ALGEBRA : JURNAL PENDIDIKAN, SOSIAL DAN SAINS

ISSN : 2807-663X e-ISSN : 2807-629X
Journal Homepage : https://algebra.yana.or.id/index.php/ejournal/index
Volume 6 Number 2 (2026)

DEVELOPMENT OF HIGHER ORDER THINKING SKILLS (HOTS) QUESTIONS ON METABOLISM MATERIAL TO EMPOWER CRITICAL THINKING OF GRADE XII SENIOR HIGH SCHOOL STUDENTS

Vioni Jilia Eka Putri[1], Evi Suryawati[2], Fitra Suzanti[3]
[123]Program Studi Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pengetahuan, Universitas Riau, Pekanbaru, Indonesia
Correspondence Email: vioni.jilia0195@student.unri.ac.id

ABSTRACT

This study aims to produce a feasible Higher Order Thinking Skills (HOTS) test instrument to measure the critical thinking abilities of Grade XII senior high school students on metabolism material. The type of research used in this study is Research and Development (R&D) with qualitative and quantitative approaches. The R&D research method is used to develop Higher Order Thinking Skills (HOTS) test items on metabolism material to empower the critical thinking of Grade XII senior high school students. The development model used in this study is the Borg & Gall (1983) model. According to Borg & Gall, development research is oriented toward developing and validating products used in education. The product to be developed in this study consists of Higher Order Thinking Skills (HOTS) questions on Biology material for the odd semester. The sequence of development research stages proposed by Borg & Gall consists of seven stages, namely: 1) research and information collecting, 2) planning, 3) developing a preliminary form of the product, 4) preliminary field testing, 5) main product revision, 6) main field testing, and 7) final product revision. The results of the study indicate that the developed Higher Order Thinking Skills (HOTS) test instrument on metabolism material has met the feasibility criteria as a tool for measuring the critical thinking abilities of Grade XII students. This confirms that the instrument is capable of proportionally measuring students' critical thinking skills and can be used as an evaluation tool for biology learning based on Higher Order Thinking Skills (HOTS).

Keywords: Critical Thinking; Higher Order Thinking Skills; Metabolism Material

INTRODUCTION

Biology is a branch of natural science that studies life in all its aspects. Biology learning is not merely the mastery of a body of knowledge in the form of facts, concepts, or principles, but also a process of discovery. Biology learning at the senior high school level is expected to serve as a medium for students to study themselves and the surrounding environment, as well as the prospects for further development in its application to daily life (Febrianti et al., 2018).

Twenty-first century learning is directed toward the development of competencies consisting of four abilities that students must possess, commonly referred to as the 4C skills: critical thinking, communication, collaboration, and creativity. Through these competencies, students are expected to be able to solve problems encountered in the world of education as well as in daily life logically and critically (Melawati, 2024). The main objective of 21st century education is no longer limited to teaching reading, writing, and arithmetic, but rather to fostering thinking skills, including problem-solving skills, decision-making skills, critical thinking skills, creative thinking skills, and Higher Order Thinking Skills (HOTS).

Based on data from the Programme for International Student Assessment (PISA), Indonesian students' critical thinking abilities are still relatively low. This is consistent with Indonesia's PISA results, which indicate a decline in educational scores in 2018 and 2022 compared to 2015. In 2022, PISA results showed a global decline in learning achievement influenced by the COVID-19 pandemic. Nevertheless, Indonesia's ranking in PISA 2022 increased by 5–6 positions compared to 2018. From 2000 to 2022, Indonesia consistently ranked among the bottom ten countries in scientific literacy during its participation in the PISA assessment (Wandari et al., 2024:19). This reflects the low level of higher-order critical thinking skills among students aged 15 years and above. Learning that develops 21st century skills encourages prospective biology teachers to master subject matter, learning strategies, and learning outcome assessments (Aulia, 2019).

