

Research Article

HOTS assessment in circulatory system learning: Validity, reliability, and item quality



Yosi Laila Rahmi^{a,1*}, Isnaini Nur Habibah^{a,2}, Z. Zulyusri^{a,3}, Rahmawati Darussyamsu^{a,4}

^aBiology Department, Faculty of Mathematics and Natural Science, Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar, Padang, West Sumatera 25131, Indonesia

¹yosibio@fmipa.unp.ac.id*; ²isnaininurhabibah5@gmail.com; ³zulyusri0808@gmail.com, ⁴rahmabio@fmipa.unp.ac.id

* Corresponding author

ARTICLE INFO

Article history

Received: 8 February 2021

Revised: 15 March 2021

Accepted: 17 March 2021

Published: 31 July 2021

Keywords

Biology learning
Circulatory system
HOTS assessment

ABSTRACT

Biology learning, based on the curriculum-2013, requires students to be skilled at analyzing, evaluating, and creating to achieve Higher-Order Thinking Skills (HOTS). Thus, HOTS assessment instrument is crucial to be made. This study aimed to produce HOTS assessment instrument focused on circulatory system materials for XI graders. This research was conducted by using Research and Development (R&D) method with 4-D model. The subjects of this study were: two biology lecturers from Faculty of Mathematics and Science of Universitas Negeri Padang, two biology teachers, and 33 XI graders of Public Senior High School 1 Nan Sabaris. The data were collected using questionnaires which were validated by experts and analyzed using ANATES 4.09. The validity results reached as high as 87.54% (very valid) in which the empirical validity value was 88%. Meanwhile, the instrument was stated as very reliable (0.78), has a moderate difficulty, good differentiating power, and excellent option quality. Thus, as the **recommendation, this instrument can be used to foster students' HOTS.**



Copyright © 2021, Rahmi et al
This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



How to cite: Rahmi, Y. L., Habibah, I. N., Zulyusri, Z., & Darussyamsu, R. (2021). HOTS assessment in circulatory system learning: Validity, reliability and item quality. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 7(2), 171-178. doi: <https://doi.org/10.22219/jpbi.v7i2.15513>

INTRODUCTION

The curriculum plays an important role in improving the success of education. Therefore, it is a necessary curriculum that can make learners have high thinking and analysis power to be more active and creative during learning. The application of the 2013 curriculum in the current education system aims to encourage students to observe, ask questions, reason, and communicate what they get in school (Fanani, 2018). The level of understanding of learners is necessary to assess student ability. Assessment is collecting and processing information to measure the achievement of learners' learning outcomes and aims to monitor and evaluate the process, the progress of learning, and improvement of learners' learning outcomes on an ongoing basis (Hairun, 2020; Permana et al., 2021).

To carry out the assessment and reach the stage of analyzing and interpreting the information obtained, learners must have a high level of thinking skills during the learning process (Paidi et al., 2020; Permana & Setyawan, 2020). Higher-Order Thinking Skills (HOTS) is an ability to provide more information, foster critical thoughts, evaluate, have the ability to think metacognition, and solve problems. Bloom's taxonomy revised by Bloom's high-level thinking assessment has three indicators: analyzing, evaluating, and creating.

The high level of thinking ability of learners in Indonesia is still relatively low, as seen from the results of several international studies that measure the high level of thinking ability of learners, namely Trends in Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA). This study was conducted in the scope of countries incorporated in the Organization for Economic Cooperation and Development (OECD). Based on TIMSS 2015, Indonesia is ranked below the science field score, which is still below the international average of 397 out of 500. This puts Indonesia in 45th place out of 50 countries.

In the PISA test results in 2018 released on December 3, 2019, Indonesia experienced a decrease in rank compared to 2015; Indonesia was ranked 74th out of 79 countries that took the survey. Indonesia's science performance capability scored 396 out of a top score of 590. From the results obtained, it can be seen that the high level of thinking ability of Indonesian learners is still very low (Thomson et al., 2019). The low value of TIMSS and PISA shows a lack of analytical power, understanding and reasoning in learners. This is due to a lack of thoroughness in understanding the problem (1) transformation errors due to miss-determining formulas, (2) process skill errors, and (3) errors in determining the final answer, due to writing down conclusions but not precisely (Fazzlah et al., 2020).

