Development of Rubric of Higher Order Thinking Skills Assessment on Mathematics Learning

Ega Gradini
Mathematics Education, Institut Agama Islam Negeri Takengon, Indonesia
ega.gradini@iain-takengon.ac.id

ABSTRACT

This research aims to develop a rubric of Higher-Order Thinking Skills Assessment on mathematics learning that is valid, reliable, and practical. The Plomp's developmental method, the Generic model for educational design, was deployed to develop the HOTS assessment rubric. The method consisted of 4 phases, namely; (1) preliminary investigation, (2) design, (3) development, and (4) implementation and evaluation. The rubric was designed and developed to measure students' higher-order thinking skills in mathematics learning. The rubric was developed by integrating HOTS characteristics in mathematics: critical thinking, problem-solving, mathematical understanding, mathematical modeling, proof and reasoning, mathematical representation, and mathematical communication. The HOTS assessment rubric's quality was examined using Plomp's product/prototype quality criteria: (1) validity, (2) reliability, and (3) practicality/usability. The rubric of the HOTS assessment was validated by two validators that are experts on HOTS in mathematics. The practicality/usability of the rubrics was examined by five mathematics teachers from different Junior High schools. The validity and reliability were measured using Gregory's Expert Agreement Index (EAI), while the rubric's practicality was analyzed using practicality product criteria. The research shows that the validity of the rubric of HOTS assessment on mathematics learning is 0.82 (valid), and the reliability is 0.79 (reliable). Meanwhile, the practicality of the rubric is 81.02 (Good). In conclusion, the rubric of the HOTS assessment on mathematics learning is valid, reliable, and practical to use to measure students' higher-order thinking skills in mathematics learning.

———————————— ◆ ————————————

A. INTRODUCTION

Assessment of higher-order thinking skills is still challenging for mathematics teachers in Indonesia. A study found that 79% of elementary teachers have some obstacles in designing and implementing HOTS-based evaluation (Rapih & Sutaryadi, 2018). This finding is supported by research that found teachers have difficulties conducting HOTS-oriented lesson plans and assessment formats (Gradini, 2021; Jelatu et al., 2019; Retnawati et al., 2017; Sujadi et al., 2020). Thus, it affects students' ability to solve HOTS problems (Gradini et al., 2018; Ichsan et al., 2019; Rahmawatiningrum et al., 2019; Sa'Dijah et al., 2020; Santoso et al., 2021). The HOTS content in textbooks and assessment tools is essential and significantly affects students' achievement (Pratama & Retnawati, 2018). A study showed that the test constructed by most teachers does not measure the top-three level of Bloom Taxonomy, which is the higher-order thinking skill level (Abosalem, 2016). Therefore, Malik found that teachers need a ready-to-use assessment instrument at the HOTS level (Malik et al., 2015).

Many terms define Higher-Order Thinking Skills (HOTS). Higher and lower-order thinking skills have been clearly described by many researchers and have become a trend in educational research in recent years (R. Collins, 2014; V. Collins, 2010; Conklin, 2012; Dewey & Bento, 2009; Ercikan & Seixas, 2020; Heong et al., 2011; Madhuri et al., 2012; Marshall & Horton, 2011; Marzano et al., 1988; Marzano & Kendall, 2007; Newmann, 1990; Preus, 2012; Yee et al., 2015). However, LOTS and HOTS have been initiated since 1956 by some educational experts through their critical thinking studies (Halpern, 1999; Miri et al., 2007; Resnick, 1987), the higher cognitive level of educational objectives (Anderson et al., 2001; Bloom et al., 1956; Marzano et al., 1988), creative thinking (Hyerle & Alper, 2013; Lewis et al., 2009), problem-solving (Kruger, 2013).

This study relied on HOTS defined by (Brookhart, 2010) that used three terms in defining HOTS, namely; (1) HOTS is a transfer process, (2) HOTS is critical thinking, and (3) HOTS is problem-solving. Two of the most critical educational objectives are to promote retention and transfer (which, when it happens, indicates significant learning). Students must remember what they have learned, whereas transfer necessitates remembering and making sense of and being able to apply what they have learned (Anderson et al., 2001). In other words, HOTS as a transfer process in learning is emerging meaningful learning, namely the ability of students to apply what students have learned into new situations with or without direction. As critical thinking, Brookhart retrieves the idea from (Norris & Ennis, 1989) that asserted critical thinking as a reasonable and reflective process. This theory is supported by a study that proposed critical thinking is essential in mathematics problem-solving skills (Peter, 2012).

Problem-solving is an activity that can help students hone and develop their Higher Order Thinking Skills (HOTS) in mathematics (Abdullah et al., 2015). Meanwhile, in developing HOTS as problem-solving, Brookhart refers to (Brookhart & Nitko, 2011) and (Bransford et al., 2005). HOTS as a problem-solving is a process to make students able to solve real problems in real life, which are generally unique so that the completion procedures are also unique and not routine. Consequently, to assess HOTS in mathematics learning, the teacher must examine the transfer process, critical thinking, and problem-solving skills.

