Advances in Social Science, Education and Humanities Research, volume 640
Proceedings of the 5th International Conference on Current Issues in Education (ICCIE 2021)

The Rasch Model for Analysing the Indonesian Language Examination Instrument

Muhammad Fajar Suminto*, Sri Wening

Universitas Negeri Yogyakarta, Indonesia
*Corresponding author. Email: muhammadfajar.2020@student.uny.ac.id

ABSTRACT
This study aimed to analyze instrument for evaluation in the Indonesian Language Test, based on its difficulty level. The Item Response Theory [IRT] was applied in this study. We investigated the examinee's responses to the instrument parameter. The respondents were 259 Vocational High School students of the 10th graders. We used Rasch Model to evaluate the difficulty level and create five categories from very difficult to very easy questions. Employing the R-Studio program for analysis, the result showed that out of 40 question items, one item was categorized as very difficult, one was categorized as difficult, twenty-nine items were classified as moderate, five items were categorized as easy, and four items were categorized as very easy. We grouped the difficult, moderate, and easy question items into the average difficulty category. As a result, 72.5 % of the items used in year-end assessment were considered moderately difficult, while 34 items, or 85% of the total items, match the Rasch model. The percentage of the items to be included in the question banks were the items that have moderate difficulty levels and match the Rasch Model, which was 60 %. It suggests that Rasch Model is applicable for instrument evaluation.

Keywords*: item response theory, rasch model, difficulty.*

1. INTRODUCTION

Assessment is a process that cannot be separated from learning activities. An exam or test is one of the evaluation procedures that can be carried out by an educator to assess the knowledge and skills of students. The test instrument is one of the mandatory components in an evaluation [1]. The instrument that can be used to measure the knowledge and abilities of the students is developed through several stages, including the validation and empirical validation stages. Many teachers, however, reported that they have not carried out these stages in developing the instrument for learning assessment. The preliminary study results indicate that most of the test instruments tested on students have not gone through the analysis test stage, the stage to determine the characteristics of the test itself [1]. This information was obtained by interviewing teachers in the city of Bandung. Another study was also conducted in Yogyakarta, where the teachers reported they had not carried out the empirical testing for their instruments that the outputs had not been able to describe the actual abilities of the students [2].

A good instrument, which can measure the ability of students, can be tested through a characteristic test of the instrument. A typical instrument testing can be identified by conducting item analysis and overall testing to determine whether it is considered good or not. A good instrument is a tool that can produce and provide the correct information so the results can describe students' actual abilities. In the measurement process, a measuring instrument is needed.

In general, the measurement of student achievement is done using a test instrument. A test is a form of instrument used to make measurements, namely collecting information on the characteristics of an object [7]. A test could also be defined as a number of questions that must be responded to measure a person's level of ability or reveal certain aspects of the person being tested [8].

Year-end assessment questions are forms of test instrument to measure learning achievement in schools, both for students and for teachers. The year-end assessment results can describe the achievement of student competency standards and the quality of learning applied by the teacher so that the test instrument used must have good and representative item characteristics in measuring every aspect of the actual student achievement [9].

Almost all exams or tests generally use a scoring approach to explain student achievement. The use of scores as a measure of achievement has weaknesses, for example, raw scores. The raw score is not essentially the result of a measurement. The raw score also has a weak quantitative meaning. In addition, raw scores cannot indicate a person's ability to do a task. The percentage of correct answers in the raw score is not always linear. Based on these problems, a different approach is needed to use the raw scores, namely using the item response theory approach.

One model of the item response theory approach is the Rasch model. Using this Rasch model, we measure the number of correct answers that students get and calculate the probability of odds ratio for each item.

This item response theory can create a hierarchical relationship between the examinee (person) and the items used so as to produce the same interval scale with the same logit unit for the person and item. Then, this can be directly compared, which results in complete information on the tests carried out with the abilities of students who work on them [10]. There are many studies on item analysis using the Rasch model. One of them is Fitri Alfarisa's research [6] with the title Analysis of Final Semester Test Items for High School Economics Using the Rasch Model, which states that 15% of the items analyzed are in the difficult category, or 72.5% was categorized as moderate, and 10% as easy, and 2.5 % was very easy. This result illustrates that almost 30% of the test instruments used were unable to describe the actual abilities of the students.

