Jurnal Biolokus:
Jurnal Penelitian Pendidikan Biologi dan Biologi
Volume 5, Nomor 1, 2022

p-ISSN: 2621-3702
e-ISSN: 2621-7538

DEVELOPMENT OF HOTS-BASED QUESTIONS ON BIOLOGY LEARNING

Adi Hartono*, Ashar Hasairin, Diky Setya Diningrat
Universitas Negeri Medan
*adi.hartono@uinsu.ac.id

ABSTRACT

According to observations on the Lower-Level Organisms Taxonomy course, it showed that the questions used in the course have not been presented at the higher-order thinking skill (HOTS) level. HOTS-based questions are one of the assessments in the cognitive domain that function to train HOTS and solve complex problems well. Moreover, it can train critical thinking skills as one aspect of the skills needed in the 21st century. Therefore, this study aims to develop HOTS-based questions on Monera topic to improve students' critical thinking skills. The research was conducted from January to February 2022. Development was carried out using ADDIE approach. Data collection techniques were carried out through the questionnaires and validation instruments by expert validators. The data were analyzed in two ways, namely content validity and construct validity. The results showed that the validity of the questions by expert validators was categorized as very good. Meanwhile, empirical validity shows that there were 10 questions declared valid from a total of 20 questions. The results of reliability test were 0.780 valid with a high interpretation. The test of difficulty level results showed that there were 8 moderate questions and 2 difficult questions. The results of differentiating power test showed that there were 1 very good question, 6 good questions, 2 sufficient questions and 1 weak question. Thus, it can be concluded that from a total of 20 HOTS-based questions that were developed only 10 questions were declared empirically feasible to improve students' critical thinking skills on Monera topic. It is recommended to use these questions in Lower-Level Organisms Taxonomy lectures.

Keywords: *Critical Thinking Skill, HOTS-Based Questions, Monera*

INTRODUCTION

High Order Thinking Skill (HOTS) is a high-level thinking ability that directs students to manipulate certain information and ideas, so that it can provide new implications (Fanani, 2018). This ability is closely related to the process of synthesizing a number of facts and information, assembling a variety of information, and reconstructing information into a goal of solving problems that are difficult to solve (Indah, 2020). This process can train the sharpness of critical thinking, namely the process of thinking logically, analytically, and reflectively by students (Sole & Anggraeni, 2020).

HOTS has a close relationship to improve one's critical thinking skills. This is because HOTS can hone one's skills in disposition of knowledge that has been acquired and develop it in new situations by means of mathematical and reflective reasoning (Andriyani & Saputra, 2020; Feriyanto & Putri, 2020). This is in accordance with the indicators of the achievement of Critical Thinking Skills, which provide opportunities for students to be skilled in generalizing concepts/facts/information, applying cognitive strategies, proving the validity of theories based on the problems presented, and evaluating actions in the right structure of consideration (Widana, et al., 2018; Puspitasari & Hidayatullah, 2020).

Critical thinking skills is one of the skills needed to be ready to face challenges in the 21st century. This skill facilitates students to be skilled in complex and systematic thinking to find solutions to a problem (Warsah, et al.,

2021). This has implications for increasing one's skills in correlating information to problems and reconstructing it in the form of a solution. These skills are one of the basic assets to develop a competitive and adaptive generation in the global era (Fanani, 2018; Rahayu, et al., 2020). One of the strategic efforts that can be made to improve critical thinking skills is through the development of HOTS-based questions (Riswanda, 2018).

HOTS-based questions can be interpreted as cognitive assessments that are developed based on reasoning power of analysis, evaluation, and creation. If an analogy with Bloom & Krathwool's taxonomy, these HOTS-based questions are at levels C4, C5, and C6 (Lestari, et al., 2016). In line with this, this assessment facilitates students in developing intellectual power that is oriented towards mastering facts and concepts and applying them in a problem solving (Fanani, 2018). Thus, students become trained to examine facts and concepts comprehensively and correlate them with alternative problem solving objects (Rohim, 2019).

The results of the PISA research reveal that the critical thinking skills of Indonesian students are still low, ranking 72 out of 78 countries with a score of 379 (Rosmalinda, et al., 2021). According to the questionnaire given to students of the Department of Biology, FMIPA, State University of Medan, they still have low critical thinking skills (65.50%) and need to be improved. In addition, they also explained that they had difficulty developing critical thinking skills, especially on Monera material because learning and questions were still presented conventionally.

Referring to the results of previous research, the development of HOTS-based questions on biological material has been carried out by previous researchers, namely on Ecosystem material (Nadifatinisa & Sari, 2020), Respiratory System material (Afrita & Darussyamsu, 2020), and Movement System (Rini & Budijastuti, 2022). However, the development of HOTS-based questions on Monera material for university students has never been carried out by any researcher.