Numerous studies have been conducted on the link between assessment and higher-order thinking skills. Those studies showed that increased student accomplishment was linked to the use of tasks and exams involving intellectual work and critical thinking. For example, Pogrow developed a program that deploys HOTS for educationally disadvantaged kids, students with learning difficulties. The program focuses on four different types of thinking abilities: (1) metacognition, or the ability to think about thinking; (2) inferences; (3) transfer or generalizing ideas across contexts; and (4) information synthesizing (Pogrow, 2005). Another study found that metacognitive training and instruction, both domain-general and domain-specific characteristics, have improved children's performance in various fields (Zohar & Barzilai, 2015). Like Pogrows' program, The Mathematics Learning Discourse (MLD) project reported fostering higher-order thinking and academic language in urban mathematics classrooms (Staples & Truxaw, 2010). Assessing HOTS increases students' thinking skills, achievement, and motivation (Brookhart, 2010). Meanwhile, Widana et al. (2018) found that HOTS assessment effectively increases students' critical thinking in mathematics (Widana et al., 2018). According to Ercikan and Seixas, developing assessments

that provide meaningful information are essential (Ercikan & Seixas, 2020). Higgins, Hall, Baumfield, and Moseley conducted a meta-analysis of studies on student cognitive, success, and attitude interventions of thinking skills.

They found that there is a strong effect of the implementation of higher-order thinking skills as an approach on (1) verbal and non-verbal reasoning; (2)reading, mathematics, and science tests; (3) students' attitude and motivation (Higgins et al., 2005). Thus, student improvement in thinking, content area achievement, and motivation can all be aided by thinking-skills interventions. To hone students' capability to analyze, evaluate, and create, teachers have to choose an appropriate learning model, develop good material, and use an appropriate assessment (Rosidin et al., 2019). The programs reported to assess students' higher order thinking skills using a series of tests and rubrics.

Generally, HOTS assessment measures the metacognitive dimension, not just the factual, conceptual, or procedural dimensions. The metacognitive dimension describes the ability of students to connect several different concepts, interpret, problem-solve, deploy problem-solving strategies, find new methods, reasoning, and make the right decisions. So that, in constructing an assessment instrument on HOTS, the teacher should consider the following characteristics: (1) transfer of one concept to another; (2) process and apply information; (3) looking for links from different kinds of information; (4) use information to solve problems; and (5) critically examine ideas and information. In constructing the assessment of HOTS, Brookhart suggested following these principles, namely; (1) using introductory material that is novel and allows students to gather information, and (2) managing the cognitive complexity and difficulty separately to overcome the misconception on level of difficulty and level of thinking.

The rubric has been believed as an assessment tool among teachers to examine students' higher-order thinking skills. Some studies found that rubric is effective in assessing students' higher-order thinking skills, e.g., Marzano Rubric (Marzano & Kendall, 2007), Simulation Thinking Rubric (Doolen, 2015), and Assessment Evaluation Rubric (Tractenberg, 2020). However, the study of the effectiveness of rubric in measuring students higher-order thinking comprehensively in mathematics learning is limited, although there are problem-solving rubrics (Blyman et al., 2020; Di Leo et al., 2019; Gallagher et al., 2000), critical thinking rubric (Saxton et al., 2012), and performance assessment rubric (Borko et al., 1997; Lane et al., 1994) used. Therefore, to measure students' HOTs in mathematics learning, the teacher needs to develop an appropriate rubric.

This study aims to develop a Higher-Order Thinking Skills (HOTs) assessment rubric on mathematics learning. Particularly, this study focused on the rubric to examine students' higher-order thinking skills with the criteria for the developed rubrics are valid, reliable, and practical. The HOTS assessment rubric is essential to develop since teachers have a high demand (Surya et al., 2020). The teacher also tends to evaluate students' understanding of Bloom's taxonomy's three-bottom level (Abosalem, 2016) due to their lack of knowledge and access to an instrument of assessment (Ahmad et al., 2018). It is essential to develop a rubric that teachers can use to measure/examine students' Higher-Order Thinking Skills (HOTS) in the mathematics classroom. This study contributes to filling the gap of effective HOTs assessment rubric in mathematics learning.

B. METHODS

This study is the developmental research. The Plomp's developmental method, the Generic model for educational design, was deployed to develop the HOTS assessment rubric. The method consisted of 4 phases namely, (1) preliminary investigation, (2) design, (3) realization, and (4) implementation and evaluation (Nieveen & Folmer, 2013; Tjeerd Plomp, 2000). The preliminary investigation was conducted in five activities; front-end analysis, student condition analysis, material analysis, task analysis, and specification of learning objectives. The critical element in this phase is defining the problem. In the design phase, the blueprint of the rubric is designed by generating all the parts of the solution, comparing and evaluating the various alternatives then producing the best design choice of the rubrics. The rubric was designed and developed to measure students' higher-order thinking skills in mathematics learning. In the realization phase, the rubric was constructed using the HOTs aspect defined by Brookhart as follows: (1) the top-three level of Bloom's Taxonomy (analyze, evaluate, and create); (2) logical reasoning; (3) problem-solving; (Brookhart, 2010). The rubric also adapted a rubric on problem-solving (Kennedy High School, 2006). In the implementation and evaluation phase, the HOTS assessment rubric's quality was examined using Akker's product/prototype quality criteria: (1) content validity, (2) reliability, and (3) practicality/usability (Nieveen & Folmer, 2013; Van De Akker et al., 2006).

