International Conference on Mathematics and Science Education
of Universitas Pendidikan Indonesia
Volume 3, 2018 | P-ISSN 2655-2361, E-ISSN 2655-3252

Interface web development for analysis of item response theory with mixed model approach and application on bank soal MGMP

T C P Utama[13*], I Made Tirta[1] and M Fatekurrahman[12]**

[1]Department of Mathematics, University of Jember
[2]Laboratory Statistics Department of Mathematics, University of Jember
[3]Senior High School PGRI 5 Tegaldlimo, Banyuwangi

[*]**clarinta.utama@gmail.com**,
[**]**itirta.fmipa@unej.ac.id**

Abstract.The development of the world of education today is progressing very rapidly, the era of technology is increasingly modern, so that teachers must have adequate competence in every process of teaching and learning activities. One type of measurement often done in education are measurement of the students' performance both for cognitive and effective aspects. These measures are extremely important therefore must use a good measuring tool and the results also easy to interpret. The measurement of students performance mostly use tests. Items response theory have evolved from traditional one to modern theories to apply more realistic models which are known as item response theory. However the use of modern test theory much rely on availability of the computer software. In this paper we report the development of a web-GUI interface that can be used to analyze polytomous responses, using Hierarchical Generalized Linear Models which will also contains theories and interpretations of the results. This web-GUI interface is expected to help teachers to understand and to do the analysis of polytomous responses more easily.

1. Introduction

The development of the world of education today is progressing very rapidly, so that teachers must have adequate competence in every process of teaching and learning activities. Therefore the teacher is required to develop on himself so as to be able to carry out the teaching and learning process that is appropriate and varied from each teaching material. By having sufficient competence then a teacher is able to teach and evaluate well. It is not easy, therefore the teachers have a container called MGMP (Subject Teachers' Meeting) to help develop the existing teacher's kopentensi is no exception in Banyuwangi District. The MGMP function here is to accommodate teachers 'aspirations and teachers' difficulties in preparing and checking or testing good test items. Because the test item is very important as one measure of student ability.

Preparation of tests performed starting from the design, assembly or manufacture and analysis of test items using various approaches, either using classical test theory or modern theory known as the grain response theory. Development of the test with the grain response theory approach has many advantages. The grain response theory model yields independent grain parameters of the test participants and test-participant parameters independent of the test set (Xitao, 1998). Since the logistics model is based on the logistical response function, and random effects are assumed to be normally distributed in this closely modeled model often referred to as normal logistic models, especially in latent model literature. But there is not much available software to analyze the problem utilizing the grain response theory. On the other hand there is open source software R one of them by making the web interface available freely, but the way its use is relatively difficult because based on programming or script, therefore needed additional program which can make analysis of grain response easy to do with R. Cara this has been done by some researchers for other needs by utilizing R-shiny. (Tirta, 2015)

From some of the above explanations there needs to be software development to measure items with IRT that are easy to use and apply them to analyze the grain of mathematics problem bank created by MGMP team and renewal in bank question test made by MGMP in Banyuwangi. Given the importance of testing the test questions of the bank it needs updating by using Item Response Theory and presents it in the form of web interface.

In addition, the question bank test must have a reliable index. Tests are often used in the world of education das psychology to get the data you want to know from certain individuals or groups, such as selection of taking individuals who have the desired criteria. Can also sekagai profile person to know where the individual is able to position himself in a work. As a matter of judgment, it is important to continuously develop to see the progress or potential of a person in a particular field. For that the presentation of quality questionnaire must be valid, objective and efficient.

