Available online at: http://journal.uny.ac.id/index.php/jpe

Jurnal Prima Edukasia, 9 (2), 2021, 179-192

Developing HOTS questions for the materials of human and animals respiratory organs for grade V of elementary school

Aulia Latifah, Ika Maryani *

Elementary School Teacher Education, Universitas Ahmad Dahlan
Jalan Ki Ageng Pemanahan No.19, Sorosutan, Yogyakarta 55281, Indonesia
* Corresponding Author. E-mail: ika.maryani@pgsd.uad.ac.id

Received: 2 September 2020; Revised: 17 March 2021; Accepted: 25 March 2021

Abstract: The study aims (1) to know the resulting development HOTS questions for the materials of human and animal respiratory organs for grade V of elementary school, and (2) to test the quality of questions for the materials of human and animal respiratory organs for grade V of elementary school. This study is a Research and Development (R&D) with ADDIE methods (analysis, design, development, implementation, and evaluation). The research subjects were 90 grade V elementary school students. Observation data through interviews, documentation, questionnaires, and tests. The data analysis technique uses qualitative data from Miles and Huberman's models, while quantitative data uses descriptive and inferential statistical data to calculate the result of the expert test (material and evaluation experts) and empirical test (validity, reliability, Difficulty index, discriminant index, and deception). The study results stated that (1) the validity of the first test that was classified as valid was 13 questions and 32 questions were invalid, while the second test, which was classified as valid, was 44 questions, and 1 question was invalid. The first reliability has sufficient and low criteria, while the second test has high and sufficient criteria. The Difficulty index of the first and second test items has difficult, medium, and easy criteria. The discriminant index of the first test has the criteria of being insufficient, sufficient, good, and some questions do not have a discriminant index, while the second test has the criteria for being insufficient, sufficient, and good. The effectiveness of the first test distractor had 48,89% functional distractor and 51,11% non-functioning, while the second test had 75,55% functional distractor and 24,44% non-functioning, and (2) quality of question development HOTS from the results of material expert validation and evaluation obtained a value of 84,52% with very feasible criteria. Based on these assessment results, the question development is HOTS is feasible to be applied in school.
Kata Kunci: HOTS instruments, Human and Animals Respiratory Organs, Elementary School.

How to Cite: Latifah, A., & Maryani, I. (2021). Developing HOTS questions for the materials of human and animals respiratory organs for Grade V of elementary school. *Jurnal Prima Edukasia, 9*(2), 179-192. doi:https://doi.org/10.21831/jpe.v9i2.34172

Introduction

21st-century life makes the educational position becoming more crucial to equip people to be able to compete globally. The governments have applied various attempts to create innovation and renewal. One of them is by improving the curriculum structure. Indonesia had changed the curriculum frequently. Curriculum changes have been started from Curriculum 1947, Curriculum 1964, Curriculum 1968, Curriculum 1973, Curriculum 1975, Curriculum 1984, Curriculum 1994, Curriculum 1997, Curriculum 2004, Curriculum KTSP, and the current curriculum is Curriculum 2013 (Muhammedi, 2016). The curriculum improvement is to improve in the educational field. Curriculum 2013 is expected to form the students' character and competence of knowledge, skills, and behavior under a contextually learned concept (Kusumawati & Rulviana, 2017). Therefore, implemented curriculum 2013 is expected to complete previous curriculum deficiency and fulfill the educational success for preparing the next qualified generations. The opportunity to implement Curriculum 2013 is consequently to train and improve students' HOTS (*Higher Order Thinking Skill*). HOTS is a thinking skill that is more than remembering and memorizing and the ability to analyze, combine, and estimate something (Yuniar et al., 2015). HOTS proficiency is thinking and assuming skills in overcoming moral problems to acquire knowledge (Putra & Agustiana, 2021). Previously mentioned descriptions show that HOTS is a critical and comprehensive thinking skill for using the knowledge to solve problems.

