HOTS MATHEMATICS TEST FOR JUNIOR HIGH SCHOOL STUDENTS: CONTENT VALIDITY ANALYSIS

Lovieanta Arriza*[1)], Zafrullah Zafrullah[2], Heri Retnawati[3], Debi Pranata[4], Rama Angger Wibowo[5], Ilham K.[6]
[1,2,3,4,5]*Universitas Negeri Yogyakarta, Indonesia*
[6]*Universitas Gadjah Mada, Indonesia*
[1*)]lovieanta0341pasca.2022@student.uny.ac.id

Received: 1-12-2024; **Accepted**: 5-4-2025; **Published**: 30-04-2025

Abstract. This study aims to analyze the content validity of a HOTS-based mathematics test designed for seventh-grade students at junior high school. The test instrument was developed based on the Core Competencies and Basic Competencies, and aligned with Higher-Order Thinking Skills (HOTS) indicators derived from Bloom's taxonomy, specifically targeting the levels of analyzing (C4), evaluating (C5), and creating (C6). Five descriptive test items were constructed to reflect these cognitive levels. The research employed a descriptive analysis method, using Aiken's V formula to assess content validity. Validation data were obtained from six experts in the fields of mathematics education, educational assessment, and junior high school mathematics teaching. The results showed that all five items were "valid," with four items scoring an Aiken's V index of 1.00 and one item scoring 0.96. These findings indicate that the developed test items have strong content validity and are suitable for assessing the HOTS abilities of seventh-grade students. This instrument can be utilized as a reliable tool in formative or summative assessments to support the development of students' higher-order thinking in mathematics.
Keywords: *Instrument, HOTS, Mathematics, Content Validity*

1. INTRODUCTION

The 21st century requires an education that creates a creative, innovative and competitive generation [1], [2]. The learning process must develop students' ability to find solutions to problems in global life maturely through a set of skills called Higher-Order Thinking Skills (HOTS) which include critical thinking, problem solving, metacognition, communication, collaboration, information literacy, and creative innovation [3], [4]. The term 21st century skills, which includes a set of knowledge, skills, behaviors and characters, is very important for learners to master in order to survive in a rapidly changing world [5], [6]. With HOTS, students will be able to distinguish various ideas clearly, can design hypotheses, argue well, can construct explanations, understand complex things more clearly, and be able to solve problems. [7], [8]. This is expected to have an impact on learners who are able to live an ideal life through academic life and a brilliant career in the future.

Teachers have an important role to play in helping learners develop HOTS skills [9], [10] through various methods or strategies [11], [12], [13]. To support this, the learning process requires a test instrument as a measuring tool for the level of students' HOTS abilities [14], [15]. The Ministry of Education and Culture of the Republic of Indonesia also emphasizes the importance of HOTS integration in the assessment process [16]. For

ISSN: 2620-4129
E-ISSN: 2621-3885

example, in mathematics learning, teachers need to develop tests that refer to non-routine problems as an effective step in stimulating students' HOTS skills [17], [18].

There are several HOTS concepts that are well known in education. HOTS in Bloom's taxonomy which was later refined by Anderson and Krathwohl, includes the competencies of analyzing (C4), evaluating (C5), and creating (C6) [19]. Meanwhile, Marzano divides higher-order thinking skills into 13 categories, including comparing, classifying, inducing, inferring, and problem-solving [20]. In this study, the HOTS instrument developed refers to Bloom's taxonomy, focusing on analyzing, evaluating, and creating competencies to match curriculum demands and measurable cognitive levels.

The use of HOTS instruments in mathematics assessment trains students to solve problems not through memorization, but through analysis, synthesis, and creativity—skills that are crucial in the 21st century [21]. Getting learners accustomed to HOTS-based assessments will help them tackle novel problems, adapt to new environments, and make decisions in unfamiliar situations [22], [23], [24].

However, in practice, the availability of mathematics HOTS test instruments that meet strong validation standards, particularly content validity, remains limited. The preparation of mathematics HOTS instruments must fulfill the element of validity to be usable. Validity refers to the degree of accuracy supported by empirical facts and theoretical reasoning [25], [26], [27]. An instrument is said to have high validity if it performs its measurement function according to its intended purpose [28], [29].