Therefore, this study aims to develop HOTS-based questions on Monera material in universities. This research is expected to increase students' critical thinking skills as one of the skills needed in the 21st century. In addition, the implications of this research are also expected to be able to facilitate students to develop their intellectual potential as an implementation effort in solving problems in a solution and accurately.

METHOD

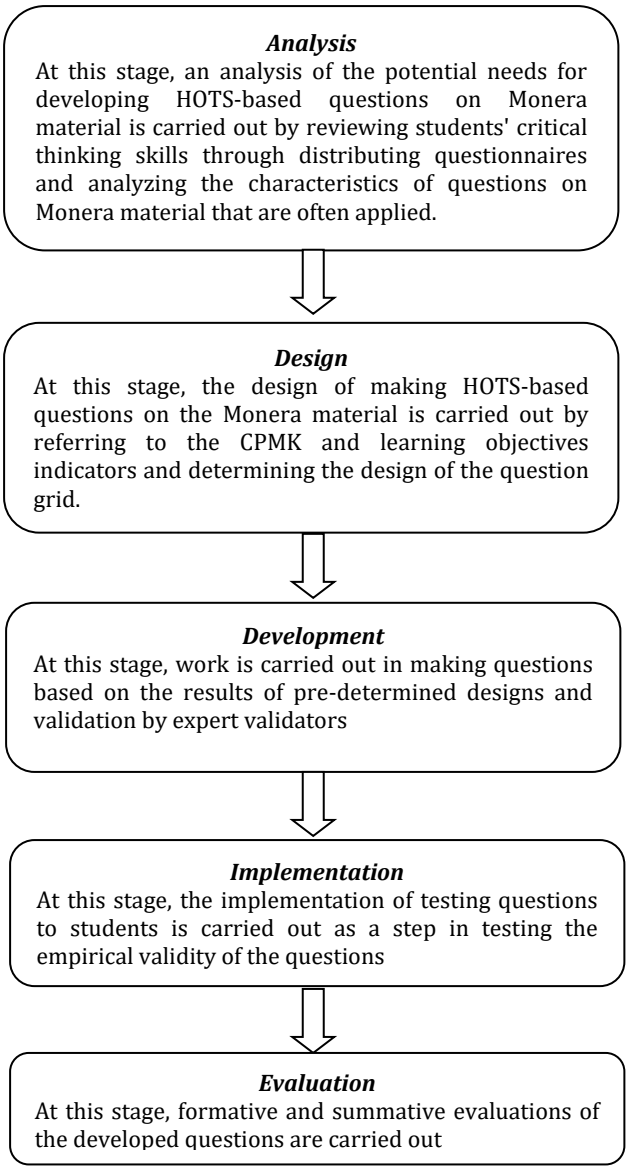

Figure 1. Research Systematic Flow

The development of HOTS-based questions is carried out following the ADDIE (Analysis, Design, Development, Implementation, Evaluation) approach (Rayanto & Sugianti, 2020). The series of research activities carried out in each stage of the approach can be seen in Figure 1.

Data was collected by using a questionnaire technique. The instrument uses a validation sheet and a questionnaire sheet. Analysis of the level of validation of the questions by expert validators uses the following formula and interpretation scale.

$$P = \frac{\sum X}{\sum Xi} X\ 100\%$$

Information:
P = Percentage of validity
∑X = Number of validator scores
∑Xi = Total number of ideal scores
(Sudijono, 2009)

Table 1. Criteria for Instrument Validity

Persentase (%)	Kriteria
81 – 100	Very Valid
66 – 80	Valid
56 – 65	Quite Valid
41 – 55	Less Valid
0 – 40	Invalid

(Source: Ariani *et al*., 2016)

The empirical validation test of the questions was carried out with the help of Ms. Excel 2007. Each empirical validation test was analyzed using the following formula.

1. Test the validity of the questions

This test is carried out using the following product moment correlation formula.

$$r_u = \frac{\sum xi.xt}{\sqrt{\sum x_i^2 . x_t^2}}$$

Information:
r_u = correlation coefficient between item scores and total scores
$\sum x_i$ = sum of squares of deviation scores xi
$\sum x_t$ = sum of squares of deviation scores xt

The correlation coefficient value is compared with the correlation coefficient value in the r-table with = 0.05. If the item score correlation coefficient is greater than the r-table correlation coefficient, then the item is declared empirically valid (Matondang, 2009).

2. Reliability test

The reliability test was carried out using the following Kuder Richardson (KR) formula.

$$KR = \frac{k}{(k-1)}\left(\frac{s_t^2 - \sum p_i q_i}{s_t^2}\right)$$

Information:
KR = Kuder Richardson coefficient
k = the number of question items in the instrument
p_i = the proportion of the number of subjects who answered each question item
q_i = 1- p_i
s_t^2 = total variance
(Yusup, 2018)

Interpretation is carried out with the following relationship level coefficient coefficient interval criteria (Table 2).

