

## Development of HOTS Problem-Based Test Instruments to Measure Level 4 Numeracy Capabilities Using Rasch Model

Veni Saputri<sup>1</sup>, Hafsa Adha Diana<sup>2</sup>

<sup>1,2</sup>Program Studi Pendidikan Matematika, Sekolah Tinggi Keguruan dan Ilmu Pendidikan Media Nusantara Citra, Jakarta

Email: <sup>1</sup>[veni.saputri@stkipmnc.ac.id](mailto:veni.saputri@stkipmnc.ac.id), <sup>2</sup>[Hafsah.adha@stkipmnc.ac.id](mailto:Hafsah.adha@stkipmnc.ac.id)

### Abstract

Numeracy ability is one of the abilities measured in the Minimum Capability Assessment (AKM). Numeracy ability is very important to be improved because numeracy is not only doing mathematical calculations, but also a basis of knowledge and increases confidence to apply it practically. This research uses qualitative and quantitative approaches. This research method is research and development (R&D). The product developed in the form of a level 4 numeracy test instrument based on HOTS. The problem developed in the form of multiple choice questions and complex multiple choices is a number of 40 questions. The variables used as reference are validity, difficulty level, and reliability. This analysis was done with the help of Winsteps software. Based on winsteps program output obtained results according to rasch model with average values outfit MNSQ for persons and items respectively 0.89 and 0.91. The ZSTD Outfit values for persons and items are 0 and -0.01, respectively. While the reliability of the instrument expressed in alpha cronbach is worth 0.87.

**Keywords:** Numeracy, High Order Thinking Skill (HOTS), Rasch Measurement Model (RMM).

### Abstrak

Kemampuan numerasi merupakan salah satu kemampuan yang diukur dalam Asesmen Kemampuan Minimum (AKM). Kemampuan numerasi sangat penting ditingkatkan karena numerasi bukan hanya sekedar melakukan perhitungan matematika, namun juga merupakan dasar pengetahuan serta menambah kepercayaan diri untuk mengaplikasikannya secara praktis. Penelitian ini menggunakan pendekatan kualitatif dan kuantitatif. Metode penelitian ini adalah penelitian dan pengembangan (*Research and Development/R&D*). Produk yang dikembangkan berupa instrumen tes kemampuan numerasi level 4 berbasis soal HOTS. Soal yang dikembangkan berupa soal pilihan ganda dan pilihan ganda kompleks berjumlah 40 soal. Variabel yang dijadikan acuan adalah validitas, tingkat kesukaran, dan reliabilitas. Analisa ini dilakukan dengan bantuan *software Winsteps*. Berdasarkan output program *Winsteps* diperoleh hasil sesuai model Rasch dengan nilai rata – rata Outfit MNSQ untuk person dan item masing – masing 0.89 dan 0.91. Sedangkan nilai Outfit ZSTD untuk person dan item masing-masing 0 dan -0.01. Sedangkan reliabilitas instrumen yang dinyatakan dalam alpha cronbach bernilai 0.87.

**Kata Kunci:** Numerasi, High Order Thinking Skill (HOTS), Rasch Measurement Model (RMM).

### INTRODUCTION

21st century education is an education that develops one's numeracy skills in the face of the demands of the times. The Indonesian government will conduct a Minimum Capability Assessment (AKM) in 2021 to prepare students to face the challenges of the 21st century (Kemdikbud, 2020). The AKM includes an assessment of reasoning skills using language (reading literacy) and mathematics (numeracy). AKM is conducted nationally and in class. The national AKM does not provide reports for the individual/student level, but teachers can access classroom AKM to design learning that suits the needs of students.

Numeracy is the ability of individuals in formulating, applying, and interpreting mathematics in various contexts of daily life (Maulidina & Hartatik, 2019; Tyas & Pangesti,

2018). Ojose (2011) explained that numeracy is not just doing mathematical calculations, but also a basis of knowledge and increases confidence to apply it. A person who mastered numeracy skills will more easily adjust to the changing times. However, the numeracy ability possessed by students in Indonesia is still relatively low. This can be seen in the 2018 PISA results (OECD, 2019) which showed that 71% of students were below the minimum competence for mathematics.

