EduMa: Mathematics Education Learning And Teaching July 2023, Vol 12 No 1 Page 021 – 038 https://syekhnurjati.ac.id/jurnal/index.php/eduma/eduma/article/view/9468 p-ISSN: 2086-3918, e-ISSN: 2502-5209 Published by Tadris Matematika IAIN Syekh Nurjati cirebon



EduMa

MATHEMATICS EDUCATION LEARNING AND TEACHING

Higher Order Thinking Skills (HOTS) Test Instrument : Validity and Reliability Analysis With The Rasch Model

Rivo Panji Yudha¹

1 Faculty Of Teaching And Education, Universitas 17 Agustus 1945 Cirebon

*Corresponding author: <u>rivoyudha@yahoo.co.id</u>

article info	abstract
How to cite this article:	This article describes the instruments that can be used to measur
Panji Yudha (2023). Higher Order Thinking Skills (HOTS) Test Instrument: Validity and Reliability Analysis With The Rasch Model, 12(1), 21 – 38. doi: <u>http://dx.doi.org/10.24235/eduma.v12i1.9468</u>	students' higher order thinking skills in learning mathematics fo low, medium and high difficulty levels. This study aims t estimate the validity and determine the reliability of the highe order thinking test assessment instrument in Mathematic learning based on the Rasch model. The research was conducte through a quantitative descriptive approach in two SMP Negeri
Article history:	Gebang which had 100 respondent. The research developmen model used is the ADDIE model. This article describes the stage of development. The instrument used is a high-level mathematica
Received: 12 01, 2022	thinking test assessment instrument which contains 20 question and expert validation observation sheets. Mathematical highe
Accepted: 01 08, 2023	order thinking test questions were presented to three materia
Published: 01, 2023 EduMa: Mathematics Education Learning and	experts. The validity used is content validity and construc validity. Reliability was tested through the Rasch model approach The results showed that the HOTS assessment instrument in th form of HOTS test questions consisting of 20 items of description from the material, construction, and linguistic aspects and th appearance was declared to be constructively valid and suitabl for use. The results of the validity can be seen and analyzed with the Winsteps program in the Out fit order table to see th suitability of items that function in the normal category to be used as a measurement of student misconceptions. The results showe that the content validity through the Rasch model approach usin Winstep software, where the peak of the graph approached th value 0, the item reliability value was 0.88, item separation wa 1.16 and response separation was 2.65. The 90% test question with moderate criteria.
Teaching Copyright © 2023 under the <u>Creative</u> <u>Commons Attribution 4.0 International</u> <u>License</u> .	Keywords: Instrument; Higher Order Thinking Skills; Mathematics



Open Access

INTRODUCTION

Entering the 21st century the world is growing rapidly, technology is developing rapidly. These developments touch and affect various aspects of human life, ranging from economic, legal, transportation, to education aspects. The development of information and communication technology in the Industrial 4.0 era has a great influence on the teaching and learning process. The ease of access to technology has been used by teachers to improve the quality of education. Information technology can be accepted as a medium for carrying out the educational process, including helping the teaching and learning process, which also involves finding references and sources of information.

Assessment of learning outcomes in the mathematics subject group is carried out through observations of knowledge and skills abilities to assess students' psychomotor and cognitive development (BSNP, 2010). The assessment carried out by the teacher covers all student learning outcomes, namely cognitive or thinking abilities, psychomotor abilities or practical abilities, and affective abilities, but the emphasis on each domain is not the same so that the characteristics of the subjects to be measured must be considered (Adayani & Mardapi, 2012).

Assessment of learning outcomes at a higher cognitive level (higher-order thinking), requires a test (task) that requires students to use knowledge and skills in new or novel situations. Thus, students are not only required to understand, but to be able to analyze, evaluate and be creative (Schraw Ed. & Robinson Ed., 2011).

Mathematical problems are given to students to train themselves in using thinking skills, as well as to find out the position of the thinking level of each student, but mathematical problem solving is strongly influenced by the level of thinking skills possessed by students. Thinking ability is the ability to process information mentally or cognitively starting from low level to high level. Each student is directed to have the ability to think to the highest level so that higher order thinking is the ultimate goal in improving thinking skills. For this reason, information on the level of higher-order thinking skills possessed by each student is needed as a first step in an effort to improve thinking skills (Purbaningrum, 2017).

Research conducted by Rofiqoh, A (2014) shows that: 1) The composition of the cognitive process dimensions of competency test questions on the subject of transformation, statistics, and opportunities in the mathematics textbooks of SMP class VII 2013 curriculum is 0% considering (C1), 34 % understand (C2), 61% apply (C3), 5% analyze (C4), 0% evaluate (C5), and 0% create (C6). 2). The composition of the knowledge dimensions of competency test questions on the subject of transformation, statistics, and opportunities in the mathematics textbooks of SMP class VII 2013 curriculum is 39% conceptual knowledge, 61% procedural knowledge, 0% factual knowledge, and 0% metacognitive knowledge (Rofiqoh, 2014)

Rufiana, I. S (2016) said that the questions in the class VII 2013 curriculum student books for mathematics subjects were mostly questions of understanding as much as 68.01%, the proportion of questions of presentation and interpretation (23.67%) is larger when compared to questions of reasoning and proof (1.45%), the small percentage of questions on the form of reasoning and proof causes students to be unfamiliar with solving questions in this form (Rufiana, 2015). Rahmah, AN, & Muharni, LP J (2019) conducted a study that aims to describe the problems in the seventh grade math book based on cognitive aspects in order to achieve basic competencies and core competencies where the object of research is practice questions in the seventh grade math book semester I chapter linear equations and inequalities of one variable, the results of the study obtained information that the percentage of questions for each cognitive level was C2 (31.70%), C3 (56.09%), C4 (12.19%), C5 (0 0.0%), and C6 (0.0%) so it is recommended that the questions in the book should be improved in order to improve students' higher-order thinking skills, so that basic competencies and core competencies are achieved. This is reinforced by the results of research conducted by Rofiqoh, A (2014), Imanuddin, TN F (2015), Rufiana, I. S (2016), and Rahmah, AN, & Muharni, LP J (2019) which can be concluded that the questions that measure higher order thinking skills have a very low percentage (Rahmah & Muharni, 2019).