Table 2. Interpretation Criteria for Question Reliability Test

Result Interval	Criteria
0,00-0,20	Very low
0,21-0,40	Low
0,41-0,60	Currently
0,61-0,80	High
0,81-1,00	Very high

(Source: Yusup, 2018)

3. Test the level of difficulty of the questions

The difficulty level of the questions is calculated using the following formula.

$$P = \frac{B}{JS}$$

Information:
P = proportion or difficulty index
B = the number of students who answered the question correctly
JS = total number of students

The difficulty index of the questions is categorized in Table 3.

Table 3. Interpretation of the Difficulty Level of Questions

Range	Interpretation
0,00 – 0,30	hard
0,31 – 0,70	currently
0,71 – 1,00	easy

(Source: Riyani *et al*., 2017)

4. Test the power of different questions

The discriminatory power test is carried out by calculating the discriminant value through the following formula.

$$D = \frac{\sum XA}{n_A} - \frac{\sum XB}{n_B}$$

Information:

D = Discrimination Index (differentiating power of questions)

$\sum X_A$ = The number of test takers who answered the upper group correctly

$\sum X_A$ = The number of test takers who answered the bottom group correctly

n_A = Number of participants in the upper group test

n_B = Number of lower group test takers

Discriminant values are interpreted according to the criteria in Table 4.

Table 4. Interpretation of Differential Power of Questions

Range	Interpretation
0,70 – 1,00	*Excellent*
0,40 – 0,69	*Good*
0,20 – 0,39	*Satisfactory*
0,00 – 0,19	*Poor*
Negative sign	ugly as hell

(Source: Susanto *et al*., 2018)

RESULTS AND DISCUSSION

At the analysis stage, the results showed that as many as 65.50% of students of the Department of Biology, FMIPA, Universitas Negeri Medan claimed to have critical thinking skills that still needed to be improved and as many as 72.38% stated that there was a need for additional treatment in the learning system that could elevate critical thinking skills. Based on a review of the characterization of the questions in the Lower-Level Taxonomy of Organisms course, it was found that the questions tested to students in the course were still arranged in levels C2 and C3 (84.37%) and C4 (15.63%). This indicates that the LOTS standard questions are more dominant to be tested in these courses.

LOTS-based questions are not yet oriented towards skills in connecting ideas and facts as well as analyzing and synthesizing information to solve problems (Yuliadini & Respati, 2019). The same thing was also reported by Harta (2017) that the essence of deductive and inductive thinking in exploring and identifying a problem has not been contained in LOTS-based questions as a necessary condition in developing Critical Thinking Skills. Therefore, it can be indicated that there is a need to develop HOTS-based questions to improve critical thinking skills in the Lower-Level Organisms Taxonomy course, especially in Monera material.

At the design stage, it was found that the design of HOTS-based questions on the Monera material followed the Course Learning Outcomes (CPMK) and Learning Objectives Indicators. Then, it is aligned with a grid of questions that lead to mastery of cognitive levels C4, C5, and C6. The CPMK and Learning Objective Indicators can be seen in Table 5. Meanwhile, the HOTS-based question grid on the Monera material can be seen in Table 6.

Table 5. CPMK and Indicators of Learning Objectives for Monera Materials

Course Learning Outcomes	**Learning Objectives Indicator**
Have sufficient knowledge and understanding of the principles of classification, nomenclature, identification, taxon description, habitat, breeding methods, and role in the life of Monera	1. Students are able to understand the classification of bacteria based on the type of nutrient supply well 2. Students are able to understand the use of bacteria in everyday life well

(Source: Syllabus of Taxonomy of Lower-Level Organisms, Department of Biology FMIPA Unimed, 2021)

Nurlita (2016) asserts that the preparation of a question grid needs to be done as a basis for measuring the abilities and skills that have been obtained by students after participating in learning. The question grid needs to be synchronized with the indicators of the learning objectives to be achieved in

following the course. In line with this, Suhady *et al*. (2020) also revealed that a system and evaluation tool must be clearly formulated so that the aspects to be measured can be in accordance with the expected final achievement indicators. This is stated in the form of a grid which becomes the measurement parameter for the learning outcomes. Thus, the preparation of the question grid at this stage is used as the basis for developing HOTS-based questions on the Monera topic.

