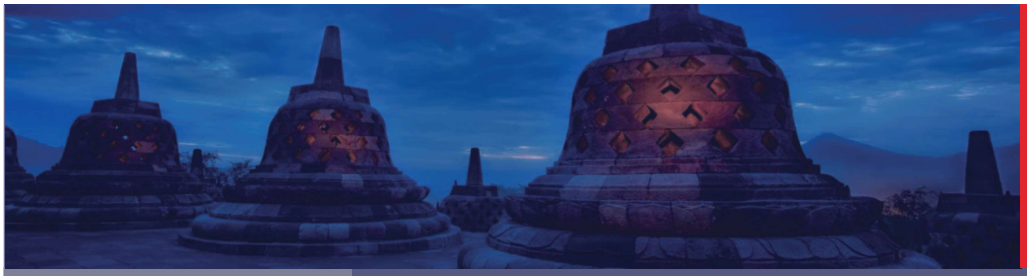

ISSN: 2407-1501

PROCEEDING

INTERNATIONAL CONFERENCE ON EDUCATIONAL RESEARCH AND EVALUATION (ICERE)

"Assessment for Improving Students' Performance"

May 29 – 31 2016
Rectorate Hall and Graduate School
Yogyakarta State University
Indonesia

Organized by:
Study Program of Educational Research and Evaluation
Graduate School, Yogyakarta State University
in Cooperation with Indonesian Educational Evaluation Association (HEPI),
and Center for Educational Assessment (PUSPENDIK) Ministry of Education and Culture

Proceeding
International Conference on Educational Research and Evaluation (ICERE) 2016

Publishing Institute
Yogyakarta State University

Director of Publication
Prof. Djemari Mardapi, Ph.D.

Address
Yogyakarta State University
ISSN: 2407-1501

Foreword of the Chairman

Assalamualaikum wr. wb.

Good morning ladies and gentlemen.

Praise be to Allah who has given abundant blessings so that we can hold this international conference.

This conference is aimed at improving the quality of assessment implemented in schools and other institutions. The quality of assessment determines students' ways of learning, so that it is hoped that the quality of education improves. Besides, this conference is a means of information exchanges in the forms of seminars dealing with results of research in educational assessment and evaluation. The expectation is that there is always improvement in educational assessment and evaluation methods, including in it is the instrument – both cognitive and noncognitive instruments.

The participants of this conference are the lecturers and teachers who teach educational assessment and evaluation, practitioners of assessment and evaluation, and researchers of assessment and evaluation. This conference can be held in cooperation with the Graduate School, Yogyakarta State University, Association of Educational Evaluation of Indonesia (HEPI), and Centre for Educational Research, Ministry of Education and Culture of Indonesia, supported by the Australian Council for Educational Research (ACER), Intel, Intan Pariwara Publisher, and many other institutions. For this reason, on behalf of the Organizing Committee, I would like to thank the Rector of Yogyakarta State University, Prof. Dr. Rochmat Wahab, M.Pd., M.A., and the Director of Graduate School, Yogyakarta State University, Prof. Dr. Zuhdan Kun Prasetyo, M.Ed., and all other institutions for their assistance and contribution that have made this conference possible. I would like to thank HEPI's Local Coordination Unit and all sponsors for supporting this conference and also all the audience for participating in this conference.

To the committee members, both in Jakarta and Yogyakarta, I would like to thank them for the hard work they have performed and for the togetherness so that this conference can be held.

Last but not least, we apologize for all the incoveniences you might encounter during this conference. Please enjoy the conference.

Wassalamu'alaikum wr. wb.

Prof. Djemari Mardapi, Ph.D.

Foreword of the Chairman of Himpunan Evaluasi Pendidikan Indonesia (HEPI)

Assalamu'alaikum Wr. Wb.

Indonesian Association for Educational Evaluation (HEPI) is a professional organization in education holding in the high esteem the principles of professionalism and knowledge development in the field of educational and psychological measurement, assessment, and evaluation. HEPI was established in November 19, 2000 in Yogyakarta, with a vision to become a professional organization that excels in the field of evaluation and measurement in education and psychology in Indonesia. Its mission is to develop up-to-date methodologies of evaluation, assessment, measurement, and data analysis in education and psychology, as well as studies of policies and technical implementation of the field for improving Indonesian education quality.

