Developing and Testing HOTS-Based Evaluation Packages for Metabolism with Science Literacy Skills Aspects

Caecelia Pascella A. Sepe[1], Hendra Michael Aquan[2]
[1,2]Biology Education, Sanata Dharma University, Yogyakarta, Indonesia.
caeceliaapascella@gmail.com[1], hendra.aquan@usd.ac.id[2]

ABSTRACT

During learning at school, difficulties were encountered in learning evaluation, such as low-level questions, lack of question variety, student dishonesty, and poor reading comprehension and understanding of questions. These factors reduce students' reasoning abilities and learning outcomes. Therefore, it is necessary to carry out learning and giving questions that emphasize HOTS abilities and science literacy skills. The purpose of learning evaluation is to measure students' higher-order thinking skills (HOTS) and science literacy skills. This kind of learning and evaluation will be able students improve their higher-order thinking skills, reading comprehension, and understanding of questions, and be able to act honestly in every implementation of learning evaluation. Therefore, this study aims to develop and test the feasibility of an evaluation package that emphasizes HOTS and science literacy skills in the topic of metabolism. Metabolism is a difficult topic for students because it covers a wide range of complex and abstract concepts. This study used the Research and Development (R&D) method with the ADDIE model, which was limited to the Development stage. The research stages consisted of needs analysis, design and development, as well as product validation and revision. Data collection was done through interviews and validation questionnaires and analyzed using quantitative and qualitative data. The results of this study were two packages of HOTS questions (A and B) with aspects of science literacy skills, each containing 20 questions with 100% HOTS proportion and five different question formats. The questions were printed on A4 paper and intended to be used as daily test assessments. The results of the validation showed that both packages were "very feasible" with a percentage of 91.6% (Package A) and 92% (Package B). Therefore, the evaluation package developed was suitable for limited testing after revision.

———————————— ◆ ————————————

A. INTRODUCTION

Ideal learning requires the development of quality learning aligned with 21st-century skills. This skill consists of 4 components include creative thinking, critical thinking and problem-solving, communication, and collaboration (Septikasari & Frasandy, 2018). These four skills can be possessed by students with an effective and efficient learning strategy prepared by the teacher. Students will be directed to be actively involved in solving problems in learning that adhere to discovery learning and meaningful learning (Masgumelar & Mustafa, 2021).

Skills in the 21st century can be applied to biology learning (Ardelia & Juanengsih, 2021). Biology learning emphasizes the components of scientific processes, scientific products, and scientific attitudes. Therefore, creative thinking, critical thinking and problem-solving, communication, and collaboration are needed to understand concepts and principles through biological events or problems in real-life. Students need to be guided and directed to ascertain biology concepts through higher-order thinking skills and equipped with qualified biology literacy to solve problems that arise in their surroundings (Irwandi, 2020).

The success of the learning process can be seen from the achievement of predetermined indicators imposed through evaluation. Learning evaluation is one of the processes to determine the value of learning conducted through measurement activities and assessment learning. Learning evaluation is considered as a salient thing to appoint the achievement of student learning outcomes based on predetermined indicators (Ratnawulan & Rusdiana, 2015). Through learning evaluation, the teacher can find out the ability of students to receive material and apply it in everyday life by solving problems that occur.

Based on the needs analysis interview results at five schools in Yogyakarta, it was found that the five schools experienced difficulties in learning evaluation. The complicacy experienced is that the teacher still gives questions with a low level of cognition with the form of questions that have not varied so that students' HOTS abilities have not been honed to the fullest. The reason for the low usage of HOTS questions is due to teachers' difficulties in designing good HOTS questions, understanding of HOTS questions that are still not qualified, lack of time to make questions, and intricacy in creating stimulus questions that are suitable for students to read and examine. In addition, in the evaluation of online learning, there were several obstacles to network access and interference, the unavailability of cellphones or laptops, and the lack of teacher skills in operating online media. The other interview results stated that the dishonesty of students in working on questions online and offline effect the decrease in students' high order thinking abilities. Therefore, they were not able to work on questions at a high level of cognition. In addition, poor reading comprehension provoked students only guess and answer randomly, as well as duplicate answers from friends or the internet. The obstacles in this evaluation are encountered in metabolism materials which have broad material, conceptual, and are still perceived as abstract.

This lack of higher-order thinking skills is evidenced by a survey from the OECD using the 2018 PISA test that education in Indonesia remains extremely low, ranking 71 out of 79 countries (Schleicher, 2018). Based on this, the teacher's role is needed to lead and guide students to hone thinking skills at a higher level (Baidlowi et al., 2019). Consequently, providing HOTS evaluation questions with various types of questions can be a solution to improve students' higher-order thinking skills. The HOTS capabilities developed are also aligned with the goals of independence and 21st-century life skills (Setiawati et al., 2019). The HOTS cognitive levels in the 2013 Curriculum based on the Revised Bloom's Taxonomy consist of C4 (analyze), C5 (evaluate), and C6 (create). Giving HOTS questions is one of the factors that increase scientific literacy (Thahir et al., 2021). This result is in line with the needs analysis interview results stated that students were still lacking in understanding the reading of the questions and were less scrupulous when reading the questions or the stimulus given, thus consequences in students' low cognitive achievement. For that reason, higher-order thinking skills and literacy skills ought to be improved to foster students' cognitive abilities in understanding science and applying them in real life wisely (Wasis et al., 2020)

Providing HOTS questions and scientific literacy in variety will help students to understand concepts, improving reasoning ability, communication, and problem-solving skills (Wasis et al.,

2020). The provision of those questions can be accommodated by having a package of questions that contains diverse HOTS questions with aspects of scientific literacy skills on an ongoing basis. The question packages provided can help students get used to working on HOTS questions so they can improve their high-level reasoning ability and scientific literacy. According to Limbong & Taufik (2017), providing evaluation packages will help provide a variety of questions with good quality. Teachers can also use it as a reference for their requisite in designing questions and as a medium to ascertain students' abilities in solving HOTS questions and scientific literacy. Furthermore, the package of questions provided will help to supervise students directly (face to face) in working on evaluations during the limited face-to-face period hence students can not commit fraud which can reduce reasoning abilities and scientific literacy.

