HOTS Test Construction Based on Orientation to The 2013 Curriculum Assessment Standards

1st Andi Saparuddin Nur
Mathematics Education Department, Faculty of Teacher Training and Education
Universitas Musamus
Merauke, Indonesia
andisaparuddin@unmus.ac.id

2nd Hariani Fitrianti
Mathematics Education Department, Faculty of Teacher Training and Education
Universitas Musamus
Merauke, Indonesia
harianifitrianti@unmus.ac.id

3rd Murni Sianturi
Mathematics Education Department, Faculty of Teacher Training and Education
Universitas Musamus
Merauke, Indonesia
murnisianturi@unmus.ac.id

4th Muhammad Sunu Queen Irianto
SMA Negeri 2 Merauke
Merauke, Indonesia

***Abstract*—Technological developments that are increasingly rapid in the 21st century require various abilities of higher order thinking skills (HOTS) including in mathematics learning. The implementation of the 2013 curriculum is a means to prepare students to face competition in the era of globalization by prioritizing HOTS principles. The purpose of this research was to obtain HOTS-based problem construction with an orientation to the 2013 curriculum assessment standards. This research was quantitative research with factor analysis design. The number of population in this research was 257 students with a sample of 200. The research instrument was in the form of questions in the linear and matrix program material amounted to 50 items with the construction of knowledge dimension variables and HOTS levels. Data analysis techniques used GFI measurement models that met unidimensional test criteria, contract reliability (CR), and variance extracted (VE). The results showed that all items were valid and reliable and met the assumptions of multivariate normality and linearity. Factor analysis obtained indicator construction that could explain well and consistently every latent variable and GFI fit model with non-congeneric model multi-factor design. It showed that the model did not meet the unidimensional test so that a re-study of the items forming HOTS C5 level indicator was needed.**

Keywords: HOTS Problem Construction, 2013 Curriculum, Factor Analysis

I. INTRODUCTION

Mathematical learning does not only focus on memorizing concepts, but rather prioritizes the ability of higher order thinking skills (HOTS). Based on the results of research conducted by PISA in 2015, Indonesia was ranked 63 of 70 countries in terms of mathematical abilities of 15-year-old students [1]. Data from the National Examination (UN) of High School (SMA) students in 2017 also experienced a decline in the average mathematics score [2]. The instrument used in the PISA research was a matter that measured HOTS abilities such as mathematical literacy, problem solving, and mathematical communication. In the 2017 National Examination, 40% of the questions used were problems that measured HOTS. Factors that cause low ability of students in HOTS include the learning process of mathematics that is less accustomed to students in solving HOTS problems. Teachers tend to use assessment instruments that only measure factual and memorization knowledge [3]. Moreover, students' perceptions and attitudes towards mathematics have a negative tendency to give a poor assessment of mathematics [4][5].

According to *Permendikbud* (Decree of Minister of Education and Culture) 2016 Number 20 concerning competency standards at the high school level consists of three dimensions, namely the dimensions of knowledge, attitude, and skills [6] In the dimension of knowledge, students must have factual, conceptual, procedural, and metacognitive knowledge. The ability of students in 4 dimensions of knowledge must be at a technical level, specific, detailed, and complex and able to link that knowledge in a broader context. Mathematical learning should not be separated from its applicative context in everyday life as well as its assessment instruments. Assessment of the success of mathematics learning is not only related to memorizing the formula, but furthermore students must have the ability to think critically and solve problems, think creatively, communicate, and collaborate in accordance with the demands of 21st-century skills.

The 2013 curriculum that has undergone revision is a manifestation of the government's seriousness in preparing students who are able to compete in the 21st century. The 2013 revised curriculum in 2017 emphasizes the strengthening of character values according to Presidential Regulation No. 87 of 2017 which includes religiosity, nationalism, independence, mutual cooperation and integrity. By strengthening the character values, it is expected to be able to encourage students to have the skills needed in the 21st century [7]. Learning mathematics according to the demands of the 2013 curriculum is student-oriented learning with contextual learning scenarios, creating a collaborative learning atmosphere, and developing 21st-century skills. The learning is intended to enable students to develop HOTS and actualize dimensions of knowledge in a wider context.