As a professional organization, HEPI brings together experts, practitioners and interested persons in the field of evaluation, assessment, and measurement of education, psychology and other social sciences. HEPI is open to anyone who has the interest the field with no restriction in terms of educational background and working experiences. Hopefully, through HEPI, members of the association can sustainably develop themselves as professionals. The existence of HEPI is also expected to contribute to the improvement of the quality of national education through research, consultancy, seminar, conference, publication, and training for members of the organization and for public audiences.

HEPI organizes annual workshop and conference in cooperation with the Regional Chapter of HEPI and universities. In 2016, for the first time HEPI organized **International Conference on Educational Research and Evaluation: Assessment for Improving Student's Performance** in May 29-30 2016 in Yogyakarta. This conference is jointly organized by HEPI and Yogyakarta State University and supported by the Center for Educational Assessment the Ministry of Education and Culture, Australian Council for Educational Research (ACER), INTEL Indonesia, and Intan Pariwara Publisher.

It is important to note that the choice of the HEPI 2016 conference theme is driven by the fact that the quality of our national education is still under expectation as shown by the results from School National Exam and international surveys conducted by some international agencies. HEPI believes that a number of factors contribute to the low quality of national education, including low teacher's knowledge and skills in classroom and school assessment. Therefore, improving the competence of teachers in classroom and school assessment is urgently required. In this context HEPI as a professional organization and individual members of the organization have to play an active role in improving teachers' competence in quality learning assessment.

In line with 2016 conference theme, HEPI invited two respected guest speakers , namely, Professor Geofferey Masters, Ph.D., Director of the Australian Council for Educational Research (ACER), who presented a paper on Assessment to Improve Student Competency and Professor Frederick Leung, Ph.D., from the University of Hong Kong, who delivered a paper on the International Assessment for Improving Classroom Assessment.

As a tradition, in 2016 conference HEPI organized two pre-conference workshops. The first workshop is on the conceptual introduction of Rasch model by Jahja Umar, Ph.D., senior lecturer at the Faculty of Psychology, State Islamic University Jakarta and the second workshop was delivered by Heru Widiatmo, Ph.D., researcher at American College Testing (ACT) Iowa, United States on Measuring Higher Order Thinking Skills (HOTS).

On behalf of HEPI, I would like to express my heartfelt gratitude to Rector of the Yogyakarta State University, invited speakers, resource persons, HEPI regional chapters, sponsors, speakers, participants, invited guests, and organizing committee who have worked hard in making this international coneference a success. Thank you very much for your participation and support and we are looking forward to seeing you in the next conference.

Last but not least, we hope that all of us get much benefit from this conference for enhancing Indonesian quality education through quality assessment.

Wassalamualaikum wr. wb.

Chairman,
BAHRUL HAYAT, Ph.D.

Table of Contents

Analisys Item Information Function on the Test of Mathematics

Rukli
Muhammadiyah University of Makassar
Hp. 085299135009
Email: a_rukli@yahoo.co.id

Abstract - This paper aims to reveal: 1) the items fitted the three parameter logistic model; 2) the item information function on the parameters of discrimination the test of mathematics; 3) the item information function on the difficulty level of the test of mathematics; (4) the item information function on the guessing chance level of the test of mathematics. The research data using examinee responses on national final examination of the math. Measurement model using three-parameter logistic model. Analysis of the response data using ASCAL program. The results showed that: 1) there are about 25 of the 35 items fitted the three parameter logistic model; 2) the function of each information about the smaller 1 except item 15; 3) the value of the function of discrimination parameter information the highest and lowest, respectively 816,226 and 1,026; 4) value function parameter information difficulty level is above 100 by 46%, while below 100 by 54%; 5) value function parameter information guessing chance level are in the top 50 as much as 20%, while under 50 as much as 80%.

