
Assessment Instrument Model for Natural Science Material Based on Research Results and Quranic Verses: An Effort to Measure Students' Higher-Order Thinking Skills

Antomi Saregar¹, Fredi Ganda Putra¹, Adyt Anugrah¹, Megawati Ridwan Fitri¹
¹Universitas Islam Negeri Raden Intan Lampung, Lampung, Indonesia

Coressponding Author Email:
antomisaregar@radenintan.ac.id

Abstract

High-order thinking skills (HOTS) are essential abilities students must possess in the 21st Century. Understanding students' HOTS is crucial, necessitating measures to gauge the extent of their HOTS. This research aims to determine the feasibility and validity of a science assessment instrument based on integrated research results and Quranic verses. The researchers employed the R&D method with the ADDIE design. The validation was carried out by three material experts and three language experts. The assessment instrument was reliable and valid, had good discriminative power, and had a proportional difficulty level. It can be concluded that the assessment instrument is suitable for measuring students' higher-order thinking skills. This research provides a pioneering approach by integrating Al-Qur'an verses into the science assessment model, highlighting the intersection of religious texts and scientific methodologies. Additionally, the developed instrument measures higher-order thinking skills and offers educators a culturally relevant tool in predominantly Muslim educational settings, potentially enhancing students' engagement and comprehension. This integrative approach offers a fresh perspective on educational assessment tools, urging educators and curriculum developers to view religious texts as valuable resources in shaping contemporary educational paradigms.

Keywords: Assessment, Instrument Model, Higher-Order Thinking Skills

How to cite this article :

Saregar, A., Putra, F., Anugrah, A., & Fitri, M. (2024). Assessment Instrument Model for Natural Science Material Based on Research Findings and Quranic Verses: An Effort to Measure Students' Higher-Order Thinking Skills. *IJIS Edu : Indonesian Journal of Integrated Science Education*, 6(2). doi:<http://dx.doi.org/10.29300/ijisedu.v6i2.2552>

INTRODUCTION

Higher-order thinking skills (HOTS) are among the thinking processes students require to meet the demands of the 21st century (Ichsan et al., 2019). HOTS prepares students to evaluate knowledge, not simply memorize it verbally, but also to analyze, synthesize, and associate to draw conclusions that lead to the generation of creative and innovative ideas (Annuuru et al., 2017). HOTS can promote creativity, critical thinking, collaboration, communication (4C) (Rati et al., 2023), social awareness (Setyarini et al., 2023), and students' learning process (Ibrahim et al., 2020). Furthermore, HOTS will help students solve problems as they think critically (Ajizah et al., 2023; Ichsan et al., 2019; Miri et al., 2007).

The measurement is done to identify the lowest measurable indicators, making finding solutions to these problems easier. As a result, a proper assessment is required to measure HOTS. HOTS instruments have been widely distributed (Kusuma et al., 2017; Widana et al., 2018) and utilized as evaluations in the form of quizzes (Suyatna et al., 2020), multiple-choice questions, and essays (Maryani et al., 2021). However, the available instruments are only suited for measuring HOTS in specific lessons. More innovative HOTS instruments must be developed.

HOTS instruments based on research results are the appropriate option to construct. Instruments based on research findings are typically more difficult to answer and require critical thinking skills. The application of research findings provides more accurate data. They can help students improve their analytical, evaluative, and creative thinking skills. Furthermore, they enable students to comprehend subjects more deeply rather than depending on superficial information. Research-based instruments will also assist students in developing the ability to face and overcome various obstacles in everyday life. Research-based questions are undoubtedly applicable to HOTS instruments.

The instrument integrated with Quranic verses will be ideal for Islamic-based educational institutions, such as madrasas and Islamic religious universities (PTKI). This assessment allows us to see the depth of students' understanding of the subject matter and their ability to read the Qur'an. The Quran is the source of all knowledge, and the subject of the study is not restricted to religious matters but encompasses all elements of human existence, including science. Science is one of the areas in which natural phenomena are the focus of knowledge, allowing it to be related to *ukhrawi* values (Latifah, 2015). The nature of natural science is to connect logical-material aspects with spiritual features that are consistent between the two and to incorporate pertinent Qur'anic verses into science.