Types of validity can be broadly classified into content validity, construct validity, and criterion-related validity [30], [31]. Among these, content validity is fundamental, especially in the initial stages of instrument development [32]. It refers to the alignment between test content and the construct being measured, typically judged by experts [33], [34], [35] . Since it is judgmental, experts may differ in their evaluations, but their level of agreement can be quantified [36].

Content validity can be achieved when the instrument items align with curriculum content, representing a sample of the learning outcomes intended by the curriculum objectives [37], [38]. Instrument items should effectively measure the specific skills and knowledge outlined, just as appropriate tools are needed to measure specific characteristics [39]. This relevance ensures that the instrument accurately represents the intended learning targets[40]. Thus, to ensure the credibility and relevance of HOTS assessment in mathematics, a rigorous analysis of content validity is essential. This study aims to fill this gap by empirically analyzing the content validity of HOTS-based mathematics test items for junior high school students, using expert judgment quantified through Aiken's V. The findings are expected to support the development of high-quality assessment tools that align with both curriculum goals and 21st-century learning demands.

ISSN: 2620-4129
E-ISSN: 2621-3885

2. RESEARCH METHOD

The research employed a descriptive method. The object of the study was a mathematics test instrument based on Higher-Order Thinking Skills (HOTS) designed for seventh-grade junior high school students. The subjects involved in the content validity assessment consisted of six experts from various fields, including specialists in mathematics assessment and measurement, mathematics education lecturers, and junior high school mathematics teachers who teach seventh grade. These experts were asked to evaluate each test item in terms of content, test construction, and language. Data were collected using questionnaires that allowed for quantification, as well as qualitative instrument review sheets.

The content validity of the HOTS-based mathematics test instrument for seventh-grade junior high school students was determined using Aiken's V coefficient. Aiken's V is a content validity index formulated to quantify the degree of agreement among a panel of n experts regarding how well each item represents the construct being measured. The Aiken's V statistic is calculated using the following formula:

$$V = \frac{\Sigma s}{n(c-1)} \quad (1)$$

Keterangan:

V : Koefisien V Aiken's

S : r – lo

r : Figures given by an expert

c : The highest number (in this case = 5)

lo : The lowest number (in this case = 1)

The results of the calculation and analysis using Aiken's V formula were interpreted based on the categorization of content validity. The categorization refers to the criteria for the content validity index proposed by [41], as presented in Table 1 below.

Table 1. Content Validity Criteria Index

No.	Criteria Indexs	Description
1.	< 0,70	Disappeared
2.	0,70 – 0,79	Need Revision
3.	≥ 0,80	Maintained

Based on Table 1, items with an index of ≥ 0.80, as assessed by more than three experts, are considered to have strong evidence of content validity and should be retained. Items with a content validity index ranging from 0.70 to 0.79 require revision, while those with an index below 0.70 are considered to lack sufficient validity and should be eliminated.

ISSN: 2620-4129
E-ISSN: 2621-3885

3. RESULTS AND DISCUSSION

3.1 Results

Proving the content validity of the instrument uses a qualitative and quantitative analysis approach. From the qualitative side, the experts gave many suggestions for improvement in the aspect of word writing. This can be seen in Table 2 below.

Table 2. Experts' Suggestions on the Developed HOTS Instrument

Expert	Suggestion
Expert 1	Overall, the questions are very good and can be used, but some parts need improvement, such as the use of terms that are less precise and ambiguous.
Expert 2	The questions provided already include HOTS indicators.
Expert 3	In general, the questions are good and can be used to measure HOTS capabilities.
Expert 4	Overall, the questions are good and can be used for assessment activities in accordance with the taught material.
Expert 5	Replace certain terms to better align with the context of the questions, ensuring they are generally understandable by all students.
Expert 6	1. Improve the image illustrations to better match the questions.
	2. The question indicators are too specific, making it difficult to develop questions flexibly. Create less specific indicators.

Meanwhile, quantitatively, the content validity coefficient was calculated based on Aiken's concept. The results of the analysis using Aiken's V formula are presented in Table 3.

Table 3. Results of Aiken's V calculation

Item Number	Expert						Aiken's Index
	1	2	3	4	5	6	
1	5	5	5	5	5	5	1,00
2	5	5	5	5	5	5	1,00
3	5	5	5	5	5	5	1,00
4	5	5	5	5	5	4	0,96
5	5	5	5	5	5	4	1,00

The results of the quantitative content validity analysis showed that the five items that had been compiled all met high content validity and supported the overall test content that had been reviewed in terms of material (content), test construction, and language. The details of the content validity of the instrument consist of four items having an Aiken's index of 1.00 and the remaining one item has an Aiken's index coefficient of 0.96.