Keywords*: parameter,* item, information, *logistic*

I. INTRODUCTION

System of entrance exams and final exams at all levels of education are still testing both the shape and dynamics of the implementation according to community needs. This regard, would require a data and processes in order to maximize results in accordance reference environment. Data, processes, and the environment is a factor reference inseparable from the quality of the output of a system including the system of evaluation.

Now, the evaluation system in Indonesia of college entrance examination system can be divided into three namely SNMPTN, SBMPTN, and Independent. Entrance exam of basic education and secondary education until now has not been of special concern to the government except for the final exam. The exam has undergone a metamorphosis, but sometimes only a change of form, but the essential problem is sometimes secondary. In fact, these changes are sometimes too abstract and theoretical political over curriculum changes.

Curriculum changes inseparable variation example Ebtanas final exam and the UN in order to test the system more credible, relevant and contemporary. The third can be used as a reference rate the system a good test. Credible and relevant to the present forecast accuracy while associated with the speed and quality of information the test results to the user.

Linkages credible and relevant, the core evaluation system lies in the prediction problem. Furthermore, the prediction accuracy of estimating the related problem of latent abilities, which revolves around the problem of how do so predictions are not far away or right on the measuring object. Sumadi Suryabrata (1989) suggested that the problem of prediction generally revolve around the issue of how pressing kemelesetan become smaller.

The ability of the test examinees unique and latent that assessment not free from error. The ability to be in a continuous variable that needs to be defined operationally. Response test examinees misses the most extreme enough mistake to mistake normal estimates. It concerns the coefficient of reliability of the test.

Function test information in item response theory has a similarity coefficient reliabiltas tests. However, the reliability of the test has limitations compared to the test information function.

Samejima (Hulin et. al., 1983) suggests reliabiltas is a dead concept theoretically because it differs from one group against another, generalizations are limited and very narrow. Function test information is an accumulation function item information. Accumulation is not found in reliability because there is no information about the reliability of the item. It shows one disadvantage compared to classical test theory item response theory.

Item information functions play a role in the development and selection of appropriate test items informative ability of the test taker. That is, with the test item information, it can be seen how much the items to uncover the optimal parameters of ability and capability where the item could be used because the estimator.

Furthermore, the value of the item information functions should not be constant on the proficiency scale. Information on the value of the ability depends on the number and attributes of the items used to assess ability. For example, the low value of the function information concerning the ability of the test taker at a high skill level given the item easily, but the item will elaborate or to inform a number of things substasial at low skill levels. How big are the different power parameters, the parameters of difficulties, and parameter guesses opportunities simultaneously affect the value of the function of each item of information about the particular characteristics of the parameters on the math test item nasionoal exam? Characteristics of test items require in-depth study of descriptive and third parameters of test items.

A. Three Parameter Logistic Model

Three- parameter logistic model in IRT has some logistics model that is one parameter logistic model, two- parameter logistic model, and a three-parameter logistic model. Furthermore, it has developed a four parameter logistic model where the fourth parameter is the parameter of time but has not been implemented optimally (Hamblenton, 1989). Three-parameter logistic model has three parameters, namely the different power parameters item, item difficulty levels, and opportunities guesses items. These three parameters are called independent variables. These three parameters either individually or simultaneously affect the ability parameters as the dependent variable. Equation three- parameter logistic model as follows.

$$P_i(\theta) = c_i + (1 - c_i)\frac{\exp(Da_i(\theta - b_i))}{1 + \exp(Da_i(\theta - b_i))} \quad \text{..... (1)}$$

Different power items are variables in the three parameter logistic model. These variables affect the variable capability. In the item characteristic curve, different power expressed by the tendency of a tangent curve at the inflection point of the curve, ie the inclination offensive line curve at the inflection point. The greater the value the greater the inclination to form different power curve is getting steeper. In theory, the difference between the power value range of negative infinity to positive infinity, but in general the practice of different power values are ranging from 0 to 2.

The level of difficulty of items to note the abscissa at an inflection point where the item characteristic curve inflection point is a point on the curve with the ordinate (1 + ci)/2. In theory, the range of item difficulty level between negative infinity to positive infinity but in practice, in general, the level of difficulty ranges from negative three to positive three. Difficulty level value of zero to the right more difficult, otherwise the value of the difficulty level from zero to the left is getting easier.