1.1 Classical Test Theory

The test is a measurement technique designed as a procedure. We propose a Response Theory Item (IRT) approach because it assesses some of the psychometric features of individual item items. Compared, Classical Test Theory (CTT) emphasizes on the combined score sekala.IRT is a non-linear probabilistic modeling technique to develop and evaluate the psychological measurements of the scales. For example, it can be assumed that scale items are designed to assess certain psychological attributes (eg, perception ability) that support higher values on items indicating a strong psychological basis (eg, stronger perceptions of security). If the respondent provides non-discriminatory support for an item when they are indeed different in terms of the underlying psychological attributes, the item should be considered inappropriate as a psychological measure of the attribute. For this purpose, the IRT calculates the respondent's probability of supporting a particular response option of each item and the approximate scale of each item's ability to distinguish respondents, which can be used to summarize strategically from a long psychological scal.

Classical theory is called Classical Test Theory (CTT), while modern test theory is called Item Response Theory (IRT). (Sudijono, 2001) that to measure classically there are programs based on Ms. Excel but has the disadvantages of the limited number of student data entry and item items that will be tested in a single implementation is only able or load 50 on each item item.

1.2 Modern Test Theory (IRT)

To overcome the weaknesses of the classical theory, the measurement experts sought to find alternatives. Modern test theory or commonly referred to as the grain theory (Item Response Theory) was developed by measurement experts in the field of education and psychology as an effort to minimize the deficiencies that exist in the classical test theory. Calculations in grain analysis based on this theory can be done using the help of web interface program. The grain response theory model yields independent grain parameters from the test participants and the test participant parameters that are independent of the set of grains tested (Xitao, 1998). Just as in classical theory, the grain response theory is also based on the basic postulate.

Opportunity students with $z_m$ability level answer correct item to i which have difficulty parameter b_i, discrimination parameter a_i, and guess parameter c_i. The function g is the link for the logistics model, ie logit or probit. For logit link.

$$g\left(ai(z_m - b_i)\right) = \frac{1}{1+e^{-a_i(z_m-b_i)}} \tag{2.6}$$

As is usually the case (model fitting there is a measure of the suitability of the model inspection model done by using the Akaike information criterion (AIC) which calculates the balance between the magnitude of the likelihood and the number of variables with the model The magnitude of AIC is calculated by the following formula.

$$AIC = -2l(\theta) + 2q, \qquad (2.7)$$

Good item criteria are given below (Anggreyani.2009)

The probability of answering correctly is in the range of 0 to 1 and this prevents data from being expressed as an interval scale. Raw sour generated from this way is difficult to declare as a scale. To overcome this problem, logistical transformations can be used, which involves the natural logarithm of odds.

$$\ln \frac{\theta_n}{\Delta_i} = \ln \left(\frac{P_{ni}}{1-P_{ni}}\right) \qquad (2.9)$$

which is worth of

$$\ln \theta_n = \ln \Delta_i = \ln \left(\frac{P_{ni}}{1-P_{ni}}\right) \qquad (2.10)$$

The form of equations is better known in the measurements for this model, which may be called the Rasch model (Hambleton, et al (1991)):

$$P_i\ (\theta) = \frac{e^{(\theta-b_i)}}{1-e^{(\theta-b_i)}} \text{ ,dengan i : 1,2,3,...}n \qquad (2.11)$$

$P_i\ (\theta)$: the probability of a test participant having θ ability randomly selected can answer item i correctly.
θ : level of subject ability (as independent variable)
b_i: difficulty index of the ith item
e : natural numbers whose value is close to 2,718
n : the number of items in the test

Approach Using HGML in IRT

In addition to traditional IRT methods for detecting DIF, some of the latest findings use alternative methods to detect DIF. HGML modeled hierarchical data. If the result is categorical data, such as nominal or ordinal data (Raudenbush and Bryk 2002). This model is an extension of the General Linier Model (GLM) to the multilevel data (McCullagh and Nelder 1989; Kamata 1998). The level-1 model in the two-tiered HGML consists of a sampling model, a function link, and a structural model. According to Raudenbush and Bryk (2002), Binomial sampling and logit links are used when the results are binary. Based on binomial distribution, expected value and Y_{ij}variance for level 1 The sampling model for model 2 can be written as follows:

$$E(Y_{ij}|\varphi_{ij}) = \varphi_{ij,}\, dan\, Var\ (Y_{ij}|\varphi_{ij}) = \varphi_{ij}\ (1 - \varphi_{ij,})$$