The rubric was validated by two validators that are experts on HOTS in mathematics. If the validity coefficient is high (R > 75%), it can be stated that the HOTS rubric is valid. If this is not the case, it is necessary to make revisions based on suggestions from the validators or by reviewing aspects that have less value. Then it is re-validated and then re-analyzed until it meets the criteria. The content validity was measured using the interrater agreement of experts, as follows (Gregory, 2011):

Table 1. Interrater Agreement Model for Content Validity

| | | Expert Judge #1 | |
		Weak Relevance (item rated 1 or 2)	Strong Relevance (item rated 3 or 4)
Expert Judge #2	Weak Relevance (item rated 1 or 2)	A	B
	Strong Relevance (item rated 3 or 4)	C	D

$$Content\ validity = \frac{D}{(A+B+C+D)} \qquad (1)$$

Thatcher states that reliability is the extent to which an experiment, test, or many measurement procedures produce the same results on repeated trials (Thatcher, 2010). Reliability measured by using Kuder-Richardson Formula 21(Brown, 2014), as follow:

$$r = \frac{K}{K-1}\left[1 - \frac{K*p(1-p)}{\sigma^2 x}\right] \qquad (2)$$

Where K is test items, p is the proportion of correct responses to each test item, and q is the proportion of incorrect responses to each test item.

Table 2. Criteria of Reliability Coefficient (University of South Florida, 2021)

Reliability Score	Reliability Criteria
$0.80 \leq r \leq 1.00$	Very Good Reliability
$0.50 \leq r \leq 0.79$	Good Reliability
$0.00 < r \leq 0.49$	Not reliable

The practicality/usability of the rubrics was examined by 15 mathematics teachers from different Junior High schools. The practicality questionnaire consisted of 5 Likert-scales then analyzed using practicality product criteria.

Table 3. The Practicality Criteria of Product

Interval Score	Category
$x > \bar{x}_i + 1{,}8\,SD_i$	Very practical
$\bar{x}_i + 0.6SD_i < x < \bar{x}_i + 1{,}8\,SD_i$	Practical
$\bar{x}_i - 0.6SB_i < x < \bar{x}_i + 0{,}6\,SB_i$	Quite Practical
$\bar{x}_i - 1{,}8\,SBi < x < \bar{x}_i - 0{,}6\,SB_i$	Less Practical
$x \leq \bar{x}_i - 1{,}8\,SB_i$	Not Practical

Where $\bar{x}_i$ is mean, SD_i is standard deviation, and x is empirical score.

C. RESULT AND DISCUSSION

Blooms' Taxonomy is a commonly used taxonomy in Indonesia since its usefulness in categorizing the learning objectives and assessment into the level of cognitive. The first aspect of the rubric is assessing students' ability to analyze, evaluate, and create. Student ability in the analysis is examined from their abilities to divide or structure information into smaller parts to identify patterns or relationships, recognize and distinguish the causes and effects of a complex scenario, and identify/formulate questions. Meanwhile, the evaluation is examined from student ability to assess solutions, ideas, and methodologies using suitable criteria or existing standards to ensure their effectiveness or benefits, make hypotheses, criticize and test, and accept or reject a statement based on predetermined criteria. The creation level is measured from the student's ability to generalize an idea or perspective on something, design a way to solve the problem and organize elements or parts into a new structure that has never existed before. The second aspect of the HOTS rubric is logic and reasoning. These aspects are measured by students understanding of generating mathematical models, the quality of the mathematical model created, solution construction, the conclusion drawn, and justification (judgment). The third aspect is problem-solving. The aspect is measured by (1) students' understanding of the topics, (2) effective and appropriate problem-solving strategies and producing correct answers, and (3) written mathematical communication. The student's higher-order thinking was measured by using a four-scale rating; exemplary (4), proficient (3), develop (2), and emerging (1). The critical thing to note is that teachers need not use all the aspects of rubrics, but it depends on the level of complexity of the HOTS problem. In assessing the top-three level of Blooms, the rubrics used to depend on the problems assigned to the cognitive level. The HOTS rubric that has been constructed is described as follow:

Table 4. Rubric of Higher-Order Thinking Skills in Mathematics

HOTS Aspect	Criteria			
	Exemplary	**Proficient**	**Develop**	**Emerging**
	4	**3**	**2**	**1**
Mathematical understanding	Students have a clear and accurate understanding of the material/topic	Students have a functional understanding of the material/ topic	Students understand some of the material/ topics fairly	Students do not or have little understanding of the topic.
Analysis, evaluation, and creation level	Student answers are clear, complete, well explained, and accurately reflect knowledge of the topic.	Student answers are clear, complete but not well explained, and do not accurately reflect knowledge of the topic.	Student answers are partially clear and accurate.	Student answers are not clear, not accurate, and irrelevant.
Logical and reasoning	Students have a clear and accurate understanding of producing accurate/correct mathematical models.	Students have an adequate understanding of producing mathematical models that are primarily correct/accurate.	Students have a partial understanding which shows some difficulties in making mathematical models.	Students show inaccurate and incomplete understanding and produce incomplete/ inaccurate models.
	How the evidence supports the answer /premise/ thesis is clear, logical, and well explained.	How the evidence supports the answer/premise/the sis is mostly clear and logical. Some explanation is given.	How the evidence supports the answer/premise/ the thesis is partially clear and logical, although some explanation is given.	How the evidence supports the answer/premise/ thesis is unclear and not logical. No explanation is given.
	Students create models to simplify complex situations and identify the limitations of models	Students create models to simplify complex situations	Students create limited models to simplify complex situations	Students do not make mathematical models of the given problem
	Students construct logical, correct, complete solutions with justification and identify the source of the error.	Students construct logical, correct, complete solutions with justification	Students provide partially correct solutions with justification or correct solutions without logical steps	Students provide partially correct or incorrect solutions without justification
Problem-solving	Students analyse all the information; constraints, objectives, definitions, and implied assumptions on mathematical problems	Students analyze most of the information; constraints, objectives, and definitions given to mathematical problems	Students analyze a small amount of information; constraints and/or objectives, given to mathematical problems	Students ignore the information given to mathematical problems
	Students use/ implement effective and appropriate problem-solving strategies, and	Students use/ implement correct problem-solving strategies. However, some strategies are	Students use/ implement problem-solving strategies that result in partially	(i) Student does not deploy clear or correct problem-solving strategies; or

HOTS Aspect	Criteria			
	Exemplary	**Proficient**	**Develop**	**Emerging**
	4	**3**	**2**	**1**
	produce correct answers.	not needed or are not necessary, even though they produce the right answer.	incorrect answers.	(ii) all student answers are wrong; or (iii) student does not solve the problem.
Mathematical Communication (written)	Students use correct or appropriate mathematical and language terms to communicate their answers.	Students use mathematical terms and language that are correct or appropriate to communicate their answers.	Students provide misleading and confusing explanations to communicate their answers.	Students provide incomplete and/or inaccurate explanations.
	Students use clear and accurate mathematical representations to explain their answers.	Students use the correct mathematical representation to explain the answer.	Students use mathematical representations that are partially correct.	Students use incorrect or incomplete mathematical representations.

Table 5. Interrater Agreement of Validity of HOTS Rubric – validity 1

		Expert Judge #1	
		Weak Relevance (Item rated 1 or 2)	**Strong Relevance (Item rated 3 or 4)**
Expert Judge #2	Weak Relevance (item rated 1 or 2)	2	1
	Strong Relevance (item rated 3 or 4)	1	12

The first validation shows that the product has a validity coefficient of 0.706 with ~~a~~ reliability of 0.809. If the coefficient of validity is high (RVI > 75%), then it can be stated that the rubric is valid. Despite the reliability being very reliable, the prototype of the HOTS rubric in mathematics is not valid, yet it needs revision and validation. After doing some revision, two judges were asked to validate the rubric. The result is shown in Table 6.

Table 6. Interrater Agreement of Validity of HOTS Rubric – validity 2

		Expert Judge #1	
		Weak Relevance (Item rated 1 or 2)	Strong Relevance (Item rated 3 or 4)
Expert Judge #2	Weak Relevance (item rated 1 or 2)	0	3
	Strong Relevance (item rated 3 or 4)	0	13

The second validation shows that the product has a validity coefficient of 0.81 with a reliability of 0.89. Since the coefficient of validity is high (RVI > 75%) and the reliability is also high, then it can be stated that the HOTS rubric in mathematics is valid and reliable.

The practicality measurement is described based on the product practicality classification formula, where the assessment for the practicality of product development consists of 18 question items with a rating scale consisting of 5 categories, namely Very Good (5), Good (4), Fair (3), Poor (2), and Bad (1). By applying the practicality criteria of the product in Table 3, the practicality criteria of the rubric are shown in Table 7.

Table 7. Practicality Criteria of The Rubric

Score Interval	Category
$x > 75{,}6$	Very practical
$61{,}2 < x \leq 75{,}6$	Practical
$46{,}8 < x \leq 61{,}2$	Quite Practical
$32{,}4 < x \leq 46{,}8$	Less Practical
$x \leq 32{,}4$	Not Practical

The practicality of the rubric score is 75.08. The HOTS rubric in mathematics is practical. Furthermore, Higher-Order Thinking Skills (HOTS) rubric is valid, reliable, and practical to deploy in mathematics learning. During the trial, the respondent's positive and negative comments were used as a revision. They made positive comments on overall "user-friendliness." They made negative comments in the construction of the rubric, such as it needs to reduce some wordiness and bolding the word that shows the score. However, the respondents were still confused about the scientific terms regarding mathematical reasoning and representation. Mostly, they did not understand the "justification," "representation," and "mathematical model" terms. This problem is solved by giving defining each mathematical term below the rubrics. The respondents and validators also suggest that the rubric is the one-page format, including the column for scoring and student higher-order thinking skills conversion score. Therefore, the revision was made to make the rubric easier to use. The final version of the HOTS rubric in mathematics is shown as follows.