Based on the explication of this introduction, research was conducted with the title "Development of Evaluation Packages of HOTS-Based Metabolism Material with Science Literacy Skills Aspects". The research conducted is supported by previous research carried out by Musayaroh et al. (2021) who developed HOTS-based scientific literacy questions, Pratiwi et al. (2022) and Setiawan & Mufassaroh (2019) evolved inquiries using the science literacy skills aspect, Muhibbuddin et al. (2022) and Harta (2017) who developed HOTS questions, as well as Avina & Winarsih (2020) and Sasongko et al. (2016) who advanced the question package with Avina & Winarsih, developed the HOTS question package and Sasongko et al. develop a package of literacy questions. The research conducted presented two packages of HOTS questions with science literacy skills aspects, consisting of 10 multiple-choice questions, 2 short answer questions, 3 complex multiple-choice questions (true-false), 2 matching questions, and 3 essay questions with proportions HOTS 100%. This study aims to develop a package of HOTS-based metabolism material evaluation questions with science literacy skills aspects and to test the feasibility of evaluation packages about HOTS-based metabolism material with science literacy skills aspects.

B. METHODS

The method conducted is Research and Development (R&D). The Research and Development (R&D) research method is a class of research to produce particular products and examine the validity and effectiveness of implementing the resulting products (Hanafi, 2017). The ADDIE model has five stages; Analysis, Design, Development, Implementation, and Evaluation (Paidi, 2012). The stages of this research are limited to only three sections, including Analysis, Design, and Development. This research was only carried out until the Development stage because the researcher did not aim to determine the effectiveness of the product that had been developed. In summary, the research and development stages as shown in Figure 1.
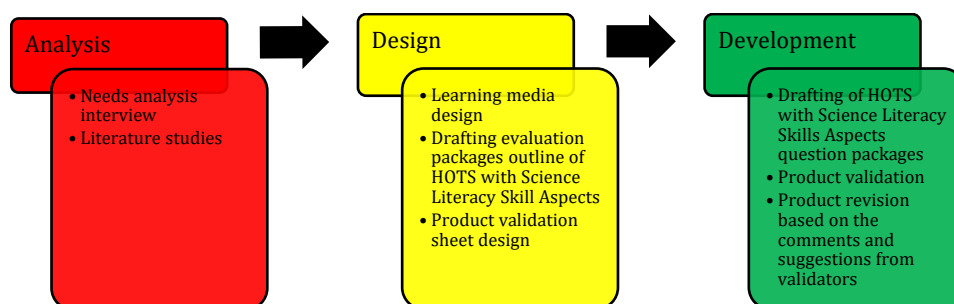

Analysis
- Needs analysis interview
- Literature studies

Design
- Learning media design
- Drafting evaluation packages outline of HOTS with Science Literacy Skill Aspects
- Product validation sheet design

Development
- Drafting of HOTS with Science Literacy Skills Aspects question packages
- Product validation
- Product revision based on the comments and suggestions from validators

Figure 1. Research and Development ADDIE Model Stages

1. Analysis

The analysis stage is carried out by looking for actual problems in the field to provide alternative solutions and linking them to literature studies from various relevant sources. Interviews were conducted directly with high school biology teachers from 5 schools in Yogyakarta City and Sleman Regency. The interview questions proposed regarding learning included learning materials, learning media, learning methods and models, learning evaluation, and development priorities. The interview results accomplished were collected and analyzed to determine alternative solutions to the current precedence problems that occurred. Moreover, literature studies were carried out from books, journals, and other relevant supporting references. This literacy study was conducted to observe references and sources in selecting and determining product development.

2. Design

Product design is made as a solution to the findings of the problems that have been analyzed previously. At this stage, researchers designed a general idea of the developed product, which is an evaluation package about HOTS-based metabolism material with science literacy skills aspects. This design begins with making learning kits, including a syllabus, Lesson Plan, and student worksheet. Afterward, the designing of the question outlines and two packages of HOTS questions with science literacy skills aspects along with the answer key or assessment rubric.

The HOTS cognitive or thinking process dimension category based on Revised Bloom's Taxonomy by Anderson and Krathwohl, is the Cognitive level of analyzing (C4), evaluating (C5), and creating (C6) (Fanani, 2018). Assessment of science literacy skills or abilities, in science literacy learning evaluation in Biology, is adapted from PISA 2018 (OECD, 2019), namely; 1) Explaining phenomena scientifically; 2) Evaluating and designing scientific investigations; 3) Interpreting data and evidence scientifically. The design of the questions content outline in packages A and B is equated in concept, cognitive domain, indicators and form of science literacy skills, as well as type and number of questions. In addition, a validation sheet was designed to be used to validate the question product.

3. Development

Product development is conducted to complement the product design so it becomes a complete product and is ready for validation. The product developed is an evaluation package of HOTS-based metabolism material with science literacy skills aspects. The final evaluation package consists of two packages, packages A and B, printed with each package consisting of 20 questions with details of 10 multiple choice questions, 2 short answer questions, 3 true-false choice questions, 2 matchmaking questions, and 3 essay questions. Overall, there are 2 evaluation packages with 40 questions printed on 80-gram HVS A4 paper. Packages A and B have similarities in concept, HOTS cognitive domains (C4, C5, C6), indicators of science literacy skills, and type of science literacy skills based on PISA 2018, as well as the number and type of questions. The distinction in the evaluation packages is only in the question indicators and the stimulus given. The developed questions are applied as an evaluation on daily tests.

The validation carried out in this study was the validation of the test items and the test material. The validation was implemented by two lecturers and two Biology teachers as the expert in learning evaluation (test item) and test material. The validation stage is conducted for the questions made can be declared very feasible, and their validity is guaranteed. After validation,

the product will be emended according to comments and suggestions from the validator. Data collected was accomplished using interview techniques to obtain needs analysis data and questionnaires to attain data on product validation results developed. The instruments used in collecting data were a list of interview questions for needs analysis and questionnaires for product validation.

a. Interview

An interview is a direct data collection technique to perceive the problems of the object under study by conducting questions and answers to the under-studied thing or through intermediaries (Ramadhani & Bina, 2021). The interview conducted is the structured interview type managed directly with the informant. The informants were high school biology teachers from 5 schools in Yogyakarta City and Sleman Regency. Interviews were undertaken to ascertain the learning process actualized in schools encompassing learning materials, learning models/methods, learning media, learning evaluations, and priorities wanted to be evolved in future learning. The interview results that have been acquired become the basis for consideration in developing products.

b. Questionnaires

Questionnaires are a technique for collecting data to attain information from the object under study using a checklist or list of questions (Ramadhani & Bina, 2021). The questionnaire was used to get data related to product validation results in questions form answered by the validator. The validation questionnaire was filled by one (1) test material expert, one (1) learning evaluation expert (test items), and two high school Biology teachers as practitioners who would become test material experts and learning evaluation experts (test items). The product validation results in the form of comments and suggestions for improvement will be the advisement in perfecting the deficiencies of the products designed previously. The research results were analyzed with quantitative data and qualitative data.