Where $\varphi_{ij,}$ is the probability of checking that gives the correct response to item *i*, level 1 of the logit link function in HGML can be written as follows:

$$\eta_{ij} = \log\left(\frac{\varphi_{ij,}}{1-\varphi_{ij,}}\right), \eta_{ij} = \log\left(\frac{\varphi_{ij,}}{1-\varphi_{ij,}}\right),$$

Where η_{ij} represents the log of the probability of responding to the correct response item *i*, it can take the true value. $\varphi_{ij,}$is limited to values between 0 and 1, because that is the probability. If $\varphi_{ij,}$ equals 0.5, the probability of the correct response is equal to 1, ie 0.5 / 0,5 = 1, and logit is 0, log (1) = 0. If $\varphi_{ij,}$is more small from 0.5, then the logit is negative; if $\varphi_{ij,}$ is greater than 0.5, then logit is positive. level 1

$$\eta_{ij} = \beta_{oj} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} = \beta_{0j} + \sum_{h1}^{k} \beta_{hj}X_{hij,}$$

Where $X_{hij,}$ is the dummy variable of the h - item indicator for teste *j*, with value 1 when $h = i$ and 0, when h ≠ *i*, for item i corresponding to the coefficient β hj, where a $h = 1,...,k$. $\beta_{0j}$is the prevention of structural equations. For the two-level HGLM model, which is algebraically similar to the Rasch model, the GLM and HLM frameworks can produce Rasch models (Kamata 2001, 2002).

2. Experimental Method

The research was conducted in 3 schools by spreading 50 randomly taken questions while web development was done in Statistics Laboratory of Jember University faculty of FMIPA by following the process of research and test of bank validation analysis about MGMP. Problems analyzed are part of the problem banks compiled by the MGMP Team including the authors themselves, covering the scope, type of questions and the number of questions. Sample problem in this research take sample problem from 150 problem of multiple choice, which owned by bank about MGMP Mathematics in Banyuwangi Regency. Problems used in the test of this study amounted to 50 multiple choice questions. Before performing the test the researcher gives the lattice of the problem first. The number of questions provided at each school institution was 50 multiple choices with 120 minutes, in each question consisting of 10 easy-weighted questions, 30 medium-weighted and 10 difficult-weighted questions. With the criteria each question has a fifth answer as much as 5 are: A, B, C, D and E.

The subjects in this study came from high school students in Banyuwangidan district conducted in 3 schools namely SMA PGRI Tegaldlimo, SMA NEGRI 1 Purwoharjo, and SMA NU Genteng. The data obtained in this research from different clusters for more details, population berklaster (classified) is the population of research subjects or research respondents whose members divided into, groups according to similarity characteristics, location, or certain conditions. As mentioned earlier, the "clusters" in this study population are:

a) Gender or class by sex of men and women,Group by status of state institutions and suwasta
b) Group according to parallel level XII IPA and,
c) Group by location located in Banyuwangi Regency

2.2 Preparation of Web Interface

The web interface is developed by using R-shiny in the format of the software (not the module) supported by the two main files ui.r (for programming interactive menus) and server.r to send all required commands to R (in server) . Capabilities / features covered include:

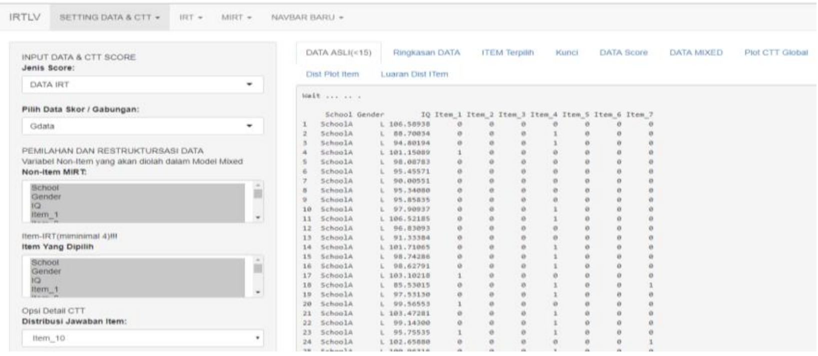
a.) Processing a rough score to a binary score
b.) Analysis of IRT (one-3 PL)
c) Multi Dimensional Analysis
d.) HGLM application in IRT

Article I. Usage

```
glmer(formula, data = NULL, family = gaussian, control =
glmerControl(),
      start = NULL, verbose = 0L, nAGQ = 1L, subset, weights,
na.action,
      offset, contrasts = NULL, mustart, etastart,
devFunOnly = FALSE, ...)
```

3.Result and Discussion

The main result of this research is web interface development for item response theory analysis with mixed model approach and its application to problem bank in Banyuwangi district with simulation of R-shiny program. In this study obtained results that aims to facilitate the teacher to analyze each item before the assessment of the test on the students. Educational assessment is not merely an assessment of learning outcomes, but covers a wider aspect of inputs or components, processes, products and educational programs. Therefore it needs a validation test tool that can be used by teachers and more efficient in the process pengerjakannya.



INPUT DATA & CTT SCORE
Jenis Score: DATA IRT

Pilih Data Skor / Gabungan: Gdata

PEMILAHAN DAN RESTRUKTURSASI DATA
Variabel Non-Item yang akan diolah dalam Model Mixed
Non-Item MIRT: School, Gender, IQ, Item_1

Item-IRT(miminimal 4)!!!
Item Yang Dipilih School, Gender, IQ, Item_1

Opsi Detail CTT
Distribusi Jawaban Item: Item_10

```
Wait ... .. .

   School Gender        IQ Item_1 Item_2 Item_3 Item_4 Item_5 Item_6 Item_7
1  SchoolA      L 106.58938      0      0      0      0      0      0      0
2  SchoolA      L  88.70034      0      0      0      1      0      0      0
3  SchoolA      L  94.80194      0      0      0      1      0      0      0
4  SchoolA      L 101.15089      1      0      0      0      0      0      0
5  SchoolA      L  98.08783      0      0      0      0      0      0      0
6  SchoolA      L  95.45571      0      0      0      0      0      0      0
7  SchoolA      L  90.00551      0      0      0      0      0      0      0
8  SchoolA      L  95.34080      0      0      0      0      0      0      0
9  SchoolA      L  95.85835      0      0      0      0      0      0      0
10 SchoolA      L  97.90937      0      0      0      1      0      0      0
11 SchoolA      L 106.52185      0      0      0      1      0      0      0
12 SchoolA      L  96.83093      0      0      0      0      0      0      0
13 SchoolA      L  91.33384      0      0      0      0      0      0      0
14 SchoolA      L 101.71065      0      0      0      1      0      0      0
15 SchoolA      L  98.74286      0      0      0      1      0      0      0
16 SchoolA      L  98.62791      0      0      0      1      0      0      0
17 SchoolA      L 103.10218      1      0      0      0      0      0      0
18 SchoolA      L  85.53015      0      0      0      1      0      0      1
19 SchoolA      L  97.53130      0      0      0      1      0      0      0
20 SchoolA      L  99.56553      1      0      0      0      0      0      0
21 SchoolA      L 103.47281      0      0      0      1      0      0      0
22 SchoolA      L  99.14300      0      0      0      1      0      0      0
23 SchoolA      L  95.75535      1      0      0      1      0      0      0
24 SchoolA      L 102.65880      0      0      0      0      0      0      1
```