Rubric of Higher-Order Thinking Skills in Mathematics

Students' Name: _____   Grade: _____   HOTS Problem no: _____   Date:_____

HOTS Aspect	Exemplary 4	Proficient 3	Develop 2	Emerging 1	Score
Mathematical understanding	Students have a clear and accurate understanding of material/ topic	Students have a functional understanding of the material/ topic	Students understand some of the material/ topic fairly	Students do not or have little understanding on the material/ topic	
Analysis, evaluation, and creation level	Student answers are clear, complete, well explained, and accurately reflect knowledge of the topic.	Student answers are clear, complete, but not well explained, and do not accurately reflect knowledge of the topic.	Student answers are partially clear and accurate	Student answers are not clear, not accurate, and irrelevant	
Logical and reasoning	Students have a clear and accurate understanding in producing accurate/correct mathematical models.	Students have adequate understanding in producing mathematical models that are mostly correct/accurate.	Students have partial understanding so that it shows some difficulties in making mathematical models.	Students show inaccurate and incomplete understanding and produce incomplete/inaccurate mathematical models	
	The way in which the evidence supports the answer /premise/ thesis is clear, logical, and well explained.	The way in which the evidence supports the answer /premise/thesis is mostly clear and logical. Some explanation is given.	The way in which the evidence supports the answer/ premise/thesis is partially clear and logical, although some explanation is given.	The way in which the evidence supports the answer/ premise/thesis is not clear and not logical. No explanation is given.	
	Students create models to simplify complex situations and identify limitations of models	Students create models to simplify complex situations	Students create limited models to simplify complex situations	Students do not make mathematical models of the given problem	
	Students construct logical, correct, complete solutions with justification and identify the source of the error.	Students construct logical, correct, complete solutions with justification	Students provide partially correct solutions with justification or correct solutions without logical steps	Students provide partially correct or incorrect solutions without justification	
Problem-solving	Students analyse all the information; constraints, objectives, definitions, and implied assumptions given to mathematical problems	Students analyse most of the information; constraints, objectives, and definitions given to mathematical problems	Students analyse a small amount of information; constraints and/or objectives, given to	Students ignore the information given to mathematical problems	

Figure 1. The snapshot of the Final version of Rubric of Higher-Order Thinking Skills in Mathematics

Recall that the rubric is used for each student, giving the higher-order thinking skill score. The teacher has to map the level of each student by following the HOTS level category.

Table 8. The HOTS Level Category

Score Interval	HOTS Level
$3.50 < x \leq 4.00$	*Exemplary*
$3.00 < x \leq 3.50$	*Proficient*
$2.50 < x \leq 3.00$	*Develop*
$1.00 \leq x \leq 2.50$	*Emerging*

The HOTs level category, as shown in Table 8 used to map the level of students' higher-order thinking skills that were measured by the rubrics. The score of each aspect is measured by rubric then sum-up to gain the total student score. The average score is then calculated and converted to measure student higher-order thinking skills, as shown in Table 9. Further, the students' higher-order thinking skills are divided into exemplary, proficient, developed, and

emerging levels. This level is not in hierarchy, e.g., students need not be on the 'emerging' level first, then on the 'develop' level.

Students on the 'emerging' level are described as students who cannot solve the problems by deploying the concept and formula. Students provide incorrect or incomplete mathematical representations, explanations, and solutions at this level. They also show a lack of understanding, and the answer is unclear, not accurate, and irrelevant. Students on the 'develop' level are described as students who understand the material, although the answers are partially clear and accurate. They had some difficulties making mathematical models, delivering the evidence to support their premise, and somehow creating limited models to simplify complex situations. Students on this level cannot provide correct solutions with logical steps. They also analyze a small amount of information given to mathematical problems that affect misleading and confusing explanations in communicating their answers.

Students on the 'proficient' level have a functional understanding of the material. Therefore, although not well explained, their answers are clear and complete and do not accurately reflect knowledge. Students have an adequate understanding of producing mathematical models and delivering the evidence to support their answer on this level. They can also create models to simplify complex situations and construct logical solutions. Students can analyze most of the information given and deploy appropriate problem-solving strategies. However, some strategies are not needed or are not necessary, even though they produce the right answer. In communicating their answers, they use appropriate mathematical terms and representations.

Students have a clear and accurate understanding and produce accurate mathematical models on this highest level of HOTs, the 'exemplary' level. The evidence to support the answer is clear, logical, and well explained. Students can create models to simplify complex situations and identify the limitations of models. Moreover, they can construct logical, correct, complete solutions with justification and identify the source of the error. They can also analyze all the mathematical problems, deploy effective and appropriate problem-solving strategies, and produce correct answers. Students use appropriate mathematical terms and representations to explain their answers in communicating their answers.

