JERE 9 (2) 2020 58-69

Journal of Educational Research and Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Analysis Of Essay Test Instruments Using Higher Order Thinking Skill (HOTS) at High School Mathematics Students Using The Rasch Model

Bella Safira Fiyali Putri✉, **Kartono Kartono, Supriyadi Supriyadi**

[1.] Sekolah Menengah Atas Negeri 1 Mandirancan, Indonesia.
[2.] Universitas Negeri Semarang, Indonesia.

Article Info

History Articles
Received:
03 Jauary 2020
Accepted:
31 March 2020
Published:
23 August 2020

Keywords:
Test; instruments; HOTS, Rasch

Abstract

Learning process in the 21st century, learners are not only expected to have a high understanding of concepts, but also HOTS. Therefore, the need to be presented about HOTS as well as its analysis to find out the ability of students on these HOTS skills. This study aimed to analyze the form of HOTSbased Essay test in Mathematics lessons focused in trigonometry. This research was conducted with a descriptive quantitative and qualitative approach. Samples were taken from high school students of class X SMA Negeri 1 Mandirancan in small class trials, amounting to 36. Quantitative analysis to look at the characteristics of instruments and item with rasch model. Qualitative data analysis was conducted using source triangulation and techniques on the factual condition of instrument. Based on the analysis, the problem item that is not valid enough there is only 1 out of 10 items, that is item number 1. Item measure in this math Essay test question is classified into 4 categories, that is 1 item is very difficult, 4 items are difficult, 4 items are categorized easily and 1 category item is very easy. Based on the output of the person fit, can be seen, a total of 52.78% of students can complete the HOTS Math essay test well, but the level of person ability is lower than the quality of the item judging by the reliability value of the person and item. In conclusion, the essay test instrument HOTS Mathematics developed tested valid and reliable and has adequate characteristics.

✉Correspondence Address :
Desa Paniis, Dusun Pon, Kecamatan Pesawahan, Kabupaten Kuningan 45559
E-mail : bellasafira.fp@gmail.com

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

Permendikbud Number 24 of 2016 concerning Core Competencies and Basic Competencies of Lessons in the 2013 curriculum at the Primary and Secondary Education has changes the teaching patterns that have been compiled. Educators must adjust learning and assessment activities so that they match the competencies to be achieved. Improper assessment can obscure expected results. Evaluation and revision need to be carried out on learning systems and devices so that they are suitable for Mathematics learning in SMA to develop knowledge, attitudes, logical, systematic, critical, analytical, and social skills thinking.

The teacher knows about problem solving problems even though they are defined differently. Problem solving problems are not new for teachers (Hidayah et al, 2014). Problem solving can be seen as a goal-oriented process that requires the use of a variety of integrated higher-order thinking skills, such as generating ideas, making interpretations and judgments, and using strategies to manage the complexity of situations (Kirkwood, 2000: 511). Problem-based learning and problem solving conditions students to develop thinking skills (Sucipto, 2017). Measurement of mathematical problem solving abilities can be done by presenting contextual problem solving problems. The measurement of students' mathematical problem-solving abilities is done by giving problem (Sari, et al, 2019).

Schraw in Kusuma et al (2017: 26-32) classified thinking skills based on Bloom's taxonomy into two categories, there are Lower Order Thinking Skill (LOTS) which consists of knowledge, understanding and application. Higher Order Thinking Skill (HOTS) consists of Analyze, Synthetic and evaluation. Krathwohl (2002: 215) stated that Bloom's revised taxonomy consists of (1) Remember, (2) Understand, (3) Apply, (4) Analyze, (5) Evaluate, and (6) Create. Especially for higher students, they must have not only a lower order thinking (LOT), but also have to reach a higher order thinking (HOT) (Kusuma, et al, 2017: 1). Students with high levels of HOTS tend to be more successful (Tanujaya, Mumu & Margono, 2017: 78). HOTS is one of the main goals in education and is one of the top five variables that can improve student achievement (ha & Retnawati, 2018: 1).

Higher order thinking skills are Analyze (C4), Evaluate (C5), and Create (C6) (Trisnawati et al, 2017). Juhanda (2016) stated that the average percentage of questions that develop high-order thinking skills starting from questions C4, C5 and C6 has a low average percentage. Regarding the Revised Bloom cognitive level that developed HOTS still needs to be improved. HOTS-oriented learning is important in the learning process because it helps solve problems in everyday life (Retnawati, et al, 2018: 219).