Table 6. Grid of HOTS-Based Questions on Monera Materials

Indicators	Question Number	Cognitive Levels C4	C5	C6
Analyze the classification of bacteria based on the provision of nutrients	1,2,3,4,5, 6,7,8,9, 10	√	-	-
Creating the use of bacteria that have a positive impact	11,12,13 ,14,15	-	-	√
Evaluating bacteria that have a negative impact	16,17,18 ,19,20	-	√	-

At the development stage, questions are developed and made based on the previously formulated question grid design. The questions were made in Ms. Word with reference material content using the Taxonomy of Lower-Level Organisms module. This activity is in line with what was stated by Kristanto & Setiawan (2020) that at the stage of developing the questions, it must be adjusted to the physical context that has been formulated, so that the goals to be measured and actualized in the questions can be in accordance with the important points to be evaluated. Furthermore, Suhady, et al. (2020) also emphasized that the selection of reading source material that is used as the basis for developing questions also needs to be done properly. It is intended that the source of the material needed is relevant to the substance of the material and the context used in the problem.

After developing the questions, the questions were validated by three expert validators. Validation includes aspects: the feasibility of the content of the question, the feasibility of the construct, the component of the question, and the feasibility of the language. Each indicator in each of these aspects can be seen in Table 7. While the results of the validation by each expert validator can be seen in Table 8.

Table 7. Aspects and Indicators of Question Validation

Aspects	Indicators
Eligibility of question content	1. The relevance of the question to the material being studied
	2. Relevance of the questions to the indicators of learning objectives
	3. Relevance of questions to learning objectives
	4. The relevance of the question to the cognitive domain
	5. Main problem formulation / essence of question presentation
Feasibility of construct	6. The deepening of the material studied
	7. Student cognitive level
	8. Exploration of student's potential
	9. Variations of strategy to answer questions
Question components	10. Instructions for filling out questions
	11. Presentation of pictures/tables/ other visuals
	12. The substance of the correct answer
Language eligibility	13. Question sentence structure
	14. Grammatical sentence questions
	15. Use of spelling and punctuation

Table 8. Results of Question Validation by Expert Validators

No.	Validators	Percentage Validity	Category
1.	First expert validator	98%	Very good
2.	Second expert validator	91,6%	Very good

Based on Table 7, it can be understood that there are 15 indicators covering 4 aspects of question validation. The indicators and aspects of the validation of the questions were

developed by adjusting to the validation needs of the content of the questions. Based on the results of validation by 2 expert validators as shown in Table 7. It can be understood that the percentage of item validity reached 98% and 91.6% with very good categories. According to the validation results, it can be indicated that the questions developed have met the eligibility criteria for content validity to be tested on respondents. This is in accordance with what was stated by Miarsyah & Ristanto (2019) that the results of the validity by expert validators with a percentage of validity >90% have been declared valid according to content validity and can be tested on respondents to determine the level of empirical validity of the questions developed.

Meanwhile, some qualitative suggestions given by expert validators as input for revising the shortcomings of the developed questions can be seen in Table 9.

Table 9. Suggestions for Question Development by Expert Validators

No.	Validators	Suggestion Given
1.	First expert validator	The format for writing questions can be made to follow the 1.15 spacing to make it more efficient in writing The sentence of the question instructions should be more concise to make it easier for students to work on the questions presented Avoid sentence questions that are double negative Avoid redundancy in the use of question punctuation marks, such as a period at the end of the question sentence The answer options should be sorted in alphabetical or numerical order
2.	Second expert validator	Need to add questions related to CPMK reproduction and habitat from kingdom Monera

It can be seen that the first expert validator gave suggestions on the direction of technical improvements in question writing, spacing, use of question sentences to be arranged more efficiently, and presentation of answer options (distractors) to pay more attention to alphabetical order. Meanwhile, the second expert validator gave recommendations to add the reproductive coverage and habitat of Monera in the problem. The suggestions given aim to improve and complete the lack of questions, so that the indicators contained in the questions can be achieved properly. This is in line with Astuti, et al (2017) that validation by expert validators serves to assess the accuracy applied in a product/object developed according to the views of competent and expert people in their field. Surya (2020) also emphasized that the expert validator is tasked with assessing content validity which aims to determine the level of conformity of the questions with the indicators that have been designed. Based on this, the suggestions given by the expert validators are needed to improve the quality of the questions in order to achieve good empirical validity.

At the implementation stage, an empirical validation test is carried out by applying the questions to be answered by students. There are 30 students who have taken the Lower-Level Organisms Taxonomy course as a course that teaches Monera material. This is in accordance with what was explained by Aviory & Susetyawaty (2021) that the implementation phase of the question was carried out by a limited trial to a group of students who had followed the material tested in the question. This aims to determine the feasibility of the questions in terms of students' understanding of the concepts tested in the questions. Sari & Ermawati (2021) also emphasized that the limited trial of questions to students was a determinant to determine the ambiguity of students in answering questions. That is why the selected students come from those who have received the material being tested because they can detect errors in questions through their understanding.

At the evaluation stage, an evaluation of the overall question development is carried out.