The results of the 2019 release of the Education Assessment Center for Research and Development Kemendikbud UNBK on HOTS items for mathematics subjects nationally for algebra material with an absorption capacity of 45.45%, calculus material with an absorption capacity of 19.24%, geometry and trigonometry material with an absorption capacity of 35.10%, and statistical material with an absorption capacity of 6.50%. Based on the results above, the national average for HOTS items in mathematics in 2019 is 26.57%, which is still in the very low category.

Most of the teachers have not made the HOTS instrument in schools optimally. In addition, it was stated that the HOTS instrument in junior high schools still experienced several obstacles including, there were still misconceptions about HOTS in the respondents studied, the lack of understanding of the majority of respondents about the procedure for making the appropriate HOTS instrument, and some respondents who quite understood the HOTS instrument. Exacerbated by the relatively high teaching burden of mathematics teachers in schools.

Budiman, A., & Jailani, J (2014) argues that the problem faced by teachers is that the teacher's ability to develop HOTS assessment instruments is still lacking in addition to the unavailability of assessment instruments specifically designed to train HOTS or students' higher-order thinking skills (Budiman & Jailani, 2014). It was further revealed that the problems that occurred at school, the questions tended to test more aspects of memory which did not train students' higher-order thinking skills, the thinking ability of Indonesian children was scientifically considered to be still low, seen from the results of the TIMSS survey. because students in Indonesia are not trained in solving questions that measure HOTS, and the problem faced by teachers is that the ability of teachers to develop HOTS assessment instruments is still lacking and there is no available assessment instrument specifically designed to train HOTS, so it is necessary to develop a HOTS assessment instrument.

Wirda, MA, Berutu, N., Rahmad, R., & Rohani, R (2017) found that from the forms of tests there is a type of test that has been standardized, namely the type of test has been tested for its validity and reliability strictly and has been tested. tested for their practical use (Wirda et al., 2017). In addition, Puspendik Balitbang Depdiknas (2007) in Wardhani, D. F., & Putra, A. P (2016) suggests that standardized tests are tests where the questions have gone through a process of analysis both qualitatively and EduMa : Education Mathematics Teaching and Learning | 23

quantitatively (Wardhani & Putra, 2016). Nisrokha (2020) adds that the standard test qualifies as a good test, namely the requirements for validity, reliability, and objectivity, standardized tests can also be used for a relatively long time and can be applied to several objects covering a wide area, in addition to standard tests. have been classified according to their age level and class (Nisrokha, 2020).

Educators should be encouraged to compile and provide higher-order thinking items to students, be encouraged to study harder and get used to seeking more information about the subject matter, not only obtained from educators or from textbooks in class. In the end it will create students who are accustomed to thinking critically in dealing with problems in everyday life. One of the tools used to assess students in learning higher order thinking is a student performance assessment model. The model for developing student performance assessment based on mathematics HOTS has not yet been felt by educators. This is a challenge for researchers to contribute to the development of student performance based on HOTS.

The results of the exam analysis start from obtaining information about students' abilities from the results of the tests carried out, which are commonly referred to as test scores. There are various ways to report scores that reflect a student's ability. A common way is to add up the number of correct answers, which indicates the student's ability. Further analysis is to perform a simple statistical procedure to be able to explain more about the quality of the questions, the quality of students and the comparison of the measured attributes. Therefore, looking for other alternatives in analyzing test results is very necessary, especially with the various weaknesses of the classical test theory above. This lack of classical test theory was then corrected by item response theory (IRT) with various variations of its logistical parameters (called PL), one of which is the 1PL developed by Georg Rasch which is called the Rasch model. In contrast to classical test theory which always depends on grades, IRT does not depend on a certain sample of questions and the abilities of the people involved in the exam (Sumintono, 2018).

Aida (2017) suggests that in the implementation of the test in the form of a description, the scoring is generally done partially based on the steps that must be taken to correctly answer an item. Scoring is done step by step and the item scores obtained by the participants are found by adding up the students' scores for each step, and the ability is estimated with the raw scores. However, such a scoring model is not necessarily correct, because the difficulty level of each step is not taken into account, so an alternative approach is needed and an alternative approach that can be used is the item response theory approach for polytomic scoring (Aida et al., 2017). Sarjono, S (2015) suggests that PCM scoring is one of the polytomic scoring models, by producing more than two categories and each item can have a different number of response categories (Sarjono, 2015).

The Rasch model is a psychometric technique that was developed to improve the accuracy of the built instrument, to monitor the quality of the instrument and calculate the respondent's performance (Boone, 2016). This is the simplest model in Item Response Theory (IRT) because it is a probabilistic model that assesses the difficulty of items and a person's abilities in such a way that they can be assessed on the same continuous scale (Deane et al., 2016). The Rasch model estimates a person's probability of choosing a

particular item or category (Mahmud & Porter, 2015; Yudha, 2016). Item difficulty and people's abilities in the Rasch Model are measured on a logit scale (Runnels, 2012).