Previous research on HOTS instrument development has included the development of HOTS assessments (Yunita & Bahriah, 2020), HOTS instruments for quadratic equations (How et al., 2023), physics HOTS tests with Moodle LMS (Widyaningsih et al., 2021), HOTS for fiqh material (Mawangir, 2022), exponential material (Zaki et al., 2020), fluid material (Akhsan et al., 2020), and wave material (Masturi et al., 2023). Assessments have been developed to measure HOTS in certain materials based on previous studies. Furthermore, there is no development of HOTS based on research results and combined with verses from the Quran. Therefore, creating a research-based science instrument incorporating Qur'anic verses to assess higher-order thinking skills at Islamic educational institutions is vital.

METHODS

The research and development (R&D) method with the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) development model was employed in the product development (Sugiyono, 2017). Figure 1 provides the flowchart of the research.

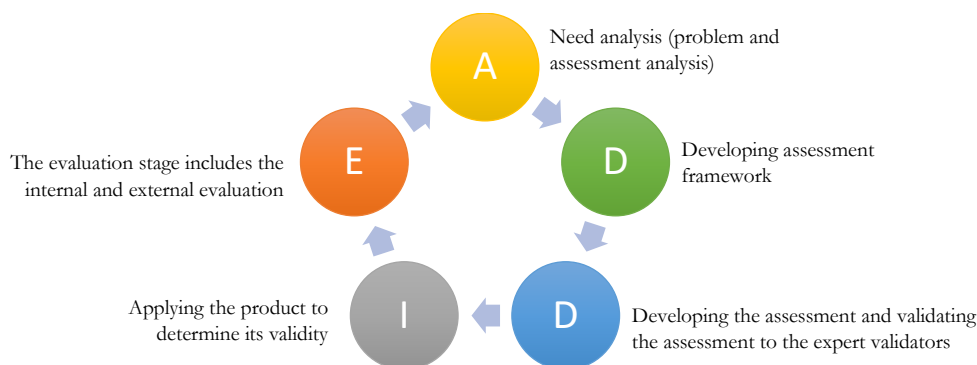


Figure 1. Research Flowchart

The research began with the analysis step, which involved assessing the needs. The design stage was completed by designing the instrument. The development stage was then completed by validating the instrument with three language experts and three material expert validators. After a team of experts validated the assessment instrument and declared it feasible, the next step was to implement the product to students to assess its validity and reliability. The implementation was carried out on students aged 17 to 19. After the product had been implemented, the following stage was to assess the results of the instrument trials by analyzing each item so that it could be used as material for improvement to perfect the product developed.

RESULTS AND DISCUSSION

1. Analysis

The first stage of this research was conducting a needs analysis to identify problems using field and literature observations, interviews with various natural science teachers about students' HOTS, and the assessment instruments utilized. According to relevant literature findings, no HOTS assessment instruments were integrated with Islamic values. Furthermore, field observations suggested that the

instruments often employed by teachers, particularly in natural science subjects, did not incorporate Islamic values and were not based on results from research. Therefore, their validity and reliability have not been assessed. Furthermore, the instruments devised by teachers did not meet the criteria for students' higher-order thinking skills (Hots). Measuring this ability is an important part of the teaching and learning process in the classroom. The science materials were tailored to the learning indicators provided by the appropriate curriculum. Based on this argument, creating a research-based assessment instrument that incorporates Islamic values is vital.

2. Design

At this stage, the researchers created the initial version of the HOTS assessment instrument for science combined with Islamic principles. The researcher created the assessment instrument based on HOTS indicators and Islamic values. The instrument designed took the form of multiple-choice questions. Figure 2 shows an example of the constructed question instrument.