It has become a necessity that the rapid development of science must be anticipated by fulfilling the skills needed in the community, including HOTS [8]. HOTS has long been an important issue in many developed countries even today it is increasingly developing in Southeast Asian countries including Indonesia. Mathematics education observers and practitioners also have a huge interest in using HOTS in learning [9]. The experts formulated HOTS as the ability to think creatively, critically, problem-solving, reflective, logical, and metacognitive, but the formulation did not easily obtain a strict definition agreement [10]. There is an approach that is easier to use to determine the specification of questions in the form of HOTS namely by referring to the taxonomy of bloom. The taxonomy of bloom is considered more relevant to the context of education in Indonesia with indicators of the ability to analyze, evaluate, and create [10].

The steps in preparing HOTS-based questions can be done by taking into account the signs of using verbs questions [7]. Another thing to note is the context and content of the questions related to real problems in daily life. Teachers need to emphasize mathematical literacy skills so that students can make mathematical models and solve the given problems. HOTS questions are generally objective tests in the form of multiple choices. Multiple choice questions have advantages such as an unbiased, easy assessment, can be used to measure broad competencies and can explore the level of mastery of students' concepts through their distractor functions. Developing multiple choice HOTS questions requires a number of mastery algorithms or alternative solutions that can stimulate students to express their opinions. HOTS questions with multiple choice forms must be able to distinguish students' abilities so that the objectives of learning mathematics according to curriculum standards can be achieved.

The learning used in high schools is still largely using traditional approaches and has not referred to the development of HOTS [4]. The implementation of the 2013 curriculum that has been implemented simultaneously by the government in 2018 is a challenge for teachers to adjust the learning process and assessment standards in schools. The purpose of the assessment is to obtain an understanding of the level of mastery and weakness of students, assessing the level of competence, evaluating material content, and other informative goals [11]. Assessment that is oriented only to mastering the concept results in students' weaknesses in aspects of mathematical literacy. Students are familiar with the mathematical context of direct objects such as knowing facts, concepts, operations, and principles. The questions that can be solved by students are more related to aspects of factual knowledge so that if students are given HOTS questions then they tend to be avoided. In addition, teachers are still limited in developing HOTS-based questions that measure all dimensions of knowledge.

The development of HOTS-based questions can be done independently and flexible depending on the achievement of core competencies for each material [12]. Conversely, students can also master HOTS through a learning process that supports critical thinking, creative, and problem-solving skills. The material that can be developed is HOTS based problem construction with 2013 curriculum assessment standards including linear and matrix programs. Linear and matrix programs are material taught in class XI with the material prerequisites for systems of equations and linear inequalities. The development of HOTS-based questions on linear and matrix program material is very good to use because it is directly related to students' daily learning experiences. HOTS questions are contextual and stimulate students to develop thinking literacy by using visualization to be easily understood [13].

The process of preparing standardized questions can be carried out in stages [14], namely, (1) planning the type of questions, (2) preparing items, (3) preparing questions, (4) preparing scoring techniques, (5) analyzing items, and (6) standardization of questions. Assessment instruments should measure objects that should be measured in order to obtain the validity of the conclusions from the measurement results. Construction of the questions is prepared to refer to basic competencies, but in the assessment the teacher cannot easily determine the achievement of each indicator of student success in learning. In linear and matrix program material there is a dimension of factual, conceptual, procedural, and metacognitive knowledge that must be mastered by students but to measure mastery the teacher needs an indicator so that the construction of questions can be relevant to the object to be measured. Therefore, HOTS-based problem construction with 2013 curriculum assessment standards on linear and matrix program material is important and becomes the object of this study.

II. RESEARCH METHODS

This research was a quantitative research with the form of developing the construction of questions on linear program material and matrix. The purpose of this research was to find the construction of latent variables from several manifest variables (indicators) to produce HOTS-based problem criteria with the orientation to 2013 curriculum assessment standards. The population in this research amounted to 257 students with 200 samples obtained through Slovin formula.

Data collection instruments used HOTS-based test sheets on linear and matrix program material. The number of questions used was 50 questions with assessment criteria at the cognitive level namely; analyze (C4), evaluate (C5), and create (C6) while the measured dimensions of knowledge were in the realm of conceptual, procedural, and metacognitive knowledge. The construction of HOTS-based questions in this study can be seen in Figure 1.