Figure 1: Web interface

Table 1. Performance of the Ms.Excel Web Program

No.	No. Item	Statistic Item			Statistic Option					Tafsiran			
		Prop. Correct	Biser	Point Biser	Opt.	Prop. Endorsing	Biser	Point Biser	Key	Gussing	Difficlt	Efektifitas Option	Problem Status
1	1	0,640	-7,019	0,120	A	0,000	-	-		Can not distinguish	medium	good	Rejected / Do not Use
					B	0,220	-	-					
					C	0,640	-	-	#				
					D	0,140	-	-					
					E	0,000	-	-					
					?	0,000	-	-					

The total number of students who took the item item test of MGMP in Banyuwangi was 82 students consisting of 39 men and 43 women from 3 institutions of origin school ie SMA PGRI Tegaldlimo, SMA NEGRI 1 Purwoharjo and SMA Nu Genteng with total items about each work students numbered 50 multiple choice questions.

There are scores data containing school (Ids), gender (Gender), School Origin (Origin of School) and Indonesian Values (NilaiBhsIndo) and the value of the distribution of each item. In result of web interface analysis of IRT Model with 1 parameter with power of differentiator of constrain 1 it will generate value of AIC 5143.979, 1.702 is generated AIC 5218.159 value whereas without konstran yield AIC 5066.638 value. Thus, the smallest AIC values are obtained when the model is without constraint as the best model (meaning that each item other than having different levels of difficulty also has different discrimination levels) with their respective values. Discrimination ability of each question is the same because there is no significant influence.

Item Analysis Results.

a) Some items are not good because the level of guessing is high (indicated from students who are able to have a chance to answer true that is smaller than the students who can not afford).
b) In addition to illustrating the quality of the questions, the HGLM may also be considered other factors / variables that may influence the probability of answering the right questions (eg, Indonesian language skills, IQ and Gender) and at the same time looking at the extent to which problems are biased against a particular group (eg gender, tribe, or mother tongue, for example)

The ICC curve form which describes the shape of the characteristic question curve of each item is also presented through the graph (ICC). Item which is shown by graph resembling the letter S)

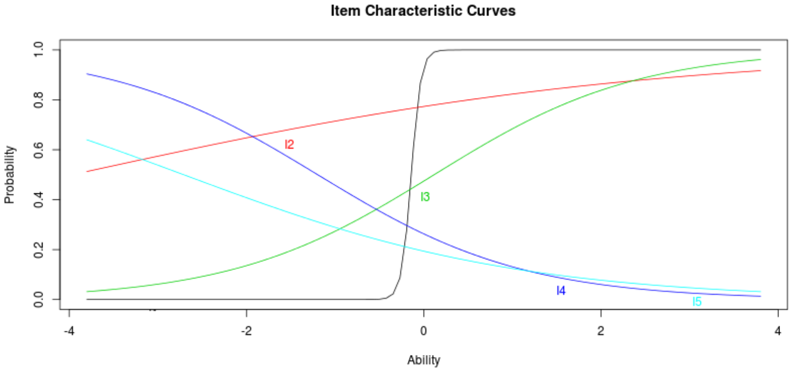


Figure 2: An example of an ICC curve that illustrates a good and poorly defined curve.

Researchers tried to enter the value of Indonesian language as a material test item but after obtained from the results of this study the value of language is not significant enough influence in this mixed approach research on the test item about MGMP bank in Banyuwangi.Maka can be concluded that from the results of this study there duahal useful that is:

1. Web Interface facilitates with the appearance of fiture - a feature that predicts each item of the problem that has a deficiency or the need for evaluation on each question that is otherwise not good. And the way that makes it easier for teachers to include a lot of this data shows web interface more efficient than the MS-based. Excel.
2. From the results of the bank trial analysis of the MGMP problem in Banyuwangi, there are 19 problems that are not good or revised on the number 2, 5, 8, 9, 10, 13, 17, 19, 21, 23, 24, 24, 29, 34, 37, 38, 40, 42 and 48.
3. With Indonesian Fixed Effects and Gender Random Effects: Schools (schools are considered random represent schools) Gender and Bhs Indonesia's contribution results are not as significant as the ability to correctly answer the items of the question.This also means non-biased (DIF) Gender Although not significant, if everything is maintained in the model then it is obtained. 1)