Opportunities guesses items expressed high ordinate at the point of intersection between the axis of the $P_i(\theta)$ item characteristic curve. These parameters are called pseudo-chance level, which is to claim much chance the correct answer by the examinees enabled an item is lower than the level of difficulty of the items.

$P_i(\theta)$ is an opportunity to test examinees θ answered correctly item-level capabilities to-i. D is a constant scaling factor with a value of 1.7.

B. Function information item

Three- parameter logistic model derived, obtained an information function Information function according to three parameters logistic model can be formulated as follows.

$$I_i(\theta) = D^2 a^2 \left(\frac{Q(\theta)}{P(\theta)} \right) \frac{(P(\theta) - c_i)^2}{(1 - c_i)^2} \quad \text{.....(2)}$$

Based on the equation 2, Hambleton and Swaminathan (1985) suggests the nature of the function information item, ie the value of the information function maximum is reached when θ it is slightly larger in value than the value of b, the value of the information function increase if the a value is growing, and the value of the information function increase if the value of c become smaller .

The function of the maximum information is obtained when $\theta = \theta_i$ where

$$\theta_i = b_i + \frac{1}{Da_i} \ln \frac{1 + \sqrt{1 + 8c_i}}{2} \quad \text{..... (3)}$$

Maximum function item information is an indication that the item matches the ability of the test taker. The way it is often used to establish adaptive tests. According to Hambleton, et al. (1991) and Lord (1980) adaptive test is a test of the level of difficulty of matching items with the level of ability of the test taker. Nevertheless, the model test as many constrained in terms of the network without wires so much in the testing phase in addition to understanding the test still lay so socialization is very important to the user (Rukli, 2010 and 2012).

Furthermore, the function item information item is a strength measuring the ability of the test taker. It plays a role in assessing the condition of an item is functioning optimally or not the maximum. This becomes further reference to the item worthy or not worthy entry in the test mainly tests nationwide.

II. RESEARCH METHODS

A. Types of research

This research uses descriptive exploratory study. The goal is that the characteristics or properties of one or more variables can be described. Characteristic parameters of the items comprising different power item, item difficulty levels, and opportunities guesses item will be described. Furthermore, these characteristics be considered to determine the function of the information.

B. Object of research

The research object is an item of national exams in mathematics consisting of 3 (three) package which package 1 from 8 rayon , package 2 from the 7 rayon , and package 3 from 8 package.

C. Item analysis

Analysis items using ASCAL to produce an estimate item parameters, namely the different power parameters of items (a) , the parameter level of difficulty of items (b), and the parameters opportunities guesses item (c) and Chi-squared statistic for each item. Estimated parameters of items will be stopped when the difference between two consecutive estimates of less than 0001 and said to be estimated accurately item parameter estimates between iterations so-called convergent estimates. Generally accurate assessment takes 5 to 10 rounds, but this study uses the number of iterations as many as 20 rounds.

The results of this analysis also result in a match against the item response models. Furthermore, the calculation of each parameter item information functions using Visual Basic.

D. Function Item Information

Calculation function information item using the item parameter different power parameters, level of difficulty, and opportunities guesses. The function calculation procedure has several stages of the test examinees response data to making programs with the following details: (1) Data UN test examinees' responses obtained from the database in the area of the object of study. (2) Data response cleared of disability data. (3) The data processing using ASCAL program. (4) 4. Examination of model fit every item on the response data. (5) 5. Examination matches the second stage until all items meet models. (6) 6. Tabulates and mendeskrifsikan each parameter item. (7) Make a program to compute the function parameter information item.

III. RESEARCH RESULT

A. Estimated Parameter Item

UN test item data analysis showed round stop on lap 15 of the 20 rounds are maximized, meaning that at the round reached a level of accuracy estimates the highest among the existing lap . Round stop on lap 15 shown in Figure 1 wherein the difference in lap 14 and lap 15 smaller 0.001.