According to Rayanto & Sugianti (2020) this evaluation includes formative and summative evaluations. Furthermore, Scriven (1967) in Selegi (2017) suggests that formative evaluation is closely related to information gathering aimed at improving the developed model/product. Meanwhile, summative evaluation is more closely related to evaluation for final decision making regarding the feasibility of the developed model/product. Based on this, the formative evaluation in this study was carried out from the results of content validation analysis by expert validators and empirical validation based on 4 aspects of the test, namely validation test, reliability test, difficulty level test, and different power test. These kinds of tests were carried out with the aim of knowing the feasibility of the questions being used as an evaluation tool in learning.

Table 10. Result of Question Validity Test

Question Number	R_{count}	r_{table} (Significance Level 5%)	Question Category
1	0,154	0,361	Invalid
2	0,277		Invalid
3	0,498		Valid
4	0,418		Valid
5	0,225		Invalid
6	0,190		Invalid
7	-0,15		Invalid
8	-0,018		Invalid
9	0,330		Invalid
10	0,582		Valid
11	0,544		Valid
12	0,385		Valid
13	0,440		Valid
14	0,652		Valid
15	0,306		Invalid
16	-0,062		Invalid
17	0,348		Invalid
18	0,382		Valid
19	0,548		Valid
20	0,745		Valid

The results of the validation test questions developed can be seen in Table 10. Based on the calculation of the validity of the questions, it can be seen that from the 20 HOTS-based questions that were developed, there were 10 items that were declared valid. This amount covers half of the total items developed. The questions were tested on 30 students. The rtable value for the number of 30 respondents at the 5% significance level is 0.361. The validity of the items is determined if $r_{count} > r_{table}$. This is in line with Zulyusri, et al. (2017) that the validity of a question shows the level of validity of a question to measure students' understanding precisely about the content of the material contained in the question. This is done by calculating the validity aspects of the questions to be compared with the rtable criteria at a certain significance level. Ramadhani, et al. (2015) also stated that questions that were declared valid had met the further test criteria to determine the level of reliability. Based on this, the items tested for reliability are items that are declared valid according to calculations.

The results of the reliability test which were declared valid according to the calculation were 0.780 with a high interpretation. This is in accordance with Amelia (2016) that the results of the reliability test questions that are in the range of 0.71-0.90 are in the high category. Ipin (2018) also revealed that questions with high reliability indicate that the questions are able to give the same results if tested on the same group at different times. Based on this, it can be understood that the HOTS-based questions developed have high consistency in measuring students' critical thinking skills on the Monera material.

Table 11. Test Results of Question Difficulty Level

Question Number	Difficulty Coefficient	Interpretation
3	0,50	Currently
4	0,33	Currently
10	0,26	Hard
11	0,23	Hard
12	0,40	Currently
13	0,33	Currently
14	0,60	Currently
18	0,46	Currently
19	0,56	Currently
20	0,36	Currently

Furthermore, the results of the test difficulty level of questions that are declared

valid can be seen in Table 11. It can be understood that there are 2 difficult questions and 8 moderate questions. Problems that are difficult have a coefficient of difficulty level between 0.00-0.30. Meanwhile, questions that are moderate have a coefficient of difficulty between 0.31-0.70. Rahmasari & Ismiyati (2016) explained that difficult questions are not good questions to use, because students have difficulty answering these questions, so they need to be revised so that the questions can be accepted. Meanwhile, Kholis (2017) also stated that questions with a medium level are good to use because the proportion of students' abilities is qualified to solve problems in these questions. Based on these considerations, it can be understood that there are 8 questions that are qualified to use, because students do not find it too difficult or easy to answer these questions. Meanwhile, 2 other questions that are categorized as difficult should not be used in learning, but need to be revised and evaluated so that they are not too burdensome for students to answer these questions.

Table 12. Results of the Differentiating Power of Questions

Question Number	Discrimination Index	Interpretation
3	0,33	*Satisfactory*
4	0,40	*Good*
10	0,53	*Good*
11	0,33	*Satisfactory*
12	0,40	*Good*
13	0,40	*Good*
14	0,80	*Excellent*
18	0	*Poor*
19	0,46	*Good*
20	0,60	*Good*

The results of the test of differentiating power of questions that are declared valid can be seen in Table 12. It can be seen that there are 6 questions in the good category (good), 2 questions in the satisfactory category (satisfactory/enough), 1 question in the excellent category (very good), and 1 question in the poor category (weak). According to Solichin (2017) questions categorized as good (good) indicate that the question is able to distinguish between high and low capable students well. On the other hand, questions in the poor (weak) category are very irrelevant to discriminate against the gradations of students' abilities. Furthermore, Magdalena, et al. (2021) also explained that the questions with the excellent category (very good) are the proportion of items that can determine the disparity in the ability of students between high and low abilities very well. Meanwhile, Iskandar & Rizal (2018) asserted that questions with a satisfactory discriminant index (satisfactory/enough) are questions that can function to measure differences in the abilities of students well, but need to be revised. Based on these considerations, it can be understood that from the total that was declared valid there was only 1 question that was less qualified in measuring differences in student competencies.