https://doi.org/10.21831/jpe.v9i2.34172

According to Anderson et al. (2001), HOTS reviewed from Boom includes several competencies; analyzing (C4), evaluating (C5), and creating (C6) (Aini & Sulistyani, 2019). The categories of higher thinking skills consist of (1) *analysis, evaluation, and creation*; (2) *logic and reasoning*; (3) *judgment and critical thinking skill*; (4) *problem solving*; and (5) *creativity and creative thinking* (Brookhart, 2010). Meanwhile, other arguments stated that the categories of higher thinking skills consist of (1) *retrieval*; (2) *comprehension*; (3) *analysis*; (4) *knowledge utilization*; (5) *metacognition*; and (6) *self-system thinking* (Marzano & Kendall, 2010). HOTS activity will become the students' provision to face the challenges in the future. Higher thinking skills help students solve problems, make decisions, and act in daily life (Triyuni et al., 2019). The students develop thinking skills to understand and learn more to obtain the right solutions. The way to train students' skills is through learning the problems in the natural science materials associated with daily life.

The implementation of HOTS in natural science learning has an important role in optimizing elementary school student's ability and potential. Natural science learning can allow elementary school students to develop their ability to question and search for the answer to be evidence of scientific thinking (Desstya, 2016). Students will be more actively involved in the learning activities as a manifestation of their effort to build an understanding of taught materials. Elementary school students under the concrete operational stage have a huge curiosity for learning something new. Natural science learning not only teaches students to have an understanding of taught materials, but it also concerns the ability of thinking (Utami et al., 2017). However, the fact that natural science learning only emphasizes memorizing aspects. Consequently, students assume that the materials taught are just for remembering without along with the activity of solving the problems. Besides, the lack of teacher ability to develop HOTS's questions addressed students at the end of learning results in the questions concerning LOTS and MOTS categories.

The impact of the teacher's ability deficiency in developing HOTS questions makes students inactively search for the information. The learning will be meaningful if the students are trained to build higher-order thinking (Julia et al., 2018). Connected to the problems, a teacher has a vital role in teaching and train the students to have higher-order thinking. As a result, it is seen from the average evaluation values of human and animal respiratory organs is above the minimum completeness criteria. It shows that students have the ability of HOTS if HOTS's question category supports them. Moreover, it is needed to develop HOTS questions of human and animal respiratory organs for students' grade V.

Methods

The type of data in this study is qualitative and quantitative data. Qualitative data were obtained from the comment and advice of the validator. In comparison, quantitative data was obtained from the questionnaire assessment result using a Likert scale of intervals 1-5. The research was conducted on May 22-28, 2020; June 3, 2020; and July 2, 2020. Research place was in Muhammadiyah Danunegaran elementary school, Muhammadiyah Demangan elementary school of Demangan, Sinungrejo 2 public elementary school, and Ambalresmi 2 Public elementary school. There are all at Yogyakarta Regency, Indonesia. The research target was to (1) know the development result of HTS questions of human and animal respiratory organs for students' grade V, and (2) examine the quality of HOTS questions of human and animal respiratory organs for students of grade V. The research subjects were 90 students in grade V.

The research procedure was conducted through the steps of ADDIE. The first step is *the analysis phase*, whereby the writer analyzes the necessary to collect the information about what needed in the learning process focusing on human and animal respiratory organs in Muhammadiyah Demangan elementary school. Furthermore, this research's analysis result was found through observation, interview, documentation, and further evaluation. The second step is the *design* phase that started by deciding the learning outcome and continued by developing question grids. There are 10 multiple choice questions and 5 matching questions of HOTS question type developed in this research. Furthermore, making foreword, table of contents, learning outcome, and indicator, manual, answer sheet, answer key, score manual, and cover. Moreover, in the design step, an initial product will be created from HOTS questions' development. The third step is a *development*, whereby the initial effect created from the development of HOTS questions will be examined by the experts and through an empirical test. All of the feedback, notes, and comments received from the material experts and the evaluation obtained through received

questionnaire distribution is received and revised. After that, the development of HOTS questions will be given to the students. After the expert's validation, the next process is an empirical test to measure the validity, reliability, Difficulty index, discriminant index, and effectiveness of the distractor. This process is aimed to acquire the development of HOTS questions with good context and construct. The revision in this step creates the final product that is qualified HOTS questions development. The fourth step is an *implementation* conducted in the next research, whereby the researcher can continue this step into implementing and evaluating. The last step is an *evaluation*, which runs a formative evaluation of the development of HOTS questions of human and animal respiratory organs. A formative evaluation is carried out in every development step of ADDIE.