3	Observe Surah Fatir verse 9 and the following information on wind speed distribution data: شُورَالَّذِيكَ مُوْتِيَهَا بَعْدَ الْأَرْضِ بِهِ فَأَحْيَيْنَا مَمَيَّتٍ بَلَدٍ إِلَى فَسَقْنَاهُ سَخَابًا فَتُنْبِئُ الرِّيحُ أَرْسَلَ الَّذِي وَاللَّهِ Wind is caused by the uneven heating of the Earth's surface by sunlight. The warmer air expands, becomes lighter, and moves upward. Below is presented the data on the distribution of wind speed over one year.
---	--

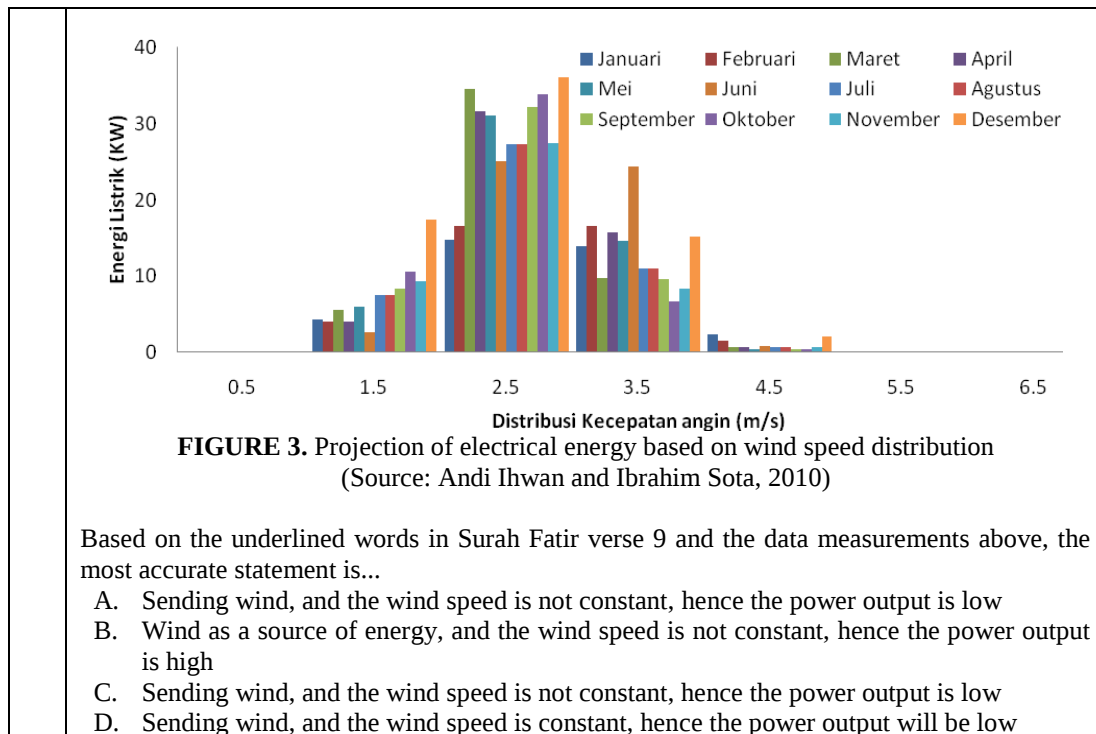


Figure 2. Example of Research-Based HOTS Instrument Design Integrated with Quranic Verses

The graphs used in the problem are based on research data from a specific publication. Furthermore, incorporating Al-Quran verses into the question aims to instill Islamic values in students through memorizing the Quranic verses.

3. Development

The product that had been designed was then developed as 25 multiple-choice questions. The

questions developed met the HOTS characteristics and were consistent with Islamic values. Two expert groups (three material experts and three language experts) subsequently validated the product. The validation data were acquired using the expert validation questionnaire. Table 2 shows the results of the expert validation.

Table 2. Expert Validation Results

No	Expert Group	Aspect	Percentage Score			Average	Criteria
			1	2	3		
1	Material Experts	Construction	89%	85%	87%	87%	Excellent
		Content	88%	83%	85%	85%	Excellent
2	Language Experts	Language	92%	87%	87%	87%	Excellent

The expert validation questionnaires covered nine aspects of construction. The average total score of the three validators for completing the construction aspects is 87% in the excellent category. The validation questionnaire covered 15 aspects of the

content. The average total score of the three validators for fulfilling the content aspect is 85% in the excellent category. The average total score of the three validators for fulfilling the language aspect is 87% in the excellent category. Based on expert validation, the

collected data was used to identify discrepancies or errors in the assessment instrument. The first product revision was carried out in line with expert validation notes and suggestions for improvement.