Summative evaluation in this study relates to the selection of questions that really deserve to be used as an evaluation tool in learning based on content and construct validity tests. The questions that are really feasible are questions number 3, 4, 10, 11, 12, 13, 14, 18, 19, and 20. According to Lestari (2020), HOTS-based questions are important to be applied to learning because they can train reasoning power high level, namely skills in analyzing, synthesizing, evaluating, and creating something as an effort to solve problems.

Pohan & Hasibuan (2019) also added that HOTS skills have positive implications for the preparation of students to face the challenges of the 21st century which is full of the use of technology, information, and communication. These implications relate to skills in solving complex problems from various fields. Ritonga, et al. (2020) also emphasized that the development of critical thinking skills through the implementation of HOTS is a need for today's students, because these skills can strengthen students' mental attitudes in criticizing, solving problems, and applying thinking in complex and varied needs. This is

carried out in a process of knowledge that is qualified to utilize high-level abilities. Based on this, it can be understood that the developed HOTS-based questions have implications for increasing high and qualified reasoning power to hone students' skills in solving problems as one of the aspects needed in the 21st century.

Specifically, this HOTS-based question is aimed at and adapted to the learning indicators in the Lower-Level Organisms Taxonomy course. However, in more complicated fields, such as Biotechnology and Genetics, these HOTS-based questions have a position and function to develop students' scientific reasoning power at a higher level of thinking. The material tested is also a bridge in opening the horizons of knowledge in this field. This is in line with Linda, et al (2019) that HOTS-based questions can hone high-level reasoning power and the material tested substantively can be an introduction to other relevant material. Therefore, the higher order thinking skills acquired by students can be applied in solving critical problems in other more complex fields of Biology.

CONCLUSION

Based on the research that has been done, it can be concluded that from a total of 20 questions on the HOTS-based Monera material that was developed, only 10 questions are suitable for use based on content validity and empirical tests. The questions that are declared eligible are numbered 3, 4, 10, 11, 12, 13, 14, 18, 19, and 20. These questions have been empirically tested to improve students' critical thinking skills on Monera material.

This study only focuses on developing HOTS-based questions on Monera material to improve students' critical thinking skills based on validity tests. This study has not studied in depth the follow-up to the development of questions into the learning mechanism as an instrument of learning evaluation. Therefore, it Is highly recommended for other researchers who want to develop HOTS-based questions on other materials to test the questions developed in learning to find out the synchronization of strategies and learning methods that are relevant to HOTS-based questions developed to achieve good quality critical thinking skills.

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.

Amelia, M. A. (2016). Analisis Soal Tes Hasil Belajar *High Order Thinking Skills* (HOTS) Matematika Materi Pecahan untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian*, *20*(2), 123-131.

Andriyani, R., & Saputra, N. N. (2020). Optimalisasi Kemampuan *Higher Order Thinking Skills* Mahasiswa Semester Awal melalui Penggunaan Bahan Ajar Berbasis Berpikir Kritis. *Al-Khwarizmi: Jurnal Pendidikan Matematika dan Ilmu Pengetahuan Alam*, *8*(1), 77-86.

Ariani, D., Saptaningrum, E., & Siswanto, J. (2016). Instrumen Penilaian Keterampilan Kerja Ilmiah pada Pembelajaran Fisika Berbasis *Inquiry. Jurnal Penelitian Pembelajaran Fisika*, *7*(2), 109-117.

Astuti, I. A. D., Sumarni, R. A., & Saraswati, D. L. (2017). Pengembangan Media Pembelajaran Fisika *Mobile Learning* Berbasis Android. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *3*(1), 57-62.

Aviory, K., & Susetyawati, M. E. (2021). Kualitas Soal HOTS (*High Order Thinking Skill*) pada Siswa SMP Kelas VII. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *10*(2), 639-647.

Fanani, M. Z. (2018). Strategi Pengembangan Soal HOTS pada Kurikulum 2013. *Edudeena: Journal of Islamic Religious Education*, *2*(1), 57-76.

Feriyanto, F., & Putri, R. O. E. (2020). Developing Mathematics Module Based on Literacy and Higher Order Thinking Skills (HOTS) Questions to Train Critical Thinking Ability of High School Students in Mojokerto. *Journal of Physics: Conference Series, 1594*(1), 012014.

Harta, J. (2017). Pengembangan Soal Esai Berbasis HOTS untuk Menyelidiki Keterampilan Pemecahan Masalah Siswa SMA. *Jurnal Penelitian USD. 21*(1), 62-69.