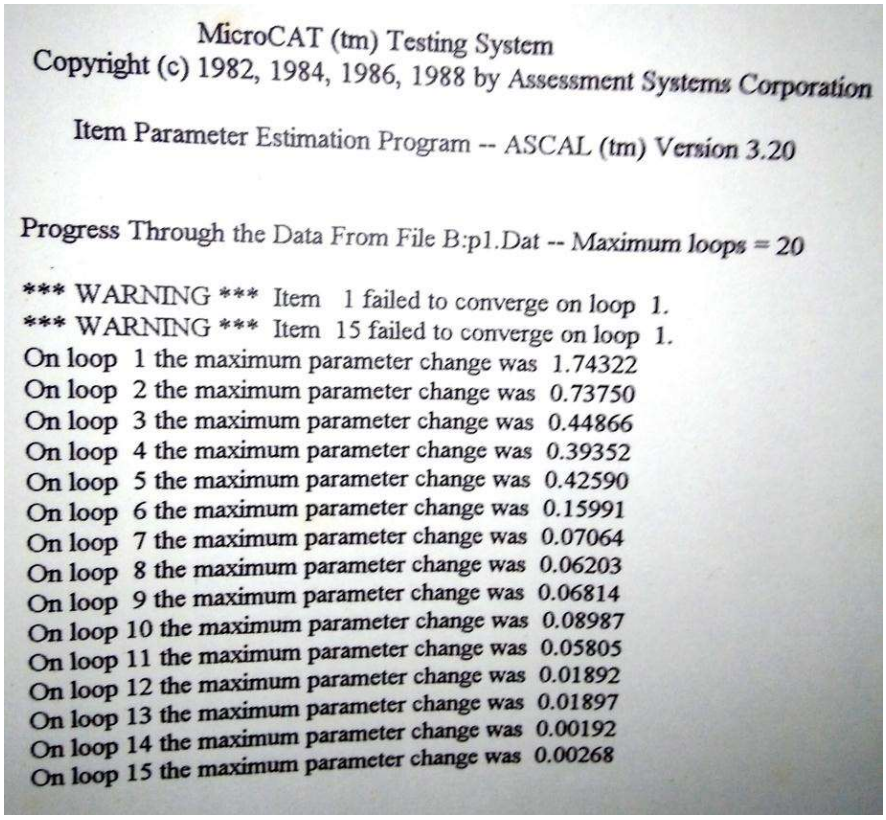

Figure 1. The Print Screen Total Estimated Stop Round

Parameter estimation different power item, item difficulty levels, and opportunities guesses items on lap 15 can be seen in Table 1.

Table 1. Item Parameter Estimates Three Parameter Logistic Model

Item number	Parameter Estimates			Item number	Parameter Estimates		
	(*a*)	(*b*)	(*c*)		(*a*)	(*b*)	(*c*)
1	1.110	3.000	0.200	19	1.008	1.862	0.170
2	0.583	2.700	0.010	20	0.720	0.146	0.000
3	0.691	0.674	0.000	21	0.682	1.467	0.100
4	0.695	0.590	0.070	22	0.497	1.373	0.210
5	0.681	0.550	0.000	23	0.793	1.463	0.150
6	0.400	0.613	0.360	24	0.888	2.267	0.110
7	0.474	0.916	0.170	25	0.659	2.105	0.100
8	0.467	1.582	0.080	26	0.778	2.296	0.150
9	0.774	0.445	0.000	27	0.558	-0.050	0.020
10	0.762	-0.044	0.010	28	0.587	1.446	0.050
11	0.950	1.759	0.210	29	0.692	2.777	0.270
12	0.618	1.687	0.110	30	0.466	1.425	0.020
13	0.4000	0.272	0.270	31	0.895	3.000	2.290
14	1.103	2.745	0.250	32	0.747	2.400	0.350
15	1.571	2.723	0.260	33	0.472	2.936	0.280
16	0.664	1.791	0.230	34	0.416	2.225	0.030
17	1.055	1.027	0.070	35	0.971	2.962	0.120
18	0.944	1.249	0.140				

Table 1 shows that the characteristics of the item has a different power greater than 1 there are five items or 14 %, ie items 1, 14, 15, 17, and 19. The items have different power less than 1 there were 20 or 86 %. Characteristics of the items have no negative different power while the item has a difficulty level equal to 3 No 2 item or 6 %, ie items 1 and 31. Further characteristics of the item has a difficulty level of negative 2 item or 6 %, ie, items 10 and 27. Characteristics items have greater opportunities than 0.300 guess there are 2 or 6 %, ie, items 6 and 32.