The analysis of the Rasch model can inform researchers about the reliability of person and item, the separation of items and persons, and the value of Cronbach's alpha. While the construct validity of an instrument can be assessed through item and fit, variable map and undimensionality. Thus, the key concepts mentioned above will be used by researchers to establish evidence for the reliability and validity of the HOTS instrument using Rasch analysis.

METHODS

Reseach Design

This type of research uses a quantitative method approach, through a Research and Development model. The research development model used is the ADDIE model. The ADDIE model is an effective and efficient development model to use (Gustafson & Branch, 2002; Trust & Pektas, 2018). By using 5 stages of research (Branch, 2009; Branch, 2013; Sharifah & Faaizah, 2015; Muruganantham, 2015; Hess & Greer, 2016; Zulyadaini, 2017) namely: 1) Analyze, 2) Design, 3) Development, 4) Implementation, and 5) Evaluation. This article is at the development stage. This research is a quantitative research at SMPN 1 Gebang class VII with a number of classes, namely three classes with a total sample of 100 respondents. This research is part of the research on developing HOTS assessment in mathematics learning. The instrument used is a description of 20 items. Considerations for selecting schools based on accreditation criteria A and using the 2013 Curriculum. With these considerations, it is believed that accredited A schools are ensured to have complete facilities and support this article. Sampling based on the characteristics of special characteristics according to the purpose of the study is called a non-probability sampling technique which does not provide equal opportunities for each element or member of the population to be selected as a sample (Etikan et al., 2016). This research is part of the research on the development of the HOTS assessment in mathematics learning. The instrument used is a description of 20 items. Some researchers always use the number of samples and the number of items, for example: 1) the minimum length of the test is 20 items, the sample size is between 100-250 students for the one-parameter model (Linacre, 1994).

Frame Work Flow

Instruments that have been designed are consulted with experts to later prove the validity of their contents. Expert test or Validation, carried out with respondents from instrument or product design experts. This validation process is called Expert Judgment using expert observation sheets. The instruments that have been produced are evaluated, whether the resulting format is feasible or not, and how appropriate the content of the learning assessment material is. If the instrument is not feasible, then the instrument is revised again so that the instrument becomes feasible to be tested. There are four separate aspects of validating the HOTS assessment instrument. Instruments that have been designed are consulted with experts to later prove the validity of their contents. Expert test or validation, carried out with respondents from instrument or product design experts. This activity is carried out for initial product reviews to provide input for further instrument improvements. This validation process is called Expert EduMa : Education Mathematics Teaching and Learning | 25

Judgment using the Delphi technique. The instruments that have been produced are evaluated, whether the resulting format is feasible or not, and how appropriate the content of the learning assessment material is. If the instrument is not feasible, then the instrument is revised again so that the instrument becomes feasible to be tested. There are four separate aspects of HOTS instrument validation. The first aspect is the suitability of aspects with indicators, aspects of writing, aspects of language, and aspects of appearance. It is expected that validators will vary in their abilities. The purpose of this examination is to reliably distinguish according to the expert whether the instrument is feasible or not for use in the field. It is expected that validators will vary in their abilities. The purpose of this examination is to reliably distinguish according to the expert whether the instrument is feasible or not for use in the field.

Data Collection and Analysis

The validity of the measurement results of the instruments used is carried out through validation by experts (expert judgment). Two mathematics education experts and two evaluation experts, 3 experts from lecturers and 1 expert from senior mathematics teachers were asked to validate the performance assessment instrument. This validation includes: first, evidence related to content, and construct validation. Content validity analysis was carried out by analyzing the results of content validation by experts using the Rasch model approach. This is because using the Rasch model is a solution to the problem of validity where the Rasch model is able to provide statistics and offers an opportunity to investigate the validity of test instruments based on the responses of research subjects. The results of the validity can be seen and analyzed with the Winsteps program in the Out fit order table to see the suitability of items that function in the normal category to be used as a measurement of student misconceptions.

In this study, the estimation of the reliability of the instrument using the Rasch model approach is used to estimate the reliability of the test instrument. The instrument used is in the form of questions that have been developed and analyzed using the Rasch model. The analysis was carried out by utilizing the Winsteps application to determine the suitability of the items with the measurement model used, namely the Rasch model. The conditions that must be met in the Rasch model are the suitability of the assessment instrument model with the INFIT MNSQ acceptance limit of 0.77 to 1.30 (Adams & Kho: 1996), the level of difficulty is in the ability range of -2 to +2 (Hambelton & Swaminathan: 1985).

The results of the reliability analysis can be seen using the Winsteps program in the Summary Statistics table. The table can provide overall information about the quality of student response patterns, the instruments used, and the relationship between students and the items.

RESULT AND DISCUSSION

Expert Validity

At this stage the instruments that have been designed are consulted with experts. The instruments that have been produced are evaluated, whether the resulting format is feasible or not, and how appropriate the content of the learning assessment material is.

	Table 1. Expert General Information						
No.	Expert	Knowledge field	Last Education				
1	Ι	Mathematics	Doctor				
2	II	Learning Evaluation	Doctor				
3	III	Instruments	Doctor				
4	IV	Mathematics	Bachelor of Education				

In the trial, the HOTS instrument was validated by 4 evaluation experts. The following are the validation results from 4 evaluation experts.