Hidayat, A. A. (2021). *Menyusun Instrumen Penelitian & Uji Validitas-Reliabilitas*. Health Books Publishing.

Indah, P. (2020). Development of HOTS (High Order Thinking Skill) Oriented Learning Through Discovery Learning Model to Increase the Critical Thinking Skill of High School Students. *International Journal of Chemistry Education Research*, *4*(1), 26-32.

Ipin, A. (2018). Pengembangan Soal-soal Pilihan Ganda untuk Mengukur Kemampuan Berpikir Kritis Siswa pada Konsep Sistem Regulasi Manusia untuk Jenjng SMA. *Jurnal Mangifera Edu*, *3*(1), 26-39.

Iskandar, A., & Rizal, M. (2018). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(1), 12-23.

Kholis, R. A. N. (2017). Analisis Tingkat Kesulitan (*Difficulty Level*) Soal pada Buku Sejarah Kebudayaan Islam Kurikulum 2013. *Jurnal Pendidikan Agama Islam*, *14*(2), 305-315.

Kristanto, P. D., & Setiawan, P. G. F. (2020). Pengembangan Soal HOTS (*Higher Order Thinking Skills*) Terkait dengan Konteks Pedesaan. *PRISMA, Prosiding Seminar Nasional Matematika 3* 370-376.

Lestari, A., Saepulrohman, A., & Hamdu, G. (2016). Pengembangan Soal Tes Berbasis HOTS pada Model Pembelajaran Latihan Penelitian di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *3*(1), 74-83.

Lestari, I. D. (2020). Kemampuan Mahasiswa Calon Guru Pendidikan Biologi UNTIRTA dalam Pembuatan Soal *Higher Order Thingking Skill* (HOTS). *Prosiding Seminar Nasional Pendidikan FKIP* 3(1) 126-128.

Linda, T., Ismail, I., & Wiharto, W. (2019). Pengaruh Penerapan Model Pembelajaran *Discovery Learning* terhadap Kemampuan Siswa Menyelesaikan Soal-soal Biologi Berkategori HOTS di SMA Negeri 1 Tana Toraja. *Seminar Nasional Biologi*.

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Jurnal Bintang*, *3*(2), 198-214.

Matondang, Z. (2009). Validitas dan Reliabilitas suatu Instrumen Penelitian. *Jurnal Tabularasa*, *6*(1), 87-97.

Miarsyah, M., & Ristanto, R. H. (2019). Memberdayakan Keterampilan Mengembangkan Soal HOTS pada Guru Biologi di Kabupaten Bekasi. *BAKTIMAS: Jurnal Pengabdian Pada Masyarakat*. *1*(4), 151-159.

Nadifatinisa, N., & Sari, P. M. (2021). Pengembangan Lembar Kerja Peserta Didik (LKPD) Berbasis *Higher Order Thingking Skill* (HOTS) pada Pembelajaran IPA Materi Ekosistem Kelas V. *Jurnal Pedagogi dan Pembelajaran*, *4*(2).

Nurlita, M. (2015). Pengembangan Soal Terbuka (Open-Ended Problem) pada Mata Pelajaran Matematika SMP Kelas VIII. *PYTHAGORAS: Jurnal Pendidikan Matematika*, *10*(1), 38-49.

Pohan, H. M., & Hasibuan, A. I. (2019). Peningkatan Kemampuan Berpikir Kritis Siswa melalui Model Pembelajaran Kooperatif *Think Pair Share* di Kelas X SMA Negeri 4 Padangsidimpuan. *Jurnal Biolokus: Jurnal Penelitian Pendidikan Biologi dan Biologi*, *2*(2), 228-233.

Puspitasari, N. A., & Hidayatullah, S. (2020). Peningkatan Wawasan Penerapan Model Pembelajaran Berorientasi HOTS (*Higher Order Thinking Skill*) Bagi Guru MTs dan MA. *Jurnal Inovasi Hasil Pengabdian Masyarakat (JIPEMAS)*, *3*(2), 128-135.

Rahayu, S., Suryana, Y., & Pranata, O. H. (2020). Pengembangan Soal *High Order Thinking Skill* Untuk Meningkatkan Kemampuan Berpikir Tingkat Tinggi Matematika Siswa Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *7*(2), 127-137.

Rahmasari, D., & Ismiyati, I. (2016). Analisis Butir Soal Mata Pelajaran Pengantar Administrasi Perkantoran. *Economic Education Analysis Journal*, *5*(1), 317-330.

Ramadhani, D. K., Susanti, R., & Zen, D. (2015). Pengembangan Soal Keterampilan Proses Sains pada Pembelajaran Biologi SMA. *Jurnal Pembelajaran Biologi: Kajian Biologi dan Pembelajarannya*. *2*(1), 96-108.