ISSN: 2620-4129
E-ISSN: 2621-3885

3.2. Discussion

The initial product produced in this study is a HOTS-based description form mathematics test instrument for seventh-grade junior high school students. The description test instrument was used because the description test is a form of test that is more appropriate to measure students' thinking and problem-solving skills [42]. The description test form also has the potential to be used in developing HOTS instruments. In other words, the preparation of HOTS instruments will be easier to do by using the description test form [43]. The test items were prepared based on the Basic Competencies (KD) on the subject matter of Integers. The test items were then developed into five question indicators. The material tested on this test instrument is the arithmetic operations of addition and subtraction of whole numbers.

The test items that have been developed based on KD and HOTS Indicators are then reviewed by the researchers themselves to check whether the test instrument items are in accordance with the aspects to be measured and check whether there are still errors in writing and so on. The next assessment was conducted by a panel of experts consisting of mathematics education experts, mathematics measurement and assessment experts, and educational practitioners, namely teachers who teach seventh grade mathematics. The experts provided an assessment of the quality of the test instrument items in terms of material coverage, construction and language used. This second stage aims to prove the content validity of the HOTS-based mathematics test instrument that has been developed. Content validity relates to the extent to which the feasibility or relevance of the instrument content covers the entire material being measured [44]. The questions or items in the instrument are analyzed and matched with the subject matter that has been given, whether the answers to the questions are related to the subject matter or not [45].

Proving the content validity of a good instrument needs to use qualitative and quantitative analysis approaches [46]. Qualitative review is also known as theoretical assessment. This theoretical analysis is carried out by considering aspects of material coverage, construction, and language used in the instrument. The technique for analyzing test items was carried out using the panel technique. The theoretical review from experts resulted in input to revise several aspects of instrument writing, such as changing some words or diction, changing the context of problems that are not in accordance with the cognitive development of students, and shortening certain sentences so that the problems in the test items become more rational and easily understood by students. The validation sheet also includes a revision column for each item, indicating whether or not the indicator or rubric, question formulation, options, pictures or tables, answer key, and scoring rubric need to be improved. The experts then filled in the form with general suggestions, which can be seen in Table 2.

The use of appropriate terminology in assessment instruments is very important as it affects clarity, understanding and fairness in the assessment process. Clear and easy-to-understand language helps ensure that each learner understands the intent of the question, so

ISSN: 2620-4129
E-ISSN: 2621-3885

that they can answer correctly according to their ability and knowledge, which will reflect their true ability [47]. If the language used is confusing or overly complex, it can lead to misinterpretation, which ultimately results in an inaccurate assessment of the student's ability.

Meanwhile, quantitative content validity analysis is carried out by calculating the content validity coefficient based on Aiken's concept. There are three types of quantitative content validity coefficients offered by Aiken's, namely the V, R and H coefficients [48]. In this study, quantitative analysis was used using the V coefficient formula. In proving content validity, researchers can choose to determine the number of rating category scales to be used. This study used a 5-category rating scale with six experts as assessors.

The experts' assessment of all test items that have been developed shows the quality of the instrument with very good criteria. This indicates that all items can be maintained and can be used to measure the mathematical HOTS skills of seventh-grade junior high school students on the sub-material of arithmetic operations of addition and subtraction of whole numbers. The results of the Aiken's formula V analysis are presented in Table 3.

The magnitude of Aiken's V index in Table 3 proves that the instrument has been valid [49] and can be used to measure the HOTS ability of seventh grade junior high school students. This is in line with the research [50], [51] and [52] which states that content validity generates confidence that instruments that are compiled and meet high content validity are instruments that are considered to be in accordance with the indicators and aspects to be measured.

Content validity is the basis that must be met in developing instruments [53]. Proof of the content validity of a test instrument will determine how far the content in the instrument is able to reflect all indicators and levels of knowledge to be measured [54]. It also emphasizes a careful process of item development and content validity analysis to improve the quality of instruments to be used in empirical studies and is key to supporting the construct validity of such instruments [55], [56]. It should also be noted that the result of the content validity analysis is an assessment based on the decision of a panel of experts (subjective element). The more experts involved, the better the quality of the instrument for the purpose of measuring students' HOTS levels [57].