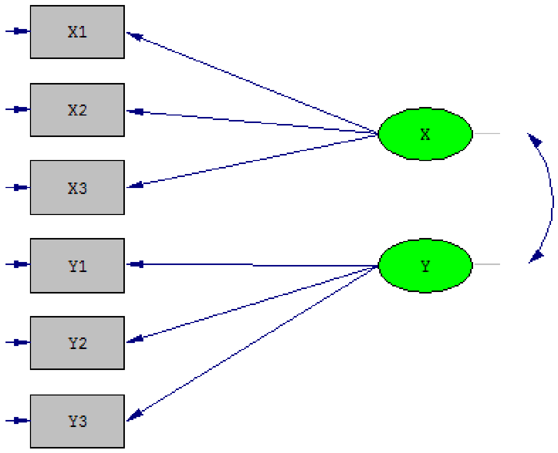

Figure 1.
Construction of HOTS-Based Problem Factor Diagram

Description:

X: Knowledge	X1: Procedural	Y1: Analyzing
Y: HOTS Level	X2: Conceptual	Y2: Evaluating
	X3: Metacognitive	Y3: Creating

The measurement equation based on figure 1 can be written as follows:

$X_i = \lambda_{xi} X + \delta_i$ dan $Y_i = \lambda_{yi} Y + \delta_i$ for i=1,2, and 3 (1)

The construction of the questions was made by using the knowledge dimension and HOTS level as a latent variable, while the items that measured each indicator of known dimensions and HOTS levels acted as manifest variables. The distribution of questions according to knowledge dimensions and HOTS levels is shown in table 1.

Table 1. Problem Indicator and Item Distribution

Knowledge dimension	**HOTS Level**			**Item number**
	C4	**C5**	**C6**	
Procedural	1,6,12,37,46,50	19,25,27,45	7,20,29	13
Conceptual	5,9,14,16,26,30,38,47	3,8,11,24,35,40,42	4,18,21,33,48	20
Metacognitive	2,10,23,39,41,43	13,15,31,34,49	17,22,28,32,36,44	17
Item number	20	16	14	50

Data analysis in this study used factor analysis to determine the conformity of manifest variables (indicators) that form latent variables [15]. The construction of questions was based on the dimensions of knowledge formed by conceptual, procedural, and metacognitive knowledge while HOTS-based questions consisted of three cognitive levels namely; C4, C5, and C6. The construction of the items produced into references was obtained by HOTS-based questions that were able to measure the dimensions of students' knowledge and mastery of HOTS levels in linear and matrix program materials. The measurement model in factor analysis is said to be feasible to use if it meets the unidimensionality test, construct reliability (CR), variance extracted (VE) and the goodness of fit indicates (GFI) as shown in table 2 [15].

Table 2. *The Goodness of Fit Indicates*

GFI Measurement	**The Goodness of Fit Model Criteria**
$P - Value$	≥ 0.05
RMSEA	≤ 0.08
CFI	≥ 0.90
GFI	≥ 0.09
AGFI	≥ 0.09
NNFI	≥ 0.09
NFI	≥ 0.09
Chi-square	≥ 0.00
Df	≥ 0.00

III. RESULTS AND DISCUSSION

Data that has been collected from the sample was analyzed descriptively. The results of the data processing are shown in table 3.

Table 3. Descriptive Statistics of Knowledge Dimensions and Students' HOTS Levels

	X1	**X2**	**X3**	**Y1**	**Y2**	**Y3**	**X**	**Y**
Mean	7.66	11.99	10.16	11.86	9.54	8.37	29.81	29.78
Median	8	12.5	10	12	10	9	31	30.50
Mode	9	12	10	14	13	10	33	30
Std. Deviation	2.47	3.804	3.157	3.482	3.045	2.857	8.042	8.084
Skewness	-.550	-.361	-.433	-.411	-.493	-.393	-.540	-.504
Kurtosis	-.257	-.484	-.116	-.434	-.249	-.201	-.157	-.193
Minimum	1.00	2.00	1.00	3.00	1.00	1.00	9.00	9.00
Maximum	13.00	20.00	17.00	19.00	16.00	14.00	46.00	46.00