Rayanto, Y. H. & Sugianti (2020). *Penelitian Pengembangan Model ADDIE dan R2D2: Teori dan Praktek.* Lembaga *Academic & Research Institute.*

Rini, A. F., & Budijastuti, W. (2022). Pengembangan Instrumen Soal HOTS Untuk Mengukur Keterampilan Pemecahan Masalah Pada Materi Sistem Gerak Manusia. *Berkala Ilmiah Pendidikan Biologi (Bioedu)*, *11*(1), 127-137.

Riswanda, J. (2018). Pengembangan Soal Berbasis *Higher Order Thinking Skill* (HOTS) Serta Implementasinya Di SMA Negeri 8 Palembang. *Didaktika Biologi: Jurnal Penelitian Pendidikan Biologi*, *2*(1), 49-58.

Ritonga, N., Gultom, H. S. B., & Nazliah, R. (2020). Peningkatan Hasil Belajar IPA melalui Pendekatan Keterampilan Proses. *Jurnal Biolokus: Jurnal Penelitian Pendidikan Biologi dan Biologi*, *3*(1), 293-297.

Riyani, R., Maizora, S., & Hanifah, H. (2017). Uji Validitas Pengembangan Tes Untuk Mengukur Kemampuan Pemahaman Relasional Pada Materi Persamaan Kuadrat Siswa Kelas VIII SMP. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *1*(1), 60-65.

Rohim, D. C. (2019). Strategi Penyusunan Soal Berbasis HOTS Pada Pembelajaran Matematika SD. *Jurnal Riset Dan Konseptual*, *4*(4), 436-446.

Rosmalinda, N., Syahbana, A., & Nopriyanti, T. D. (2021). Analisis Kemampuan Berpikir Kritis Siswa SMP dalam Menyelesaikan Soal-soal Tipe PISA. *Transformasi: Jurnal Pendidikan Matematika dan Matematika*, *5*(1), 483-496.

Selegi, S. F. (2017, December). Model Evaluasi Formatif-Sumatif terhadap Motivasi Belajar Mahasiswa pada Mata Kuliah Perencanaan Pengajaran Geografi. *Prosiding Seminar Nasional Program Pascasarjana Universitas PGRI Palembang*.

Sari, I. P. M., & Ermawati, F. U. (2021). Instrumen Tes Diagnostik Konsepsi Lima Tingkat pada Materi Gerak Lurus: Pengembangan, Uji Validitas dan Reliabilitas serta Uji Coba Terbatas. *PENDIPA Journal of Science Education*, *5*(2), 152-162.

Sole, F. B., & Anggraeni, D. M. (2020). Analysis Of High Order Thinking Skill (HOTS) In Joint Midterm Examination At Yapnusda Elementary School. *Journal Of Physics: Conference Series 1440(1)* 012102.

Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirasat: Jurnal Manajemen dan Pendidikan Islam*, *2*(2), 192-213.

Sudijono, A. (2009). *Pengantar Statistik Pendidikan.* Rajawali Pers.

Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan Soal untuk Mengukur *Higher Order Thinking Skill* (HOTS) Siswa. *Jurnal Gantang*, *5*(2), 143-150.

Surya, R. (2020). Pengaruh Model *Discovery Learning* Berbantuan Multimedia terhadap Keterampilan Proses Sains, Berpikir Tingkat Tinggi dan Keterampilan Bertanya pada Materi Sistem Pernafasan di SMA Negeri 5 Langsa. *Jurnal Biolokus: Jurnal Penelitian Pendidikan Biologi dan Biologi*, *3*(1), 234-244.

Susanto, H., Rinaldi, A., & Novalia, N. (2015). Analisis Validitas Reliabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas XII IPS di SMA Negeri 12 Bandar Lampung Tahun Ajaran 2014/2015. *Al-Jabar: Jurnal Pendidikan Matematika*, *6*(2), 203-218.

Warsah, I., Morganna, R., Uyun, M., & Afandi, M. (2021). The Impact of Collaborative Learning on Learners' Critical Thinking Skills. *International Journal of Instruction*, *14*(2), 443-460.

Widana, I. W., Parwata, I., & Sukendra, I. K. (2018). Higher Order Thinking Skills Assessment Towards Critical Thinking on Mathematics Lesson. *International Journal of Social Sciences and Humanities*, *2*(1), 24-32.

Yuliandini, N., Hamdu, G., & Respati, R. (2019). Pengembangan Soal Tes Berbasis *Higher Order Thinking Skill* (HOTS) Taksonomi Bloom Revisi di Sekolah Dasar. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *6*(1), 37-46.

Yusup, F. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian Kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan*, *7*(1), 17-23.

Zulyusri, Z., Sumarmin, R., & Miswati, M. (2017). Pengembangan Soal Biologi Berbasis Literasi Sains untuk Siswa SMA Kelas X Semester 1. *Jurnal Bioeducation*. *1*(1), 88-94.