Based on observations made by researchers at Public Senior High School 1 Nan Sabaris Padang Pariaman, the West Sumatera Province, on September 12, 2019, researchers conducted a trial on high-order thinking skills. From the results of the trial question, it is known that the high order thinking skills of learners are still relatively low. They are evidenced by the average value of learners only 38.45 according to Prasetyani et al (2016) that the value in the range of 21-40 is included in the criteria less.

Several things influence the low ability of learners in answering questions, there are: (1) Teachers have not fully designed problems in high cognitive levels and difficulty turning problems with low cognitive levels (C1-C3) into problems with higher cognitive levels (C4-C6) so that students are not used to working on problems with high levels of difficulty. The usual problem is still in the cognitive realm of C1-C3. This can be seen from the analysis of daily replay questions compiled by biology subject teachers at Public Senior High School 1 Nan Sabaris, where problems with cognitive levels of C2 as much as 50%, C3 40%, and C4 10%. (2) Learners are not used to doing problems with the help of Tables, graphs, and images, so that students have difficulty analyzing the data contained in Tables and images. In line with Harisman (2020) to develop high-level thinking skills, learners are influenced by the precepts and professionalism of teachers in developing instruments of high-level thinking ability.

One of the efforts to develop high-level thinking skills is to use appropriate assessments where assessment is an integral part of learning (Rahmi & Alberida, 2017). So for that, it is necessary to develop high-order thinking skills, assessment instruments that are valid, reliable, and have good quality items to train learners to have high-level thinking skills according to the demands of the 2013 curriculum (Otty & Milton, 2019).

Previous research on HOTS assessment shows that educator skills are needed in developing HOTS assessments (Widana, 2019). Research on HOTS assessment in physics subjects shows that HOTS assessment instruments effectively train student HOTS skills (Kusuma et al., 2017; Serevina et al., 2018). In addition, research on instrument HOTS in the subjects of the circulatory system shows that instruments can be used to access student HOTS abilities (Madang et al., 2017).

Based on these previous studies, it shows that no one has conducted research related to HOTS assessment, especially for subjects of the circulatory system. So that the novelty of this HOTS assessment research is the subjects used, namely the subject matter of the circulatory system. This research activity will provide many benefits, including this HOTS assessment can be used by educators to access student HOTS skills.

In this article, researchers develop a high-level assessment instrument of thinking ability about circulatory system material. Circulatory system material is material with a 3.6 for odd semester XI class with analyzing verb (C4). The purpose of this research is to produce High Order Thinking Skills instruments about circulatory system materials for students of grade XI Public Senior High School that are valid, reliable, and have good quality items.

METHOD

This type of research is research and development. This research aims to produce products in the form of high-level thinking ability assessment instruments on circulatory system materials for grade XI Public Senior High School students that are valid, practical, reliable, and have good item quality by using a 4-D development model. At the definition stage, problem analysis, students, curriculum, and learning objectives are performed. Then in the design stage is done creating grids, problem design, cover, and instructions of question work in general and specifically. At the stage of development, a question experiment has been done, and then an analysis of the quality of the item is carried out to obtain a valid instrument. However, due to time constraints and costs, the researchers did not perform the disseminate stage.

The study subjects consisted of experts. Two lecturers majoring in biology at Universitas Negeri Padang and two of biology subject teachers at Public Senior High School 1 Nan Sabaris and the test subjects were two teachers of biology subjects at Public Senior High School 1 Nan Sabaris and 33 students of grade XI mathematics and science 1 at Public Senior High School 1 Nan Sabaris. The object of this research is a research instrument of high order thinking skills about circulatory system material for students of grade XI Public Senior High School in the form of 50 questions. This research was conducted at Universitas Negeri Padang and Public Senior High School 1 Nan Sabaris Padang Pariaman. The resulting product is a high-order thinking skills research instrument about circulatory system material for Public Senior High School grade XI learners from November 26, 2020, to January 22, 2021.