Based on data processing in Table 3, it can be seen that the dimensions of conceptual knowledge (X2) were slightly more controlled by students than other dimensions of knowledge. It was indicated by the average value, median, mode and greater standard deviation. Items in each dimension of knowledge can generally be mastered by students marked with maximum values for X1, X2, and X3 which were the number of items in that aspect. On the HOTS level indicator, the C6 cognitive aspect was the most difficult aspect mastered by students. The average value, median, mode and standard deviation at C6 (Y3) were the smallest among other cognitive aspects. In general, the level of mastery of students solving HOTS-based questions on linear and matrix program material was quite well marked by median and mode values that were greater than average. It indicated that the students' scores converge was slightly above the average.

After obtaining the data description, then a factor analysis was performed to determine the construction of the problem that has met the measurement model of knowledge dimensions and HOTS levels. To conduct factor analysis, a prerequisite test was needed, namely multivariate normality test using Mahalanobis distance. Based on the results of the data from 200 samples, Mahalanobis distance the sample dataset is smaller than the required value of 26.12. This shows that there was no data outliers from the data set so that the analysis can proceed. The next test was done by determining the linearity between variables of the proposed model. The linearity test results showed that all manifest variables of knowledge dimension and HOTS level had p-value of > 0.05 and sig. of <0.05, which means that the construction of latent variables could be explained linearly by items that formed manifest variables.

Based on this, the analysis continued with factor analysis to obtain an overview of the suitability of the problem construct produced in explaining the latent variables. The results of the measurement of the initial stage factor analysis based on GFI are shown in Table 4.

Table 4. Results of GFI factor analysis on early stage

GFI measurement	Criteria	Statistical value	Conclusion
$P - Value$	≥ 0.05	0.01312	Model was not fit
RMSEA	≤ 0.08	0.0842	Model was not fit
CFI	≥ 0.90	0.992	Model was fit
GFI	≥ 0.09	0.97	Model was fit
AGFI	≥ 0.09	0.921	Model was fit
NNFI	≥ 0.09	0.985	Model was fit
NFI	≥ 0.09	0.987	Model was fit
Chi-square	≥ 0.00	19.34	*Over identified*
Df	≥ 0.00	8	*Over identified*

Table 4 showed that the measurement model was not fit with the p-value and RMSEA that did not meet the goodness of fit model. The cause of the measurement model that was not fit can be traced through the results of the parameter estimation. factor weight coefficient. CR and VE values. A summary of the results of factor analysis can be seen in table 5.

Table 5. Summary of parameter Estimation Results

Measurement model	λ	T	R^2	Errorvar	CR	VE
X					0.806	0.584
$X_1 \rightarrow X$	0.72	12.43	0.52	0.48		
$X_2 \rightarrow X$	0.81	14.23	0.65	0.35		
$X_3 \rightarrow X$	0.76	13.10	0.57	0.43		
Y					0.823	0.608
$Y_1 \rightarrow Y$	0.80	14.04	0.64	0.36		
$Y_2 \rightarrow Y$	0.78	13.72	0.62	0.38		
$Y_3 \rightarrow Y$	0.76	13.12	0.57	0.43		

Based on the results of the analysis in Tables 4 and 5, the following conclusions are obtained:

1. Test the suitability of the model is not fit so that the proposed model has not been able to generalize the phenomenon under study
2. All factor weight coefficients of the measurement model are statistically significant
3. CR and VE values so that each indicator meets consistency and composite criteria in measuring the latent variables studied. but all indicators have an estimate of less than 70% which indicates that there is no dominant indicator forming latent variables so that there is a correlation between indicators or there are variables that can be explained by other variables forming indicators in the model.

After analysis, in the diagram shown that the Y2 indicator can explain the latent variable X, so that the proposed model becomes a non-congeneric multi-factor [15]. Efforts to improve are done by correlating Y2 and X so that parameter estimation can be obtained as shown in Figure 2.