Based on the results of a pre-survey conducted through interviews with teachers, it was found that the implementation of Higher Order Thinking Skills (HOTS) questions on metabolism material in Grade XII senior high schools is considered very important to encourage students to think more critically and to analyze problems in depth. Although teachers have attempted to apply HOTS-based questions, their implementation has only reached approximately 50% due to students' limited abilities, as they often respond randomly without in-depth analysis. Another challenge faced is the lack of teacher experience in implementing a learning ecosystem that supports Higher Order Thinking Skills (HOTS), considering that at the senior high school level students are not yet accustomed to this approach. Nevertheless, the influence of HOTS questions on the development of students' thinking patterns has been quite positive, although currently only about 2% of students demonstrate critical thinking abilities after their implementation. Therefore, it is necessary to further develop and implement Higher Order Thinking Skills (HOTS) questions more intensively to empower students' critical thinking abilities in the future.

One strategic effort that can be made to improve critical thinking skills is through the development of Higher Order Thinking Skills (HOTS)-based questions. Higher Order Thinking Skills (HOTS) refer to the ability to critically and creatively connect, manipulate, and transform previously acquired knowledge and experience. These skills are essential in decision-making and problem-solving in new situations. Higher-order thinking skills are basically divided into critical thinking, creative thinking, and metacognitive thinking, each of which has different characteristics. Higher Order Thinking Skills (HOTS) involve the ability to understand and find solutions to problems in varied ways, differing from usual approaches, and from different perspectives according to each student's abilities. Therefore, HOTS-based questions are needed to train students' skills in problem-solving, creative thinking, and critical thinking. According to Lismaya (2019), critical thinking is an intellectual process aimed at believing in something in order to take action by evaluating information obtained through observation, experience, reasoning, or communication.

The success of Higher Order Thinking Skills (HOTS) as learning outcomes is determined by the assessment conducted. In fact, teachers already possess sufficient knowledge about HOTS-based questions. However, difficulties in developing appropriate stimuli cause teachers to include only a limited number of HOTS questions in daily assessments (Lestari, 2020). When students successfully acquire these skills, they become critical, logical, reflective, metacognitive, and creative—characteristics of Higher Order Thinking Skills (HOTS). This has implications for improving students' abilities to correlate information with problems and reconstruct it into solutions. These skills serve

as fundamental capital for developing a competitive and adaptive generation in the global era (Hartono et al., 2022).

Research by Amin et al. (2022) shows that the average total Higher Order Thinking Skills (HOTS) of biology students at the levels of analyzing, evaluating, and creating indicate that further efforts are still needed for students to achieve adequate HOTS competencies. Research by Rahmawati and Trimulyono (2021) shows that the reliability of the instrument using percentage agreement was categorized as reliable, with results of 99.53% for multiple-choice items and 100% for essay items. The validity score of the HOTS assessment instrument was 3.94 for multiple-choice items 1–4 and a score of 4 for multiple-choice items 5–15 and essay items 1–5, categorized as very valid in terms of content, item construction, language, and Higher Order Thinking Skills (HOTS). Based on these results, it can be concluded that the HOTS assessment instrument on biodiversity material is valid and reliable based on expert validation and can therefore be used to measure students' abilities in the learning process.

The selection of metabolism material as the focus for the development of Higher Order Thinking Skills (HOTS) questions is based on the urgency and relevance of this topic in the context of education and real life. Metabolism is not only an integral part of the senior high school Biology curriculum but also plays an important role in ecosystem sustainability and human well-being. By developing HOTS-based questions on this material, students are expected to develop critical, analytical, and creative thinking skills in understanding complex concepts. These questions are expected to support meaningful learning and enable students to engage in higher-level cognitive thinking, as well as serve as materials for student assessment (R. Azizah et al., 2023).