Based on the expert validation, the validators made several suggestions for the overall improvement of the assessment tool, as shown in Table 3.

Table 3. Inputs and Suggestions from the Experts

Validators	Inputs	Revision
Material Expert 1	- Questions should have a good proportion (hard, medium, easy). - Use representative images. - Each question should consist of more than one indicator.	The question instrument has been revised based on input from the validators.
Material Expert 2	- The number of questions is adjusted to the learning duration. - Questions must have a good proportion (hard, medium, easy).	
Material Expert 3	- There should be conceptual and operational definitions in the instrument's specification. - In the rubric section, the scale must be defined.	
Language Experts 1	- Use language according to proper Indonesian spelling (EYD). - Do not use conjunctions at the beginning of sentences.	
Language Experts 2	- Avoid using ambiguous language. - Add punctuations and adjust them.	
Language Experts 3	- Pay attention to punctuation. - The description contained in the picture must be legible.	

4. Implementation

The implementation stage was designed to assess the response of students and lecturers to the developed assessment instrument, as well as to determine the instrument's validity and reliability. The first step was to give the instrument to the students to complete, followed by a questionnaire to assess

their response to the instrument. A positive response from students is one of the criteria required for the instrument's feasibility. The test results were then analyzed for validity and reliability. Table 4 contains the results of the calculation of the validity test of the assessment instrument.

Table 4. Item Validity Test Results

Item Number	$r_{xycritical}$	$r_{xyobserved}$	Criteria
1	0,396	0,519	Valid
2	0,396	0,476	Valid
3	0,396	0,185	Invalid
4	0,396	0,726	Valid
5	0,396	0,271	Invalid
6	0,396	0,498	Valid

Item Number	$r_{xycritical}$	$r_{xyobserved}$	Criteria
7	0,396	0,385	Invalid
8	0,396	0,695	Valid
9	0,396	0,519	Valid
10	0,396	0,583	Valid
11	0,396	0,784	Valid
12	0,396	0,716	Valid
13	0,396	0,715	Valid
14	0,396	0,467	Valid
15	0,396	0,586	Valid
16	0,396	0,509	Valid
17	0,396	0,715	Valid
18	0,396	0,458	Valid
19	0,396	0,343	Invalid
20	0,396	0,414	Valid
21	0,396	0,669	Valid
22	0,396	0,743	Valid
23	0,396	0,603	Valid
24	0,396	0,467	Valid
25	0,396	0,356	Invalid

Table 4 indicates that five items were invalid; therefore, they were deleted, leaving 20 questions that can be used. In addition to calculating the validity, the instrument developed was also tested for reliability, with results indicating that $r_{observed} = 0,89 > r_{critical} = 0,396$. These findings indicate

that the instrument developed met the reliability criteria. The assessment instrument, which had been validated and reliable, was then tested for discriminating power and difficulty level. Table 5 shows the results of the discriminating power (DP) test and the level of difficulty of each item.

Table 5. Discriminating Power and Level of Difficulty

Item	Discriminating Power	Interpretation	Level of Difficulty (%)	Interpretation
1	0,533	High	53,33%	Moderate
2	0,467	High	50,00%	Moderate
4	0,533	High	40,00%	Moderate
6	0,4	High	53,33%	Moderate
8	0,733	Excellent	36,67%	Moderate
9	0,4	High	66,67%	Moderate
10	0,533	High	60,00%	Moderate
11	0,733	Excellent	56,67%	Moderate
12	0,667	High	53,33%	Moderate
13	0,467	High	23,33%	Difficult
14	0,4	High	53,33%	Moderate
15	0,467	High	56,67%	Moderate
16	0,533	High	60,00%	Moderate
17	0,467	High	23,33%	Difficult
18	0,333	Moderate	43,33%	Moderate
20	0,333	Moderate	50,00%	Moderate
21	0,6	High	56,67%	Moderate
22	0,733	Excellent	56,67%	Moderate
23	0,4	High	26,67%	Difficult
24	0,4	High	53,33%	Moderate

Based on Table 5. The 20 valid and reliable questions had good discriminating index and difficulty levels and corresponded to the students' HOTS indicators. The developed assessment instrument was then assessed next. The assessment was conducted by a team of eight teachers

from Islamic institutions. The instrument sheets used to evaluate the product were in the form of a statement consisting of two aspects: material feasibility and language feasibility. Table 6 shows the results of the instrument product assessment.