Numeracy skills can be developed through hots problem-based mathematics learning (Hera & Sari, 2015; Tyas & Pangesti, 2018). HOTS problems not only measure mathematical ability in conceptual and procedural terms, but also connect several problem concepts, problem solving processes, and interpret up to decision making. Giving hots

questions can be done by providing contextual and actual questions (Tyas and Pangesti, 2018; Kamsurya and Saputri, 2020). Contextuality will provide motivation for students in solving problems because of the curiosity they have.

Improving the quality of learning in the classroom can not be separated from the assessment / assessment carried out by a teacher. According to Chan, et. al. (in Uli Sihombing, Rosita; S. Naga, Dali; Rahayu, 2019), a good assessment process must be able to measure learning outcomes accurately and produce a proper analysis of the efforts to improve ability measured. Therefore, the instrument used must be of good quality. Good instruments can be developed using the Rasch Measurement Model (RMM) approach. RMM aims to develop an objective measurement (Safihin & Hamdani, 2019). According to Supriyati, Raihanati and Nilawati (2020), RMM not only explained the quality of the problem item but also the quality of respondents in giving their responses. So that the data obtained is more precise and has a smaller error rate.

Numeracy skills can provide a capital for students to be able to face the future. AKM in Indonesia will also be implemented soon as a replacement for the National Examination in 2021. However, until now there are no instruments to measure numeracy capabilities that are in accordance with the development design of AKM problems. Currently, the national AKM is in the process of development while the class AKM can be developed by teachers in every school. So in this research, will be developed numeracy proficiency test instruments for students level 4 (grades 7-8) with the Rasch Measurement Model (RMM) approach. This research is expected to be a source of information in developing numeracy capabilities. In addition, the results of this study can also be used in AKM level 4 class. This research scheme is the research of beginner lecturers, so the expected externality of this research is in the form of publication of research results in scientific journals.

## METHOD

This research uses qualitative and quantitative approaches. This research method is research and development (R&D). The research

and development model used is a modification of the Borg & Gall model. There are seven stages of development:

- Phase 1: Literature study on the development of AKM and analysis of the needs of teachers and students in dealing with AKM in school.
- Phase 2: Planning for instrument development is done by preparing a schedule and describing the core activities that must be done.
- Phase 3: The initial product is developed in accordance with the conditions in the field and the development design of the AKM problem. Furthermore, the product will be validated by several mathematical education experts.
- Phase 4: The original product that has been valid or revised will be piloted in the initial stage to a limited sample.
- Phase 5: The results of the trial will serve as a basis for improvement to produce problems that are suitable and reliable.
- Phase 6: The revised initial product (final product) will be piloted in the field to find out the feasibility of the instruments that have been made.
- Phase 7: The final product will be revised based on the analysis of the trial results. If the instrument is in accordance both in terms of the quality of the problem and the ability of the student, it can be continued with conclusions.

The instruments used in this study are problem validation sheets, HOTS-based numeracy tests, suspension rubrics. The problem developed is BASED ON HOTS and in accordance with the form of national assessment problems consisting of multiple choices, complex multiple choices, matchmaking, short stuffing, and description. The number of problems developed in the initial product is as many as 50 questions.

The population of this study is the students of class VIII junior high school in the city of East Jakarta. The number of study samples was adjusted to the sample size for rasch modeling. The study used a rasch model with a stable aitem calibration value of  $\pm 0.5$  logit and a 95% confidence level of 64-144. (Sumintono & Widhiarso, 2014).

This study uses rasch modeling which is one of the Item Response Theory (IRT) models that can explain the interaction between students' abilities and test item/ item. Data analysis is done with the help of Winstep software. The data generated by rasch modeling has met objective measurement criteria. The results of the data analysis obtained include the quality of students, the quality of instruments, and the interaction between the two. The criteria used to see the suitability of test items (Sumintono & Widhiarso, 2014) are as follows:

- Mean Square Outfit Value (MNSQ) received:  $0.5 < \text{MNSQ} < 1.5$
- Z-Standard Outfit Value (ZSTD) received:  $-0.2 < \text{ZSTD} < +0.2$
- Point Measure Correlation Value (Pt Mean Corr):  $0.4 < \text{Pt Mean Corr} < 0.85$ .

## RESULT AND DISCUSSION

The numeracy capability test instrument developed is a hot problem-based problem that uses reference development of Minimum Capability Assessment (AKM). The domains of numeracy developed include geometry, algebra, data and certainty.