1) Qualitative Data

Qualitative data were obtained from the interview results, and comments and suggestions for improvement get from the feasibility validation questionnaire by learning evaluation experts (test items) and test materials experts. Interview data are utilized to acquire initial data to analyze product development needs. The interview data were attained through direct interviews with Biology teachers at 5 high schools in Yogyakarta City and Sleman Regency. The interview results were written on the interview results sheet and recorded using a recorder (cell phone). The recording results obtained are played back, then transcribed into a script on the interview results sheet. The writing of the interviews results then checked back with the recordings. After that, the data were summarized by taking vital points according to the aspects asked. Furthermore, the data from the interview were recapitulated into a table. The recapitulation table contains learning aspects, interview questions for each aspect, and a summary of interview results at schools. The next step was analyzing the data in the table. The data from the analysis become the consideration in developing product research development. The data is then described descriptively and summarized to commit product development.

Data from the questionnaire results were obtained from 4 validators (two high school biology teachers and two lecturers) as learning evaluation experts (test items) and test materials experts. The validators filled in the data in the comments and suggestions for

improvement. The comments and suggestions for improvement given are assessed one by one. After review, the comments and suggestions are combined and sorted into a number of groups, namely comments/suggestions for learning media, question outlines, test items, and answer keys or assessment rubrics. After sorting, improvements are made according to comments and suggestions. Moreover, sorting is also conducted on parts that can be revised and cannot be revised. Revisions are made under the suggestions or comments given and the ability of researchers to make product revisions.

2) Quantitative Data

Quantitative data were obtained from validation results by learning evaluation experts (test items) and test materials experts consisting of 2 lecturers and 2 Biology teachers. Quantitative data acquisition is used to determine the feasibility or validity of the product being developed. Quantitative data obtained from validation results are in numerical form, so the data must be converted into qualitative data in interval form. This step was conducted in order that the data obtained in numbers may be translated or described. Quantitative data can be converted into qualitative data by adapting Arikunto's formula (Pratiwi, 2019) to find the percentage of validity by dividing the total number of answers in all items (Σx) by the total ideal score in all items (Σxi) multiplied by a constant (100). Mathematically, this formula can be written as follows:

$$P(\%) = \frac{\sum X}{\sum Xi} \times 100 \tag{1}$$

The description of the percentage validity formula: P(%) is validity percentage, $\sum x$ is the total number of answers in all items, $\sum xi$ is the total ideal score in all items and 100 is constant. The overall average validation value of each validator (evaluation (test items) and material) can be discovered using the following formula:

$$P_{final} = \frac{\sum P}{Total\ validators} \tag{2}$$

The description of the percentage validity formula: P_{final} is the final validity percentage and $\sum P$ is the total of validity percentage of all validators. Based on the results of the validation percentage calculation, the feasibility of the evaluation instrument product can be categorized through the following criteria in Table 1:

Table 1. Product Feasibility Criteria

Percentage(%)	Feasibility Criteria
76-100	Very feasible
51-75	Feasible
26-50	Not feasible
≤ 25	Very unfeasible

C. RESULT AND DISCUSSION
1. Evaluation Packages

The product resulting from the research is two evaluation packages of HOTS-based metabolism material with science literacy skills aspects. This product is made as a solution based on the problems found from the needs analysis results. The main problem that gets the highlight is the learning evaluation. Difficulties discovered from the evaluation are teacher still gives questions with a low level of cognition with questions type which has no variations, so students' HOTS abilities are not honed maximally. The reason for the low usage of HOTS questions is due to teachers' difficulties in designing good HOTS questions, lack of understanding about HOTS questions, lack of time to make questions, and difficulties in creating stimulus questions that are suitable for students to read and examine. The dishonesty of students in the evaluation also caused a decline in their high-level thinking skills. For that reason, they were not able to work on questions at a high level of cognition. In addition, lack of reading comprehension makes students only guessing and answering randomly, and duplicating answers from friends or the internet.

Difficulties in giving questions with a high level of cognitive level can be overcome by providing a variety of questions with various types of questions. HOTS is a high-level thinking skill which is a skill to think logically, critically, creatively, and solve problems independently (Setiawati et al., 2019). The development of HOTS questions will be able to help students to enhance their thinking skills so as students do not only memorize readings but also relate them to real life around them so they can understand easily. The importance of HOTS questions was realized by Avina & Winarsih (2020), who developed a sample HOTS package in the form of multiple choices on environmental pollution, and it can be used to measure students' abilities in the C4-C6 domain. Moreover, Harta (2017), in his study, revealed that the HOTS questions developed were able to measure students' problem-solving skills in acid-base solution material.

The dishonesty of students in doing the test can be overcome by randomizing the questions by giving different question packages. According to Limbong & Taufik (2017), providing evaluation packages will help provide a variety of questions with good quality. Various HOTS question packages with good quality will help students hone their reasoning skills with numerous types of questions. It is also in accordance with the development suggestions from Bagus et al. (2016) research, which states that more than one evaluation package must be made in order that students are not able to cheat because they have different questions. This provision is stated in the test instructions, namely, students with odd roll numbers work on package A, while students with even roll numbers work on package B. In addition, the packages provided will supervise students directly (face-to-face) in working on questions so that students can not cheat which can reduce their reasoning abilities.

Insufficient reading comprehension inflicting in students only guessing and answering randomly and duplicating the answers from friends or the internet. This problem can be surmounted by providing stimulus reading questions that are able to present phenomena that occur in real life and can be applied in daily life. The stimulus questions given are in line with science literacy skills that are currently needed to deal with scientific problems that occur in daily life (Irwandi, 2020). The importance of science literacy was discerned by Pratiwi et al. (2022) who developed science literacy skills questions on ecosystem material and stated that the questions prepared could measure students' science literacy in learning Plantae and Animalia.