Table 6. The Assessment Results of the Developed Instruments

No.	Aspect Assessed	EX Per Aspect	Max Score	Percentage	Category
1.	Material Feasibility	211	225	93,77%	Highly feasible
2.	Language Feasibility	269	300	89,67%	Highly feasible
Total		480	525		
Average Percentage				91,72%	Highly feasible

Table 6 shows that the material feasibility aspect, consisting of three assessment items, received 211 points, with a maximum score of 225 and a percentage of 93.77%. Furthermore, the language feasibility aspect, consisting of three assessment items, received 269 points with a maximum score of 300 and an accuracy of 89.67%. The average assessment score is 91.72%, with the category "Highly Feasible."

5. Evaluation

Evaluation was carried out at each of the previous stages. At this stage, the entire development process (analysis, design, development, and implementation) was evaluated to determine whether the completed stages were feasible for the next stage. The evaluation process differed at each stage. The researchers interviewed the educators at the analysis stage to evaluate the product. At the design stage, the evaluation was based on the product's suitability to the desired criteria. In the third stage (development), the evaluation was carried out by experts who became validators. At the implementation stage, the evaluation was carried out using the results of the questionnaire distribution

as well as validity and reliability tests. Based on the stages, it is determined that the research-based science instrument combined with Quranic verses on students' HOTS is appropriate for use.

Based on the overall data analysis findings, the main problem is the lack of enrichment instruments that measure all high-level cognitive domains. One of the potentials for the institution is to develop HOTS test instruments that cover all levels of higher-level cognition and measure the dimensions of scientific knowledge integrated with Quranic verses. In general, the development of the assessment instrument was divided into five stages.

The Analyze stage analyzes students' needs and problems using field observations, literature studies, and discussions with lecturers. Students require learning instruments that promote higher-order thinking skills. According to observations and literature reviews, HOTS assessment instruments focused on Islamic values were not commonly found in Islamic-based educational institutions, particularly in the natural science subjects. The available instruments were only routine questions that were not integrated with Islamic values (Akhsan et al., 2020; Masturi et al., 2023; Widyaningsih et al., 2021; Zaki et al., 2020), so they are unable to

assess students' abilities in higher-level thinking.

The Design stage serves to create the initial product of the assessment instrument, which measures students' HOTS ability in science subjects while also incorporating Islamic values. Based on expert validation results, the HOTS assessment indicators have been developed by the revised Bloom Taxonomy, which measures students' HOTS: the ability to analyze (C4), evaluate (C5), and create (C6) (Umami et al., 2021). This statement is consistent with the development of HOTS instruments in previous studies, which include C4, C5, and C6 (Damayanti et al., 2020; Ndiung & Jediut, 2020; Nurhasanah et al., 2023).

At the development stage, the instrument is developed based on the results of the analysis and design stages. The assessment instrument consisted of 25 multiple-choice HOTS questions that integrated Islamic values. The primary goal of this development research is to create HOTS assessment instruments and validate them with experts. The product was validated by two groups of experts: material and language experts, each with three members.

Implementation is a simple trial stage to see how students and educators respond to developing assessment instruments based on students' HOTS. At this stage, the instrument's validity, reliability, discriminating power, and difficulty level will be tested to ensure its feasibility. This is consistent with the theory that item analysis, including testing validity, reliability, difficulty level, and discrimination power, is necessary in developing and analyzing test items (Moran, 2023). Of the 25 items, 20 were valid, and five were invalid. Following what was stated by Flake et al. (2017), validity determines the extent to which a measurement instrument is accurate and reliable in performing its measuring function. Therefore, the items declared invalid were removed. The reliability test results show that the items met the reliable criteria ($r_{critical} = 0.396$).