The problem developed as many as 50 problems consisting of the cognitive realm of knowledge (C1), the cognitive realm of understanding (C2), and the cognitive realm of application (C3). The developed problem was tested to junior high school students as many as 2 times trials. The first trial was conducted on a small group of 32 students, while the second trial was conducted on a large group of 78 students. This aims to determine the suitability of the model which includes the validity of the problem, the difficulty level of the problem, and the reliability of the problem.

Based on data analysis using Winstep software, there are 40 questions that are fit with rasch model and 10 problems that misfit with Rasch model. These results in full can be presented as follows.

**Table 1.** Summary of Statistical Data

	Information	Value
<b>Reliability</b>	Person Reliability	0.76
	Item Reliability	0.71
	Alpha Cronbach	0.87
<b>Outfit MNSQ</b>	Person	0.89
	Item	0.91

Outfit ZSTD	Person	0
	Item	-0.01

Table 1 shows that the Mean Square Outfit Value (MNSQ) obtained is 0.89 for person and 0.91 for items, this means that both values are located in the range of  $0.5 < \text{MNSQ} < 1.5$  so it can be concluded that the test instrument developed is in accordance with the model to measure level 4 numeracy capabilities. In addition, the Z-Standard Outfit Value (ZSTD) is 0 for persons and -0.01 for items. Both values are located in the range of  $-0.2 < \text{ZSTD} < +0.2$ . This shows that the data obtained is a rational value and can be stated that the problem item developed is in accordance with the model and can be used as a test instrument for level 4 numeracy capabilities.

Table 1 also shows an item reliability value of 0.71, person reliability of 0.76, and alpha cronbach value of 0.87. The value shows that the level of consistency of shiva in answering questions is quite high and the quality of the problem item is quite good. In addition, cronbach's high alpha value indicates that there is a significant interaction between the person and the item as a whole.

The distribution of misfit problem items with models is as follows.

**Table 2.** Misfit Distribution Data with Rasch Models.

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MODEL MEASURE	MODEL S.E.	MISFIT MNSQ	MISFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR=AL CORR	EXP	EXACT ORES	MATCH EXPS	DIFF
3	38	50	-28	.47	.92	-1.3	1.97	1.47	A .41	.45	92.5	92.5	.53
21	34	50	-51	.51	.97	1.27	2.22	1.58	B .37	.43	87.0	86.0	.521
15	18	50	2.03	.34	1.25	1.37	1.38	1.27	C .29	.36	77.5	86.6	.525
27	32	50	-07	.42	.91	1.51	.89	.04	D .05	.48	85.5	82.3	.527
4	19	50	1.91	.34	.89	.87	.76	.97	E .21	.49	58.5	70.2	.54
33	38	50	-78	.53	.73	-.65	.37	-.45	F .58	.41	90.0	83.0	.533
11	35	50	-78	.53	.73	-.65	.37	-.45	G .59	.41	90.0	83.0	.521
8	29	50	-.49	.40	.57	-2.17	.42	-1.83	H .56	.42	70.0	70.2	.55
23	34	50	-51	.51	.95	-1.02	.52	-.61	I .62	.47	93.5	89.0	.523
19	34	50	-51	.51	.95	-1.02	.52	-.61	J .65	.47	93.5	89.0	.519
MEAN	30.3	50	0.41	0.46	0.83	-0.06	0.96	0.56			83.8	83.16	
P.SD	8.13	0	0.91	0.08	0.19	1.19	0.62	0.92			11.08	7.15	

## CONCLUSION

Test instruments developed to measure level 4 numeracy capabilities have been in line with rasch's model. This is indicated by an item reliability score of 0.71, person reliability of 0.76, and alpha cronbach value of 0.87. In addition, the Mean Square Outfit Value (MNSQ) obtained is 0.89 for persons and 0.91 for items. The Z-Standard Outfit Value (ZSTD) is 0 for people and -0.01 for items. So it can be concluded that the problem item that is fit with the Rasch model is as many as 40 points of the problem.

## SUGGESTION

The author realizes that it still has flaws that the author needs to correct. This is due to the lack of knowledge of the author and limited time in the implementation of research. Therefore, constructive criticism and advice from readers is very much expected as an evaluation material for the future.

## ACKNOWLEDGMENTS

The author thanked KemenristekBrin for providing appropriate material support. Thank you also spoken to STKIP Media Nusantara Citra who has provided support for the development of this research.