The material that becomes an obstacle for most schools is metabolism material. Therefore, the research focused on giving HOTS questions on metabolism material with aspects of science literacy skills. The study by giving HOTS questions has also been carried out by Muhibbuddin et al. (2022) who developed 100 HOTS-based multiple-choice questions on metabolism. Consequently, HOTS questions were created in the form of packages with 20 questions in each

package with additional science literacy skills aspects on metabolism material. The addition of science literacy skills aspects and HOTS follows previous research conducted by Musayaroh et al. (2021), who developed a literacy instrument integrated with HOTS questions on acid-base titration material. The product evolved is used as a reference for teachers in developing HOTS questions and science literacy skills on metabolism material. HOTS question is questions that contain skills to think logically, critically, creatively, and solve problems independently (Setiawati et al., 2019). In addition, the development of this problem can improve the science literacy skills of students in supporting 21st-century skills. Science literacy skill is the ability to show students' science literacy which consists of explaining scientific phenomena, evaluating, and designing scientific investigations, and interpreting data and scientific evidence (OECD, 2016). Providing HOTS questions with the competency aspects of science literacy will help students hone their higher-order thinking skills continuously so that they are able to solve existing problems through readings given to the stimulus questions.

The questions developed are used in daily assessments (daily tests) on metabolism material for class XII held at the end of class. This is adjusted with the purpose of making questions to be given at the daily tests. A daily test is a type of formative evaluation carried out at the end of a lesson or module to revise and improve the learning process in the classroom (Ratnawulan & Rusdiana, 2015). The daily test questions developed will help the teacher review students' capabilities in mastering metabolism material.

The development of HOTS questions with science literacy skills aspects was carried out in several stages, namely analyzing basic competencies, compiling questions outline according to the cognitive domain and indicators of science literacy skills, formulating stimuli, writing question items, and making assessment guidelines or question answer keys. The stages of making questions are adjusted to the steps for making questions from Puspendik (Isbandiyah & Sanusi, 2019). The cognitive levels used in compiling HOTS questions are C4 (analyzing), C5 (evaluating), and C6 (creating) (Fanani, 2018), while the science literacy skills that are evaluated consist of 3 indicators according to PISA 2018, namely explaining scientific phenomena, assessing and designing scientific investigations, and interpreting data and scientific evidence. The evaluation packages are packed up in print (offline) in the form of 2 evaluation packages, namely packages A and B. The evaluation packages are wrapped up according to the needs of teachers who want assessments to be carried out entirely face-to-face (offline) without any problems with internet access while still paying attention to honesty aspects in working on the questions. The parts of the evaluation package consist of; 1) Cover page which contains writing material, subject, education level, developer name, university name, and type of package; 2) Identity page and work instructions containing subjects, class/semester, material, time, number of questions, type of questions, type of packages, and instructions for working on the questions; and 3) Daily test items. Both packages have the same question package parts. The parts of the question package as shown in Table 2.

Table 2. Parts of Evaluation Package

No	Parts of Evaluation Package	Information

| 1 | | The cover page consists of writing material, subject, education level, developer name, university name, and type of package. |
| 2 | | The identity page and work instructions contain subjects, class/semester, materials, time, number of questions, the form of questions, type of package, and work instructions. |
| 3 | The question page contains test items consisting of multiple-choice (a), short answer (b), true-false (c), matching (d), and essay (e). |

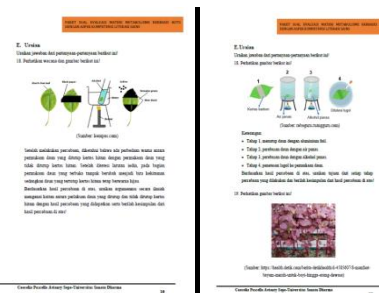(a) Multiple Choice Questions for Packages A and B

(b) Short Answer Questions for Packages A and B
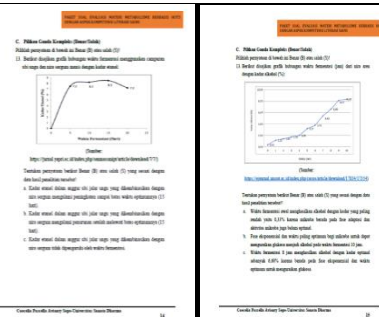
(c) Complex Multiple Choice Questions
(True-False) in Packages A and B
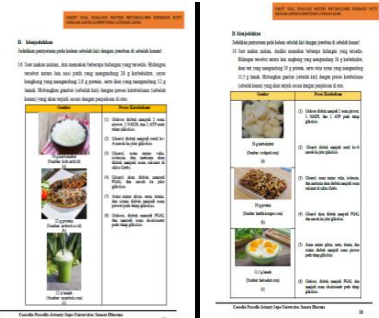
(d) Matching Questions for Packages A
and B

(e) Essay Questions for Packages A and
B

The evaluation packages developed follow the packages developed by Harta (2017), who evolved 20 HOTS-based essay questions (4 question packages consisting of 5 items in each package), and the evaluation packages developed by (Sasongko et al., 2016), who elaborated two packages of mathematical literacy questions with 12 questions for each package. The evaluation packages made in this study were modified by having 40 questions consisting of 20 questions in package A and 20 questions in package B. The two packages are generalized in order that those have parity in concept, cognitive domain, indicators and scientific literacy skills aspects, and the number of questions (Fortuna R et al., 2013).

The questions form used consisted of ten multiple choice questions (1, 2, 3, 4, 5, 6, 7, 8, 9, and 10), two short answer questions (11 and 12), three complex multiple choice (true-false) questions (13, 14, and 15), two matching questions (16 and 17), and three essay questions (18, 19, and 20). The selection of these five forms of questions is based on the reference to the variations in the forms of questions used in the Minimum Competency Assessment, which consists of multiple choice, complex multiple-choice, matching, short answer, and essay (Pusat Asesmen dan Pembelajaran, 2020). The development of various questions aims to provide more detailed and

comprehensive information about students' abilities so that the assessment is carried out objectively (Widana, 2017).