Mathematics learning must be given to students in order to be able to solve a problem, think logically, analytically, systematically, critically, creatively, be able to communicate properly (communicative) and be able to work well together (Ahmad, et al, 2017: 1). In implementing mathematics learning, it is expected that the students should be able to feel the usefulness of learning math (Widana, et al, 2018: 26). Therefore, high-level thinking is needed, but in fact based on Nusarastriya's research (2013: 24) stated that only 1% of students in Indonesia have advanced thinking skills. Students in Indonesia have a lower order thinking skills by 78%. The mathematical problem solving abilities of Indonesian students are still low because mathematics is not a subject of interest to students (Nidya, 2015; Simamora & Siragih, 2019: 62). The results of this survey showed that Indonesia is still far below the students of Taiwan, South Korea, Singapore, Hong Kong and Japan who have high-level thinking skills of students above 40%. This data is supported by the results of research from 30 students spread across several junior high schools in

Jember Regency. It was found that 18 students were able to perform logic and reasoning skills, analysis, evaluation, and creation well. 12 students were not able to perform analytical, evaluation, creative, logic and reasoning skills well in solving all the questions (Kurniati et al, 2016). The teacher can see the higher order thinking skills of students using test techniques. Therefore, in order to face the 21st century, students are not only expected to have a high understanding of concepts, but also have high-order thinking skills / HOTS (Rusminati et al, 2015). In this research, it is important to do this by developing and analyzing the form of HOTS-based Essay test questions.

Research on HOTS has been conducted by Sadieda, Indayati, & Faizien (2018: 1). This study aimed to develop a mathematical HOTS instrument that integrates the Islamic context and identifies the Islamic context which is the problem for students in answering HOTS test questions. Another study was conducted by Ichsan, et al (2019: 935). The results showed that learning must be oriented towards increasing HOTS through various media development learning, learning materials, learning models, and strategies. This study concludes that the HOTS score of students was still very low and needs to be improved. Research by Suhirman, et al (2020: 183) aimed to investigate the effect of problem-based learning with character emphasis toward the students' higher-order thinking skills (HOTS) and characters. The research data were collected using essay tests and student character data were collected through self-assessment sheets, and analyzed using MANOVA. Handayani, Hartono & Lestari (2019) conducted an analysis of the development needs of HOTS questions. The result is that 20 item statements filled in by 8 teachers in the Sunan Kalijaga area, Gumuh District were declared valid with an index of 0.707 which indicates that the item statement is greater than the r-table. The needs analysis conducted by Handayani, Hartono & Lestari (2018) can be used as a guide for current research. Similar research was also conducted by Subay, Kartono & Sulhadi (2019) which measured the validity and reliability of HOTS questions in mathematics.

Tests are systemic procedures, containing samples of behavior and measuring behavior. The items in the test are arranged according to certain methods and rules, the test administration procedure and scoring of the results must be clear and specified in detail. The person who takes the test must receive the same items in comparable conditions. Regardless of the length of a test, the items in it will not be able to cover the entire content of the material that may be asked and the feasibility of a test depends on the extent to which the items in the test represent the area (domain) of behavior being measured. The items in the test require the subject to show what is known or what the subject has learned by answering questions (Azwar, 2010: 3).

The researcher conducted a preliminary study by conducting interviews and looking at the test documents used. The researcher found that there were only limited question document and answer keys in the questions. Researchers did not find any clear scoring guidelines to categorize students' answers. The teacher arranges based solely on the questions in the textbook and only measures low-level thinking skills. The teacher does not categorize each item to measure the cognitive level in its preparation. The teacher assessed students' answers subjectively so that the assessment is quite difficult for other teachers to do because only those who make the items know the meaning and purpose of the items. The teacher also gives a bonus value for each answer without any particular category.

Therefore, it is necessary to develop a HOTS essay test instrument mathematics for high school. Student achievement results regarding HOTS need to be evaluated to determine the extent to which students' HOTS are in the learning process (Hadi, et al, 2018: 521). After the test is used by the test taker, it is necessary to know whether the test is of

good or poor quality so that a test analysis is necessary. This was to help improve the test through improvements to find out information on whether students have mastered the material taught by the teacher (Amalia & Widayati, 2012: 3). This instrument was a previous research and this article will focus on further research on the analysis of the HOTS essay test instrument in SMA Mathematics.