RESEARCH METHODS

The type of research used in this study is Research and Development (R&D) with qualitative and quantitative approaches. The R&D research method is used to produce a specific product and to test the effectiveness of that product (Sugiyono, 2019). The development model applied in this study is the model proposed by Borg & Gall (1983). According to Borg & Gall, development research is oriented toward developing and validating products used in education. The product developed in this study consists of Higher Order Thinking Skills (HOTS) questions for Biology material in the odd semester. The sequence of development research stages proposed by Borg & Gall consists of seven stages, namely: 1) research and information collecting, 2) planning, 3) developing a preliminary form of the product, 4) preliminary field testing, 5) main product revision, 6) main field testing, and 7) final product revision.

The development research was conducted with Grade XII students. The research subjects were selected on a large scale and consisted of 37 students from Grade XII Science 2 at SMA N 14 Pekanbaru. In addition, the participants included biology subject teachers or teachers who teach metabolism material, teachers who have experience in developing Higher Order Thinking Skills (HOTS) questions, and teachers who were willing to participate in the study.

The research instrument used in this study consisted of Higher Order Thinking Skills (HOTS) test items. Validation sheets were used to determine the validity level of the test items, which were evaluated by two expert validators. The HOTS test items were also used to determine reliability, difficulty level, distractor effectiveness, and discrimination index after being tested. The data collection technique in this development research was a written test, with data collection instruments in the form of multiple-choice and essay questions. The techniques and instruments for data collection are summarized in the table below.

Table 1. Data Collection Techniques and Instruments

No	Validity	Technique	Instrument
1	Content	Preparation of the test blueprint for Biology HOTS items	Biology HOTS test blueprint sheet
			Test item validation sheet
2	Construct	Test	Initial product of Biology HOTS questions
			Final product of Biology HOTS questions

(Source : Winanda 2023)

The data analysis techniques used in this study include qualitative and quantitative analyses. The validation sheets were analyzed qualitatively, while the Higher Order Thinking Skills (HOTS) test

items were analyzed quantitatively. Qualitative analysis was used to determine the theoretical quality of the HOTS questions, whereas quantitative analysis was used to determine the empirical quality of the product.

Item difficulty refers to the measure of how easy or difficult a test item is, determined by the proportion of test takers who answer the item correctly. The higher the percentage of correct responses, the easier the item; conversely, the lower the percentage, the more difficult the item.

Item discrimination is the ability of a test item to distinguish between students in the upper group, namely high-ability students, and those in the lower group, namely low-ability students. In small-group testing, the discrimination index is calculated based on dividing the group into two parts: the upper group and the lower group.

Distractor effectiveness refers to the pattern showing how test takers choose answers based on the options available for each test item. It reflects the extent to which distractors function to direct test takers toward incorrect but plausible answers. If distractor effectiveness is poor, the discrimination index will be low and the item difficulty level will also be low. Conversely, if the distractors in a test item function effectively, the discrimination index will be good and the item difficulty level will be high. All distractors (incorrect options) should be selected by at least some test takers.

RESULTS AND DISCUSSION

The product titled "Development of Higher Order Thinking Skills (HOTS) Questions on Metabolism Material to Empower the Critical Thinking of Grade XII Senior High School Students" was validated by experts in their respective fields, namely three biology teachers with experience in teaching at the senior and junior high school levels. The expert review resulted in the following conclusions:

Teacher Validation

The validation was conducted by three teachers who are competent in the field of biology to assess the feasibility of the developed Higher Order Thinking Skills (HOTS) test instrument on metabolism material. This validation process aimed to ensure that each test item was aligned with learning indicators, assessment constructs, and proper test construction principles. The validators provided assessments based on two aspects, namely content and construct aspects. The evaluation results from the three validators are presented in Table 2 below.

Table 2. Validator Assessment Results

Assessed Aspect	Validator 1	Validator 2	Validator 3
Content	96,6%	96,6%	93,3%
Construct	96,6%	90%	93,3%
Language	100%	100%	96%
Total	97,6%	95,2%	94%
Category	Very Good	Very Good	Very Good

(Source: Processed Data (2025)

Based on Table 2, it can be seen that the assessment results from the three validators show that the first validator obtained a total score of 97.6%, the second validator 95.2%, and the third validator 94%. All three validators fall into the *very good* category. These values indicate that the Higher Order Thinking Skills (HOTS) test instrument on metabolism material has a high level of feasibility in terms of content, construct, and language aspects, and therefore can be considered suitable for use without the need for substantial revision.