2. Product Validation Results

The development product was validated by four validators consisting of one test material expert lecturer, one learning evaluation (test items) expert lecturer, and two Biology teachers. The evaluation items validation was carried out by three learning evaluation experts, including one Biology education lecturer and two high school Biology teachers. Validation of the test items was accomplished to see the feasibility and quality of the HOTS-based evaluation items with the science literacy skill aspects that had been compiled. The validation assessment of the test items in packages A and B focuses on four aspects that are material, construction, science literacy competence, and language. Each form of question has several statements on the four aspects validated. Multiple choice items have 31 statements, short answers consist of 26 statements, 27 true-false statements, 31 matching statements, and 24 essay statements. The evaluation items validation was conducted by 3 test material experts, including one Biology education lecturer and two high school Biology teachers.

Validation of the test material was carried out to see the accuracy and correctness of the material from HOTS questions with aspects of science literacy competence in the Metabolism material that had been prepared. The validation assessment of the question material in package A and package B focuses on four aspects, namely the suitability of the material, the accuracy of the material, the language, and the presentation containing 17 statements for all types of questions. Product validation in the form of developing HOTS question packages with aspects of science literacy skills was carried out by validators I, II, III, and IV, so as the recapitulation results of final product validation for package A were 91.6% and package B were 92% classified in the criteria of "very feasible". It indicates that the two evaluation packages are categorized as 'very feasible' and can be tested further with revisions or improvements provided by the validator. Details of the final product validation results as shown in Table 3.

Table 3. Final Product Validation Results

Aspects	Final Validation Percentages (%)	
	Package A	Package B
Evaluation (Test items)	93,2	92,6
Test Materials	90	91,3
Total	183,2	183,9
Average	91,6	92
Validation Criteria	Very Feasible	Very Feasible

Besides assessing, the validator gives comments and suggestions for improvements to the products made. The test items correction was done in package A and package B. Questions revised in package A are numbers 1, 2, 3, 6, 7, 8, 12, 13, 14, 15, 16, 17, 18, 19, 20. Questions revised in package B are numbers 1, 2, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20. Revisions were made in both Packages A and B, containing several aspects that aim to make the questions provided better and used in learning.

The first revision is regarding the improvement of the question outlines. The question outline is information in a matrix type that presents a reference for assembling questions. The question outline consists of Basic Competencies, Materials, Competency Achievement Indicators, Question

Indicators, Cognitive Levels, Science Literacy Competency Indicators, Skill Types, Question Types, Question Numbers, and Answer Keys. Revisions in the question outlines are based on the validator's comments and suggestions given. It is made in accordance with one of the requirements for compiling outlines which are having detailed components, clear, and easy to understand (Isbandiyah & Sanusi, 2019).

The second revision is the improvement of the question indicators. This revision found in package A are numbers 13 and 17, and in package B are numbers 1, 2, 7, and 13. The question indicator is one of the components to measure valid and reliable learning achievement. Question indicators can be signs or references for giving stimulus questions, so the question indicators and question stimuli must be continuous (Isbandiyah & Sanusi, 2019). The indicators are described based on the competency achievement indicators. According to Isbandiyah & Sanusi (2019), the question indicators must be clear and able to create according to the form of the questions that have been assigned. Therefore, questions that still have deficiencies in the question indicators need to be completed and cleared, so those are sync with the questions stimulus and the type of questions given. For example, in question number 1 B, which previously had an indicator question "Presented a discourse and picture about the α-amylase enzyme in making bread, students can assess the structural components of the α-amylase enzyme" was changed to "Presented a discourse and picture about the α- amylase enzyme in making alternative gelatin from cassava for making marshmallows, students can assess the structural components of the α-amylase enzyme" by taking a different stimulus but made equivalently and in question 13 A which has the previous indicator "Presented a graph of the relation between fermentation time using a mixture of sweet potato and sweet sorghum juice with ethanol content, students can develop hypotheses stated the causative factors and results of anaerobic respiration (fermentation)" changed to "Presented a graph of the relation between fermentation time using a mixture of purple sweet potato and sweet sorghum juice with ethanol content, students can examine the results of the fermentation obtained" because the presentation of the questions in the form of graphs of research results is more suitable for analyzing or reviewing.

The third revision is improving the cognitive level. Revisions in cognitive level are found in package A number 13, and package B number 13. The question indicators are more in line with C4 (analyzing) with the Operational Verb "analyze" than C6 (creating) with the Operational Verb "compiling". It was caused by the question indicators used are in the form of graphical stimulus presentations from research results that must be analyzed or studied. It is in accordance with what was written by Setiawati et al. (2019), which states that the provision of operational verb domains is strongly influenced by what thought processes are needed to answer questions.

The fourth revision is the improvement in the presentation of the stimulus questions. The revisions are found in package A numbers 1, 12, 16, and 20, and in package B numbers 1, 2, 8, 9, 10, and 18. Stimulus is a reference for understanding contextual and interesting information Isbandiyah & Sanusi (2019). Improvements to the stimulus questions include improving the resolution of images and tables to make them clearer, providing information on images, replacing stimuli according to the question indicators, and adding explanations about plants that are not familiar. Improvement of image resolution to be clearer was carried out on numbers 1 A and 1 B, and image replacement was done on number 12 A. The revision was done to help students see the pictures clearly and neatly in order that students could answer the questions given correctly. Table revisions are made in number 10 B by replacing the numbers in the table sequentially. Giving picture descriptions is conducted at 1 B, 8 B, 18 B, and 1 A to aid students in understanding

the meaning of the pictures given to the stimulus questions to make it easier to analyze and answer questions.

The change of stimulus questions is carried out according to the changes in the question indicators. The changes created were based on the validator's comments concerning the similarity of questions in packages A and B. Therefore, changes were needed to the indicators and stimulus questions as was done in numbers 1 B, 2 B, 7 B, and 12 B. This step was taken to make the variation on stimulus questions but still had the same indicators and science literacy competence type and cognitive level with package A. In addition, the stimulus questions were changed to match the indicators as in question 16 A by changing the type of food, portions, and the food nutrient content in that it became more defined and no overlap in the nutrient content of food. Additional explanations to the stimulus questions were carried out in numbers 20 A and 9 B. It was done because the plants used in the stimulus questions were not yet familiar or known to students, so additional explanations needed to be given in the form of Indonesian names, characteristics, and pictures to make it easier for students to imagine the plant. Improvements to the stimulus questions are made for students to be able to transfer things that have been learned easily in order that they can create a positive attitude, appreciate, and appraise things that have been learned, and are able to use higher-order thinking skills (Tim Pusat Penilaian Pendidikan, 2019).