4. CONCLUSION

Based on the results of the research and discussion, it can be concluded that the developed mathematics HOTS test instrument for seventh-grade junior high school students meets the elements of high content validity based on the assessment of the expert panel consisting of mathematics education lecturers, mathematics measurement and assessment experts, and also teachers who teach junior high school mathematics. The developed test instrument has been able to measure what it is intended to measure, which in this case is the HOTS ability of seventh-grade students. This can be seen from the content elements that are in accordance with the subject matter, basic competencies, HOTS indicators, and the stage

ISSN: 2620-4129
E-ISSN: 2621-3885

of cognitive development of students. The review of the test construction also showed that the instrument has illustrated HOTS ability, encouraged students to think at a higher level, and also the items have used operational verbs based on Bloom's taxonomy levels. The expert assessment also agreed that the instrument was appropriate based on the language aspect, where the use of words and language in the items was formulated straightforwardly and clearly, and did not cause multiple interpretations.

Based on the research results and conclusions presented, the researcher suggests that understanding and ability in developing test instruments is very important to master so that the instruments developed meet the elements of high content validity and are suitable for use. Although the instrument is based on theory and expert panel assessment, the instrument development process should be continued by proving its construct validity and reliability.

REFERENCES

[1] B. Thornhill-Miller *et al.*, "Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education," *J Intell*, vol. 11, no. 3, p. 54, Mar. 2023, doi: 10.3390/jintelligence11030054.

[2] L. I. González-Pérez and M. S. Ramírez-Montoya, "Components of education 4.0 in 21st century skills frameworks: Systematic review," *Sustainability*, vol. 14, no. 3, p. 1493, Jan. 2022, doi: 10.3390/su14031493.

[3] L. Greenstein, *Assessing 21st Century Skills: A guide to evaluating mastery and authentic learning*. California: Corwin, 2012.

[4] R. H. Mardhiyah, S. N. F. Aldriani, F. Chitta, and M. R. Zulfikar, "Pentingnya keterampilan belajar di abad 21 sebagai tuntutan dalam pengembangan sumber daya manusia," *Lectura : Jurnal Pendidikan*, vol. 12, no. 1, pp. 29–40, 2021, doi: 10.31849/lectura.v12i1.5813.

[5] K. Larsson, "Understanding and teaching critical thinking: A new approach," *Int J Educ Res*, vol. 84, pp. 32–42, 2017, doi: 10.1016/j.ijer.2017.05.004.

[6] M. Miterianifa, A. Ashadi, S. Saputro, and S. Suciati, "Higher order thinking skills in the 21st century: Critical thinking," in *Proceedings of the 1st International Conference on Social Science, Humanities, Education and Society Development*, A. Zainudin, Y. Prihatin, Suyono, P. Susongko, and Taufiqullah, Eds., EAI, Jan. 2021. doi: 10.4108/eai.30-11-2020.2303766.

[7] W. I. Redhana, "Mengembangkan keterampilan abad ke-21 dalam pembelajaran kimia," *Jurnal Inovasi Pendidikan Kimia*, vol. 13, no. 1, 2019, doi: 10.15294/jipk.v13i1.17824.

[8] D. Darmaji, A. Astalini, D. A. Kurniawan, and N. Prameswari*, "Analysis science process skills, arnguing ability and digital literacy of MAN 5 Batanghari students based on gender differences," *Jurnal Pendidikan Sains Indonesia*, vol. 11, no. 2, pp. 221–235, Apr. 2023, doi: 10.24815/jpsi.v11i2.27678.

[9] I. R. Suwarma and S. Apriyani, "Explore teachers' skills in developing lesson plan and assessment that oriented on higher order thinking skills (HOTS)," *Journal of Innovation in Educational and Cultural Research*, vol. 3, no. 2, pp. 106–113, Jan. 2022, doi: 10.46843/jiecr.v3i2.66.