## REFERENCES

- 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*, 1(2), 139. <https://doi.org/10.21831/jrpm.v1i2.2671>
- Dinni, H. N. (2018). HOTS ( High Order Thinking Skills ) dan Kaitannya dengan Kemampuan Literasi Matematika. *Prisma*, 1, 170–176.
- Hanifah, N. (2019). Pengembangan instrumen penilaian Higher Order Thinking Skill (HOTS) di sekolah dasar. *Conference Series*, 1(1), 1–8. Retrieved from <http://ejournal.upi.edu/index.php/crecs/article/view/14286>
- Hera, R., & Sari, N. (2015). Literasi Matematika: Apa, Mengapa dan Bagaimana? *Seminar Nasional Matematika Dan Pendidikan Matematika UNY 2015*, 713–720.
- Kamsurya, R., & Saputri, V. (2020). Influence of Auditory Intellectually Repetition (AIR) and Self Efficacy Learning Models on HOTS Problem-Based Problem Solving Ability. *Jurnal Ilmiah Mandala Education*, 6(2).
- Kemdikbud. (2020). *Desain Pengembangan Soal Asesmen Kompetensi Minimum 2020* 1.
- Kusuma, M. D., 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 (IOSRJRME)*, 07(01), 26–32. <https://doi.org/10.9790/7388-0701052632>
- Leder, G., & Forgasz, H. (2014). Learning from Assessment : NAPLAN and Indigenous Students. *Curriculum in Focus: Research Guided Practice (37th Annual Conference of the Mathematics Education Research Group)*, 389–396
- Maulidina, A. P., & Hartatik, S. (2019). Profil Kemampuan Numerasi Siswa Sekolah Dasar Berkemampuan Tinggi Dalam Memecahkan Masalah Matematika. *Jurnal Bidang Pendidikan Dasar*, 3(2), 61–66. <https://doi.org/10.21067/jbpd.v3i2.3408>
- OECD. (2019). PISA 2018 Results. Combined Executive Summaries. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Ojose, B. (2011). Mathematics literacy : are we able to put the mathematics we learn into everyday use? *Journal of Mathematics Education*, 4(1), 89–100.
- O’Keeffe, L., & Paige, K. (2021). Re-Highlighting the Potential of Critical Numeracy. *Mathematics Education Research Journal*, 33(2), 285–299. <https://doi.org/10.1007/s13394-019-00297-8>
- Rose, J., Low-Choy, S., Singh, P., & Vasco, D. (2020). NAPLAN Discourses: a Systematic Review after the First Decade. *Discourse*, 41(6), 871–886. <https://doi.org/10.1080/01596306.2018.1557111>
- Safihin, M., & Hamdani, H. (2019). Pengembangan Tes Menggunakan Model Rasch Materi Gaya Untuk SMA. *Jurnal Pendidikan Dan Pembelajaran*, 8(6), 1–11. Retrieved from <http://jurnal.untan.ac.id/index.php/jpdp/article/view/33424/75676581548>

- Sihombing, R. U., Naga, D. S., & Rahayu, W. (2019). A Rasch Model Measurement Analysis on Science Literacy Test of Indonesian Students: Smart Way to Improve the Learning Assessment. *Indonesian Journal of Educational Review*, 6(1), 44–55.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Rasch Model untuk Penelitian Ilmu-Ilmu Sosial (Revisi)*. Cimahi: Trim Komunikata Publishing House.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan (Cetakan I)*. Cimahi: Trim Komunikata Publishing
- Supriyati, Y., Raihanati, R., & Nilawati, W. (2020). The Development of Horizontal Anchor Items Test Tool by Rasch Model for Physics National Examination using Macromedia Flash. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 6(1), 37–50.  
<https://doi.org/10.21009/1.06105>
- Tyas, F., & Pangesti, P. (2018). Menumbuhkembangkan Literasi Numerasi Pada Pembelajaran Matematika Dengan Soal Hots. *Indonesian Digital Journal of Mathematics and Education*, 5, 566–575.
- Uli Sihombing, Rosita; S. Naga, Dali; Rahayu, W. (2019). A Rasch Model Measurement Analysis On Science Literacy Test Of Indonesian Student : Smart Way To Improve Learning Assessment. 6(1), 44–55.
- Yunita, L., Supriyati, Y., & Hariwibowo, H. (2020). Assessment Of Higher Order Thinking Skills ( Hots ) For Chemistry Pre-Service Teacher Using Computer Based Testing ( CBT ).  
<https://doi.org/10.4108/eai.30-9-2019.2291179>