RESULTS AND DISCUSSION

The logical validity of high-level thinking ability research instruments aims to prove the validity and representative of all possible grains made in the developed instrument. Two biology lecturers carried out this logical validity from the Mathematics and Natural Sciences Faculty, Universitas Negeri Padang, and two biology subjects at Public Senior High School 1 Nan Sabaris Padang Pariaman using validation questionnaires. During the validation process, suggestions were given by the experts to improve the quality of the instrument improvements can be seen in Table 1.

Table 1. Suggestions were given by the experts

No	Suggestion	Picture						
1	Improved Structure of Option	3.6.2 Menganalisis mekanisme pembekuan darah	Diberikan sebuah pertanyaan, peserta didik diminta menganalisis jawaban yang benar	C4	<p>Respons inflamasi melibatkan peningkatan aliran darah ke area tubuh yang mengandung sel-sel sistem kekebalan tubuh melawan patogen atau benda asing. Respons inflamasi adalah contoh dari loop umpan balik positif. Berdasarkan ini, apa yang akan menyebabkan respon inflamasi berhenti di daerah tertentu dari tubuh?</p> <p>A. Sel darah putih menghancurkan semua patogen/benda asing. B. Sel darah putih menyebar jauh dari situs. C. Sel darah putih tidak menyebar jauh dari situs. D. Lebih banyak sel darah putih tiba untuk melawan patogen/benda asing. E. Lebih sedikit sel darah putih yang diperlukan untuk menyelesaikan memerangi infeksi</p> <p><i>handai!!</i></p>	A	12	<p><i>Perhatikan urutannya!</i> <i>Ajar lebih banyak!</i> <i>bagus!</i> <i>Revisi & cek</i></p>
2	Increasing Cognitive level in the problem	3.6.5 Menganalisis perbedaan systole dan diastole pada tekanan darah manusia	Diberikan narasi, peserta didik diminta menganalisis jawaban yang paling benar	C4	<p>Baca narasi berikut untuk menjawab soal nomor 10 - 11</p> <p>Pada suatu hari Budi mengalami gangguan penyakit dengan gejala sering muntah, pusing, lemas dan tidak berenergi. Waktu diperiksa, tekanan darah Budi yang ditunjukkan oleh sphygmomanometer adalah 100/60 mmHg. Maka hal yang tepat menyatakan keadaan Budi adalah.</p> <p>A. Tekanan systole dibawah normal B. Tekanan diastole dibawah normal C. Otak jantung bekerja terlalu lemah D. Tekanan systole dan diastole normal E. Tekanan systole dan diastole di bawah normal</p> <p><i>+ alasan!</i></p>	E	10	
		3.6.9 Menganalisis kelainan dan penyakit pada organ peredaran darah	Diberikan ciri-ciri suatu gangguan pada system sirkulasi, peserta didik diminta	C4	<p>Berdasarkan ciri-ciri pada narasi diatas, gangguan apakah yang dialami oleh Budi?</p> <p>A. Stroke B. Leukimia C. Hipertensi D. Hipotensi E. Jantung Koroner</p> <p><i>+ alasan</i></p> <p><i>↳ kalau spt itu, masih C2!</i></p>	E	11	

No	Suggestion	Picture						
3	Turn the question sentence into the form of stuffing	3.6.9 Menganalisis kelainan dan penyakit pada organ peredaran darah	Diberikan sebuah wacana, peserta didik diminta menganalisis kelainan yang terjadi dalam wacana	C4	Tono mengalami beberapa gejala penyakit seperti pusing, cepat lelah, wajah pucat, dan tidak memiliki nafsu makan. Untuk memastikan, tono menjalani tes darah di laboratorium. Hasil tes menunjukkan jumlah eritrosit = 3,5 juta/mm ³ , kadar hemoglobin = 6, serta penurunan tekanan darah (90/60 mmHg). Berdasarkan data tersebut maka tono didiagnosa menderita: A. Tifus B. Anemia C. Leukimia D. Hemofilia E. Demam berdarah	B	20	20
		3.6.1 Menganalisis struktur dan fungsi komponen	Diberikan sebuah wacana, peserta didik diminta	C4	Berdasarkan teks diatas, berapakah seharusnya jumlah eritrosit normal pada pria dan wanita? A. 4,2 - 5,4 juta/mm ³ dan 4,0 - 5,5 juta/mm ³ B. 4,7 - 6,1 juta/mm ³ dan 4,2 - 5,4 juta/mm ³ C. 4,0 - 5,5 juta/mm ³ dan 4,7 - 6,1 juta/mm ³	B	21	21