Based on table 1, the researcher chose 4 experts from different points of view and with different criteria based on the wishes of the researcher but homogeneous according to their interests and their relationship to the variables to be validated from academics, practitioners, and content, to find the selected variables. From the 4 experts, comments or suggestions will be obtained in the form of sentences on research variables, addition and subtraction of the number of variables, data processing, and so on. The following are experts who meet the requirements of researchers.

The results of the four assessors' assessment of the HOTS test assessment instrument in mathematics learning are summarized in Table 2. The aspects of the assessment include: the suitability of aspects of performance assessment with existing indicators, aspects of conformity with indicators, writing, language aspects, and appearance aspects physical.

	Table 2.								
	Assessment Instrument Validation Results								
No.	Rated	Number	Expert		Average	Aspect	Presentase		
	aspect	Item	1	2	3	4		Average	
1	Aspect	1	5	5	5	4	19		
	$\operatorname{conformity}$							0.95	95%
	with	2	4	5	5	5	19	0,00	5570
	indicators								
2	Writing	3	4	4	5	4	17		
	Procedure	4	5	4	4	5	18	0,9	90%
		5	4	5	5	5	19		
3	Language	6	4	5	4	4	17		
	Usage	7	4	5	4	5	18	0,87	87%
		8	5	4	4	4	17		
4	Physical	9	5	5	4	4	18	0.005	00 50/
	appearance	10	5	4	5	5	19	0,925	92,5%
Total Score				44	43	43	181	3,645	90,5%

In Table 2, it appears that the general validation results of the four assessors of the HOTS test assessment instrument can be seen from the value of expert judgment, which

gets the highest average score for the aspect of conformity with the indicator, with a percentage of 95%.

Some of the main points that became input from the four assessors were 1) The procedure for writing language is still not quite right, for example combining or separating sentences; 2) Question number 5 does not match the HOTS indicator; 3) The language of assignment in item items is still ambiguous; 4) The instrument should measure the specific competencies that are in accordance with the HOTS test; 5) Assessment rubrics should be combined if the types of questions are the same to be more efficient. The assessment rubric is quite clear but it will take a lot of time (time consuming) to conduct a detailed assessment. Furthermore, the above inputs are used to revise the HOTS test assessment instrument for further confirmation to the validator as the input provider. The results of the four expert validations stated that the HOTS test assessment instrument was feasible to use

Characteristics of Items and Respondents

Based on the results of the analysis using the Winsteps program, it provides information, both in terms of items and respondents, showing differences in the items and students analyzed using the Rasch model, indicating the occurrence of misconceptions for some students. The test results in the form of scores were analyzed using Winsteps software. From the output of the Winsteps software, several parameter items are obtained that match the Rasch model. In addition, the value of Cronbach's alpha is obtained which is the result of the overall item reliability test. The Rasch model uses a probabilistic response distribution as a logistic function of the person and item parameters to determine the unidimensional latent trait (Yudha, 2020). The description of the distribution of the ability of 100 students and the distribution of the difficulty of the items on the same scale.

Another valuable information with Rasch modeling is being able to see the difficulty level of an item with the model or item measure for short. Item size is related to the probability of correct answer. The greater the difficulty level of the item, the higher the respondent's ability to answer correctly is needed. The greater the difficulty level of the item, the more to the right of the item's characteristics.

When we analyzed people and items at the same scale on the person-item map (Figure 2), the ability level of the person on the left side, represented by #, was related to the difficulty of the item on the right side. The students' level of ability (more abilities at the top of the left side of the map) was higher than the average difficulty of the items (the most difficult items at the top of the right), indicating that the majority of children mastered the knowledge in the test items. The difference is close to 1-logit, indicating that the average ability level of the respondents is in the average item difficulty level. All items correspond to the child's level of competence. HOTS items lie in the range of 1 logit unit, from -1-logit to +1-logit. Question P2 is the most difficult item (1 SD above the average item difficulty). Item P16 is the easiest item, all of which are 1 SD below the average. Other items P15, P17, P18, P19, P20, P7, P8, P10, P13, P14, P8, P12, P4, P5, P6, P1, P11, P12 are easy items, all of which are 0 SD on average. This result is in accordance with Boopathiraj and Chellamani (2013) who say that questions that have

high discriminating power are questions where students who have low test scores cannot answer the questions correctly (Boopathiraj, 2013)

The previous explanation proves that there are still some students who have misconceptions about the HOTS test instrument in mathematics learning, as a mathematics teacher, it is better if things like this can be minimized by understanding the actual mathematical concepts according to most experts so that students in the future will no longer have misconceptions. to mathematical materials.



Figure 1 Variable Maps

.Based on the results of data analysis, it shows that there is a valid level on the instrument construct of the questions developed so as to obtain a level of conformity between student responses and the test instrument. The following table shows how the items developed can be said to be normal or not in a measurement of student misconceptions about the HOTS test instrument.

Instrument Reliability

From the analysis, it was found that the dissociation index of 2.65 and the reliability score of 0.88 indicated that this test might be useful when applied to other test takers from the same population (Wright, 1996). Furthermore, the separation index of 2.65 and reliability of 0.88 indicates that this rater shows sufficient inter-item reliability. Although the separation index of 2.64 can be considered low, being greater than 1.00 indicates that each student item is close to heterogeneous data. The Person measure value of 2.20 logit shows the average value of all persons in working on the item. The average value is less than the logit value of 0.0, which means that the tendency of students' abilities is more than the item difficulty level.