Furthermore, the discriminating power test is used to distinguish between students with high and low abilities, and the items' difficulty level indicates whether the items are easy,

moderate, or difficult. Items that are too easy cannot help students improve their problem-solving abilities. In contrast, too difficult questions will cause students to become discouraged when attempting to solve problems. The results indicate that the discriminating power is high, and the difficulty level is proportional. The proportional difficulty levels will assist teachers in distinguishing students with deeper understanding and better analytical skills, as stated by (AlKhuzaei et al., 2021). Additionally, according to research conducted by (Ebert et al., 2024) creating questions with proportional difficulty levels will avoid the floor effect and ceiling effect. After meeting the instrument feasibility requirements, 20 items were obtained, which served as the final product of the test questions and met the HOTS indicators. Finally, the Evaluation stage is the final stage, during which each previous stage will be evaluated. Meeting the eligibility criteria will help ensure the accuracy and reliability of the test (Ntumi et al., 2023; Rezigalla et al., 2024).

The appropriate instrument is required to identify HOTS in students (Hadiati et al., 2021; Melawati et al., 2022; Ningsih & Kamaludin, 2023). The developed instrument can measure the extent of students' HOTS. This HOTS instrument also allows for the exploration of Al-Qur'an literacy skills. The novelty of this product is the data presented for the instrument in the form of actual research results. It will undoubtedly foster students' critical thinking skills because they solve problems using valid and real data. So far, many instruments have focused on the level of remembering, which is ineffective in fostering HOTS in students (Abosalem, 2016). HOTS uses open-ended questions to allow students to practice critical thinking and express their opinions (Setyarini et al., 2018).

CONCLUSION

The ADDIE stage successfully developed research-based science assessment instruments integrated with Quranic verses. Based on the analysis and discussion findings, it is possible to conclude that the research-based science assessment instrument on HOTS, which

incorporates Quranic verses, has been developed and validated in terms of face and content. This assessment instrument is highly reliable and valid, with good discriminating power and proportional difficulty. As a result, the items in this assessment instrument are reliable and can be used as a research data collection instrument, particularly for students' HOTS. Based on the findings, future researchers should use this assessment instrument to measure students' higher-order thinking skills in science materials in the classroom. For educators in educational institutions, particularly those based on Islam, a good assessment instrument will improve the learning process in the classroom; thus, this research-based science assessment instrument is appropriate for use in class.

ACKNOWLEDGMENT

Research can be conducted on PTKIN research assistance funding in UIN Raden Intan Lampung based on the Decree of the Rector of UIN Raden Intan Lampung Number 172 of 2022, Higher Education Development Cluster.

REFERENCES

- Abosalem, Y. (2016). Assessment Techniques and Students' Higher-Order Thinking Skills. *International Journal of Secondary Education*, 4(1), 1–11. <https://doi.org/10.11648/j.ijisedu.20160401.11>
- Ajizah, R., Karim, K., & Suryaningsih, Y. (2023). Development of higher order thinking skills based math problems with wetland contexts for junior high school students. *Math Didactic: Jurnal Pendidikan Matematika*, 9(1), 134–145. <https://doi.org/10.33654/math.v9i1.2030>
- Akhsan, H., Wiyono, K., Ariska, M., & Melvany, N. E. (2020). Development of HOTS (higher order thinking skills) test instruments for the concept of fluid and harmonic vibrations for high schools. *Journal of Physics: Conference Series*, 1480(1), 012071. <https://doi.org/10.1088/1742-6596/1480/1/012071>
- AlKhuyaey, S., Grasso, F., Payne, T. R., & Tamma, V. (2021). A Systematic Review of Data-Driven Approaches to Item Difficulty Prediction BT - Artificial Intelligence in Education. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 29–41). Springer International Publishing.
- Annuuru, T. A., Johan, C. R., & Ali, M. (2017). Peningkatan kemampuan berpikir tingkat tinggi dalam pelajaran ilmu pengetahuan alam peserta didik sekolah dasar melalui model pembelajaran treffinger. *EDUTCEHNOLOGIA*, 3(2), 136–144. <http://www.ejurnal.unisri.ac.id/index.php/widyawacana/article/download/1483/1307%250Ahttp://repositori.usu.ac.id/handle/123456789/4596%250Ahttps://www.liputan6.com/global/read/4126480/skor-terbaru-pisa-indonesia-merosot-di-bidang-membaca-sains-dan-matemat>
- Damayanti, N., Hartono, Subali, B., Nugroho, S. E., & Sureeporn, K. (2020). Items analysis of physics assessment based on cognitive level of high order thinking skills in bloom taxonomy. *Journal of Physics: Conference Series*, 1521(2), 022022. <https://doi.org/10.1088/1742-6596/1521/2/022022>
- Ebert, K. D., Pham, G. T., Levi, S., & Eisenreich, B. (2024). Measuring children's sustained selective attention and working memory: validity of new minimally linguistic tasks. *Behavior Research Methods*, 56(2), 709–722. <https://doi.org/10.3758/s13428-023-02078-5>
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>
- Hadiati, S., Sukadi, E., & Pramuda, A. (2021). Development of Higher Order Thinking Skills Assessment Refer the Theory of Bloom in Laboratory Work. *AL-ISHLAH: Jurnal Pendidikan*, 13(3), 2106–2113. <https://doi.org/10.35445/alishlah.v13i3.1075>
- How, R. P. T. K., Zulnaidi, H., & Rahim, S. S. B. A. (2023). Development of Higher-Order Thinking Skills test instrument on Quadratic Equation (HOTS-QE) for secondary school students. *Pegegog Journal of Education and Instruction*, 13(1), 379–394. <https://doi.org/10.47750/pegegog.13.01.41>