ISSN: 2620-4129
E-ISSN: 2621-3885

[10] M. M. Huda and P. Rais, "Improving arabic language learning based on higher order thinking skills (HOTS) in excellent senior high school," *Fenomena*, vol. 20, no. 2, pp. 283–296, Nov. 2021, doi: 10.35719/fenomena.v20i2.68.
[11] G. Carel, N. Jusniani, and E. Monariska, "Kemampuan higher order thinking skills dalam pembelajaran metakognitif ditinjau dari persepsi siswa," *Pythagoras: Jurnal Pendidikan Matematika*, vol. 16, no. 2, pp. 204–216, Dec. 2021, doi: 10.21831/pythagoras.v16i2.37926.
[12] H. Hamzah, M. I. Hamzah, and H. Zulkifli, "Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules," *Sustainability*, vol. 14, no. 2, p. 813, Jan. 2022, doi: 10.3390/su14020813.
[13] P. Janson Silaban, B. Sinaga, and E. Syahputra, "Mathematics education based on Toba Batak culture learning model to improve the HOTS capabilities of prospective elementary school teachers," *The Effectiveness Of Developing The Realistic Mathematics Education Based On Toba Batak Culture Learning Model To Improve The HOTS Capabilities Of Prospective Elementary School Teachers*, vol. 30, no. 5, pp. 5625–5644, 2024, doi: 10.53555/kuey.v30i5.3184.
[14] H. Tambunan and T. Naibaho, "Performance of mathematics teachers to build students' high order thinking skills (HOTS)," *Journal of Education and Learning (EduLearn)*, vol. 13, no. 1, pp. 111–117, Feb. 2019, doi: 10.11591/edulearn.v13i1.11218.
[15] J. Jihannita, Z. K. Prasetyo, and I. Wilujeng, "How to prepare HOTS to face the 21st century?," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 8, pp. 486–492, Aug. 2023, doi: 10.29303/jppipa.v9i8.2847.
[16] C. P. Tanudjaya and M. Doorman, "Examining higher order thinking in Indonesian lower secondary mathematics classrooms," *Journal on Mathematics Education*, vol. 11, no. 2, pp. 277–300, 2020, doi: 10.22342/jme.11.2.11000.277-300.
[17] N. Azid, R. Md. Ali, I. El Khuluqo, S. E. Purwanto, and E. N. Susanti, "Higher order thinking skills, school-based assessment and students' mathematics achievement: Understanding teachers' thoughts," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 11, no. 1, pp. 290–302, Mar. 2022, doi: 10.11591/ijere.v11i1.22030.
[18] M. N. Kholid, M. H. Mahmudah, N. Ishartono, F. G. Putra, and B. Forthmann, "Classification of students' creative thinking for non-routine mathematical problems," *Cogent Education*, vol. 11, no. 1, Dec. 2024, doi: 10.1080/2331186X.2024.2394738.
[19] L. W. Anderson *et al.*, *A taxonomy for learning teaching and assessing: A revision of Bloom's taxonoy of educational objectives*. Boston: Addison Wesley Longman, 2001.
[20] Y. M. Heong, W. B. Othman, J. B. M. Yunos, T. T. Kiong, R. Bin Hassan, and M. M. B. Mohamad, "The level of marzano higher order thinking skills among technical education students," *International Journal of Social Science and Humanity*, pp. 121–125, 2011, doi: 10.7763/ijssh.2011.v1.20.
[21] S. D. Harahap, M. Ahmad, and J. W. Pasaribu, "The development of HOTS questions based on ethnomathematics to measure students' mathematical problem-solving abilities," *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran*, vol. 10, no. 2, pp. 553–564, Jun. 2024, doi: 10.33394/jk.v10i2.11511.