The model developed to analyze the items can use the Rasch model. Rasch modeling is an analytical tool that can test the validity and reliability of the research instrument in detail, by testing the suitability of persons and items simultaneously (Anindita & Cahyadi, 2020: 223). The one-parameter model or the Rasch model has several advantages, which is it can identify error responses, can predict missing data scores, can distinguish the ability of respondents with the same raw score, and can identify indications of guesses and cheats (Sumintono & Widhiarso, 2015: 44-45). The aimed of this research was to analyze the characteristics of the essay test instrument items that measure the high-order thinking skills of class X mathematics developed using the Rasch Model.

METHODS

Qualitative data analysis was carried out using triangulation of sources and techniques regarding the factual conditions of the HOTS essay assessment instrument on mathematics subject matter trigonometry equipped with a question grid and scoring guidelines. The competency achievement indicators tested are 2: (1) Explain the trigonometric ratios (sine, cosine, tangent, co-secant, secant, and cotangent) in right triangles and (2) Generalize the trigonometric ratios for angles in various quadrants and angles. related angles. The research was conducted in class X SMA Negeri 1 Mandirancan. The results of observations, interview transcripts and documentation were matched and analyzed about their suitability in order to obtain conclusions about the factual conditions of the HOTS essay test assessment instrument on mathematics which was carried out at SMA Negeri 1 Mandirancan. In addition to triangulating sources and techniques, qualitative analysis in research was also carried out on the initial instruments that had been prepared. The analysis of the instruments by experts is a qualitative analysis which is meant to see the input, suggestions, and considerations and decisions of the experts. Quantitative data analysis was performed by calculating validity and reliability. Apart from calculating the validity and reliability, the researcher also analyzed the item fit level, item measure difficulty, person fit, and DIF analysis using the Rasch model analysis.

RESULTS AND DISCUSSION

A set of Mathematics essay test questions consisting of 10 questions is given to students to analyze, prove, and provide descriptions of the extent to which students can answer HOTS questions. The main skills that must be possessed in the context of the 21st century are learning and innovation skills. HOTS will direct students in constructing appropriate and effective arguments to make rational decisions or solutions (Syadiah & Hamdu, 2020: 145). High-order thinking skills are needed by students to follow the demands of competence in the 21st century, one of which is the assessment model (Aliyah & Fajriyah, 2020).

After the test activities are carried out on students, then an analysis is carried out so that the results of student work can be evaluated more comprehensively, one of which is by using Rasch modeling. Objective measurements, as well as the application of the Rasch Model in educational assessments using software designed for the Rasch model application. (Bond, 2013) In the context of the Rasch model, this 'settled' scoring pattern is none other than a test dependent scoring; then what must be done in quantitative research in educational assessment is an objective measurement (Sumintono, 2015). Benefit from

the results of the analysis, a teacher will find it easier to pay attention according to the authenticity of the students. In addition, the teacher will find it easier to make corrections to the questions that have been made so as not to make questions with the same mistakes, so the analysis of the questions using the Rasch modeling provides many pictures that can be obtained by anyone.

Analysis of the Rasch model in this study was carried out using Winsteps software. The characteristics of the items analyzed in the Rasch model are able to reveal interpretations in terms of items, persons, and instruments (Lia, Rusilowati & Isnaeni, 2020). The stage that must be passed in the instrument development procedure for the Rasch model is verification of the assumptions of unidimensionality and local independence (Sumintono, 2014, p. 5). On the Rasch model

Using the Winsteps program, unidimensionality is found in the Item function: dimensionality and item accuracy with the model (in fit-out fit) and its location (measure) can be seen in Item: measure and Item: fit order (assuming local independence will be explained in point 4.4). Unidimensionality testing also serves to see the validity of the construct.

Before looking at item functions, the first step to see the construct validity with the Rasch model is through the Polarity Item Diagnosis Output (Hayati & Lailatussaadah, 2016, p. 173). As shown in the output of the polarity item diagnosis, all items have a positive Point Measure Correlation (Pt. Mea-Corr). Based on the test using the item: dimensionality on Winstep to obtain construct validity with the criteria that it is stated to have the ability to measure range variables or measure all respondents if Raw Variance Explained at least 20% (Novinda, Silitonga & Hamdani, 2019, p. 8).