Another study conducted by Srika Ningsih Pasi [11] describes the results of the analysis of the Indonesian language items made by the teacher in which 40% was categorized as good, and the rest of it was categorized as poor. These previous studies and the lack of item analysis in the Indonesian language study field using the Rasch model item response theory became the background of the preparation of this research. The author feels that the analysis of the PAT items for the Indonesian Language of Class X is needed. Therefore, this study aimed to analyze proper instruments in the Indonesian Language Test based on its difficulty level.

2. METHOD

2.1. Source of Data

The data used in this study was the secondary data from the year-end assessment results taken from one of the Vocational High Schools in Sleman, Yogyakarta. The number of student samples was 259 students. At the same time, the number of items was 40 questions.

2.2. Method of Analysis

The analysis used in this study was the Item Response Theory using the Rasch model. The Rasch model was first introduced by a mathematician from Denmark, Dr. Georg Rasch, in 1950. Rasch developed a mathematical model that can measure the probabilistic relationship between a person's ability and the level of problem difficulty using logarithmic functions to produce measurements with equal intervals. The result is a new unit called logit (log odds unit), which shows the student's ability and problem difficulty. The logit value concluded that the level of success of the students in solving the problem is very dependent on the level of ability and the level of difficulty of the item.

Beside the level of difficulty, the compatibility test of the items with the Rasch model is carried out in this study. Items that match the model mean that the item has behavior that is consistent with what is expected by the Rasch model. When an item does not fit, there will be misconceptions among students about the item [3]. In conducting the item parameter analysis in this study, the writer used the RStudio software program.

3. RESULT AND DISCUSSION

3.1. Estimation of the Difficulty Level of Items using RStudio Program

The difficulty index usually ranges from -2.0 logit to +2.0 logit. The values that are closer to -2 logit indicate having easy characteristics, while values that are closer to +2 logit indicate having more difficult characteristics. In the 1-PL model, the value of b represents the level of ability (θ) with a 50% chance of answering correctly. Therefore, if b = 0 logits, the probability of the correct answer will be equal to 0.5 on the ability level = 0 logits. The one-parameter logistic model was built from the following equation

$$\rho_{i(\theta)} = \frac{e(\theta - b_i)}{1 + e(\theta - bi)} \quad (1)$$

Pi(θ) is the probability of a randomly selected respondent with the ability to answer item i correctly, bi is the difficulty parameter of item i, and e is a value of 2.718. The parameter bi is a point on the capability continuum where the probability of a correct response is 0.5. The greater the value of the bi parameter, the greater the ability required for the respondent to have a 50% chance to answer the item correctly [5]. The division of criteria for a more detailed level of difficulty adapted from Hidayatulloh [13] is as follows: b>2 is very difficult, 1<b≤2 is difficult, - 1 < b ≤ 1 is moderate, - 1 < b ≥ -2 is easy and b < - 2 is very easy [13].

Table 1. analysis result of item difficulty level using rstudio program.

Item	Result of Item Difficulty				
	Level of difficulty	*Category*	*Item*	*Level of difficulty*	*Category*
1	-1.796	Easy	21	0.278	Moderate
2	-0.899	Moderate	22	1.927	Difficult
3	-0.734	Moderate	23	-1.182	Easy
4	-0.119	Moderate	24	0.576	Moderate
5	0.769	Moderate	25	-0.182	Moderate
6	-2.149	Very Easy	26	-0.008	Moderate
7	-0.824	Moderate	27	-0.663	Moderate
8	-1.161	Moderate	28	-0.843	Moderate
9	-2.369	Very Easy	29	0.326	Moderate
10	-0.957	Moderate	30	-0.295	Moderate
11	0.198	Moderate	31	-0.119	Moderate
12	-0.788	Moderate	32	-2.149	Very Easy
13	0.645	Moderate	33	2.580	Very Difficult
14	0.627	Moderate	34	0.425	Moderate
15	-1.387	Easy	35	-0.475	Moderate
16	-0.751	Moderate	36	-0.788	Moderate
17	-0.938	Moderate	37	-0.135	Moderate
18	-0.327	Moderate	38	-2.579	Very Easy
19	0.102	Moderate	39	-0.977	Moderate
20	-1.077	Easy	40	-0.680	Moderate