The fifth revision is about the main questions formulated. Improvements were made to package A numbers 13 and 17 and package B numbers 13 and 17. The main questions formulated were changed according to the question indicators. In questions numbers 13 A and 13 B, the main questions asked about the hypothesis in accordance with the research results were replaced with statements that were in accordance with the research results. In numbers 17 A and 17 B, changes occurred in the direction of the previous question "Match the photosynthesis process that will occur (on the left) to the plant (on the right)!" becomes "Match the photosynthetic process that occurs (on the left) with the correct type of plant and type of photosynthesis (on the right)!" and changing the column statement of plants to type of plants and type of photosynthesis and changing pictures to text with other plant alternatives that are more familiar or known to students.

The sixth revision is about indicators and forms of science literacy skills. Science literacy competence is needed as an indicator to show students' science literacy, consisting of explaining scientific phenomena, evaluating, and designing scientific investigations, and interpreting data and scientific evidence (OECD, 2019). This improvement was carried out on package A questions number 13 and 20 and package B numbers 13 and 20. The improvement on question number 13 in packages A and B was carried out in indicators and forms of science literacy skills which previously had science literacy competency indicators "Explaining scientific phenomena" to "Interpret data and scientific evidence" and the skills form "Give clear hypotheses" to "Analyze and interpret data and draw conclusions appropriately". This was changed because it was inappropriate with the stimulus questions given in the form of research data which should no longer be hypotheses but more to analyze and interpret the data and draw conclusions from the results acquired. This also applies to 20 numbers in packages A and B changing the form of science literacy skills from "Transforming data from one form of representation" to "Analyzing and interpreting data and drawing appropriate conclusions". Changes in indicators and forms of science literacy skills are adapted to the 2018 PISA skills form.

The seventh revision is the improvement of answer choices, answer keys and assessment guidelines. Improvements to the answer choices were carried out on questions in package A number 17 and numbers 1, 2, 7, 13, and 17 in package B. Changes in answer choices were based on alterations in the indicator formulation and stimulus questions. The answer keys and

assessment guidelines revisions were conducted in questions numbers 18, 19, and 20 in Package A and numbers 7, 11, 18, 19, and 20 in Package B. The answer keys for questions 7 B and 11 B were corrected due to a mistake in giving the correct answer key. Improvements to the assessment guidelines for questions 18, 19, and 20 are related to the essay questions so that the scoring rubric is more operationalized and replaced in the score.

The eighth Revision is the refinement of the question form in the type of true-false questions. The revision was stated by one of the validators that the question type should be grouped into complex multiple-choice. Therefore, the researcher includes true-false questions in complex multiple-choice while still including true-false, so it becomes complex multiple-choice (true-false). Complex multiple-choice is a form of question aimed to test students' understanding of continuous issues related to one statement and another which has the form of questions such as true-false and yes-no (Widana, 2017). That is the reason why the researcher still put true-false after complex multiple-choice becomes complex multiple-choice (true-false).

The ninth revision is the improvement in the word writing or sentences that are not correct or typing errors. The replacements were carried out on package A numbers 2, 3, 6, and 8, and package B numbers 6, 8, and 11. This improvement was based on the existence of some incorrect word writing or typos and less effective sentences. Writing errors will interfere students when reading and understanding the questions. Consequently, the use of language which is in accordance with the guidelines and rules of Indonesian is required in order to be communicative and easily understood by students. This is aligned with the statement of Isbandiyah & Sanusi (2019) the language used in the formulation of questions must be in line with the rules of Indonesian and use communicative sentences.

Another mistake is in question number 7 in package A. That is because question card number 7 is wrong and erroneous since the question card should be the question for number 5. This mistake can be corrected by following a pre-existing question outline and questions guidelines. The indicators and science literacy skills type, question indicators, and cognitive level follow the questions outline number 7. The changes made are to match the question cards, the stimulus, and the question items. In addition to the test items, the assessment rubric requires several improvements and has to be completed for all questions. Changes occur in the short-answer score assigning from the total score of 3 to 2, and the essay score is based on changes in the scoring guidelines and the addition of main points. Revisions in scoring short-answer questions and essays are based on the condition of the answers required in those questions (Arifin, 2012). That makes the total score achieved changed to 60 from the previous 50.

Improvement of learning tools is needed based on suggestions given by validators who ask for adjustments between giving questions that use science literacy and the learning steps that are carried out. That happens because there is a mistake in using the learning model that is used to support the questions giving and learning activities held at school. Therefore, changes were made to the syllabus and lesson plans. Replacements made comprise the modification to learning steps and learning models used. The learning steps are replaced with rotational learning or a hybrid between online and offline meetings to adjust to the provision of evaluation packages to students provided in printed form. The learning model is changed by adding learning variations that can support science literacy skills consisting of Discovery Learning, Inquiry Learning, and PBL. That is aligned with the application of constructivism learning theory managed to help develop and increase motivation, critical thinking skills, and willingness to learn independently (Sugrah, 2019). In addition, cutting was carried out into five meetings from the previous six meetings and added science literacy skills aspects (indicators and type of science literacy skills) to the lesson plan.

Adding science literacy skills aspects can help students to grow science literacy skills so they are able to create higher-order thinking skills or skills (HOTS) (Utama & Rahman, 2020).

Learning applications carried out can improve HOTS abilities and science literacy skills which can be seen in the learning steps and student worksheets created. The first and second meetings used the discovery learning model, the third meeting used inquiry learning, the fourth meeting used problem-based learning, and the fifth meeting is daily tests. The learning steps applied provide conditions for students to be active in experiencing real problems that will be solved by making plans for solving problems, determining the tools or materials needed, collecting, analyzing, processing, and discussing data or information, and formulating conclusions to evaluate the problem-solving quality.

Student worksheets used in learning present HOTS and science literacy. It can be viewed in the application of discovery worksheets, study guides, and practicum. The discovery worksheets are in worksheets 5A and 5B requiring students to provide hypotheses, design investigations to interpret data, and draw conclusions. The practicum worksheet is perceived in worksheet 4 which requires students to identify and determine tools/materials and work procedures in scientific investigations, collect and process data and determine the right conclusions. The study guide worksheets are found in worksheets 1, 2, 3, and 6 by presenting stimuli in the form of discourse, pictures, graphics, and videos in order that students can explain scientific phenomena that occurred, which are written in the answer form and then analyzed based on evidence, assumptions, and scientific arguments and draw conclusions. Activities carried out by students will foster HOTS skills in the form of critical, logical, and creative thinking, as well as evaluative and solutive, in order that the problems can be solved (Wasis et al., 2020). In addition to HOTS, the learning activities carried out can develop science literacy skills starting from making hypotheses and designing investigations to conclude. The application of HOTS learning and science literacy is also implemented through practicum activities.