Figure 2 provides information that questions numbered 1 to 20 are questions with a moderate level of difficulty, which ranges from -1.0 to +1.0. The last is the reliability of the mathematical modeling assessment instrument. Based on the results of the analysis, the reliability coefficient of Cronbach's alpha is 0.88. If it is adjusted to the interpretation of the reliability coefficient in Figure 2, the reliability of the assessment instrument developed is in the good category.

1	TOTAL			MODEL	IN	IFIT	OUT	FIT
	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	82.6	20.0	2.20	.46	.98	13	.96	21
SEM	.8	.0	.16	.01	.08	.22	.08	.22
P.SD	7.4	.0	1.49	.10	.75	2.12	.75	2.11
S.SD	7.5	.0	1.50	.11	.76	2.13	.75	2.12
MAX.	99.0	20.0	6.60	1.03	3.43	5.67	3.47	5.35
MIN.	61.0	20.0	56	.24	.03	-3.79	.02	-3.89
REAL	RMSE .53	TRUE SD	1.40 SE	PARATION	2.65 Per	son REL	IABILIT	Y .88
MODEL	RMSE .47	TRUE SD	1.42 SE	PARATION	3.02 Per	son REL	IABILIT	Y .90
S.E.	OF Person M	EAN = .16						i

Figure 2 Level of Appropriateness of Items

The requirement to know that the items can be categorized as acceptable or not by looking at the MNSQ scale with a range of 0.5 < MNSQ < 1.5. If you look at Figure 2, it shows that of the 10 items developed, they are in the good or accepted category, so it can be concluded that there are no misconceptions from students about these items. The ZSTD value scale is categorized as acceptable or not with a range of -2.0 < ZSTD < +2.0, indicating that from Figure 2 the items are included in the category of meeting the criteria for good items. The point measure correlation (Pt Mean Corr) value scale is categorized as acceptable or not with a range of 0.4 < Pt Mean Corr < 0.85. If the items in the three criteria (OUTFIT MNSQ, OUTFIT ZSTD, and Point Measure Correlation) are not met, it can be ascertained that the item is not good enough so it needs to be dropped (Sumintono & Widhiarso, 2015). If one or two criteria are not met, then the item can still be used for measurement and the Point Measure Correlation value is met if it is positive (Chan et al., 2020). Based on Table 4.13 above, all items meet one of the requirements of the Outfit MNSQ, Outfit ZSTD, and Point Measure Correlation values. So it can be concluded that the 20 items of the HOTS-based student performance task are fit for the

Rasch model. Indicating that from the figure 2 the questions are included in the eligible category, which means that the test instrument is accepted (Chan et al., 2020; Janssen et al., 2017; Said, 2016; Sumintoro, Bambang, Widhiarso, 2013; Tabatabaee-Yazdi, 2018).

Another valuable information with Rasch modeling is being able to see the student's ability or level of suitability of students (person fit order). Person fit order explains whether students' ability to respond to items is functioning normally or not. Some of the person fit order data can be seen in Table 3.

Measure	Count	Score	Modlse	In.Msq	In.Zstd	Out.Msq	Out.Zstd	Ptma
0.37	100	413	0.19	1.1349	0.7011	1.3002	1.3913	0.5556
0.54	100	408	0.18	1.5533	2.4316	1.419	2.2615	0.5423
0.8	100	400	0.17	1.4948	2.2915	1.4329	2.3815	0.5736
0.17	100	418	0.2	0.6886	-1.6293	0.7354	-1.3193	0.6742
0.13	100	419	0.2	0.717	-1.4493	0.6846	-1.6193	0.7113
0.17	100	418	0.2	1.1545	0.7812	1.0732	0.4111	0.6348
-0.16	100	426	0.21	0.9056	-0.3991	0.8272	-0.7992	0.6547
0.09	100	420	0.2	1.1997	0.9812	1.0108	0.121	0.5862
-0.16	100	426	0.21	0.7606	-1.1992	0.7939	-0.9792	0.6565
0.01	100	422	0.21	0.9315	-0.2691	0.8225	-0.8192	0.6621
-0.49	100	433	0.22	1.4242	1.9614	1.3503	1.5714	0.5432
0.29	100	415	0.19	1.0194	0.161	1.1978	0.9612	0.6443
0.05	100	421	0.2	0.6471	-1.8894	0.8074	-0.9092	0.7196
-0.03	100	423	0.21	0.8631	-0.6191	0.686	-1.5993	0.7048
-0.3	100	429	0.21	0.5676	-2.4694	0.5131	-2.7595	0.7313
-0.58	100	435	0.22	0.8037	-1.0192	0.7186	-1.4093	0.6793
-0.25	100	428	0.21	1.2248	1.0912	1.1366	0.6911	0.6415
-0.21	100	427	0.21	0.74	-1.3293	0.7122	-1.4393	0.699
-0.25	100	428	0.21	1.0203	0.171	0.8871	-0.4791	0.6729
-0.21	100	427	0.21	0.9692	-0.079	0.904	-0.3991	0.6739

Table 3Item Fit Order

Based on the results of the analysis of the HOTS instrument using the Winsteps program in Table 1, 20 items were found that were misfit, and 20 items were fit, so that the final HOTS instrument was 20 items.

Unidimensional is to see the items only measure one ability only. To see which HOTSbased items meet the unidimensional requirements, see Table 4.