After the revision, the validation questionnaire of high-level thinking ability instruments of the experts is analyzed which can be seen in the [Table 2](#).

Table 2. Results of analysis of the validity of logical instruments assessment of high-level thinking ability

No	Assessment Component	Expert			Total	Validity value (%)	Criteria
		1	2	3			
1	Material	30	29	25	84	87.5	Valid
2	Construction	35	35	35	105	87.5	Valid
3	Language	14	14	14	42	87.5	Valid
4	High-Order Thinking	15	17	18	50	81.67	Valid
Total amount of validity value						350.17	
Average value of validity						87.54	Valid

Based on the results of data analysis in [Table 1](#) obtained results, the instrument of high-level thinking ability meets the criteria and is valid. This is seen from the average validation value of the assessment instrument of 87.54 is included in the valid criteria (Ngalim, 2009). Therefore, the assessment of high-order thinking skills developed by the author is valid, both in terms of material, construction, language, and HOTS.

Analysis of question items was carried out after conducting instrument trials on 33 students of class XI mathematics and natural science 1 at Public Senior High School 1 Nan Sabaris Padang Pariaman. The analysis of the problem grains was carried out using ANATES 4.09. The analysis results obtained 44 valid questions with 88% and six invalid questions with a percentage of 12%. Thus, the number of problems with cognitive levels of C4 is as many as 34 questions, C5 as many as nine questions, and C6 1 problem. The percentage of cognitive levels can be seen in [Figure 1](#).

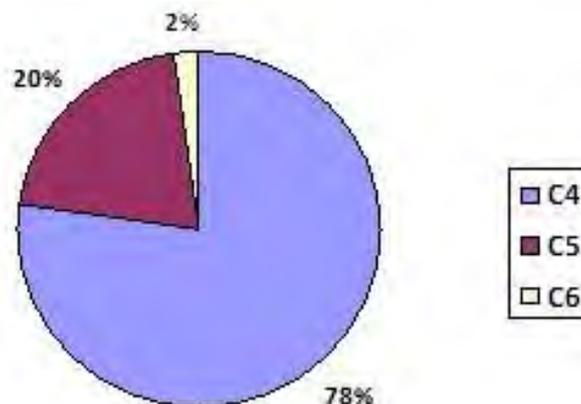


Figure 1. Cognitive rate percentage

Reliability of high order thinking skills instruments known from problem grain analysis when using ANATES version 4.09 with the formula Kuder-Richardson 21 (K-R-21). Based on the results of the study of the question, grain produced 44 items of reliable questions with a very reliable value of 0.78.

The difficulty of high-order thinking skills instruments is known from the results of problem analysis using ANATES version 4.09. Based on the results of the analysis of the problem items that have a moderate

difficulty level of 41 questions and increase the difficulty of difficult categories by nine questions. Data analysis of difficulty level can be seen in [Table 3](#).

Table 3. Data analysis of the difficulty level of the instruments of higher order thinking skills assessment

Difficulty level (%)					
Ordinary multiple choice		Cause and effect multiple choice		Association multiple choice	
Moderate	Difficult	Moderate	Difficult	Moderate	Difficult
21 questions (33,33-54,55)	9 questions (21,21-27,27)	10 questions (36,36-54,55)	-	10 questions (36,36-51,52)	-

The differentiating power of high-order thinking skills instruments is known from the analysis of the question using ANATES version 4.09. Based on the results of grain analysis of the instrument having good differentiating power as much as 52%, the differentiating power is sufficient as much as 46%, and the bad differentiating as much as 2%. Data analysis can be seen in [Table 4](#).