- Ibrahim, N. N., Ayub, A. F. M., & Yunus, A. S. M. (2020). Impact of Higher Order Thinking Skills (HOTS) Module Based on the Cognitive Apprenticeship Model (CAM) on Student's Performance. *International Journal of Learning, Teaching and Educational Research*, 19(7), 246–262. <https://doi.org/10.26803/ijlter.19.7.14>
- 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>
- 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 (IOSRJME)*, 7(1), 26–32. <https://doi.org/10.9790/7388-0701052632>
- Latifah, S. (2015). Pengembangan Modul IPA Terpadu Terintegrasi Ayat-Ayat Al-Qur'an Pada Materi Air Sebagai Sumber Kehidupan. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 4(2), 155.
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitriawanati, M. (2021). HOTS Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Turkish Journal of Science Education*, 18(4), 674–690. <https://doi.org/10.36681/tused.2021.97>
- Masturi, Khasanah, N., Anggraeni, A., & Susilawati. (2023). The development of high order thinking skills (HOTS) wave test instrument for high school students. *AIP Conference Proceedings*, 050050. <https://doi.org/10.1063/5.0125874>
- Mawangir, M. (2022). The Development of Higher Order Thinking Skill Test Instrument on the Fiqh Subject: The Case of a State Islamic Senior High School in West Bangka Regency. *Journal of Higher Education Theory and Practice*, 22(7), 167–177. <https://doi.org/10.33423/jhetp.v22i7.5281>
- Melawati, Y., Rochmiyati, R., & Nurhanurawati, N. (2022). A needs analysis of HOTS-based assessment instruments for elementary school mathematics learning. *Asian Journal of Educational Technology*, 1(2), 90–95. <https://doi.org/10.53402/ajet.v1i2.41>
- Miri, B., David, B.-C., & Uri, Z. (2007). Purposely Teaching for the Promotion of Higher-order Thinking Skills: A Case of Critical Thinking. *Research in Science Education*, 37(4), 353–369. <https://doi.org/10.1007/s11165-006-9029-2>
- Moran, V. (2023). *Item and Exam Analysis BT - Item Writing for Nurse Educators* (V. Moran (ed.)). Springer International Publishing. https://doi.org/10.1007/978-3-031-30211-4_5
- 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), 94–111. <https://doi.org/10.25273/pe.v10i1.6274>
- Ningsih, N. R., & Kamaludin, A. (2023). Development of Higher Order Thinking Skills-Based Assessment Instrument on Acid-Base Materials in High School. *Jurnal Penelitian Pendidikan IPA*, 9(1), 13–19. <https://doi.org/10.29303/jppipa.v9i1.1457>
- Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the Psychometric Properties (Item Difficulty, Discrimination and Reliability Indices) of Test Items using Kuder-Richardson Approach (KR-20). *Shanlax International Journal of Education*, 11(3), 18–28. <https://doi.org/10.34293/education.v11i3.6081>
- Nurhasanah, A., Nugraha, F. F., & Wijayatna, S. A. (2023). Analysis Hots Content in Erlangga Straight Point Series (Esps) 5Th Grade Mathematics Textbook Base on Bloom'S Taxonomy Theory. *Journal Of Educational Experts (JEE)*, 6(1), 66–75. <https://doi.org/10.30740/jee.v6i1.188>
- Rati, N. W., Arnyana, I. B. P., Dantes, G. R., & Dantes, N. (2023). HOTS-Oriented e-Project-Based Learning: Improving 4C Skills and Science Learning Outcome of Elementary School Students. *International Journal of Information and Education Technology*, 13(6), 959–968. <https://doi.org/10.18178/ijiet.2023.13.6.1892>
- AlKhuzaey, S., Grasso, F., Payne, T. R., & Tamma, V. (2021). A Systematic Review of Data-Driven Approaches to Item