Table 2. Categories of items based on the analysis results using rstudio program.

b	Difficulty level	# items	Percentage (%)	Items
b > 2	Very difficult	1	2.5	33
1 < b ≤ 2	Difficult	1	2.5	22
-1 < b ≤ 1	Moderate	29	72.5	2,3,4,5,7,10,11,12,13,14,16,17,18,19,21,24,25,26,27,28,29,30,31,34,35,36,37,39.40
-1 > b ≥ -2	Easy	5	12.5	1,8,15,20,23
b < -2	Very Easy	4	10	6,9,32,38

Based on the results of the analysis using the RStudio program, we can see that the average of the items tested for the year-end assessment has an average level of difficulty, ranging from 72.5%. In addition, 12.5% were categorized as easy, and 10% were categorized as very easy. Meanwhile, 2.5% was classified as difficult, and another 2.5% was categorized as very difficult. This result indicates that the items tested have a moderate level of difficulty. Therefore, it is considered as having a good proportion, as stated by Sudjana, one of the basic references in determining the proportion of items of a difficult, moderate, and easy category is by having a balance which is based on the direction of the normal curve [14]

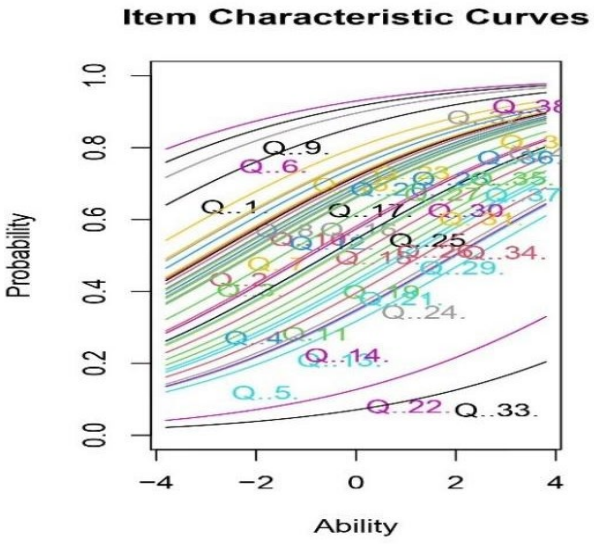

Figure 1 ICC All Item

Based on the plot of the results of the analysis of the characteristic curve of the item, it can be concluded that

items 22 and 33 or those on the right side are the items that belong to the difficult and very difficult categories. In comparison, the items on the left side are the items that belong to the very easy category, namely items 6, 9, 32, 38.

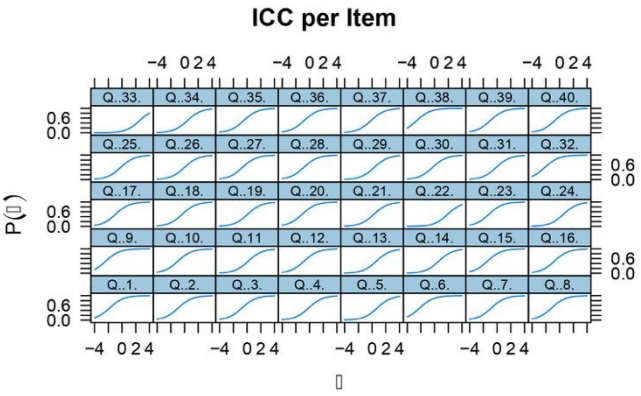

Figure 2 ICC Per Item

Figure 2 is an image of the item characteristic curve displayed per item to make it easier to interpret. A good curve tends to form the letter S. If the vertical axis is drawn (probability of answering correctly) at a point 0.5 to the right to meet on the curve and then pulled down on the axis, the category of questions will be obtained, whether the category is medium, easy or difficult. Based on the ICC image, it can be concluded that there is one item that is included in the very difficult category, namely item 33, and one item in the difficult category, namely item 22. In addition, there are four items that are included in the very easy category, namely item 6, 9, 32, 38, and there are five items that are included in the easy category, namely item 1, 8, 15, 20, 23.