Table 4. Data analysis of differentiating power of the instruments of high order thinking skills assessment

Differentiating power (%)	Ordinary multiple choice	Excellent	-
		Good	14 question (0.44-0.67)
		Sufficient	16 question (0.22-0.33)
		Bad	-
	Cause and effect multiple choice	Excellent	-
		Good	8 question (0.44)
		Sufficient	2 question (0.33)
		Bad	-
	Association multiple choice	Excellent	-
		Good	4 question (0.44-0.67)
		Sufficient	5 question (0.22-0.33)
		Bad	1 question (0.00)

Quality option of high order thinking skills instruments about the circulatory system material from the analysis of the problem using ANATES version 4.09 obtained excellent criteria 50%, both 35.5%, less good 12% and bad 2.5%. The results of the data analysis can be seen in [Figure 2](#).

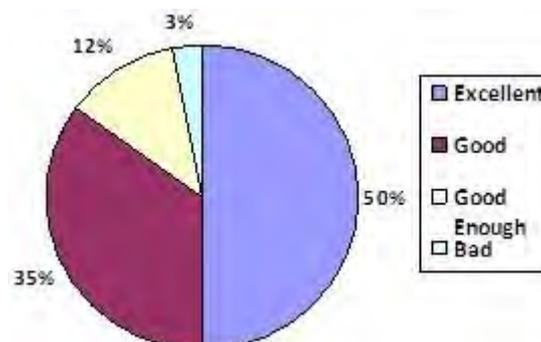


Figure 2. Quality option

Logical validity is obtained from the results of validation questionnaires filled by three experts. The criteria assessed in logical validity are four aspects, namely material, construction, language, and high level of thinking ability. According to [Arikunto \(2013\)](#) the results of data analysis showed that high order thinking skills instruments were included in valid criteria with an average value of 87.54%. Valid criteria on material aspects with a value of 87.5% indicate that the material used can improve the high-level thinking ability of learners, which is in line with [Fanani \(2018\)](#) stating that the material developed must be in accordance with the demands of core and basic competencies.

Judging from the construction aspect of the assessment instrument is in a valid category with a value of 87.5%. This indicates that the instrument has a clear problem formulation, homogeneous and logical answer options and is highly structured in accordance with the opinion of [Wahyudi et al \(2019\)](#) that validity construction is a validity that refers to the extent to which a measuring test instrument to be measured based on theoretical construction that can be used as the basis in the preparation of the instrument.

In terms of language, assessment instruments get a percentage of 87.5%. This indicates that this high order thinking skills instruments already uses language that is in accordance with the rules, easy to understand and does not give rise to double interpretation (Indra et al., 2018). This is in accordance with the opinion of Afrita & Darussyamsu (2020) that each item of the question must use clear language, easy to understand and in accordance with the enhanced spelling

Judging from the high level of thinking aspects, there is an average score of 81.67 that achieves valid criteria. This indicates that the problem is already at C4-C6 level, discourse, Tables, graphs and images work well in accordance with the context of the material, the question of using interesting stimulus (Sayan & Mertoğlu, 2020) so as to measure students' ability in analyzing, evaluating and creating. According to Arif et al (2020) high order thinking skills is a process of thinking in higher levels of knowledge develop from various cognitive concepts and methods and learning taxonomies, such as using bloom taxonomies. Judging from the above four aspects obtained an average yield of 87.54% from three experts. Based on these categories, this high order thinking skills instruments meets logical validity and can be used by learners.

Overall the high order thinking skills instruments developed meets all four aspects in the validation test so that this instrument can be used for biological assessment, especially circulatory system material. Before use, empirical tests and practicality tests must be carried out so that the instruments used are valid and can be used to develop the high level of thinking ability of learners.