The data collecting technique used in this research is observation, interview, documentation, questionnaire, and test. Observation sheet and interview guidance used to analyze the research's needs to obtain the data that includes field situation analysis, the material of human and animal respiratory organ, and student characteristic in the learning. Documentation data is a lesson plan and the material of human and animal respiratory organs, and the evaluation result is a general description supporter of the problems in the Muhammadiyah Demangan elementary school. Questionnaire guidance is used to obtain the experts' data that includes the experts of material and evaluation. The test is carried out to examine the quality of HOTS questions of human and animal respiratory organs.

Analysis technique uses qualitative data of Miles and Huberman model, while quantitative data is used summing up the assessment of expert and empirical test by calculating the inferential and descriptive statistic. Qualitative data analysis is carried out to process the data obtained from observation, interview, and documentation. Furthermore, data reduction is carried out by choosing the important dan and reduct unnecessary data, while data presentation is to collect the data according to data types, and concluding is to answer the need analysis. Whereas providing quantitative data is by the inferential and descriptive statistic. The quality of the HOTS question development of the expert test can be seen in Table 1 (Sugiyono, 2016).

Table 1. Criteria of *HOTS* Questions' Quality

Scale	Criteria
0-25	Insufficient
26-50	Sufficient
51-75	Good
76-100	Very feasible

After conducting the expert test, the following executed process is an empirical test. The empirical test analysis is undertaken to measure validity, reliability, difficulty index, discriminant index, and the distractor's effectiveness. The explanation of this process will be described below.

Validity

The formula used to measure the validity of the question items by using the biserial correlation. This formula is used to measure the validity of the question items of multiple choice and matching questions. The formula of the biserial question is as follows (Arikunto, 2010).

$$r_{pbis} = \frac{M_{p-} M_t}{SD_t} \sqrt{p/q} \quad \text{..... 1)}$$

Description:

r_{pbis} = Biserial Correlation Coefficient
M_t = Mean total
M_p = The mean total score of subjects who answered correctly
SD_t = The standard deviation of the total score
P = The proportion of students who answered correctly
Q = The proportion of students who answered incorrectly (q= 1- p)

Reliability

The formula used to determine the reliability of HOTS questions is by using *Kuder-Richardson* or KR_{20} technique. This formula used to process the data of multiple choice and matching question types. *Kuder-Richardson* formula is described as follows (Arikunto, 2005).

$r_{11} = (\frac{k}{k-1}) (\frac{V_t - \sum pq}{V_t})$ ……... 2)

Description =
r_{11} = Instrument Reliability
Q = Question Items
V_t = Variance of the questions
P = The proportion of subjects who answered the item correctly (The proportion of subjects who has score 1)
Q = The proportion of subjects who get the score (q= 1- p)

Table 2. Criteria of Reliability Test

No.	Correlation Coefficient	Qualification
1.	0,91-1,00	Very High
2.	0,71-0,90	High
3.	0,41-0,70	Sufficient
4.	0,21-0,40	Low
5.	Negatif-0,20	Very Low

Difficulty index

Difficulty index tests are used to examine the questions with easy, medium, and difficult criteria. The formula is described as follows (Arikunto, 2016).

$DI = \frac{JB}{n}$……... 3)

Description:
DI = Difficulty index
JB = Total of students answered correctly
n = Total of students

The index of Difficulty index used is described in Table 3 (Arikunto, 2016).

Table 3. Difficulty index Category

No.	Interval	Category
1.	0,00 – 0,30	Difficult
2.	0,31 – 0,70	Medium
3.	0,71 – 1,00	Easy

Discriminant index

The level of the discriminant index is used to examine every student's ability. The formula used to determine the discriminant index is described as follows.

$DP = \frac{2\ (KA-KB)}{n}$ ……... 4)

Description:
DP = Discriminant index
KA = Total of upper group students answered correctly
KB = Total of lower group students who answered correctly
n = Total of students

Classification of the discriminant index used Table 4.

Table 4. Discriminant index Criteria

Discriminant index Criteria	Category
0,00-0,20	Insufficient
0,21-0,40	Sufficient
0,41-0,70	Good
0,71-1,00	Very Good

Distractor efficiency

The formula used to calculate the index of the distractor's effectiveness is described in formula 5.