Product Testing

Product testing in this research on the development of Higher Order Thinking Skills (HOTS) questions on metabolism material was conducted with senior high school students. Small-group testing was carried out with 12 Grade XII students to evaluate language clarity, item readability, and completion time. After revisions were made based on feedback obtained during the small-group testing stage, large-group testing was conducted with 37 Grade XII students. The large-group testing aimed to obtain quantitative data regarding the quality of the test items, including validity, reliability, item difficulty, and discrimination power. All participating students were asked to serve as respondents by completing the test items.

Small-Group Testing

Small-group testing was conducted with 12 Grade XII students as an initial step to examine readability, language clarity, and the suitability of the Higher Order Thinking Skills (HOTS) questions on metabolism material before being applied on a broader scale. Data were obtained through questionnaires completed by the respondents after working on the instrument. The results of the small-group testing are presented in the following table:

Table 3. Results of Small-Group Testing

Assessed Aspect	Percentage	Category
Content	92,8%	Very Good
Construct	96,2%	Very Good

(Source: Processed Data (2025)

Based on Table 3 above, it can be seen that the results of the small-group testing indicate that the content aspect obtained a percentage of 92.8% and the construct aspect obtained 96.2%, both of which fall into the *very good* category. This indicates that, in general, the Higher Order Thinking Skills (HOTS) questions on metabolism material are clear, easy to understand, and can be used for large-group testing without significant revision.

Large-Group Testing

Large-group testing was conducted with 37 Grade XII students to obtain more comprehensive data regarding the feasibility of the Higher Order Thinking Skills (HOTS) test instrument on metabolism material. The purpose of this testing was to determine the consistency of assessment results from a larger number of respondents, as well as to evaluate the quality of the instrument in terms of both content and construct more broadly. Data were collected through questionnaires completed by all respondents after working on the test items.

Table 4. Results of Large-Group Testing

Assessed Aspect	Percentage	Category
Content	92,8%	Very Good
Construct	96,2%	Very Good

(Source: Processed Data (2025)

Based on Table 4, it can be seen that the results of the large-group testing show that the content aspect obtained a percentage of 92.8% with a *very good* category, while the construct aspect obtained 96.2% with a *very good* category. This confirms that the developed Higher Order Thinking Skills (HOTS) test instrument on metabolism material has met the feasibility requirements for use in Grade XII senior high school learning.

Validity Testing

Item Validity Test

Item validity testing was conducted using the Pearson Product–Moment correlation technique between each item score and the total score from 37 respondents. The criteria for determining validity were based on the r-table value of 0.325 ($\alpha = 0.05$; $N = 37$). An item was declared valid if r-calculated ≥ 0.325, and invalid if r-calculated < 0.325. The results of the item validity test are presented in Table 5 below.

Table 5. Results of Item Validity Testing

Description	Result
Number of Respondents	37 students
Total Score Range	20 – 27
Average Total Score	24
r table ($\alpha = 0,05$; $N = 37$)	0,325
Test Result	r calculated ≥ r table

(Source: Processed Data (2025)

Based on Table 5, the analysis using the Anates program produced data on raw scores and weighted scores from 37 test participants, which served as the basis for calculating item validity. The correlation values between each item score and the total score indicate that most test items have r-calculated values greater than the r-table value (0.325) at a 5% significance level, and therefore are declared valid. Of the total 30 items tested, 11 items were found to be valid, while 19 items need to be revised or discarded. This indicates that most of the test items have been able to measure students' higher-order thinking skills in accordance with the established indicators; however, several items still require improvement in sentence formulation and content alignment. The results of the item-by-item validity test are presented in Table 6 below.