D. CONCLUSION AND SUGGESTIONS

The higher-Order Thinking Skills (HOTS) rubric is valid, reliable, and practical to assess students' HOTS in mathematics learning. The validity coefficient is 0.81 (high validity), the reliability coefficient is 0.89 (very reliable), and the practicality score is 75.08 (practical). The final version of the rubrics is ready to be presented to a bigger group of mathematics teachers to gain some feedback. The teacher feedback is valuable to refine the rubrics. However, the rubric needs to be trialed to measure its effectiveness in assessing the students' higher-order thinking skills. Future research on the HOTS rubric is also needed to see the mathematics teachers' acceptance and usefulness.

REFERENCES

Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving Higher Order Thinking Skills (HOTS) problems for the topic of fraction. *Asian Social Science*, *11*(21), 133–142. https://doi.org/10.5539/ass.v11n21p133

Abosalem, Y. (2016). Assessment Techniques and Students' Higher-Order Thinking Skills. *International Journal of Secondary People*, *4*(1), 1–11. https://doi.org/10.11648/j.ijsedu.20160401.11

Ahmad, S., Prahmana, R. C. I., Kenedi, A. K., Helsa, Y., Arianil, Y., & Zainil, M. (2018). The instruments of higher order thinking skills. *Journal of Physics: Conference Series*, *943*(1). https://doi.org/10.1088/1742-6596/943/1/012053

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. In *Longman* (2nd ed). Pearson Education Inc. https://doi.org/10.2307/2281462

Bloom, B. S., Engelheart, M. D., Furst, E. J., Walker H, H., & Krathwohl, D. . (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1. Cognitive Domain. In *Longman, Green and Co Ltd* (1st ed.). David McKay Company, Inc. https://doi.org/10.1300/J104v03n01_03

Blyman, K., Arney, K., Adams, B., & Hudson, T. (2020). Does Your Course Effectively Promote Creativity? Introducing the Mathematical Problem Solving Creativity Rubric. *Journal of Humanistic Mathematics*, *10*(2), 157–193. https://doi.org/10.5642/jhummath.202002.09

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, *13*(3), 259–278.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2005). How people learn: Brain, mind, experience, and school. In R. R. Cocking (Ed.), *Social Sciences*. National Research Council, National Academy Press.

Brookhart, S. M. (2010). How To Assess Higher-Order Thinking Skills in Your Classroom. In *Association for Supervision and Curriculum Development*. Association for Supervision and Curriculum Development. https://doi.org/10.1177/002205741808801819

Brookhart, S. M., & Nitko, A. J. (2011). Strategies for Constructing Assessments of Higher-Order Thinking Skills. In G. Schraw & D. H. Robinson (Eds.), *Assessment of Higher Order Thinking Skills* (pp. 153–168). Information Age and Publishing. https://doi.org/10.4018/978-1-7998-2314-8.ch008

Brown, J. D. (2014). Classical Theory of Validity. In A. J. Kunnan (Ed.), *The Companion to Language Asessment: Vol. III* (1st ed.). https://doi.org/10.14361/transcript.9783839428566.toc

Collins, R. (2014). Skills for the 21st Century : Teaching Higher-Order Thinking. *Curriculum and Leadership Journal: An Electronic Journal for Leaders in Education*, *12*(14), 4–7. http://www.curriculum.edu.au/leader/teaching_higher_order_thinking,37431.html

Collins, V. (2010). Higher order thinking (HOT) program assessment plan. *Fort Worth, Texas: University of North Texas Health Science Center*.

Conklin, W. (2012). Higher-Order Thinking Skills to Develop 21st Century Learners. In *Shell Education*.

Dewey, J., & Bento, J. (2009). Activating children's thinking skills (ACTS): The effects of an infusion approach to teaching thinking in primary schools. *British Journal of Educational Psychology*, *79*(2), 329–351. https://doi.org/10.1348/000709908X344754

Di Leo, I., Muis, K. R., Singh, C. A., & Psaradellis, C. (2019). Curiosity… Confusion? Frustration!

The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology*, *58*, 121–137.

Doolen, J. (2015). Psychometric properties of the simulation thinking rubric to measure higher order thinking in undergraduate nursing students. *Clinical Simulation in Nursing*, *11*(1), 35–43. https://doi.org/10.1016/j.ecns.2014.10.007

Ercikan, K., & Seixas, P. (2020). Assessment of Higher Order Thinking: the Case of Historical Thinking. In G. Schraw & D. R. Robinson (Eds.), *Assessment of Higher Order Thinking Skills* (1st ed., pp. 153–168). Infor. https://doi.org/10.4018/978-1-7998-2314-8.ch008

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V, Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*(3), 165–190.

Gradini, E. (2021). Penyusunan Soal Tes Bermuatan Lower dan Higher-Order Thinking Skills pada Guru SMP. *JIPEMAS: Jurnal Inovasi Hasil Pengabdian Masyarakat*, *4*(3), 1–2. https://doi.org/http://dx.doi.org/10.33474/jipemas.v4i3.10121

Gradini, E., Firmansyah, B., & Noviani, J. (2018). Menakar Kemampuan Berpikir Tingkat Tinggi Calon Guru Matematika Melalui Level HOTS Marzano. *EduMa*, *7*(2), 41–48.