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --		Modeled
Total raw variance in observations	=	14.0	100.0%	100.0%
Raw variance explained by measures	=	4.0	28.5%	28.5%
Raw variance explained by persons	=	1.6	11.3%	11.3%
Raw Variance explained by items	=	2.4	17.2%	17.2%
Raw unexplained variance (total)	=	10.0	71.5% 100.0%	71.5%
Unexplned variance in 1st contrast	=	2.2	15.6% 21.8%	
Unexplned variance in 2nd contrast	=	1.9	13.7% 19.1%	
Unexplned variance in 3rd contrast	=	1.4	10.2% 14.2%	
Unexplned variance in 4th contrast	=	1.3	9.1% 12.7%	
Unexplned variance in 5th contrast	=	1.0	7.2% 10.1%	

Figure 1. Test Undimensionality

Seeing the raw variance in Figure 1 shows a figure of 28.5%. This showed that the minimum unidimensionality requirement of 20% can be met. Thus the HOTS essay test instrument developed in this study was able to measure what should be measured. Unexplained variance, respectively 2.2; 1.9; 1.4; 1.3 and 1.0. This showed that the variance that cannot be explained by the instrument is all below 10%. Unexplained variance below 10% indicates that the unidimensionality of the instrument is in the good category (Wibisono, 2014, p. 744).

After the unidimensional requirements are met, the analysis with the Rasch model can be continued. The main characteristic of the analysis with the Rasch model which is part of the IRT is that the response of the test taker to a question being tested will form a curve called the item characteristic curve. This curve is a functional formula from mathematics which states the relationship between the probability of answering correctly P (θ) and ability (θ) (Hambleton & Swaminathan, 1985 in Rusilowati, 2018). IRT with a one-parameter logistic model is usually called the Rasch model, although the Rasch

model is different from the IRT but one logistic parameter IRT has the same mathematical form as the Rasch model (Hambleton & Swaminathan, 1985). The following is the test characteristic curve on the HOTS Mathematics essay test instrument testing which is presented in Picture 2.

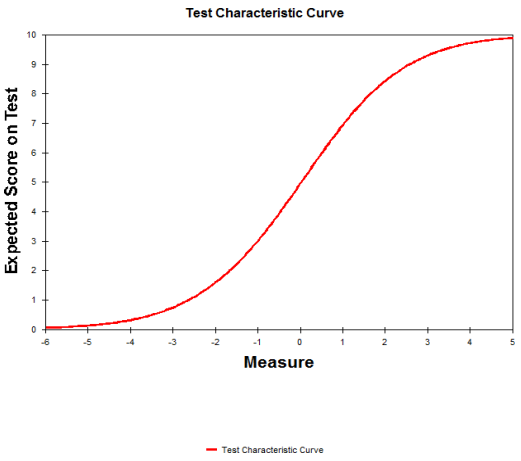

Figure 2. Characteristic Test Curve

In Picture 2, the curve of the curve is determined by the parameters of the question and the parameters of the participants who responded to the question. Because the test characteristic curve is an ascending monotonic curve, the higher the ability ( ), the greater the chance of answering a question correctly (Rusilowati, 2018, p. 8). The characteristics of the questions are constant and do not change, even though the group of participants who answer the questions changes. Also, if the group of participants has the same characteristics, the characteristics of the questions will remain, even though the questions are answered differently. This meant that the curvature of the participant's response or the question characteristic curve will not change even though the test is given to several groups of participants with different distribution of abilities (Rusilowati, 2012).

Analysis of the Rasch model in this study was carried out using Winsteps software. The characteristics of the items analyzed in the Rasch model are able to reveal interpretations in terms of items, persons, and instruments (Lia, Rusilowati & Isnaeni, 2020). The analysis performed using Winstep in this study includes Person Reability, Item Reliability, Item Fit, Item Measure, and Person Fit.

Reliability of the Rasch (Person Reliability and Item Reliability) Model

Reliability on Rasch can be determined by analyzing the value of the information function, for test trials based on graphic images, the peak of the graph is in the middle (close to zero), not much different from the information obtained during the test, that is on the chart 0.27), however, if you look carefully, the top of the chart is slightly to the right of 0 (leaning to the right). This shows that the Higher Order Thinking Skils (HOTS) assessment test in mathematics learning will provide maximum information if it is given to students who have moderate ability (Hambleton & Swaminathan, 1985, p. 93). Information graphics can be seen in Picture 3.