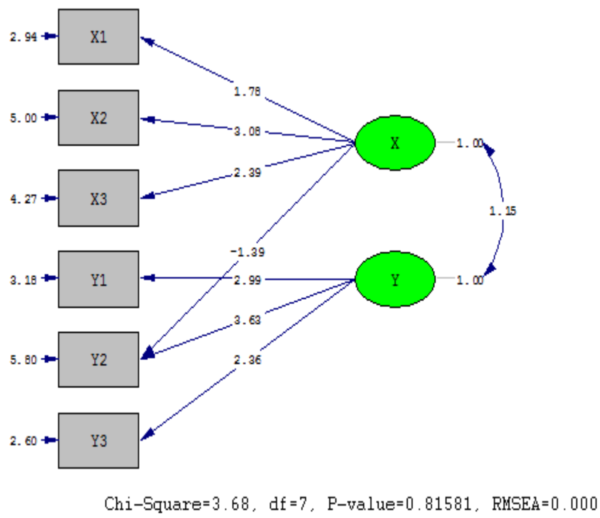

Figure 2. Parameter Estimation Results After Model Repair

In Figure 2, it can be seen that X cannot be explained unidimensionally right and consistent by the three forming indicators while Y2 could not singly explain Y, but could also explain X. The consequences of changing the measurement model caused the model specification to be re-examined based on existing theories [15]. The results of the improvement of the measurement model showed the P-value and RMSEA which means that the model was fit with the data and was suitable for use. Furthermore, a summary of the parameter estimation results after an improvement was made on the model shown in table 6.

Table 6. Summary of Parameter Estimation Results After Model Repair

Measurement model	λ	**t**	R^2	**Errorvar**	**CR**	**VE**
X					0.81	0.58
$X_1 \rightarrow X$	0.72	12.43	0.52	0.48		
$X_2 \rightarrow X$	0.81	14.23	0.65	0.35		
$X_3 \rightarrow X$	0.76	13.10	0.57	0.43		
$Y_2 \rightarrow X$	-0.46	-2.98				
Y					0.85	0.72
$Y_1 \rightarrow Y$	0.86	15.21	0.64	0.36		
$Y_2 \rightarrow Y$	1.19	7.27	0.62	0.38		
$Y_3 \rightarrow Y$	0.83	14.31	0.57	0.3		

The parameter estimation results after repairing the model were generally better with GFI meeting the goodness of fit criteria, CR, and VE. However. Unidimensional tests cannot be fulfilled which means that in the construction of the problem, there are indicators that have not been able to measure single latent variables. Knowledge dimension consisting of three indicators, namely procedural, conceptual, and metacognitive knowledge were also constructed by the ability to evaluate.

Based on the results of the study, it revealed that the ability to evaluate (C5) was related to the dimensions of knowledge they have. This is in line with the opinion [16] which mentioned the level of thinking C5 was a means of students in making judgments based on certain criteria or standards while metacognitive knowledge was general knowledge about awareness and knowledge of one's personal awareness including self-knowledge, strategic, contextual and conditional knowledge. Someone who had the ability to evaluate something would be very dependent on the knowledge of recognizing one's own abilities or strategies to be used as well as the time related to their application. C5 cognitive dimensions required a more complex construct analysis of the problem in order to reach the dimensions of students' knowledge [17]. The frequency of C5 questions was the lowest from HOTS level in revealing the dimensions of students' knowledge [11], [17]. This means it was not easy to make C5 questions that were consistently able to measure HOTS abilities without affecting the dimensions of knowledge. In line with the results of this study, Muslim & Nugraha concluded that the questions adopted from TIMSS with evaluating indicators were the character of the problem that was difficult to distinguish the dimensions of students' knowledge [18]. The problem of C5 had the ability to build all dimensions of student knowledge.

IV. CONCLUSION

The results of the study produced 50 items of linear program material and matrix that tested for validity and reliability with the results of the construction of the questions on the knowledge dimension and HOTS level met the criteria for the GFI model, CR, and VE which means that the model can be generalized to the population with each indicator can explain well the latent variables studied. However, the model obtained is not unidimensional which means that there are indicators that are HOTS C5 levels that do not consistently explain latent variables. This means that a review study on the construction of the questions is needed, especially at the HOTS C5 level, so it is recommended to the next researcher to make HOTS C5 level questions, and need to carefully formulate the questions. Efforts to minimize errors in constructing HOTS C5 level questions can be done by avoiding the sentence being evaluative to students' understanding or learning strategies because it creates interpretations that are similar to the dimensions of metacognitive knowledge.