Table 4
Unidimensional

		01 1
	Eigenvalue	Observed
Total raw variance in observation	30.7117	100.0%
Raw variance explained by measures	10.7117	48.3%
Raw variance explained by person	6.9899	22.8%
Raw variance explained by items	3.7218	28.4%
Raw unexplained variance (total)	20.000	65.3%
unexplained variance in 1 st contrast	3.8143	12.4%

EduMa : Education Mathematics Teaching and Learning | 31

unexplained variance in 2 nd contrast	2.3965	7.8%
unexplained variance in 3 rd contrast	2.2100	7.4%
unexplained variance in 4 th contrast	1.4821	4.8%
unexplained variance in 5 th contrast	1.3074	4.3%

As a general reference to determine if an item is unidimensional is to look at the value of Raw variance explained by measures of at least 40% (Ali, 2018; Conrad et al., 2017). Based on Table 2 above, the value of Raw variance explained by measures is 48.3%. Thus it can be said that the HOTS-based student performance task is undimensional.

CORREL-	ENIKY	.	ENTRY	.
ATION	NUMBER	Ite	INUMBER	Ite
	+		+	
46	11	P11	12	P12
42	15	P15	16	P16
41	19	P19	20	P20
39	8	P8	9	P8
38	18	P18	20	P20
38	14	P14	16	P16
38	18	P18	19	P19
38	7	P7	8	P8
35	6	P6	j 7	P7
46	7	P7	13	P13
41	5	P5	18	P18
40	4	P4	20	P20
38	8	P8	20	P20 İ
36	8	P8	13	P13
36	6	P6	13	P13
36	2	P2	İ 7	P7
36	8	P8	18	P18
35	3	P3	12	P12
35	1	P1	6	P6
- 34	6	P6	11	P11
		· · ·		

Figure 3

Local Independenes

Local Independence is that each item and person is independent. Thus there is no relationship between one item with another item. This also applies to respondents, each respondent is independent. Likewise, there is no relationship between one respondent and another. To see the Local Independence check, the value in the Correlin Residual column is < 0.20 (Christensen et al., 2017). So that Local Independence is fulfilled as shown in Figure 3. Based on Figure 3, all value items in the Correlin Residual column are < 0.20. Thus the HOTS-based student performance items meet Local Independence.



Figure 4 Test Information Function

In Figure 4, the X-axis shows the level of ability of the person in working on mathematical items with probability and statistical domains, while the Y-axis explains the magnitude of the information function. Person's ability starts from very low ability (far left), low ability, moderate ability (approximately Measure = 0), high ability, and very high ability (far right). Seen at a very low level of ability, the information obtained is also quite low (as well as very high capability). At the moderate level of ability, the information obtained by the measurement is very high. This shows that these items produce optimal information when given to a person with moderate ability. The conclusion from the graph above is that the 30 items given to 379 people indicate that the items are suitable for knowing a moderate level of ability. The item information function also shows the reliability of the measurements that we do, the higher the peak of the information function achieved, the higher the reliability value.

In general, the mathematical HOTS measuring instrument for seventh grade junior high school students in the first semester can be used. This instrument contains indicators of analysis, evaluation, and creation. The opinion of Arifin & Retnawati (2017) states that to make HOTS questions valid, reliable, and suitable to be used, they must include critical and creative thinking skills. Of course, the ability to think critically is clearly included in the ability to analyze and evaluate. It is further stated that the right instrument can certainly measure students' HOTS abilities. This instrument can be used to train and familiarize students with doing and dealing with HOTS questions. HOTS form questions are needed as an effort to improve students' HOTS abilities. Student activities working on or solving a problem can support the improvement of HOTS abilities. Abdullah, Abidin, Ali (2015) produced a finding that student activity in problem solving is an activity that can generate HOTS (Arifin & Retnawati, 2017).

Furthermore, it is said that the right HOTS instrument no longer contains routine questions. It is natural that the results of the difficulty level of the instrument developed by the researcher are difficult and medium levels. This means that HOTS questions are not easy questions to answer. Usually, routine questions only contain knowledge or

EduMa : Education Mathematics Teaching and Learning | 33

procedural indicators. However, this indicator does not fall within the realm of HOTS capabilities. stated that in the knowledge dimension, the level of knowledge about facts is not included in the HOTS (Pratama & Retnawati, 2018). It was further stated that the level of the task was stated that the level of the task with the HOTS character was that the steps were unpredictable, not a routine matter, containing many solutions, requiring more effort in doing it. The HOTS is based on the cognitive level, namely analysis, evaluation, and creation.

This ability must be continuously trained so that there is no confusion of students in working on math problems that are classified as high-level thinking. When students are given questions that are different from the teacher's example, they look confused to solve the problem. HOTS type questions that require high-level thinking can train students to think at the level of analysis, evaluation, and creation so that these questions must be further developed in the 2013 curriculum in order to support the improvement of students' mathematical literacy skills (Suryapuspitarini et al., 2018).

CONCLUSION AND IMPLICATION

Conclusion

The results of the general validation of the four assessors of the HOTS test assessment instrument can be seen from the expert judgment scores which get the highest average value for the aspect of conformity with the indicators, with a percentage of 95%. There are 20 items that have valid criteria with the following criteria: 1) 0.50 < MNSQ < 1.50; 2) -2.00 < ZSTD < +2.00; and 3) 0.40 < Point Measure Correlation < 0.85. If the item meets all three criteria (OUTFIT MNSQ, OUTFIT ZSTD, and Point Measure Correlation). The reliability of the HOTS instrument has good criteria with a Cronbach Alpha (KR-20) value of 0.88. The measurement information function shows respondents with moderate ability in responding to items.