Tabel 6. Results of Item-by-Item Validity Testing

Category	Item Numbers	Jumlah
Valid	1, 5, 6, 9, 10, 14, 18, 23, 24, 26, 27	11
Invalid	2, 3, 4, 7, 8, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 25, 28, 29, 30	19

(Source: Processed Data (2025)

Based on Table 6 above, it can be seen that the results of the item validity analysis show that, out of a total of 30 test items, 11 items were valid, namely items 1, 5, 6, 9, 10, 14, 18, 23, 24, 26, and 27, while 19 items were invalid, namely items 2, 3, 4, 7, 8, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 25, 28, 29, and 30. Valid items had correlation values (r-calculated) greater than the r-table value, and therefore were suitable for use in subsequent analyses such as reliability testing, item difficulty analysis, discrimination index analysis, and distractor quality analysis. The invalid items were due to several questions being unclear and difficult to understand, as well as some items having inadequate stimuli.

Reliability Testing

Reliability testing was conducted to determine the level of consistency or stability of the instrument in measuring a variable. An instrument is considered reliable if it yields relatively similar results when administered repeatedly to the same subjects under equivalent conditions. In this study, reliability testing was conducted on 30 test items. The results of the reliability test are presented in the following table.

Table 7. Results of Reliability Testing

Calculation Component	Result
Split-Half Correlation (r_xy)	0,67
Spearman–Brown Reliability Coefficient (r_{11})	0,81
Interpretation	Reliable (high)

(Source: Processed Data (2025)

Based on Table 7, the results of the reliability test using the split-half method through the Anates program show that odd and even item scores were compared to assess internal consistency among test items. The correlation between odd and even scores was then corrected using the Spearman–Brown formula, resulting in a reliability coefficient of 0.81. This value falls into the *very high* category, indicating that the test instrument has a good level of consistency and stability. Thus, the test items used are consistent in measuring students' higher-order thinking skills and are suitable for use in the subsequent data collection stage.

Item Difficulty Testing

Item difficulty testing was conducted to determine the extent to which the developed test items are easy or difficult for students. This analysis was carried out on 30 test items using data from the large-group trial involving 37 Grade XII students.

Table 8. Results of Item Difficulty Testing

Difficulty Level Category	Number of Items	Percentage
Easy	7 items	23.33%
Moderate	12 items	40%
Difficult	11 items	36.67%

(Source: Processed Data (2025)

Based on the data in Table 8 above, it is known that the results of the item difficulty analysis were obtained using the Anates program. The analysis of item difficulty using Anates showed that the percentage values for each item ranged from 23.33% to 40%, with categories ranging from difficult to easy. Overall, most of the items fall into the moderate category with 12 items (40%), followed by the easy category with 7 items (23.33%), and the difficult category with 11 items (36.67%). This indicates that most of the developed items are still relatively easy for students; therefore, in the next stage, several items need to be revised so that their difficulty level falls into the moderate category in order to more optimally measure Higher Order Thinking Skills (HOTS).

Item Discrimination Index Testing

Item discrimination testing was conducted to determine the ability of each test item to distinguish between students with high and low abilities. The discrimination analysis was carried out using the product–moment correlation between each item score and the total score. The results of the discrimination index test are presented in Table 14 below.

Table 9. Results of Item Discrimination Index Testing

Category	Number of Items
Accepted	6

Revised	18
Rejected	2

(Source: Processed Data (2025)

Based on Table 9, the results of the item discrimination analysis obtained using the Anates program indicate that the discrimination index (DI) for each test item ranged from 30% to 60%, falling into the *accepted* category. Most items were categorized as *revised* (0% to 20%), while several items fell into the *rejected* category (−10% to 10%). Items with negative values, such as items number 2 and 12, indicate that these questions were unable to effectively distinguish between upper- and lower-group students and therefore require revision. Overall, the results of this analysis show that the quality of item discrimination is fairly good; however, several items need to be improved in order to more optimally differentiate between students with high and low abilities.