ACKNOWLEDGMENT

We thank Universitas Musamus for facilities and supports. This project was supported by Belmawa Ristekdikti and the publication was supported by Universitas Musamus.

REFERENCES

[1] OECD, “PISA 2015: Results in focus,” *Pisa 2015*, no. 67, p. 16, 2018.

[2] Kemendikbud, “Konferensi Pers UN 2017 Jenjang SMP,” p. 23, 2017.

[3] T. Widodo and S. Kadarwati, “Higher order thinking berbasis pemecahan masalah untuk meningkatkan hasil belajar berorientasi pembentukan karakter siswa,” *J. Cakrawala Pendidik.*, vol. 5, no. 1, pp. 161–171, 2013.

[4] A. A. Adeyemi, “The Development and Confirmatory Factor Analysis of a Scale for the Measurement of Gifted Students Attitude towards Mathematics,” *World J. Educ.*, vol. 1, no. 1, pp. 52–62, 2011.

[5] N. A. Fitriana, I. Slamet, and I. Sujadi, “Classroom Assessment: Student Achievement in Mathematics through Formative Assessment,” in *Advances in Social Science, Education and Humanities Research (ASSEHR) INCOMED 2017*, 2018, vol. 160, no. 1, pp. 208–211.

[6] Kemendikbud, “Permendikbud No 20 Tahun 2016 Tentang SKL Pendidikan Dasar dan Menengah,” pp. 1–8, 2016.

[7] A. Lasmanawati, J. Abdi, I. Suyawan, and E. Suryana, *Modul Pelatihan Implementasi Kurikulum 2013 SMA Tahun 2018 Mata Pelajaran Matematika Kelas X*. Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah, 2018.

[8] S. Arikan, “Factors Contributing to Mathematics Achievement Differences of Turkish and Australian Students in TIMSS 2007 and

2011," *EURASIA J. Math. Sci. Technol. Educ.*, vol. 12, no. 8, pp. 2039–2059, 2016.

[9] A. Budiman and Jailani, "Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran Matematika SMP kelas VIII semester 1.," *J. Ris. Pendidik. Mat.*, vol. 1, no. 2, pp. 139–151, 2014.

[10] Z. Arifin and H. Retnawati, "Pengembangan Instrumen Pengukuran HOTS dalam Pembelajaran Matematika," in *Desain Pembelajaran Matematika untuk Melatihkan Higher Order Thinking Skill*, Yogyakarta: UNY Press, 2017, pp. 167–191.

[11] R. Rani and Anisha, "Construction and Standardization of Mathematics Achievement Test for IX th Grade Students," *Educ. Quest- An Int. J. Educ. Appl. Soc. Sci.*, vol. 8, no. 3, p. 629, 2017.

[12] R. Mursid, "The Effectiveness of Higher Order Thinking Skill (HOTS) Based Learning Model Through Using E-Learning and Interactive Multimedia on ICT Learning," vol. 104, no. Aisteel, pp. 217–222, 2017.

[13] W. Wayan, "Modul penyusunan higher order thingking skill (HOTS)," p. 40, 2017.

[14] S. Devi, "Construction of an Achievement Test for the students of VIII class in the Subject of Mathematics Education," no. 2277, pp. 2–4, 2013.

[15] Kusnendi, *Model Persamaan Struktural Satu dan Multigrup Sampel dengan LISREL*. Bandung: Alfabeta, 2008.

[16] Suwarto, "Dimensi Pengetahuan dan Dimensi Proses Kognitif dalam Pendidikan," *Widyatama*, vol. 19, no. 1, pp. 76–91, 2010.

[17] R. W. B. M. Mathumbu David, "Knowledge and cognitive process dimensions of Technology teachers lesson objectives," *South African J. Educ.*, vol. 34, no. 3, pp. 1–9, 2014.

[18] Muslim., A. Suhandi, and M. A. Nugraha, "Development of Reasoning Test Instruments Based on TIMSS Framework for Measuring Reasoning Ability of Senior High School Student on the Physics Concept," *IOP Conf. Ser. J. Phys. Conf. Ser.*, no. 1, pp. 1–5, 2017.