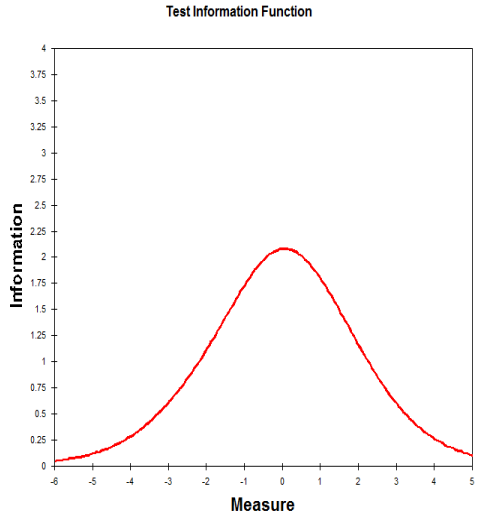

Figure3. very low low middle high very high

The graph of the measurement information function (test information function) in Picture 3 is divided into an X-axis and a Y-axis. The X-axis showed the level of student ability in working on the questions. Meanwhile, the Y axis showed the size of the information function (Sumintono & Widhiarso, 2015, p. 86).

Picture 3 explained that at a low level of ability (ability), it shows that the information

obtained from the measurement is also quite low. Likewise, the very high ability, the information obtained is still low. At moderate ability, the information obtained by measurement is very high. This shows that it produces optimal information when given to individuals when given to individuals with moderate ability.

The reliability value of the Rasch Model using Winsteps can be seen by displaying the results from the main menu Output Table, then selecting Table 1. Summary Statistics. The reliability value can be seen from the Person Reliability and Item Reliability values that appear. Based on the Winstep Guide described (Boone, Staver, & Yale, 2013: 222). Reliability output can be seen in Table 1.

Table 1. *Output* Reliability of Model Rasch

Measured Person ji				
	Infit MNSQ	ZSTD	Outfit (S MNSQ	ZSTD a
Mean	2.6			
Separation	1.08			
Person Reliability	0.54			b
Measured Item				
	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD c)
Mean	0.97	0.1	0.98	0.2
Separation	1.58			
ItemReliability	0.72			
KR-20 Test Reliability	0.84			b

Reliability in the Rasch model is described by the existence of a separation index. The reported separation index is item reliability and person reliability plus the Cronbach Alpha KR-20 reliability coefficient value is added. Based on Table 1, the Reliability Output Rasch Model is matched with the Sumintono & Widhiarso (2015: 85) criteria table, it can be explained that the value of person reliability 0.54 is categorized as weak. This figure showed that the consistency of students' answers when answering the HOTS-based Mathematics essay test questions is weak or inconsistent. The value of the item reliability of 0.72 is categorized sufficient, meaning that the quality of the items in the instrument has sufficient reliability aspects (Azura, Samsudin & Utari, 2020). The reliability coefficient value of the Cronbach Alpha KR-20 shows a number of 0.84 which is categorized as good.

MNSQ INFIT and MNSQ OUTFIT values for person can not be displayed, because the value is very extreme so it can not show the value. The values showed 0.97 and 0.98 for the value of INFIT MNSQ and OUTFIT MNSQ items. The value of measured items is a good one because the ideal value is 1 (the closer to 1 the better). The values for INFIT ZSTD and OUTFIT ZSTD, respectively, are -, -, 0.1, 0.2. Ideally, the ZSTD value is 0.0, so it can be concluded that the ZSTD value is considered ideal except for the INFIT ZSTD value for the person with extreme value.

Item Fit

The fit item in the Rasch model aims to see the quality of the items from the validity aspect, means that they meet the following criteria (umintono & Widhiarso, 2015).

MNSQ (Mean Square) Outfit Value is accepted: 0.5 <MNSQ <1.5)

ZSTD (Z - Standard) Outfit Value is accepted: –2.0 <ZSTD <+2.0 '

Pt Measure Corr (Point Measure Correlation) value: 0.4 <Point Measure Corr <0.85

Then, the fit item test was carried out on each item. The fit item output is shown in Table 2.

Based on the results in Table 2, item fit is determined by three criteria. Analysis of the items using the Rasch modeling approach is considered more accurate when the item questions meet 3 (three) criteria, there are the MNSQ Outfit value, the ZSTD Outfit value, and the Point Measure Correlation value. If all three are fulfilled, the question is considered valid (Erfan, et al, 2020). Item 6, 2, 7, 3, 8, 4, 9 meet the criteria of three categories. Items 10 and 5, the MNSQ outfit value does not meet the criteria but this data deviation can still be tolerated, therefore it is still categorized as valid. There is only 1 item that is not valid enough, that is item number 1. This is because the MNSQ and Pt Measure Corr outfit values do not meet the criteria. The treatment required for item number 1 is by revising the questions.