Based on the results of the validity, reliability, difficulty level, and discrimination index tests using the Anates program, it was found that several test items needed to be revised because they did not meet the instrument feasibility criteria. The main reasons for revision included overly long item wording, stimuli that were not aligned with the indicators, and distractors that did not function effectively. The revised items included numbers 5, 10, 23, and 24 with value ranges of 0.031–0.253. Revisions were carried out by improving contextual clarity, updating answer choices to be more proportional, and ensuring alignment between learning indicators and the higher-order thinking constructs being measured. The results of the analysis of valid items are presented in Table 10 below.

Table 10. Results of Valid Item Analysis

Description	Item Numbers	Number of Items
Accepted Items	1, 6, 9, 14, 18, 26, 27	7
Revised Items	5, 10, 23, 24	4
	Total	11

(Source: Processed Data (2025)

Based on Table 10 above, it is known that there are 7 valid and accepted items, namely items 1, 6, 9, 14, 18, 26, and 27, with score ranges from 64.86% to 67.56%. Meanwhile, there are 4 valid items that require revision, namely items 5, 10, 23, and 24, with score ranges from 43.24% to 62.16%. Based on the table above, these items will be revised for the following reasons:

Table 11. Revised Items

Item No	Description
5	Revised because the stimulus did not match the indicator
10	Revised because the distractors did not function properly
23	Revised because the wording was unclear
24	Revised because the item context did not represent a real-life situation

(Source: Processed Data (2025)

Based on Table 11 above, it is known that the four revised items are items number 5, 10, 23, and 24, which must be revised according to the identified shortcomings. After revisions were made based on validator feedback, the quality of the items improved, with better validity and discrimination power.

After analyzing validity, reliability, difficulty level, and discrimination index, the next stage was to analyze the distribution pattern of students' responses for each test item using the Anates program. This analysis aimed to examine the distribution of answer choices (a, b, c, d, e) selected by students and to identify whether the answer key functioned properly and whether the distractors worked effectively. An item is considered good if most high-ability students choose the correct answer key (indicated by double asterisks ** in the answer column), while lower-ability students tend to choose the distractor options. The complete distribution of answer choices is presented in Table 12 below.

Table 17. Distractor Quality

Test Items	A	B	C	D	E	*
1	4--	29**	3+	0--	1-	0
2	0--	0--	34**	0--	3--	0
3	28**	3+	1-	3+	2++	0
4	4+	26**	5--	2+	0--	0
5	5+	23**	4++	4++	1-	0
6	30**	4--	1+	1*	1+	0
7	29**	5--	1-	1-	1-	0

8	32**	1++	1++	1++	2-	0
9	6-	23**	4++	2+	2+	0
10	0--	34**	3--	0--	0--	0
11	29**	7--	0--	1-	0--	0
12	2+	2+	1-	6--	26**	0
13	1+	3--	31**	2+	0--	0
14	2+	1+	31**	3--	0--	0
15	0--	2--	34**	0--	1+	0
16	16**	8-	3+	5++	5++	0
17	0--	0--	4--	30**	3-	0
18	3++	1-	3++	3++	27**	0
19	2++	29**	4--	2++	0--	0
20	31**	2+	1+	1+	2+	0
21	5--	1-	1-	2++	28**	0
22	2++	29**	2++	2++	2++	0
23	3--	31**	2+	0--	1+	0
24	0--	3++	4+	26**	4+	0
25	4--	3+	28**	2++	0--	0
26	1+	0--	0--	34**	2--	0
27	1++	2-	32**	1++	1++	0
28	1++	2-	32**	2-	0--	0
29	1--	36**	0--	0--	0--	0
30	2++	0--	28**	7--	0--	0

(Source: Processed Data (2025)

Based on the results of the response distribution analysis in Table 12 above, it can be seen that most test items have answer keys that function well, as indicated by the dominance of double asterisks (**) on the correct answer choices. This shows that students were consistently able to identify the correct options. However, several items had distractors that did not function effectively because they were rarely or never chosen by students, such as items number 2, 10, and 29. Therefore, these items require revision of the distractor options to make them more attractive and proportional in measuring students' Higher Order Thinking Skills (HOTS).