Based on the results of the study, it can be concluded that the development of the HOTS test instrument using the Rasch model can detect student misconceptions on the HOTS test items with the results of data analysis showing a match between students and the test instrument used with very good quality items so that all students with appropriate abilities with an average with a logit value of +1 which means that almost all students have no misconceptions about the concept of the HOTS items being tested.

It is important to test the HOTS instrument and determine a student's ability in educational assessments. Analysis that can produce more precise measurements (producing the same interval scale) will determine the quality of the analysis results and improve the educational process to help students learn. The Rasch model can help teachers evaluate and improve the quality of the analysis performed because it applies basic principles of appropriate data processing. This is because the Rasch model addresses objective measurement requirements. The application of Rasch modeling to formative testing has many advantages because of its focus on measurement accuracy. This can be used to detect item difficulty, as well as to identify individual abilities and provide appropriate learning aids.

Implication

Based on the results of the analysis, the HOTS test assessment instrument developed has been declared valid and reliable. This instrument can be used to measure the high order thinking ability of students in learning mathematics for low, medium and high difficulty levels. Suggestions for further research is to use a larger sample in testing the HOTS test assessment instrument. This article can be used as a material to measure the HOTS ability of seventh grade junior high school students. This research can be developed further to analyze the level of HOTS of junior high school students in Indonesia. For teachers, it can be used as a reference for daily test questions in the form of HOTS to train students' HOTS activities.

ACKNOWLEDGMENTS

Thank you to the entire team of lecturers, students and teachers of SMPN 1 Gebang Cirebon Regency for their cooperation which has helped in research and discussions that were constructive and supportive so that the research was carried out to produce this paper.

REFERENCES

- Adayani, S., & Mardapi, D. (2012). Performance Assessment dalam Perspektif Multiple Criteria Decision Making. Prosiding Seminar Nasional Penlitian, Pendidikan Dan Penerapan MIPA.
- Aida, N., Kusaeri, K., & Hamdani, S. (2017). Karakteristik Instrumen Penilaian Hasil Belajar Matematika Ranah Kognitif yang Dikembangkan Mengacu pada Model PISA. Suska Journal of Mathematics Education. <u>https://doi.org/10.24014/sjme.v3i2.3897</u>
- Ali, R. (2018). Rasch Model Validation Of Instrument To Measure Students Readiness To Embedded Systems Design Course. February.
- Arifin, Z., & Retnawati, H. (2017). Pengembangan instrumen pengukur higher order thinking skills matematika siswa SMA kelas X. PYTHAGORAS: Jurnal Pendidikan Matematika. <u>https://doi.org/10.21831/pg.v12i1.14058</u>
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*. <u>https://doi.org/10.1187/cbe.16-04-0148</u>
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International journal of* social science & interdisciplinary research, 2(2), 189-193.
- Branch, R. M. (2009). Instructional Design: The ADDIE Approach. In Encyclopedia of Curriculum Studies. <u>https://doi.org/10.4135/9781412958806.n258</u>
- Branch, R. M. (2013). Robert Maribe Branch. 24–27. https://doi.org/10.13811/j.cnki.eer.2013.08.005
- BSNP. (2010). *Paradigma Pendidikan Nasional Abad XXI*. In Paradigma Pendidikan Nasional Abad XXI.

- Budiman, A., & Jailani, J. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots) Pada Mata Pelajaran Matematika Smp Kelas Viii Semester 1. Jurnal Riset Pendidikan Matematika. <u>https://doi.org/10.21831/jrpm.v1i2.2671</u>
- Chan, S. W., Looi, C. K., & Sumintono, B. (2020). Assessing computational Thinking Abilities Among Singapore Secondary Students: a Rasch Model Measurement Analysis. Journal of Computers in Education, 0123456789. <u>https://doi.org/10.1007/s40692-020-00177-2</u>
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Applied Psychological Measurement. <u>https://doi.org/10.1177/0146621616677520</u>
- Conrad, K. J., Conrad, K. M., Mazza, J., Riley, B. B., Funk, R., Stein, M. A., & Dennis, M. L. (2017). Dimensionality, Hierarchical Structure, Age Generalizability, and Criterion Validity of the GAIN's Behavioral Complexity Scale. *Physiology & Behavior*, 176(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040
- Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2016). Development of the statistical reasoning in biology concept inventory (SRBCI). CBE Life Sciences Education. <u>https://doi.org/10.1187/cbe.15-06-0131</u>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics, 5(1), 1-4.
- Gustafson, K. L., & Branch, R. M. (2002). What is Instructional Design. Trends and Issues in Instructional Design and Technology, 2, 10–16.
- Hess, A. K. N., & Greer, K. (2016). Designing For Engagement: Using The ADDIE Model To Integrate High-Impact Practices Into An Online Information Literacy Course. *Communications in Information Literacy*, 10(2), 264–282.
- Janssen, K. C., Phillipson, S., O'Connor, J., & Johns, M. W. (2017). Validation of the Epworth Sleepiness Scale for Children and Adolescents using Rasch analysis. Sleep Medicine, 33, 30–35. <u>https://doi.org/10.1016/j.sleep.2017.01.014</u>
- Linacre, J. M. (1994). Sample Size and Item Calibration or Person Measure Stability. Rasch Measurement Transactions, 7(4), 328.
- Mahmud, Z., & Porter, A. (2015). Using rasch analysis to explore what students learn about probability concepts. *Journal on Mathematics Education*. <u>https://doi.org/10.22342/jme.6.1.1937.1-10</u>
- Muruganantham, G. (2015). Developing of E-Content Package By Using ADDIE Model. 1, 52–54.
- Nisrokha. (2020). Tes Terstandarisasi. Jurnal Madaniyah, 10(1), 15-36.
- Pratama, G. S., & Retnawati, H. (2018). Urgency of Higher Order Thinking Skills (HOTS) Content Analysis in Mathematics Textbook. Journal of Physics: Conference Series. <u>https://doi.org/10.1088/1742-6596/1097/1/012147</u>