B. Item Fit Three Parameter Logistic Model

Compatibility between the data from the work of test takers and the theoretical value of the grain refers to the suitability of the items on the three parameter logistic model. If Chi- Squared value of an item is greater than the critical value in the model, which is 27.6 then the item is not compatible with the model. Conversely, if the value of an item Khi-Square is less than or equal to the critical value, which is 27.6 then the item fits with the model.

Number of items meet the logistic model three parameters as many as 25 of the 35 items, or 71%, namely, 1, 4, 6, 7, 8, 11, 12, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25 , 26, 28, 29, 31, 32, 33, 34, and 35. While the item does not meet the model as much as 10 or 29%, ie, items 2, 3, 5, 7, 9, 10, 13, 16, 20, 27 , and 30.

Incompatibility between data items with the model likely to be caused in terms of material, construction, and the language of the item. For example in terms of language, lack of clarity of the revelation by the symbol given in the item. Examples on item 2, pernyaataan "complement (A∩B)" will confuse the test taker. The statement requires the test taker to understand the meaning of meaning, complement, and sliced, where the two concepts requires different thinking.

The item does not match the model is not working properly to uncover capabilities. An item does not work properly due to the item usual bias or the ability of test takers do not fit the participant's ability to respond to the test items, such as too difficult or too easy.

The items do not match the model can be used for other purposes. Sumadi Suryabrata (1987) revealed that if statistically a characteristic item is not feasible then the item is returned to the test objectives. Furthermore, Hambleton et. al., (1991) reveals that the item does not match the three-parameter logistic model may be matched with two parameters logistic model that only has two different parameters such as power parameters item and item difficulty level. Or may be incompatible with the model of the parameters that only pay attention to the items based on level of difficulty alone.

If you notice an item has a high chance guesses there are 4 items located at the end of the test, namely items 29, 31, 32, and 33. To answer a complicated math test item takes about 4.5 minutes (Sumadi Suryabrata, 1987), while the UN test mathematics SMP 120 minutes with 40 items. Thus, such tests including speed test. According to the research Adkin and Malenkopt in different years (Sumadi Suryabrata, 1987) showed no untoward effect on the estimated parameters of the items to be taken at the end of the test if the processing time narrow test.

Based on this, most likely takers have not tried to do all of the items on the test time is up so that the test taker to guess. That is, the items are less than helpful in revealing the ability of the test taker. However, further analysis of the matter was not done in accordance limitation of this study is a quantitative analysis. Therefore, it can be concluded that the item is an item fits models have different power high, high level of difficulty, and the opportunities and guesses were high among 35 items.

C. Function Parameters Information Item

Item meets the models analyzed to calculate the value of the item information functions. Function information item only performed on the items meet the logistic model. The function of each parameter item of information contained in Table 2.

Table 2. Function Item Information

Item number	Function Item Information			Item number	Function Item Information		
	I a	Ib	Ic		I a	Ib	Ic
1	4.474	7.617	1948.87	4	149,089	106.616	1177.80
6	113.891	21.319	785.71	7	201.499	40.215	785.71
8	406.51	37.987	1619.81	11	34.378	31.759	2236.85
12	211.821	40.299	2127.83	14	4.767	8.524	2499.39
15	1.026	10.231	2521.26.	17	76.69	125.042	2909.29
18	67.072	69.666	2226.09	19	29.993	30.619	2704.69
21	181.553	51.415	2197.02	22	213.669	29.123	1280.83
23	96.267	45.944	2151.69	24	44.984	20.977	3831.41
25	170.336	26.54	30.11.95	26	60.155	16.647	2927.84
28	329.488	54.877	2279.84	29	35.665	5.692	2182.22
31	6.505	3.005	2299.29	32	26.17	6.607	1901.75
33	152.968	5.44	1973.11	34	816.226	28.174	2968.80
35	10.318	4.438	4319.35				

Table 2 shows that the value of the power parameter information functions depending highest item is item 34, the items have value power parameter information functions depending smallest item is item 15. The items have value parameter information functions difficulty level above 100 there are 11 or 44%, while the items have value parameter information functions difficulty level below 100 there are 14 or 56%. The items have value information function parameter guesses opportunities in the top 50 there are 5 or 20%, while the items have value information function parameter guesses opportunities under 50 there are 20 or 80% .