3.2. Compatibility Test of Rasch Model

Item fit with the Rasch model was obtained when the p-value > 0.05. This is in accordance with the statement of Istiyono [1], saying that an item that does not fit has a probability or opportunity of < 0.05. Based on the probability value, this study also analyzes the year-end assessment results described in table 3.

Based on the results of the analysis outlined in table 3, we can see that there are 34 items or 85% in the fit category or match the Rasch model used for the end-of-semester assessment. While 15% or the remaining six items do not match the Rasch model.

Table 3. Fit and non-fit items with rasch model

Item	*p*	*Result*	*Item*	*p*	*Result*
1	0.095	Fit	21	0.207	Fit
2	0.233	Fit	22	0.288	Fit
3	0.474	Fit	23	0.117	Fit
4	0.199	Fit	24	0.632	Fit
5	0.122	Fit	25	0.027	Non Fit
6	0.234	Fit	26	0.045	Non-Fit
7	0.148	Fit	27	0.099	Fit
8	0.474	Fit	28	0.585	Fit
9	0.613	Fit	29	0.003	Non-Fit
10	0.052	Fit	30	0.151	Fit
11	0.387	Fit	31	0.024	Non-Fit
12	0.783	Fit	32	0.089	Fit
13	0.855	Fit	33	0.318	Fit
14	0.839	Fit	34	0.001	Non-Fit
15	0.458	Fit	35	0.296	Fit
16	0.826	Fit	36	0.798	Fit
17	0.675	Fit	37	0.388	Fit
18	0.411	Fit	38	0.761	Fit
19	0.188	Fit	39	0.651	Fit
20	0.767	Fit	40	0.009	Non-Fit
12	0.783	Fit	32	0.089	Fit

The result of the item parameter estimation analysis using the Rasch model and the item compatibility test is 24 items, or 60 % of the total items can be included in the questions bank. Those are the items that have moderate difficulty and match the Rasch model.

The findings of this study support the findings of another study entitled Question Items Analysis on Year-end Assessment of Economics for General High School Using Rasch Model, which was conducted by Fitri Alfarisa and Dian Normalitasari Purnama [15]. They found that the instrument had six items of difficult level (15%), 29 items of moderate level (72.5%), four items of easy level (10%), and 1 item of very easy level (2.5%). The average difficulty of the instrument used in Economics year-end assessment for the XI class was 0.00 logit [moderate category] and 1,00 logit of standard deviation

Table 4. List of items with moderate difficulty and match the Rasch Model

Item	Result of Item Difficulty		Fit and Non-fit Items with Rasch Model	
	Level of difficulty	Category	p	Result
2	-0.899	Moderate	0.233	Fit
3	-0.734	Moderate	0.474	Fit
4	-0.119	Moderate	0.199	Fit
5	0.769	Moderate	0.122	Fit
7	-0.824	Moderate	0.148	Fit
8	-1.161	Moderate	0.474	Fit
10	-0.957	Moderate	0.052	Fit
11	0.198	Moderate	0.387	Fit
12	-0.788	Moderate	0.783	Fit
13	0.645	Moderate	0.855	Fit
14	0.627	Moderate	0.839	Fit
16	-0.751	Moderate	0.826	Fit
17	-0.938	Moderate	0.675	Fit
18	-0.327	Moderate	0.411	Fit
19	0.102	Moderate	0.188	Fit
21	0.278	Moderate	0.207	Fit
24	0.576	Moderate	0.632	Fit
27	-0.663	Moderate	0.099	Fit
28	-0.843	Moderate	0.585	Fit
30	-0.295	Moderate	0.151	Fit
35	-0.475	Moderate	0.296	Fit
36	-0.788	Moderate	0.798	Fit
37	-0.135	Moderate	0.388	Fit
39	-0.977	Moderate	0.651	Fit