Table 2. Output Item Fit

	Outfit			
No Item	MNSQ	ZSTD	Pt. Measure Corr	Remark
10	1.67	1.7	0.55	valid
6	1.15	0.5	0.45	valid
2	1.13	0.5	0.69	valid
7	1.08	0.3	0.51	valid
3	1.09	0.4	0.66	valid
8	0.99	0.1	0.64	valid
4	0.95	-0.1	0.61	valid
9	0.83	-0.5	0.74	valid
5	0.42	-0.66	0.46	valid
1	0.49	-0.66	0.92	Not enough valid

Measure items

In the Rasch model the difficulty level of the items is reviewed according to (Sumintono and Widhiarso, 2015: 70) which can be grouped into four categories based on the measure values obtained in the Rasch model analysis as follows.

Table 3. Criteria for the Difficulty Level of Rasch Items In table 3, S.D. is the Standard Deviation. The S.D value in this test is 1.04. difficulty level -S.D. logit indicates the value -1.04. The output measure items are shown in Table 4.

Table 3. Output Item Measure

No Item	Total Count	Total Score	Measure (logit)	Conclusion
5	36	3	1.89	Very difficult
6	36	6	0.88	diificult
7	36	7	0.61	diificult
10	36	8	0.36	diificult
4	36	9	0.11	diificult
8	36	10	-0.14	easy
2	36	11	-0.39	easy
3	36	11	-0.39	easy
9	36	12	-0.64	easy
1	36	17	-2.30	Very easy

Table 4 provided the information on the difficulty level output obtained by the measure item. The difficulty level of the items based on the Rasch model is reviewed based on the measure value in logit units (Susdelina, Perdana & Febrian, 2018).

The outputs presented in Table 4 are sorted according to their level of difficulty. The highest logit value indicates a high level of problem difficulty. This corresponds to the total score column, which states how many correct answers are (Kurniawan, 2018, p.36). For example, for item no. The 5 logit scores were 1.89 logits and only three people

answered correctly out of a total of 36 students. The item measure on this Math Essay test question is classified into 4 categories, there are very difficult, difficult, easy and very easy. Item number 5 is an item with a very difficult category with measure values (Source: Sumintono & Widhiarso, 2015, p. 70)

by a modification of 1.89 logit. There are 4 items that are categorized as difficult because the difficulty level value is 0 ≤ Measure logit ≤ 1.04. Item number 8, 2, 3 and 9 are categorized as easy, and there is one question that is categorized as very easy based on its logit value, that is item number 1. This can also be seen from the person who answered correctly, there are 17 out of 36 students could answer correctly.

Table 4. Analysis of Student

Difficulty Level	Criteria
Measure logit < - S.D. logit	Very easy
- S.D. logit ≤ Measure logit ≤ 0	Easy
0 ≤ Measure logit ≤ S.D. logit	Difficult
Measure logit > S.D. logit	Very Difficult

Person Fit

In addition to analyzing the items, analysis of student ability is also important because it is possible to map students' abilities based on their level of ability through person fit output (Kurniawan & Andriyani, 2018, p. 37). The determination of this test is the same as in the fit item output, there are by looking at 3 criteria (MNSQ, ZSTD, and Pt Measure Corr outfit. Based on the person fit output, it can be mapped into 3 categories, there is one student is not suitable because of the MNSQ and Pt Measure Corr values. In addition to the person fit order output, the level of student ability is strengthened from the schalogram response which can be explained that the respondent with code A019 found inappropriate responses. A number of 8 students found that their analysis results were not sufficiently suitable. MNSQ scores 2.07 and Pt Measure Correct is -0.11. If seen from the schalogram response, number 10 with a difficult level of questions can be done correctly, but numbers 1, 9, 2, 3, 8 with an easy level of questions cannot be done well. Likewise with the respondent The codes A036, A06 and A032 are also able to do difficult questions, which indicates that the test taker answered by guessing.

Based on the person fit output, it can be seen that 52.78% of students can complete the HOTS Math essay test well. It is proven by the MNSQ, ZSTD, and Pt Measure Corr values which meet the criteria, but the person's ability level is still below the item quality seen from the person and item reliability value.