$PPJ = \frac{JPJ}{n}$ …… (5)

Description:
PPJ = Answer distribution for the certain answer choice
JPJ = Total of students choosing the answer
n = Total of students

Result and Discussion

This study's development resulted in three groups of HOTS questions of human and animal respiratory organs for students grade V of elementary school. In every group, there are 10 multiple choice questions and 5 matching questions. This study used the development model of ADDIE consisting of analysis, design, development, implementation, and evaluation.

In the analysis step, the methods used are observation, interviews, and documentation. This step aims to analyze the needs of learning material of human and animal respiratory organs in Muhammadiyah elementary school of Demangan. The design step started with deciding the learning outcome and indicator according to the analysis of the needs. The basic competencies and indicators are listed in Table 5.

Table 5. Learning Outcome

Study material	Learning outcome		Indicator	
Human and animal respiratory system	3.7	Knowing the human and animal respiratory system and respiratory disease	3.7.1	Analyzing the human and animal respiratory system
			3.7.2	Identifying respiratory disease

After that, the step is to determine the HOTS indicator that consists of *problem-solving*, *logic and reasoning*, judgment, and critical thinking skills. Afterward, the indicator is developed into question grids. The question grids are listed in Table 6.

Table 6. Blueprint of *HOTS Questions*

HOTS Indicators		Question Indicators	Question Number		Stimulus	Cognitive Level
			Multiple Choice Question	Matching Question		
Problem Solving	Identifying problems	• Presented with text, students can identify the factors that cause internal respiratory diseases in humans	3	-	Text	C4
		• Presented with text, students can identify the disease according to the symptoms	4	-	Text	C4
Judgment and critical thinking skill	Analyzing Information	• Presented with a picture of the human respiratory organ, students can analyze the laryngeal function	-	3	Picture	C4

These blueprints are further applied into question form; moreover, they obtained the initial product from the HOTS question's development question listed in this design step.

Multiple Choice Question

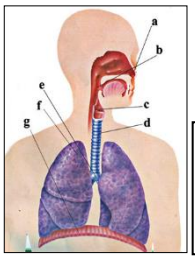
The following is the text for working on problems 3 & 4.

> One of the respiratory organs of Grandpa Hasan has an infection. When he breathes, he will be heard the sound "Ngik". When coughing, Grandpa Hasan's chest will feel tight and the phlegm is difficult to get out. The cold weather in the morning and night causes breathing problems in Grandpa Hasan. As a result, Grandpa Hasan will spray the nasal spray to overcome his out of breath. After that, Grandpa Hasan feels more comfortable breathing easily.

1. Based on the abovementioned case, the respiratory disorder that infected Grandpa Hasan is caused by the factor of...
   a. Physical
   b. Age
   c. Environment
   d. Disease
2. The respiratory disorder that infected Grandpa Hasan is...
   a. Sinusitis (Sinus Infection)
   b. Asthma
   c. TBC
   d. Emphysema

Matching Question

The function of Human Respiratory Organ



3. The respiratory organs that are located in front of the esophagus and act as a place for air to pass are...

Every question has a different stimulus such as text, picture, experiment, and data table. There are four choices in the multiple-choice question: a, b, c, and d, while in the matching question, the question and the answer must be paired. When answering the correct answer, the score is 1, and the wrong question score is 0. This score is for multiple-choice questions and matching questions.

In the developing step, the initial product is examined by the expert and through the empirical test. The experts examine the material and evaluation test. The material expert examines the quality of HOTS questions, while the evaluation expert evaluates the failures of HOTS question development. The expert's assessment consists of material, learning outcome, the construction of the question, question grids, stimulus, question item, and score manual. The assessment result from the expert is listed in Table 7.