Empirical validity aims at the level of reliability of the instruments developed (Tapsir et al., 2018). The results of empirical validity were obtained from the analysis of questions using ANATES 4.09 with the question analyzed as many as 50 questions tested to 33 students of class XI mathematics and natural science 1, Public Senior High School 1 Nan Sabaris Padang Panjang. Based on the results of the analysis, this high order thinking skills instruments is valid with the results obtained, namely 44 valid questions with a percentage of 88% and 6 invalid questions with a percentage of 12%. Percentages for each cognitive level of the problem are C4= 77.27%, C5=20.45%, and C6=2.27%. An instrument is said to be valid when the instrument elements of the assessment are relevant and represent the construction of a measuring instrument that is targeted for a specific purpose (Ihsan, 2016).

Reliability is a state that indicates a fixed (consistent) state of the instrument (Utama et al., 2020). Reliable instruments will get the same results if tested in the same group at different times. This high order thinking skills instruments is highly reliable with a reliability value of 0.78. Reliability is the consistency of measurement results shown at different times on the same subject (Ndiung & Jediut, 2020). The test is said to be reliable if the gain score has a high correlation with the total score. The validity and reliability of an instrument is influenced by the subject being shaved, the user of the instrument, and the instrument itself (Purwanti et al., 2020). So validity and reliability must always be tested before the instrument is used.

A good difficulty level for an instrument is the difficulty level ranging in the range of 0.00 - 1.00 (Arifin, 2017). The greater the difficulty of the count, the easier it will be for it to be revised. A question has $p = 0.00$ meaning that no student answers correctly and if it has $p = 1.00$ it means that all students answer correctly. The result of data analysis high order thinking skills instruments in get a problem that has a moderate difficulty level of 41 questions and increase difficulty category as much as 9 problems.

The differentiating power in high order thinking skills instruments has the aim to know and distinguish highly capable learners with low or moderate ability learners (Masitoh & Aedi, 2020). Based on the results of grain analysis of the instrument has a good differentiating power as much as 52%, the differentiating power is quite as much as 46% and the power of the bad differentiating as much as 2%. In line with Irmaya et al (2020) that good differentiating power can distinguish students who have mastered competencies with students who have not / lack of competence based on certain criteria.

The quality of options in an instrument is seen from how the learner chooses the answer to the given question. This aims to see if the phishing function is working properly or not. The quality of instrument options assessment ability to think highly about the circulatory system material from the results of analysis of the problem using ANATES version 4.09 obtained excellent criteria 50%, both 35.5%, less good 12% and bad 2.5%.

From this research, an instrument with appropriate indicators was obtained to assess the high level of thinking ability in the material of the circulatory system. The impact of the existence of this instrument for teachers is that it can be used as a means of training learners to familiarize themselves with thinking at a high level and become a collection of quality questions (Utama et al., 2020), for learners can be used to know and train high-level thinking skills (Kholis et al., 2020) and for schools this instrument can be useful to familiarize

learners using high-level thinking skills so as to improve the quality of learners (Arifin, 2018; Wilson & Narasuman, 2020).

CONCLUSION

The resulting HOTS instruments have an average logical validity value of 87.54% with very valid criteria. Practical validity value is 88% with each percentage of C4 = 77.27%, C5 = 20.45%, and C6 = 2.27%, has very reliable reliability with a value of 0.78, has a moderate difficulty, good differentiating power, and excellent option quality. To sum up, that has been produced the instruments of HOTS about circulatory system materials for students of grade XI Public Senior High School are logically valid, empirical, and reliable. **This conclusion is instruments that were designed can be used to measure learners' ability of high-level thinking.** Therefore, as a recommendation, this instrument can be used to promote students' HOTS.

ACKNOWLEDGEMENT

This research can be carried out thanks to the guidance of the experts, teachers, and students of Public Senior High School 1 Nan Sabaris and the Department of Biology, Universitas Negeri Padang.