CONCLUSIONS

Based on the results of the research and discussion on the development of Higher Order Thinking Skills (HOTS) questions on metabolism material to empower the critical thinking of Grade XII senior high school students, the conclusions drawn from the analysis of 30 test items, including item analysis, difficulty level, discrimination power, and distractor quality, are as follows: Based on item quality, of the 30 test items, 11 items (36.6%) were valid, namely items 1, 5, 6, 9, 10, 14, 18, 23, 24, 26, and 27. Meanwhile, of the 30 test items, 19 items (63.3%) were not valid, namely items 2, 3, 4, 7, 8, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 25, 28, 29, and 30. In terms of difficulty level, 11 items (36.67%) were categorized as difficult, 12 items (40%) as moderate, and 7 items (23.33%) as easy. In terms of item discrimination, the discrimination index (DI) of each item ranged from 30% to 60% and was categorized as accepted. Most items fell into the revised category (0% to 20%), while several items were categorized as rejected (−10% to 10%). Items with negative values, such as items number 2 and 12, indicate that these questions were unable to distinguish effectively between upper- and lower-group students and therefore should be discarded. In terms of distractor effectiveness, several items had distractors that did not function properly because they were rarely or never chosen by students, such as items number 2, 10, and 29.

BIBLIOGRAPHY

Amin, A. M., Karmila, F., Muna, L., Hujjatusnaini, N., Adiansyah, R., & Yani, A. (2022). Efek Pembelajaran Online terhadap High Order Thinking Skills pada Siswa Biologi pada Masa Pandemic Covid-19. *Bioedusiana: Jurnal Pendidikan Biologi*, *7*(1), 1–17.

Aulia, N. (2019). Pengaruh Media Pembelajaran Berbasis Video Terhadap Kemampuan Membaca Bahasa Inggris Siswa. *Sustainability (Switzerland)*, *11*(1), 1–14.

Azizah, R., Hernawati, D., & Chaidir, D. M. (2023). Keanekaragaman Gastropoda Air Tawar dan Analisis Trematoda di Ekosistem Situ Kota Tasikmalaya. *Biota : Jurnal Ilmiah Ilmu-Ilmu Hayati*, *8*(1), 19–29.

Borg, W., & Gall, M. (1983). *Educational Research: Introduction*. David McKay Company.

Febrianti, E. S., Karyadi, B., & Kasrina, K. (2018). Penerapan Model Kooperatif Tipe-Group In

Hartono, Hasairin, A., & Diningrat, D. (2022). Development of HOTS-Based Questions on Biology Learning. *Jurnal Biolokus*, *5*(1), 51–54.

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

Lismaya, L. (2019). Berpikir Kritis & PBL (Problem Based Learning). Surabaya: Media Sahabat Cendekia.

Melawati, Y. (2024). Pengembangan Instrumen Higher Order Thinking Skills (HOTS) Untuk Mengukur Kemampuan Berpikir Kritis Siswa Kelas V Sekolah Dasar. In *Αγαη* (Vol. 15, Nomor 1). Universitas Lampung.

Rahmawati, D. E., & Trimulyono, G. (2021). Validitas Instrumen Penilaian Higher Order Thinking Skills (Hots) pada Materi Keanekaragaman Hayati. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, *11*(1), 138–147.

Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Wandari, W., Nursamsu, & Wahyuni, A. (2024). Peningkatkan Kemampuan Berpikir Kritis Siswa Biologi Dengan Menggunakan Model PjBL Berbasis STEM (Science Technology , Engineering , Mathematics). *Biodik: Jurnal Ilmiah Pendidikan Biologi*, *10*(4), 743–754.

Winanda, K. P. (2023). Pengembangan Soal Higher Order Thinking Skills (HOTS) Pada Materi IPA Kelas VIII Semester Ganjil.