Gregory, R. J. (2011). *Psychological Testing: History, Principles, and Applications* (7th editio). Pearson.

Halpern, D. F. (1999). Thought and knowledge an Introduction to critical thinking. In *Education* (Issue 80). https://doi.org/10.1002/tl.8005

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. Bin, & Mohamad, M. M. B. (2011). The Level of Marzano Higher Order Thinking Skillsamong Technical Education Students. *International Journal of Social Science and Humanity*, *1*(2), 121–125. https://doi.org/10.7763/ijssh.2011.v1.20

Higgins, S., Hall, E., Baumfield, V., & Moseley, D. (2005). *A Meta-analysis of the Impact of the Implementation of Thinkings Skills approach on Pupils* (Vol. 338, Issue October).

Hyerle, D., & Alper, L. (2013). Thinking Maps for Meetings of the Mind. In A. L. Costa & P. W. O'Leary (Eds.), *The power of the Social Brain: Teaching, Learning, and Interdependent Thinking* (pp. 102–114). Teacher College Press, Columbia University.

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research*, *8*(4), 935–942. https://doi.org/10.12973/eu-jer.8.4.935

Jelatu, S., Mandur, K., Makur, A. P., Nendi, F., & Gunur, B. (2019). Konstruksi Tes High Order Thinking Skills (HOTS) bagi Guru-Guru Matematika SMP di Manggarai Timur. *E-Dimas: Jurnal Pengabdian Kepada Masyarakat*, *10*(2), 214. https://doi.org/10.26877/e-dimas.v10i2.3070

Kennedy High School. (2006). *Critical Thinking and Problem Solving - Mathematics*. Kennedy High School. https://www.waterbury.k12.ct.us/2/Content2/460

Kruger, K. (2013). *Higher-Order Thinking . Hidden Sparks*. New York Ic.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (1994). Reliability and validity of a mathematics performance assessment. *International Journal of Educational Research*, *21*(3), 247–266.

Lewis, A., Smith, D., & Lewis, A. (2009). Defining Higher Order Thinking. *Theory Into Practice*, *32*(March 2015), 131–137. https://doi.org/10.1080/00405849309543588

Madhuri, G. V., Kantamreddi, V. S. S. N., & Prakash Goteti, L. N. S. (2012). Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education*, *37*(2), 117–123. https://doi.org/10.1080/03043797.2012.661701

Malik, A., Ertikanto, C., & Suyatna, A. (2015). Deskripsi Kebutuhan HOTS Assessment Pada Pembelajaran Fisika dengan Metode Inkuiri Terbimbing. *Prosiding Seminar Nasional*

Fisika (E-Journal) SNF2015, *IV*, 1–4.

Marshall, J. C., & Horton, R. M. (2011). The Relationship of Teacher-Facilitated, Inquiry-Based Instruction to Student Higher-Order Thinking. *School Science and Mathematics*, *111*(3), 93–101. https://doi.org/10.1111/j.1949-8594.2010.00066.x

Marzano, R. J., Brandt, R. S., Jones, B. F., Presseisen, B. Z., & Suhor, C. (1988). *Dimensions of Thinking: A Framework for Curricullum and Instruction*. Association for Supervision and Curriculum Development.

Marzano, R. J., & Kendall, J. S. (2007). Praise for the Second Edition of The New Taxonomy of Educational Objectives. *Corwin Press*, i–ii.

Miri, B., David, B., & Uri, Z. (2007). Purposely Teaching for the Promotion of Higher-order Thinking Skills : A Case of Critical Thinking. *Research of Science and Education*, *37*(3), 353–369. https://doi.org/10.1007/s11165-006-9029-2

Newmann, F. M. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies*, *22*(1), 41–56. https://doi.org/10.1080/0022027900220103

Nieveen, N., & Folmer, E. (2013). Formative Evaluation in Educational Design Research. In T Plomp & N. Nieveen (Eds.), *Netherlands Institute for Curriculum Development: SLO* (pp. 152–169). Netherlands Institute for Curriculum Development: SLO. http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ815766

Norris, S. P., & Ennis, R. H. (1989). *Evaluating Critical Thinking. The Practitioners' Guide to Teaching Thinking Series. Box 448,* (Issue September). Critical Thinking Press and Software.

Peter, E. E. (2012). Critical thinking: Essence for Teaching Mathematics and Mathematics Problem Solving Skills. *African Journal of Mathematics and Computer Science Research*, *5*(3), 39–43. https://doi.org/10.5897/ajmcsr11.161

Plomp, Tjeerd. (2000). Development Research in / on Educational Development. *CITE Seminar Series*, *04*. http://hdl.handle.net/10722/44041

Pogrow, S. (2005). HOTS Revisited: A Thinking Development Approach to Reducing The Learning Gap After Grade 3. *Phi Delta Kappan*, *87*(1), 64–75.