REFERENCES

- Afrita, M., & Darussyamsu, R. (2020). Validitas instrumen tes berpikir tingkat tinggi (HOTS) pada materi sistem respirasi di kelas XI SMA. *Mangifera Edu*, 4(2), 129–142. <https://doi.org/10.31943/mangiferaedu.v4i2.83>
- Arif, S., Basyaruddin, B., & Yuhdi, A. (2020). Integration of high order thinking skills in research method subject in university. *Britain International of Linguistics Arts and Education (BioLAE) Journal*, 2(1), 378–383. <https://doi.org/10.33258/biolae.v2i1.207>
- Arifin, B. (2018). Meningkatkan mutu pendidikan melalui manajemen peserta didik. *FALASIFA : Jurnal Studi Keislaman*, 9(2), 1–20. <https://doi.org/10.36835/falasifa.v9i2.115>
- Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems (the Original Research of Mathematics)*, 2(1), 28–36. <https://doi.org/10.31949/th.v2i1.571>
- Arikunto, S. (2013). *Prosedur penelitian suatu pendekatan praktik*. Edisi Revisi. Rineka Cipta. <https://www.belbuk.com/prosedur-penelitian-suatu-pendekatan-praktik-p-1806.html>
- Fanani, M. Z. (2018). Strategi pengembangan soal HOTS pada kurikulum 2013. *Edudeena*, 2(1), 57–76. <https://doi.org/10.30762/ed.v2i1.582>
- Fazzlah, E., Effendi, K. N. S., & Marlina, R. (2020). Analisis kesalahan siswa dalam menyelesaikan soal pisa konten uncertainty and data. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 04, 1034. <https://doi.org/10.31004/cendekia.v4i2.306>
- Hairun, Y. (2020). *Evaluasi dan penilaian dalam pembelajaran*. Deepublish. <https://penerbitbukudeepublish.com/shop/buku-evaluasi-dan-penilaian/>
- Harisman, Y. (2020). Profesionalisme guru dalam membuat soal higher order thinking skills. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 4(1), 75. <https://doi.org/10.33603/jnpm.v4i1.2587>
- Ihsan, H. (2016). Validitas isi alat ukur penelitian konsep dan panduan penilaiannya. *PEDAGOGIA Jurnal Ilmu Pendidikan*, 13(2), 266. <https://doi.org/10.17509/pedagogia.v13i2.3557>
- Indra, R., Kustati, M., & Saregar, A. (2018). The effect of principals' leadership towards effective learning at an Indonesian secondary school. *European Journal of Educational Research*, 9(3), 1063–1074. <https://doi.org/10.12973/eu-jer.9.3.1063>**
- Irmaya, P., Ferna, F., & Sunarti, T. (2020). Validitas instrumen penilaian berbantuan google form untuk mengukur kemampuan berpikir kreatif fisika pada bahasan fluida dinamis. *IPF : Inovasi Pendidikan Fisika*, 09(02), 69–75. <https://jurnalmahasiswa.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/33346>
- Kholis, N., Kartowagiran, B., & Mardapi, D. (2020). Development and validation of an instrument to measure a performance of vocational high school. *European Journal of Educational Research*, 9(3), 955–966. <https://doi.org/10.12973/EU-JER.9.3.955>
- 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 (IOSR-JRME)*, 7(1), 1–7. <https://doi.org/10.9790/7388-0701052632>