The application of learning with practicum can improve critical thinking skills and describe and perceive as a form of student interpretation of their experiences (Lestari T et al., 2020). At meeting 1, students were given the task of analyzing the practicum video at home and making a report. At meeting 2, the practicum was carried out at home, while the practicum in meeting 3 was conducted at school. The results of the practicum that have been carried out will be processed and made in a report form. Data processing from discussions or experiments and drawing conclusions are included in indicators of skills in interpreting data and scientific evidence in the skills form in analyzing and interpreting data and drawing conclusions correctly (Irwandi, 2020). In addition, indicators and other forms of science literacy skills exist in meeting 3, namely explaining scientific phenomena; identifying questions to be explored further through scientific investigation at the stage of formulating problems and explicating scientific phenomena; providing clear hypotheses at the part of formulating hypotheses. During the practicum, students are facing real problems, so at the same time science literacy skills as described by PISA 2018, will be formed starting from discussions to provide hypotheses or direct predictions, giving scientific questions to be explored in scientific investigations, identifying the tools or materials used, collect and process data/scientific evidence, draw conclusions, communicate (oral and written), and reuse them in different situations (Irwandi, 2020).

D. CONCLUSION AND SUGGESTIONS

Based on the results of the development research that has been carried out, it can be concluded that; (1) Evaluation packages about HOTS-based metabolism material with science literacy skills aspects were developed in the form of package A and package B of 40 questions with 5 types of questions (multiple choice, short answer, complex multiple choice (true-false), matchmaking, and essay); and (2) Evaluation packages about HOTS-based metabolism material with science literacy skills aspects get a final average percentage of the feasibility of 91.6% (Package A) and 92% (Package B) with the criteria of "very feasible" and suitable for testing after revision. Suggestions for the improvement of development research in the future based on the limitations experienced are; (1) Packaging of evaluation packages can be accomplished online using the available applications and websites; (2) Products can be developed as post-test questions; (3) Question preparations need to add other skill types of science literacy indicators to reach all science literacy skills of students; (4) Questions should be adjusted to the cognitive level and indicators of science skills assigned; (5) In preparing the stimulus, it is necessary to read many references to find other contextual or culturally appropriate stimulus questions related to metabolism materials; (6) In making HOTS questions, the researcher should pay more attention in selecting Basic Competencies that use Operational Verbs; (7) Future research can use the revised items to be used in the implementation and evaluation stages; and (8) Future research can use other materials that have Basic Competencies that use HOTS operational verbs (C4-C6).

ACKNOWLEDGEMENT

REFERENCES

Ardelia, N., & Juanengsih, N. (2021). Implementasi Pembelajaran Abad 21 pada Mata Pelajaran Biologi di SMA Negeri Kota Tangerang Selatan. *Jurnal Inovasi Pembelajaran Biologi*, *2*(2), 1–11. https://doi.org/https://doi.org/10.26740/jipb.v2n2.p1-11.

Arifin, Z. (2012). *Evaluasi Pembelajaran*. Bandung: PT. Remaja Rosdakarya.

Avina, Y. P. Al, & Winarsih. (2020). Pengembangan Instrumen Penilaian sebagai Contoh Paket Soal Higher Order Thinking Skills (HOTS) Materi Pencemaran Lingkungan Kelas X SMA. *Bioedu*, *9*(1), 217–222. https://doi.org/https://doi.org/10.26740/bioedu.v9n2.p217-223.

Bagus, B., Kusairi, S., & Sutarman. (2016). Pengembangan Model Penilaian Formatif Formal Berbantuan Komputer untuk Membantu Pembelajaran Dinamika Rotasi. *Jurnal Tarbiyah (Jurnal Ilmiah Kependidikan)*, *5*(2), 7–22. https://jurnal.uin-antasari.ac.id/index.php/jtjik/article/download/981/781.

Baidlowi, M. Hisyam., Sunarmi, & Sulisetijono. (2019). Pengembangan Instrumen Soal Essay Tipe Higher Order Thinking Skills (HOTS) Materi Struktur Jaringan dan Fungsi Organ pada Tumbuhan Kelas XI SMAN 1 Tumpang. *Jurnal Pendidikan Biologi*, *10*(2), 57–65. https://doi.org/http://dx.doi.org/10.17977/um052v10i2p57-65.

Fanani, Moh. Zainal. (2018). Strategi Pengembangan Soal Higher Order Thinking Skill (HOTS) dalam Kurikulum 2013. *Edudeena: Journal of Islamic Religious Education*, *2*(1), 57–76. https://doi.org/https://doi.org/10.30762/ed.v2i1.582.

Fortuna R, D., Chandra, E., & Gloria, R. Y. (2013). Pengembangan Tes Diagnostik untuk Mengukur Miskonsepsi Siswa pada Pokok Bahasan Sistem Regulasi Manusia untuk Siswa SMA Kelas XI

Semester II. *Jurnal Scientiae Educatia*, *2*(2). https://www.syekhnurjati.ac.id/jurnal/index.php/sceducatia/article/view/514/452.

Hanafi. (2017). Konsep Penelitian R&D dalam Bidang Pendidikan. *Saintifika Islamica: Jurnal Kajian Keislaman*, *4*(2), 129–150. https://jurnal.uinbanten.ac.id/index.php/saintifikaislamica/article/view/1204.

Harta, J. (2017). Pengembangan Soal Esai Berbasis HOTS untuk Menyelidiki Keterampilan Pemecahan Masalah Siswa SMA. *Jurnal Penelitian*, *21*(1), 62–69. https://e-journal.usd.ac.id/index.php/JP/article/view/896.

Irwandi. (2020). *Strategi Pembelajaran Biologi*. Bandung: Pustaka Reka Cipta.

Isbandiyah, S., & Sanusi, A. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Biologi*. Jakarta: Direktorat Pembinaan Sekolah Menengah Atas.