Pratama, G. S., & Retnawati, H. (2018). Urgency of Higher Order Thinking Skills (HOTS) Content Analysis in Mathematics Textbook. *Journal of Physics: Conference Series*, *1097*(1). https://doi.org/10.1088/1742-6596/1097/1/012147

Preus, B. (2012). Authentic Instruction for 21st Century Learning : Higher Order Thinking in an Inclusive. *American Secondary Education*, *40*(3), 59–79.

Rahmawatiningrum, A., Kusmayadi, T. A., & Fitriana, L. (2019). Student's ability in solving higher order thinking skills (HOTS) mathematics problem based on learning achievement. *Journal of Physics: Conference Series*, *1318*(1). https://doi.org/10.1088/1742-6596/1318/1/012090

Rapih, S., & Sutaryadi, S. (2018). Perpektif Guru Sekolah Dasar Terhadap Higher Order Tinking Skills (HOTS): Pemahaman, Penerapan dan Hambatan. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, *8*(1), 78. https://doi.org/10.25273/pe.v8i1.2560

Resnick, L. B. (1987). *Education and Learning to Think*. National Academy Press. https://bit.ly/3imoWkT

Retnawati, H., Munadi, S., Arlinwibowo, J., Wulandari, N. F., & Sulistyaningsih, E. (2017). Teachers' difficulties in implementing thematic teaching and learning in elementary schools. *New Educational Review*, *48*(2), 201–212. https://doi.org/10.15804/tner.2017.48.2.16

Rosidin, U., Suyatna, A., & Abdurrahman, A. (2019). A combined HOTS-based assessment/STEM learning model to improve secondary students' thinking skills: A development and evaluation study. *Journal for the Education of Gifted Young Scientists*,

7(3), 435–448. https://doi.org/10.17478/jegys.518464

Sa'Dijah, C., Sa'Diyah, M., Sisworo, & Anwar, L. (2020). Students' mathematical dispositions towards solving HOTS problems based on FI and FD cognitive style. *AIP Conference Proceedings*, *2215*(April). https://doi.org/10.1063/5.0000644

Santoso, T., Cholily, Y. M., & Syaifuddin, M. (2021). An Analysis of Students' Errors in Completing Essay HOTS Questions Based On Watson's Criteria Viewed From The Cognitive Style Perspective. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *5*(1), 121. https://doi.org/10.31764/jtam.v5i1.3776

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, *17*(4), 251–270.

Staples, M. E., & Truxaw, M. P. (2010). The Mathematics learning discourse project: Fostering higher order thinking and academic language in urban mathematics classrooms. *Journal of Urban Mathematics Education*, *3*(1), 27–56.

Sujadi, I., Kurniawati, I., & Wulandarai, A. N. (2020). Upaya Meningkatkan Pembelajaran Berorientasi HOTS Bagi Guru Matematika SMP Kota Surakarta dengan Pemanfaatan Hasil Program AKSI for. *DEDIKASI : Community Service*, *2*(1), 1–8.

Surya, A., Sularmi, ., Istiyati, S., Wahyuningsih, T., & Sriyanto, . (2020). *Analysis of The Elementary School Teacher's Need in The Implementation of HOTS (Higher Order Thinking Skills) based on Mathematics Learning*. *ICMIs 2018*, 517–521. https://doi.org/10.5220/0008524705170521

Thatcher, R. W. (2010). Validity and Reliability of Quantitative Electroencephalography. *Journal of Neurotherapy*, *14*(2), 122–152. https://doi.org/10.1080/10874201003773500

Tractenberg, R. (2020). *The Assessment Evaluation Rubric: Promoting learning and learner - centered teaching in higher education*. https://doi.org/10.31235/osf.io/bvwhn

University of South Florida. (2021). *Reliability and Validity Classroom Assessment*. https://fcit.usf.edu/assessment/basic/basicc.html

Van De Akker, J., Graveimejer, K., & McKenney, S. (2006). *Introducing Educational Design Research* (1st Editio). Routledge Publishers.

Widana, I. W., Parwata, I. made Y., Parmithi, N. N., Jayantika, I. G. A. T., Sukendra, K., & Sumandya, I. W. (2018). Higher Order Thinking Skills Assessment towards Critical Thinking on Mathematics Lesson. *International Journal of Social Sciences and Humanities (IJSSH)*, *2*(1), 24–32. https://doi.org/10.29332/ijssh.v2n1.74

Yee, M. H., Yunos, J. M., Othman, W., Hassan, R., Tee, T. K., & Mohamad, M. M. (2015). Disparity of Learning Styles and Higher Order Thinking Skills among Technical Students. *Procedia - Social and Behavioral Sciences*, *204*(November 2014), 143–152. https://doi.org/10.1016/j.sbspro.2015.08.127

Zohar, A., & Barzilai, S. (2015). Metacognition and Teaching Higher Order Thinking (HOT) in Science Education: Students' Learning, Teachers' Knowledge, and Instructional Practice. In R. Wegerif, L. Li, & J. C. Kaufman (Eds.), *The Routledge International Handbook of Research on Teaching Thinking* (1st ed., p. 14). Routledge: Taylor & Francis Group.