- Purbaningrum, K. A. (2017). Kemampuan Berpikir Tingkat Tinggi Siswa Smp Dalam Pemecahan Masalah Matematika Ditinjau Dari Gaya Belajar. Jurnal Penelitian Dan Pembelajaran Matematika. <u>https://doi.org/10.30870/jppm.v10i2.2029</u>
- Rahmah, A. N., & Muharni, L. P. J. (2019). Identifikasi Soal Tipe Higher Order Thinking Skills (Hots) Pada Buku Matematika Materi Persamaan Dan Pertidaksamaan Linear Satu Variabel. Edu Math Journal Prodi Pendidikan Matematika.
- ROFIQOH, A. (2014). Analisa Soal-Soal Pada Buku Siswa Pelajaran Matematika Smp Kelas Vii Kurikulum 2013 Berdasarkan Taksonomi Bloom Terevisi. Jurnal Pembelajaran Dan Pengembangan Matematika (PEMANTIK). <u>https://e-journal.unmas.ac.id/index.php/pemantik/article/view/1350</u>
- Rufiana, I. S. (2016). Level kognitif soal pada buku teks matematika kurikulum 2013 kelas VII untuk pendidikan menengah. Jurnal Dimensi Pendidikan dan Pembelajaran, 2(2), 13-22. <u>https://doi.org/10.24269/dpp.v2i2.153</u>
- Runnels, J. (2012). Using the Rasch model to validate a multiple choice English achievement test. *International Journal of Language Studies*.
- Said, R. F. M. (2016). Application of Rasch Measurement Model in Evaluating Student Performance for Foundation of Computing II. In 7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings (Issue InCULT). <u>https://doi.org/10.1007/978-981-287-664-5_58</u>
- Sarjono. (2015). Penilaian Unjuk Kerja Dalam Praktikum Fisika. *Jurnal Madaniyah*, 5(1), 109–130. https://journal.stitpemalang.ac.id/index.php/madaniyah/article/view/69
- Schraw Ed., G., & Robinson Ed., D. H. (2011). Assessment of Higher Order Thinking Skills. Current Perspectives on Cognition, Learning and Instruction. In IAP -Information Age Publishing, Inc.
- Sharifah, R., & Faaizah, S. (2015). The Development of Online Project Based Collaborative Learning using ADDIE Model. *Proceedia - Social and Behavioral Sciences*, 195, 1803–1812. <u>https://doi.org/10.1016/j.sbspro.2015.06.392</u>
- Sumintono, B. (2018, February). Rasch model measurements as tools in assessment for learning. In 1st International Conference on Education Innovation (ICEI 2017) (pp. 38-42). Atlantis Press.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.
- Sumintoro, Bambang, Widhiarso, W. (2013). Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial (B. Sumintoro (ed.); II).
- Suryapuspitarini, B. K., Wardono, & Kartono. (2018). Analisis Soal-Soal Matematika Tipe Higher Order Thinking Skill (HOTS) pada Kurikulum 2013 untuk Mendukung Kemampuan Literasi Siswa. Prisma, Prosiding Seminar Nasional Matematika.

- Tabatabaee-Yazdi, M. et. al. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. International Journal of Instruction, 11(2), 129–144. <u>https://doi.org/10.12973/iji.2018.11210a</u>
- Torrey, T., & Emrah, P. (2019). Using the ADDIE Model and Universal Design for Learning Principles to Develop an Open Online Course for Teacher Professional Development. Journal of Digital Learning in Teacher Education, 34(4), 219–233. <u>https://doi.org/10.1080/21532974.2018.1494521</u>
- Wardhani, D. F., & Putra, A. P. (2016). Pengembangan Instrumen Tes Standar Kognitif pada Mata Pelajaran IPA Kelas 7 SMP Di Kabupaten Banjar The Development of Standardized Test Instrument Natural Science Subjects Seventh Grade Junior High School in Kabupaten Banjar. Proceeding Biology Education Conference, 13(1), 75–82.
- Wirda, M. A., Berutu, N., Rahmad, R., & Rohani, R. (2017). Pengembangan Tes Standar Berasis Teknologi Informasi Di Jurusan Pendidikan Geografi. *Tunas Geografi*. <u>https://doi.org/10.24114/tgeo.v6i2.8548</u>
- Yudha, R. P. (2016). Pengembangan Instrumen Unjuk Kerja Materi Bangun Ruang Di Sekolah Dasar Kota Cirebon. Eduma: Mathematics Education Learning and Teaching. <u>https://doi.org/10.24235/eduma.v5i2.1079</u>
- Yudha, R. P. (2020). Validity And Reliability Rubric Of Performance Assessment Geometry Study In Junior High School Using The Many Facet Rasch Model Approach. Eduma: Mathematics Education Learning and Teaching. https://doi.org/10.24235/eduma.v9i2.7100
- Zulyadaini, D. (2017). A Development of Students' Worksheet Based on Contextual Teaching and Learning. *IOSR Journal of Mathematics*, 13(01), 30–38. <u>https://doi.org/10.9790/5728-1301033038</u>