Lestari T, M., Ramlawati, & Yunus, S. R. (2020). Hubungan antara Keterampilan Proses Sains dengan Hasil Belajar IPA Peserta Didik Kelas VII di SMPN 40 Makassar. *Jurnal IPA Terpadu*, *3*(2), 46–53. https://doi.org/https://doi.org/10.35580/ipaterpadu.v3i2.12324.

Limbong, T., & Taufik, I. (2017). Aplikasi Pengacak Soal Ujian untuk Type Soal Berbasis Microsoft Word Menggunakan Metode Linear Congruent Method (LCM). *MEANS (Media Informasi Analisa Dan Sistem)*, *2*(1), 58–65. https://media.neliti.com/media/publications/282494-aplikasi-pengacak-soal-ujian-untuk-type-d6c430bb.pdf.

Masgumelar, N. K., & Mustafa, P. S. (2021). Teori Belajar Konstruktivisme dan Implikasinya dalam Pendidikan dan Pembelajaran. *GHAITSA: Islamic Education Journal*, *2*(1), 49–51. https://siducat.org/index.php/ghaitsa/article/view/18.

Muhibbuddin, Ulfa, M., Pada, A. U. T., Rahmatan, H., & Hasanuddin. (2022). The Development of Higher Order Thinking Skills (HOTS) Test Instrument On Metabolism Topic for Senior High School Level. *IJAEDU-International EJournal of Advances in Education*, *8*(23), 118–126. https://doi.org/https://doi.org/10.18768/ijaedu.1140976.

Musayaroh, T., Yuliana, I. F., & Fatayah, F. (2021). Pengembangan Instrumen Tes Literasi Kimia Berbasis Hots Yang Layak Ditinjau dari Validitas Isi oleh Ahli. *UNESA Journal of Chemical Education*, *10*(3), 243–351. https://doi.org/https://doi.org/10.26740/ujced.v10n3.p243-251.

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing.

Paidi. (2012). *Metodologi Penelitian Pendidikan Biologi*. Yogyakarta: UNY Press.

Pratiwi, O. (2019). *Pengembangan Media Pembelajaran Biologi Berbasis Web pada Materi Evolusi Kelas XII di SMA Negeri 1 Meulaboh*. [Unpublished Thesis]. Universitas Islam Negeri Ar-Raniry Banda Aceh.

Pratiwi, S. A., Syamsiah, & Bahri, A. (2022). Pengembangan Instrumen Penilaian Kemampuan Literasi Sains pada Pembelajaran Biologi. *Jurnal Biology Teaching and Learning*, *5*(1), 1–6. https://doi.org/https://doi.org/10.35580/btl.v5i1.33834.

Pusat Asesmen dan Pembelajaran. (2020). *Desain Pengembangan Soal AKM*. PPusat Assesmen dan Pembelajaran Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan dan Kebudayaan.

Ramadhani, R., & Bina, N. S. (2021). *Statistika Penelitian Pendidikan: Analisis Perhitungan Matematika*. Jakarta: Kencana.

Ratnawulan, E., & Rusdiana, H. A. (2015). *Evaluasi Pembelajaran dengan Pendekatan Kurikulum 2013*. Bandung: Pustaka Setia.

Sasongko, T. P. M., Dafik, & Okatvianingtyas, E. (2016). Pengembangan Paket Soal Model Pisa Konten Space and Shape untuk Mengetahui Level Literasi Matematika Siswa SMP. *Jurnal Edukasi*, *3*(1), 27–32. https://doi.org/https://doi.org/10.19184/jukasi.v3i1.4317

Schleicher, A. (2018). Paris: *PISA 2018 Insights and Interpretations*. OECD Publishing.

Septiani, D., Widiyawati, Y., & Nurwahidah, I. (2019). Pengembangan Instrumen Tes Literasi Sains PISA Aspek Menjelaskan Fenomena Ilmiah Kelas VII. *Science Education and Application Journal (SEAJ) Program Studi Pendidikan IPA Universitas Islam Lamongan*, *1*(2), 46–55. https://doi.org/https://doi.org/10.30736/seaj.v1i2.144.

Septikasari, R., & Frasandy, R. N. (2018). Keterampilan 4C Abad 21 dalam Pembelajaran Pendidikan Dasar. *Jurnal Tarbiyah Al-Awlad*, *8*(2), 107–117. https://doi.org/https://doi.org/10.15548/alawlad.v8i2.1597.

Setiawan, A. R., & Mufassaroh, A. Z. (2019). Menyusun Soal Literasi Saintifik untuk Pembelajaran Biologi Topik Plantae dan Animalia. *BIOSFER, J.Bio. & Pend. Bio*, *4*(1), 33–40. https://doi.org/https://doi.org/10.23969/biosfer.v4i1.1484.

Setiawati, W., Asmira, O., Ariyana, Y., Bestary, R., & Pudjiastuti, A. (2019). *Buku Penilaian Berorientasi Higher Order Thinking Skills*. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.

Sugrah, N. (2019). Implementasi Teori Belajar Konstruktivisme dalam Pembelajaran Sains. *Humanika, Kajian Ilmiah Mata Kuliah Umum*, *19*(2), 121–138. https://doi.org/10.21831/hum.v19i2.29274.

Thahir, R., Magfirah, N., & Anisa. (2021). Hubungan antara High Order Thinking Skills dan Kemampuan Literasi Sains Mahasiswa Pendidikan Biologi. *BIODIK: Jurnal Ilmiah Pendidikan Biologi*, *7*(3), 105–113. https://doi.org/https://doi.org/10.22437/bio.v7i3.14386.

Tim Pusat Penilaian Pendidikan. (2019). *Panduan Penulisan Soal HOTS-Higher Order Thinking Skills*. Jakarta: Pusat Penilaian Pendidikan.

Utama, C., & Rahman, A. F. (2020). Kaitan Antara Literasi Sains dan HOTS untuk Pembelajaran IPA di Sekolah Dasar. *Prosiding Seminar Nasional Pendidikan Dasar*, 250–256. http://eproceedings.umpwr.ac.id/index.php/semnaspgsd/article/view/1389.

Wasis, Rahayu, Y. S., Sunarti, T., & Indana, S. (2020). *HOTS & Literasi Sains: Konsep, Pembelajaran, dan Penilaiannya*. Jombang: Kun Fayakun.

Widana, W. I. (2017). *Modul Penyusunan Higher Order Thinking Skills (HOTS)*. Jakarta: Direktorat Pembinaan SMA.