4. CONCLUSION

Based on the results of the analysis of the item parameter estimation using the Rasch model and the item compatibility test, it can be concluded that the item parameter estimation regarding the level of difficulty in the items used for the year-end assessment test at one of the Vocational High School in Sleman district of Yogyakarta Kalasan Sleman Yogyakarta has five criteria, namely very difficult, difficult, moderate, easy, and very easy. One item was categorized as very difficult, one was categorized as difficult, twenty-nine items were classified as moderate, five items were categorized as easy, and four items were categorized as very easy. Based on this analysis, we can conclude that the instrument for the year-end examination has moderate difficulty, or 72.5% of the items have moderate criteria. The result of the item compatibility test using the Rasch Model is 34 items, or 85 % of the total items match the Rasch Model and the percentage of items to be included in the questions bank, which have moderate difficulty and match the Rasch Model was 60 %.

ACKNOWLEDGMENTS

Researchers thank Survey METER for funding this research. We also thank Yogyakarta State University.

REFERENCES

[1] P. T. R. R. Asri Andayani, "Kajian Implementasi Teori Respon Butir dalam Menganalisis Instrumen Tes Materi Fisika," Prosiding Seminar Nasional Fisika, 2019, p. 37.

[2] H. Akhtar, "Pedoman Analisis Item Menggunakan Rasch Model," 27 Juli 2017. [Online]. Available: https://www.semestapsikometrika.com/2017/07/analisis-dan-seleksi-item-menggunakan.html.

[3] H. Akhtar, "Semesta Psikometrika," 1 September 2017,https://www.semestapsikometrika.com/2017/09/mengubah-skor-ke-bentuk-skor-standar-di.html.

[4] E. Istiyono, Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika Dengan Teori Klasik dan Modern, 2018, UNY Press.

[5] H. S. H. J. R. Ronald K Hambleton, Fundamentals of Item Response Theory, Newbury Park: Sage Publication Inc., 1991.

[6] D. N. P. Fitri Alfarisa, "Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi Menggunakan Rasch Model," Jurnal Pendidikan Ekonomi, 2019, p. 367.

[7] D. Mardapi, Pengukuran, Penilaian dan Evaluasi Pendidikan, 2015, Nuha Medika.

[8] E. P. Widoyoko, Teknik penyusunan instrumen penelitian, Pustaka Pelajar, 2012.

[9] R. R. Muh. Syahrul Sarea, "Karakteristik Butir Soal : Classical Test Theory VS Item Response Theory?," DIdaktika : Jurnal Kependidikan, 2019, p. 2.

[10] K. D. K. A. Untung kurniawan, "Analisis Soal Pilihan Ganda Dengan Rasch Model," Statistika, pp. 2018, 34-36.

[11] Y. Srika Ningsih Pasi, "Analisis Butir Soal Ujian Bahasa Indonesia Buatan Guru MTsN di Kabupaten Aceh Besar," Master Bahasa, 2018, p. 195.

[12] R. S. H. R. H. Hambleton, Fundamental of Item Response Theory, Sage Publications, 1991.

[13] M. A. Hidayatulloh, "Analisis kualitas butir soal ulangan semester genap bahasa arab kelas VIII tahun pelajaran 2012/2013 dengan program quest

di SMP Muhammadiyah 3 Depok Sleman Daerah Istimewa Yogyakarta," UIN Sunan Kalijaga., 2013.

[14] N. Sudjana, Penilaian hasil Proses Belajar Mengajar, Remaja Rosdakarya, 2017.

[15] F. A. a. D. N. Purnama, "Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi SMA Menggunakan Rasch Model," Jurnal Pendidikan Ekonomi, 2019, pp. 366-374.