4.Acknowledgments

The authors would like to thank the MGMP Team in Banyuwangi especially in the mathematics of Bapak, Suparto (SMA NEGRI 1 PURWOHARJO), Nila Eka Sari (SMA NU Genteng) and Tutus Wahyudi (SMA PGRI TEGALDLIMO) who have helped to allow and use question bank and help in technically collecting data.

5.References

[1] Ackerman,T.A.,Gierl,M.J.,& Walker,C.M.(2003). Using Multidimensional item response theory to evaluate education and psycological tests. Education Measurement,Vol.22,pp.37-53.

[2] Arikunto,S.1999.Dasar –dasar Evaluasi Pendidikan. Jakarta : Bumi Aksara.

[3] Allen, M.J & W.M Yen. 1979. Introduction to measurement theory. Montere: Books/Cole Publising Company.

[4] Anggreyani,A.2009. Penerapan Teori Uji Klasik dan Teori Respon Butir dalam Mengefaluasi *Butir Soal* (Skripsi Jurnal Statistika IPB).

[5] Kilmen S & Demirtasli N.(2012). *Coparison of test equating methods based on item response theory according to the sample size and ability distribution*.SecienceDirect; www.elsevier.com/locate/jval.

[6] Knol, D.L & Berger, MP.F.(1991*). Empirical comparison between factor analysis and multidimentional item response models. Multivariate Behavioral Research,* No. 26, pp. 457-477.

[7] Kose Ibrahim A(2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test legth and sample size.SciVerse SecienceDirect; www.elsevier.com/locate/jval.

[8] Rabe- Hesketh dan Jeon Minjeong, 2012 " Profile – Likelihood Approach for Estimating Generalized Linear Mixed Models with Factor Strustures. " Jurnal of Education and Behavioral Statistic 37 (4). SAGE Publications Sage CA: Los Angeles, CA : 518 – 42.

[9] Retnawati H,(2008). Estimasi relative tes berdasarkan teori tes klasik dan teori respons butir. Disertasi. Universitas Negri Yogyakarta,tidak dipublikasikan.

[10] Retnawati H (2014). Teori Respons Butir dan Penerapannya,Yogyakarta. Email: nuhamedika@gmail.com – nuhamedika@yahoo.com.

[11] Rizopoulos D.2006. *ltm An R Package for Laten Variabel Modeling and Item Response Theory Analysis. Journal of matematical Software* vol.17(5):1-25.

[12] Reckase, M.D. (1997). *A linier logistic multidimensional model for dicotomous item response data.* In W.J Linden & R.K Hambleton (Eds), *Handbook of modern item response theory* (pp. 271 – 286). New York : Springer.

[13] Rabe- Hesketh dan Jeon Minjeong, 2012 " *Profile – Likelihood Approach for Estimating Generalized Linear Mixed Models with Factor Strustures. " Jurnal of Education and Behavioral Statistic* 37 (4). SAGE Publications Sage CA: Los Angeles, CA : 518 – 42.

[14] Tirta, IM. (2015). Pengembangan Analisis Respon Item Interaktif Online Menggunakan R Untuk Respon Dikotomus Dengan Model Logistik (1-Pl, 2-Pl-3Pl) *Prosiding Seminar Nasional Pendidikan dan Produk Akademik Universitas Jember* 30 Mei 2015 hal : 420 – 427.

[15] Youngchim P(2014). Development of a Mathematical Problem Solving Diagnostic Method: an Application of Bayesian Networks and Multidimensional Item Respond Theory. ScientDirect: www.scientdirect.com.