Table 7. The Result of Expert Test

No.	Expert Type	Score	Category
1.	Material Expert	87,37	Very Feasible
2.	Evaluation Expert	81,67	Very Feasible
	Total		169,04
	Average	84,52	Very Feasible

Based on Table 7, the assessment of material experts obtained a value of 87,37 with very feasible criteria, while the result of the empirical test obtained a value of 81,67 with very feasible criteria. The average values obtained a value of 84,52 with very feasible criteria from the HOTS questions' value calculation. Furthermore, the development of the HOTS question is very feasible to be applied in the learning activities. Afterward, HOTS questions are tested online and continued by empirical test. The empirical test assessment encompasses validity, reliability, Difficulty index, discriminant index, and Distractor efficiency. In the *implement* step, the development of the HOTS question will be continued by the next researcher. In the last step, *evaluate* step, it will be conducted a formative evaluation of HOTS question development in the steps of *analysis*, *design*, *development*, and *implementation.* The following is the assessment result of the empirical test.

Validity

Validity is a condition in which the test instrument can accurately measure what the target is (Taherdoost, 2016). Validity is to determine the suitability of measuring instruments with the objectives to be measured (Siregar, 2013). This issue can be concluded that the question items can be said valid if they can reach the measured goals. The analysis result of the first validity test is listed in Table 8.

Table 8. Result of the 1st test validity

Item	Group A		Group B		Group C	
	N:20, R_{table}: 0,444		N:20, R_{table}: 0,444		N:16, R_{table}: 0,497	
	R	Criteria	*R*	Criteria	*R*	Criteria
Multiple choices questions						
1.	0,250	invalid	0,640	Valid	0,406	invalid
2.	0,375	invalid	0,692	Valid	0,288	invalid
3.	0,205	invalid	0,449	Valid	0,379	invalid
4.	0,408	invalid	0,646	Valid	-0,019	invalid
5.	0,288	invalid	0,296	invalid	-0,046	invalid
6.	0,144	invalid	0,457	Valid	0,228	invalid
7.	0,687	Valid	0,290	invalid	0,531	Valid
8.	0,613	Valid	0,051	invalid	0,173	invalid
9.	0,190	invalid	0,368	invalid	0,046	invalid
10.	0,319	invalid	0,567	Valid	0,421	invalid
Matching questions						
1.	0,737	invalid	0,401	invalid	0,568	Valid
2.	0,250	invalid	-0,166	invalid	0,653	Valid
3.	0,489	Valid	0,117	invalid	0,228	invalid
4.	0,243	invalid	0,449	Valid	0,438	invalid
5.	-0,067	invalid	0,279	invalid	0,482	invalid

Based on Table 8, the result shows that the first validity test results were not appropriate with the standard criteria. The validity test results of group A classified as valid were 3 questions and invalid were 12 questions, in group B, the questions classified as valid were 7 questions and 8 invalid questions, and the questions in group C categorized as valid were 3 and 12 invalid questions. After that, it was examined the second time to add the respondent total. The analysis result of the second validity test is listed in Table 9.

Table 9. Result of the 2nd test validity

Soal	Group A		Group B		Group C	
	N:30, R_{table}: 0,361		N:30,R_{table}: 0,361		N:30, R_{table}: 0.361	
	R	Criteria	*R*	Criteria	*R*	Criteria
Multiple choices questions						
1	0,362	Valid	0,404	Valid	0,371	Valid
2	0,674	Valid	0,416	Valid	0,385	Valid
3	0,009	Invalid	0,425	Valid	0,462	Valid
4	0,380	Valid	0,414	Valid	0,501	Valid
5	0,397	Valid	0,562	Valid	0,509	Valid
6	0,397	Valid	0,557	Valid	0,454	Valid
7	0,439	Valid	0,380	Valid	0,398	Valid
8	0,684	Valid	0,525	Valid	0,421	Valid
9	0,397	Valid	0,425	Valid	0,407	Valid
10	0,433	Valid	0,645	Valid	0,447	Valid
Matching questions						
1	0,396	Valid	0,6533	Valid	0,410	Valid
2	0,574	Valid	0,386	Valid	0,385	Valid
3	0,664	Valid	0,503	Valid	0,384	Valid
4	0,601	Valid	0,439	Valid	0,383	Valid
5	0,433	Valid	0,530	Valid	0,398	Valid

Based on Table 9, there is an invalid question shown in group A. It is caused by the values if Rcount < Rtable, the questions could not measure achieved objectives. The questions are categorized as valid if $R_{count} > R_{table}$. The questions could measure the goals from what has been completed. The cause of invalid questions that was seen from data analysis result was from question Difficulty index, discriminant index, and language that was not understandable by the students, bad timing of the test, and other factors. Therefore, HOTS questions need to be revised. The revision of HOTS questions is listed in Table 10.