The correlation coefficient between the estimated value of different power parameters and parameter information function value of -0.623. According to Hambleton and Swaminathan (1985) the tendency of the correlation varies by level of ability. Low skill levels of negative tendencies, whereas a high ability level positive tendency. Furthermore , the correlation coefficient between the estimated value of the parameter level of difficulty and the value of the function parameter information sebasar -0761 and the correlation coefficient between the estimated value parameter guesses opportunities and the value of the information function parameters sebasar -0343

Furthermore, the value of r - table with n = 25 obtained 0396 . That is, the correlation between the correlation between the estimated value of the item parameter and function value information item only significant parameter in different power and level of difficulty. Not significant parameter estimated value of the item and the value of the function 's parameter information item guesses opportunities occur because the opportunity is less consistent guesses matching using ad hoc methods. Hulin et. al., (1983) stated that the estimated parameters chance guesses are usually suitable when using a method that involves consideration of researchers in interpreting the pattern of test-takers answer . The way it is difficult to do even can be said can not be done when national exams take place.

IV. CONCLUSION

Items that match the three-parameter logistic model as many as 25 of the 35 items, or 71%. Value function power parameter information depending highest item is item 34, the items have value different power parameter information functions smallest item is item 15. The items have value parameter information functions difficulty level above 100 there are 11 or 44%, while the items have value parameter information functions difficulty level below 100 there are 14 or 56%. The items have value information function parameter guesses opportunities in the top 50 there are 5 or 20%, while the items have value information function parameter guesses opportunities under 50 there are 20 or 80%. The functions of the items are all less infromasi 1, except item 15 has a value greater item information function 1. The low value of the item information functions is different due to the low power, level of difficulty is higher when compared to the ability of the test taker, and guess high enough opportunities.

REFERENCES

[1] Sumadi Suryabrata, ”Seleksi calon mahasiswa baru: perguruan tinggi yang sekarang dan kemungkinan untuk masa yang akan dating,” Yogyakarta: Andi Offset Yogyakarta, 1989.

[2] Hulin, H. L., Drasgow, F., and Parsons, C. K., ”Item response theory. Application to psychological measurement,” Homewood, IL: Thorsey Professional Series, 1983.

[3] Hambleton, R. K.,"*Principles and selected applications of item response theory," in R. L. Linn (Eds.), Educational Measurement, 1989,*. Macmillan, New York, *pp.147-200.*

[4] Hambleton, R. K. dan Swaminathan, H., "Item response theory, " Boston, MA: Kluwer, 1985.

[5] Hambleton, R. K., Swaminathan, H., & Rogers, H. J., "*Fundamentals of item response theory*," CA: Sage, Newbury Park, 1991.

[6] Lord, F. M., "Application of item response theory to practical testing problems," Hillsdale, NJ: Lowrence Erbaum Associates, 1980.

[7] Rukli, "Penerapan model logistik satu parameter pada *Computerized Adaptive Testing* Ujian Masuk Perguruan Tinggi;" Prosiding Seminar Nasional Himpunan Evaluasi dan Pendidikan Indonesia, Jakarta. ISBN 978-602-96343-0-3, 2010.

[8] Rukli, "Simulasi Pola Inisialisasi Kemampuan dengan Metode Penalaran *Fuzzy* Tsukamoto dalam Menidentifikasi *Cheating*," PPS Universitas Negeri Yogyakarta dan HEPI Yogyakarta, Yogyakarta. 978-602-96172-8-3. 2012.

[9] Sumadi Suryabrata, "Pengembangan tes hasil belajar," Jakarta: Rajawali Press, 1987.