CONCLUSION

The 10 items HOTS math essay test instrument was analyzed using the Rasch model. The output produced in this Rasch model analysis is validity (item fit), reliability (item and person reliability), item measure and person fit strengthened by scalogram data. There are only 1 out of 10 items that are not valid enough, that is item number 1.The item measure in this Math Essay test item is classified into 4 categories, that is 1 very difficult item, 4 difficult items, 4 items categorized as easy and 1 item categorized very easy. Based on the person fit output, it can be seen that 52.78% of students can complete the HOTS Math essay test well. Evidenced by the MNSQ, ZSTD, and Pt Measure Corr values that meet the criteria, but the level of ability of persons is lower than the quality of the items seen from the reliability value of person and item. The conclusion is that the HOTS Mathematics essay test instrument developed has been tested valid and reliable and has adequate characteristics.

REFERENCES

Ahmad, S., Prahmana, R. C. I., Kenedi, A. K., Helsa, Y., Arianil, Y., & Zainil, M. (2017, December). The instruments of

higher order thinking skills. In Journal of Physics: Conference Series (Vol. 943, No. 1, p. 012053). IOP Publishing.

Aliyah, N., & Fajriah, N. (2020). Pengaruh Pendekatan Saintifik dalam Pembelajaran Matematika Terhadap Kemampuan Berpikir Tingkat Tinggi Siswa.

Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota Yogyakarta tahun 2012. Jurnal Pendidikan Akuntansi Indonesia, 10(1).

Anindita, R., & Cahyadi, L. (2020). Aplikasi Model Rasch dalam Mengukur Komitmen Dosen. JMSP (Jurnal Manajemen dan Supervisi Pendidikan), 4(3), 220-231.

Azura, A., Samsudin, A., & Utari, S. (2020). Analisis Peta Wright Keterampilan Berpikir Level LOTs dan HOTs Siswa Kelas XI pada Materi Miopi. WaPFi (Wahana Pendidikan Fisika), 5(1), 76-83.

Azwar, S. (2010). Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar. Yogyakarta: Pustaka Pelajar.

Boone, W.J., Staver, J.R., dan Yale, M. S. 2014. Rasch Analysis in the Human Sciences. Dordrecht : Springer.

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model rasch. Indonesian Journal Of Educational Research and Review, 3(1), 11-19.

Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. Problems of Education in the 21st Century, 76(4), 520.

Handayani, F., Hartono, H., & Lestari, W. (2019). Need Analysis in The Development of HOTS- Oriented Study Project Assesment Instrument in Android-Based Science Learning. Journal of Research and Educational Research Evaluation, 8(1), 57-64.

Hidayah, N., Suyitno, H., & Junaedi, I. (2014). Analisis Kemampuan Guru Matematika SMP dalam Membuat Soal-Soal Pemecahan Masalah. Unnes Journal of Mathematics Education Research, 3(1).

W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher Order Thinking Skills from Elementary to Master Students in Environmental Learning. European Journal of Educational Research, 8(4), 935-942.

Juhanda. (2016). Analisis Soal Jenjang Kognitif Taksonomi Bloom Revisi Pada Buku Sekolah Elektronik (BSE) Biologi SMA. Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam, 21(1).

Kirkwood, M. (2000). Infusing higher-order thinking and learning to learn into content instruction: A case study of secondary computing studies in Scotland. Journal of Curriculum Studies, 32(4),509–535.

Kurniati, D., Harimukti, R., & Jamil, N. A. (2016). Kemampuan Berpikir Tingkat Tinggi Siswa Smp Di Kabupaten Jember Dalam Menyelesaikan Soal Berstandar Pisa. Jurnal Penelitian Dan Evaluasi Pendidikan, 20(2), 142.

Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan RASCH Model. Jurnal Statistika Universitas Muhammadiyah Semarang, 6(1).

Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model. REiD (Research and Evaluation in Education), 6(1), 41-50.

Merta Dhewa, K., Rosidin, U., Abdurrahman, A., & Suyatna, A. (2017). The development of Higher Order Thinking Skill (Hots) instrument assessment in physics study. IOSR Journal of

Research & Method in Education (IOSR- JRME), 7(1), 26-32.

Nusarastriya, Y. H. (2013). Permasalahan Dan Tantangan Guru PKn Menghadapi Perubahan Kurikulum (2013). Satya Widya, 29(1), 23.