- Madang, K., Huzaifah, S., & Suratmi, S. (2017). Development of higher order thinking skills (hots) instruments on human circulatory system for second grade students at senior high school. *International Conference on Science Education (ICoSEd)*. <http://icosed.conference.unesa.ac.id/ocs/index.php/icosed2017/icosed2017/paper/view/59>
- Masitoh, L. F., & Aedi, W. G. (2020). Pengembangan instrumen asesmen higher order thinking skills (HOTS) matematika di SMP kelas VII. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 4(2), 886–897. <https://doi.org/10.31004/cendekia.v4i2.328>
- Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(1), 94. <https://doi.org/10.25273/pe.v10i1.6274>
- Ngalim, P. (2009). *Prinsip-prinsip dan teknik evaluasi*. Remaja Rosdakarya. <https://rosda.co.id/pendidikan-keguruan/418-prinsip-prinsip-teknik-evaluasi-pengajaran.html>
- Otty, R., & Milton, L. (2019). Promoting soft skills with the HOT (higher order thinking) sheet assignment. *Journal of Instructional Pedagogies*, 22, 1–20. <http://files.eric.ed.gov/fulltext/EJ1216819.pdf>
- Paidi, P., Mercuriani, I. S., & Subali, B. (2020). Students' competence in cognitive process and knowledge in biology based on curriculum used in Indonesia. *International Journal of Instruction*, 13(3), 491–510. <https://doi.org/10.29333/iji.2020.13334a>
- Permana, F. H., & Setyawan, D. (2020). Peningkatan kualitas pembelajaran model flipped classroom dalam meningkatkan karakter siswa melalui lesson study. *Prosiding Seminar Nasional IKIP Budi Utomo*, 223–228. <https://doi.org/10.33503/prosiding.v1i01.893>
- Permana, F. H., Sukma, E., & Wahyono, P. (2021). The use of distance learning through whatsapp and google meeting to identify differences in biology learning outcomes. *Biosfer*, 14(1), 86–98. <https://doi.org/10.21009/biosferjpb.20094>
- Prasetyani, E., Hartono, Y., & Susanti, E. (2016). Kemampuan berpikir tingkat tinggi siswa kelas XI dalam pembelajaran trigonometri berbasis masalah di SMAN 18 Palembang. *Jurnal Gantang Pendidikan Matematika FKIP - UMRAH*, 1(1), 31–40. <https://doi.org/10.31629/jg.v1i1.4>
- Purwanti, E., Palupi, R. Z. P., Galuh, A., & Rianingsih, D. (2020). Pengembangan instrumen penilaian keterampilan abad 21. <https://ojs.unm.ac.id/semnasfisika/article/view/8704>
- Rahmi, Y. L., & Alberida, H. (2017). Improving students' higher order thinking skills through portfolio assessment on biology curriculum and textbook analysis course. *Bioeducation Journal*, 1(1), 22–33. <https://doi.org/10.24036/bioedu.v1i1.21>
- Sayan, H., & Mertoğlu, H. (2020). Equipment use in biology teaching. *Journal of Educational Issues*, 6(1), 357. <https://doi.org/10.5296/jei.v6i1.17042>
- Serevina, V., Sari, Y. P., & Maynastiti, D. (2018). Developing high order thinking skills (HOTS) assessment instrument for fluid static at senior high school. *The 2018 International Conference on Research and Learning of Physics*, 1–9. <https://doi.org/10.1088/1742-6596/1185/1/012034>
- Tapsir, R., Nikpa, N. A., & Zamri, S. N. A. B. S. (2018). Reliability and validity of the instrument measuring values in mathematics classrooms. *Malaysian Online Journal of Educational Sciences*, 6(2), 37–47. <http://mojes.um.edu.my/index.php/MOJES/article/view/12444/8051>
- Thomson, S., Bortoli, L. D., Underwood, C., & Schmid, M. (2019). *Australian council for educational research student performance programme for international student assessment PISA 2018*. <https://research.acer.edu.au/ozpisa/35/>
- Utama, C., Sajidan, S., Nurkamto, J., & Wiranto, W. (2020). The instrument development to measure higher-order thinking skills for pre-service biology teacher. *International Journal of Instruction*, 13(4), 833–848. <https://doi.org/10.29333/iji.2020.13451a>
- Wahyudi, N., Umam, K., Nugroho, Z., & Herawaty, D. (2019). Modifikasi software lisrel dengan membuat teknik analisis konstruksi validitas instrumen tes. *Jurnal Pendidikan Matematika Raflesia*, 05(01), 82–90. <https://ejournal.unib.ac.id/index.php/jpmr/article/view/10641>
- Widana, I. W. (2019). The effect of digital literacy on the ability of teachers to develop hots-based assessment. *International Conference on Mathematics and Natural Sciences 2019 (IConMNS 2019)*, 1–9. <https://doi.org/10.1088/1742-6596/1503/1/012045>
- Wilson, D. M., & Narasuman, S. (2020). Investigating teachers' implementation and strategies on higher order thinking skills in school based assessment instruments. *Asian Journal of University Education*, 16(1), 70–84. <https://doi.org/10.24191/ajue.v16i1.8991>