- Difficulty Prediction BT - Artificial Intelligence in Education. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 29–41). Springer International Publishing.
- Ebert, K. D., Pham, G. T., Levi, S., & Eisenreich, B. (2024). Measuring children's sustained selective attention and working memory: validity of new minimally linguistic tasks. *Behavior Research Methods*, 56(2), 709–722. <https://doi.org/10.3758/s13428-023-02078-5>
- Moran, V. (2023). *Item and Exam Analysis BT - Item Writing for Nurse Educators* (V. Moran (ed.)). Springer International Publishing. https://doi.org/10.1007/978-3-031-30211-4_5
- Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the Psychometric Properties (Item Difficulty, Discrimination and Reliability Indices) of Test Items using Kuder-Richardson Approach (KR-20). *Shanlax International Journal of Education*, 11(3), 18–28. <https://doi.org/10.34293/education.v11i3.6081>
- Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhoussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24(1), 445. <https://doi.org/10.1186/s12909-024-05433-y>
- Setyarini, S., Muslim, A. B., Rukmini, D., Yuliasri, I., & Mujianto, Y. (2018). Thinking critically while storytelling: Improving children's HOTS and English oral competence. *Indonesian Journal of Applied Linguistics*, 8(1), 189–197. <https://doi.org/10.17509/ijal.v8i1.11480>
- Setyarini, S., Salim, H., & Purnawarman, P. (2023). Higher-Order Thinking Skills (HOTS)-based literacy media: An innovative learning strategy to promote the secondary students' social awareness. *Forum for Linguistic Studies*, 5(2), 1706. <https://doi.org/10.59400/fls.v5i2.1706>
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif dan RnD*. Alfabeta.
- Suyatna, A., Viyanti, V., & Rosidin, U. (2020). Optimizing Computer-Based Hots Instruments: An Analysis of Test Items, Stimulus, and Quiz Setting Based on Physics Teachers' Perceptions. *Universal Journal of Educational Research*, 8(3D), 97–105. <https://doi.org/10.13189/ujer.2020.081714>
- Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student asesment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, 7(1), 57–68. <https://doi.org/10.37058/jp3m.v7i1.2069>
- Widana, I. W., Parwata, I. M. 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>
- Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The Development of the HOTS Test of Physics Based on Modern Test Theory: Question Modeling through E-learning of Moodle LMS. *International Journal of Instruction*, 14(4), 51–68. <https://doi.org/10.29333/iji.2021.1444a>
- Yunita, L., & Bahriah, E. S. (2020). The development of asesment higher order thinking skills (HOTS) through online based application. *Journal of Physics: Conference Series*, 1511(1). <https://doi.org/10.1088/1742-6596/1511/1/012035>
- Zaki, M., Amalia, R., & Sofyan, S. (2020). Development of high order thinking skills (HOTS) test instrument on exponent for junior high school students. *Journal of Physics: Conference Series*, 1521(3), 032096. <https://doi.org/10.1088/1742-6596/1521/3/032096>