Pratama, G. S., & Retnawati, H. (2018, September). Urgency of higher order thinking skills (HOTS) content analysis in mathematics textbook. In Journal of Physics: Conference Series (Vol. 1097, No. 1, p. 012147). IOP Publishing.

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The Development of an Instrument to Measure the Higher Order Thinking Skill in Physics. European Journal of Educational Research, 8(3), 743-751.

Retnawati, H., Djidu, H., Kartianom, A., & Anazifa, R. D. (2018). Teachers' knowledge about higher- order thinking skills and its learning strategy. Problems of Education in the 21st Century, 76(2), 215.

Rusminiati, N. N., Karyasa, I. W. dan Suardana, I. N. 2015. Komparasi Peningkatan Pemahaman Konsep Kimia dan Keterampilan Berpikir Kritis Siswa antara yang Dibelajarkan dengan Model Pembelajaran Project Based Learning dan Discovery Learning, EJournal Program Pascasarjana Universitas Pendidikan Ganesha, (5): 1–11.

Sadieda, L. U., Indayati, T., & Faizien, M. I. (2018, September). Developing an Assessment Instrument of Higher Order Thinking Skills in Mathematics with in Islamic Context. In Journal of Physics: Conference Series (Vol. 1097, No. 1, p. 012151). IOP Publishing.

Samo, D. D. (2017). Kemampuan pemecahan masalah matematika mahasiswa tahun pertama dalam memecahkan masalah geometri konteks budaya. Jurnal Riset Pendidikan Matematika, 4(2), 141.

Sari, N. M., & Yaniawati, P. Darhim, & Kartasasmita, BG (2019). The Effect of Different Ways in Presenting Teaching Materials on Students' Mathematical ProblemSolving Abilities. International Journal of Instruction; e-ISSN: 1308-1470; p-ISSN: 1694-609X, 12(4), 495-512.

Simamora, R. E., & Saragih, S. (2019). Improving Students' Mathematical Problem Solving Ability and Self-Efficacy through Guided Discovery Learning in Local Culture Context. International Electronic Journal of Mathematics Education, 14(1), 61-72.tanujaya

Subay, R, Kartono, K., & Sulhadi, S. (2019). Validity and Reliability of Higher Order Thinking Skills (HOTS) Test Assessment in Mathematics Learning At Seventh Grade Based on The Expert Study. Journal of Research and Educational Research Evaluation, 8(2), 141-147.

Sucipto, S. (2017). Pengembangan Ketrampilan Berpikir Tingkat Tinggi dengan Menggunakan Strategi Metakognitif Model Pembelajaran Problem Based Learning. Jurnal Pendidikan (Teori Dan Praktik), 2(1), 77.

Suhirman, S., Yusuf, Y., Muliadi, A., & Prayogi, S. (2020). The Effect of Problem-Based Learning with Character Emphasis toward Students' Higher-Order Thinking Skills and Characters. International Journal of Emerging Technologies in Learning (iJET), 15(6), 183-191.

Sumintono, B. & Widhiarso,W 2015. Aplikasi Pemodelan Rasch pada Assessment Pendidikan. Cimahi : Penerbit Trim Komunikata.

Susdelina, Perdana & Febrian (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan rasch model. Jurnal Kiprah, 6(1), 41-48.

Syadiah, A. N., & Hamdu, G. (2020). Analisis rasch untuk soal tes berpikir kritis pada pembelajaran STEM di sekolah dasar. Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran, 10(2), 138-148.

Tanujaya, B., Mumu, J., & Margono, G. (2017). The Relationship between Higher Order Thinking Skills and Academic Performance of Student in Mathematics Instruction. International Education Studies, 10(11), 78-85.

Trisnawati, W., Citrasukmawati, A., & Thohir, M. A. (2017). Self Assessment For Student Performance Based on Higher Order Thinking Skills in Physics Learning. Journal of Education and Learning, 11(4), 446–452.

Widana, I. W., Parwata, I., Parmithi, N. N., Jayantika, I. G. A. T., Sukendra, I. K., & Sumandya, I. W. (2018). Higher order thinking skills assessment towards critical thinking on mathematics lesson. International journal of social sciences and humanities, 2(1), 24-32.

Tanujaya, B. (2016). Development of an Instrument to Measure Higher Order Thinking Skills in Senior High School Mathematics Instruction. Journal of education and Practice, 7(21), 144-148.