ISSN: 2620-4129
E-ISSN: 2621-3885

[22] W. Wiyaka, E. F. Prastikawati, and A. P. Kusumo Adi, "Higher-order thinking skills (HOTS)-based formative assessment: A proposed model for language learning assessment," *Vision: Journal for Language and Foreign Language Learning*, vol. 9, no. 2, pp. 115–130, Oct. 2020, doi: 10.21580/vjv9i25859.
[23] H. Retnawati, H. Djidu, E. Apino, and R. D. Anazifa, "Teachers' knowledge about higher-order thinking skills and its learning strategy," *Problems of Education in the 21st Century*, vol. 76, no. 2, pp. 215–230, 2018, doi: 10.33225/pec/18.76.215.
[24] M. H. Yee, M. F. Mahmad, T. K. Tee, B. C. Kok, and N. Azid, "The level of marzano higher-order thinking skills among polytechnic students," *Hong Kong Journal of Social Science*, vol. 57, pp. 381–388, 2021, Accessed: Dec. 14, 2022. [Online]. Available: http://hkjoss.com/index.php/journal/article/view/444/0
[25] M. T. Kalkbrenner, "A practical guide to instrument development and score validation in the social sciences: The MEASURE Approach," *Practical Assessment, Research, and Evaluation*, vol. 26, no. 1, pp. 1–18, 2021.
[26] J. Lee, E.-H. Lee, and D. Chae, "eHealth literacy instruments: Systematic review of measurement properties," *J Med Internet Res*, vol. 23, no. 11, p. e30644, Nov. 2021, doi: 10.2196/30644.
[27] Y.-Y. Wang and Y.-S. Wang, "Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior," *Interactive Learning Environments*, vol. 30, no. 4, pp. 619–634, Apr. 2019, doi: 10.1080/10494820.2019.1674887.
[28] R. Radeswandri, A. Budiawan, R. Vebrianto, and M. Thahir, "Developing instrument to measure the use of online comic as educational media," *Journal of Education and Learning (EduLearn)*, vol. 15, no. 1, pp. 119–126, Feb. 2021, doi: 10.11591/edulearn.v15i1.18961.
[29] A. Falk, A. Becker, T. Dohmen, D. Huffman, and U. Sunde, "The preference survey module: A validated instrument for measuring risk, time, and social preferences," *Manage Sci*, vol. 69, no. 4, pp. 1935–1950, Apr. 2023, doi: 10.1287/mnsc.2022.4455.
[30] E. Septianingsih and M. A. Jerusalem, "Developing instrument of academic potential test analogy verbal ability for undergraduate students," *Journal of Education and Learning (EduLearn)*, vol. 15, no. 2, pp. 234–241, May 2021, doi: 10.11591/edulearn.v15i2.14220.
[31] L. Qiao *et al.*, "Development and validation of the coronary heart disease scale among the system of quality of life instruments for chronic diseases QLICD-CHD (v2.0) based on classical test theory and generalizability theory," *Int J Gen Med*, vol. Volume 17, pp. 1975–1989, May 2024, doi: 10.2147/IJGM.S447752.
[32] V. H. Sheth, N. P. Shah, R. Jain, N. Bhanushali, and V. Bhatnagar, "Development and validation of a risk-of-bias tool for assessing in vitro studies conducted in dentistry: The QUIN," *J Prosthet Dent*, vol. 131, no. 6, pp. 1038–1042, Jun. 2024, doi: 10.1016/j.prosdent.2022.05.019.
[33] S. Ruel, J. El Baz, D. Ivanov, and A. Das, "Supply chain viability: Conceptualization, measurement, and nomological validation," *Ann Oper Res*, vol. 335, no. 3, pp. 1107–1136, Apr. 2024, doi: 10.1007/s10479-021-03974-9.
[34] W. A. D. Silva, T. R. de Sampaio Brito, and C. R. Pereira, "COVID-19 anxiety scale (CAS): Development and psychometric properties," *Current Psychology*, vol. 41, no. 8, pp. 5693–5702, Aug. 2022, doi: 10.1007/s12144-020-01195-0.
[35] L. Zhang, J. D. Basham, and R. A. Carter, "Measuring personalized learning through the Lens of UDL: Development and content validation of a student self-report

instrument," *Studies in Educational Evaluation*, vol. 72, p. 101121, Mar. 2022, doi: 10.1016/j.stueduc.2021.101121.
[36] S. Azwar, *Reliabilitas dan validitas*, IV. Yogyakarta: Pustaka Pelajar, 2021.
[37] T. M. Alqahtani, F. D. Yusop, and S. H. Halili, "Content validity of the constructivist learning in higher education settings (CLHES) scale in the context of the flipped classroom in higher education," *Humanit Soc Sci Commun*, vol. 10, no. 1, p. 268, May 2023, doi: 10.1057/s41599-023-01754-3.
[38] B. Becerra, P. Núñez, C. Vergara, D. Santibáñez, D. Krüger, and H. Cofré, "Developing an instrument to assess pedagogical content knowledge for evolution," *Res Sci Educ*, vol. 53, no. 2, pp. 213–229, Apr. 2023, doi: 10.1007/s11165-022-10042-0.
[39] Asrul, R. Ananda, and Rosnita, *Evaluasi pembelajaran*. Bandung: Citapustaka Media, 2014.
[40] Inanna, Rahmatullah, and M. Hasan, *Evaluasi pembelajaran: Teori dan praktek*. Klaten: Tahta Media Group, 2021.
[41] D. F. Polit, C. T. Beck, and S. V. Owen, "Is the CVI an acceptable indicator of content validity? Appraisal and recommendations," *Res Nurs Health*, vol. 30, no. 4, pp. 459–467, Aug. 2007, doi: 10.1002/nur.20199.
[42] E. P. Widoyoko, *Teknik penyusunan instrumen penelitian*. Pustaka Pelajar, 2012.
[43] A. Yani, *Cara mudah menulis soal HOTS (higher order thinking skills): Suatu pendekatan "jarak nalar" yang dilengkapi dengan pembelajaran berorientasi keterampilan berpikir tingkat tinggi*. Refika Aditama, 2019.
[44] A. Qodir, *Evaluasi dan penilaian pembelajaran*. Yogyakarta: K-Media, 2017.
[45] Haryanto, *Evaluasi pembelajaran: Konsep dan manajemen*. Yogyakarta: UNY Press, 2020.
[46] H. Retnawati, *Validitas reliabilitas dan karakteristik butir*. Pramana Publishing, 2016.
[47] T. Stoeckel, S. McLean, and P. Nation, "Limitations of size and levels tests of written receptive vocabulary knowledge," *Stud Second Lang Acquis*, vol. 43, no. 1, pp. 181–203, Mar. 2021, doi: 10.1017/S027226312000025X.
[48] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educ Psychol Meas*, vol. 45, no. 1, pp. 131–142, Mar. 1985, doi: 10.1177/0013164485451012.
[49] Munali, W. Rahayu, and H. Haribowo, "Analysis of the content validity of assessment for learning based on HOTS in high school," in *AIP Conf. Proc*, 2023, p. 040026. doi: 10.1063/5.0126298.
[50] A. Susanta, E. Susanto, and H. Sumardi, "The development of HOTS-based mathematical problem using local context for senior high school," *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, vol. 11, no. 4, pp. 3334–3347, Dec. 2022, doi: 10.24127/ajpm.v11i4.6219.
[51] M. Gabriela and M. Susana, "Content validity of a questionnaire to assess parental involvement in education," *European Journal of Psychology and Educational Research*, vol. 4, no. 2, pp. 83–95, Dec. 2021, doi: 10.12973/ejper.4.2.83.
[52] P. Hartini, H. Setiadi, and E. Ernawati, "Cognitive domain analysis (LOTS and HOTS) assessment instruments made by primary school teachers," *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 25, no. 1, pp. 16–24, Jun. 2021, doi: 10.21831/pep.v25i1.34411.

ISSN: 2620-4129
E-ISSN: 2621-3885

[53] Kurniawati, "Analisis validitas isi instrumen tes berpikir kritis ips kelas v sd kota yogyakarta," *Pelita: Jurnal Penelitian dan Karya Ilmiah*, vol. 21, no. 1, pp. 130–140, 2021, doi: 10.33592/pelita.v21i1.1396.
[54] B. St. Marie, A. Jimmerson, Y. Perkhounkova, and K. Herr, "Developing and establishing content validity of vignettes for health care education and research," *West J Nurs Res*, vol. 43, no. 7, pp. 677–685, Jul. 2021, doi: 10.1177/0193945920969693.
[55] M. Suhaini, A. Ahmad, and N. M. Bohari, "Assessments on vocational knowledge and skills: A content validity analysis," *European Journal of Educational Research*, vol. 10, no. 3, pp. 1529–1540, Jul. 2021, doi: 10.12973/EU-JER.10.3.1529.
[56] E. M. Panggabean, F. Haryati, and S. Wahyuni, "Development of mathematics assessment instruments for high school students based on higher-order thinking skills (HOTS)," *AL-ISHLAH: Jurnal Pendidikan*, vol. 14, no. 4, pp. 5393–5400, Sep. 2022, doi: 10.35445/alishlah.v14i4.1411.
[57] S. Masfuah, F. Fakhriyah, I. Wilujeng, and D. Rosana, "The content validity of scientific literacy-based diagnostic assessment," in *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, 2021, pp. 684–691. doi: 10.2991/assehr.k.210305.099.

ISSN: 2620-4129
E-ISSN: 2621-3885