Tabel 10. Revision of the 2nd Validity Test

No.	HOTS Indicator		Before Revision		After Revision	
			Question Indicator	Question	Question Indicator	Question
3.	*Problem-solving*	Identifying Problems	Presented with text, students can identify the cause of the internal respiratory disease in humans (C4)	Sarah came home from school on foot along the edge of the highway. Sarah coughed as she inhaled the air laced with motor vehicle fumes. Conversely, when inhaling the air, Sarah coughed severely that her voice was hoarse. At night, Sarah could not sleep, her body was hot, and she sneezed. Eventually, Sarah's parents brought her to the hospital for treatment. 1. The caused factor that interferes with Sarah is...	Presented with text, students can detect human respiratory triggers (C4)	One afternoon Sarah and mom and dad were sitting on the porch. Their terrace overlooks the garden in front of the house. The garden grows a variety of flower gardens. In the corner of the park also grows mango trees. The atmosphere on the terrace of the house was cold. Sarah suddenly coughed. Sarah smelled the smoke. It looked as if someone was burning a litter of leaves that had not dried up yet. The burnt garbage smoke was so dense that it made it difficult for Sarah to breathe. 3. The caused factor that interferes with Sarah's respiratory is...

Reliability

Reliability is a situation in which the test instrument has a stable index when conducting the measurement of more than one (Herwin & Nurhayati, 2021; Siregar, 2013) Reliability shows the measured test consistency always has unbroken stability (Bajpai & Bajpai, 2014). The study used *Kuder-Richardson* or KR_{20} is to measure reliability. This formula is used to process the data under the question of multiple choice and matching questions. The analysis results of the 1st and 2nd reliability tests are listed in Table 11.

Table 11. The Results of the 1st and 2nd Reliability Test

Reliability Test	Group	Reliability	Criteria
1st	**A**	0,420	Sufficient
	B	0,568	Sufficient
	C	0,340	Low
2nd	**A**	0,725	High
	B	0,750	High
	C	0,660	Sufficient

Based on the analysis result above, the results show that the first test was not a good question, while the second results showed improvement. The analysis result of the first reliability test of group A was 0,420 with sufficient criteria, 0,568 with sufficient criteria in group B, 0,340 with low criteria in group C. Meanwhile, the analysis result of the second reliability test of group A was 0,725 with high criteria, group B was 0,750 with high criteria, and group C with sufficient criteria. The instrument can be said reliable if a reliable index is more than 0,70 or categorized high reliable, and if less than 0,70 is not considered reliable or categorized as low reliable (Rahayu & Djazari, 2016). Factors that affect validity and reliability are instruments, researchers who take measurements, and respondents (Sugiyono, 2014). Furthermore, the high-reliability factor is caused by the number of samples, knowledge of each

individual, and the environment's atmosphere (Alwi, 2015). Therefore, the addition of students as research subjects can affect the increase in reliability. However, validity becomes an important factor compared to reliability when test questions lead to higher-order thinking skills. It shows that the reliability results that have a low index will not affect validity.

Difficulty index

The question Difficulty index is defined as the easy or difficult level questions examined by respondents (Azis, 2016). The Difficulty index is aimed to investigate simple, medium to complex questions (Chauhan et al., 2015). It shows that the question Difficulty index in the test instrument must be propositional. The analysis result of the first Difficulty index can be seen in Table 12 and the second test in Table 13.

Table 12. Result of the 1st Difficulties Index Test

Group	Criteria					
	Difficult	%	Medium	%	Easy	%
A	2, 10 (multiple choices); 3, 4, 5 (matching)	33%	1, 4, 5, 6, 7, 8, 9 (multiple choices) 1, 2 (matching)	60%	3 (multiple choices)	6,67%
B	1, 7, 8 (multiple choices)	20%	2, 3, 4, 6, 9, 10 (multiple choices) 1, 2, 3, 5 (matching)	66,67%	5 (multiple choices) 4 (matching)	13,33%
C	8, 9 (multiple choices) 4 (matching)	20%	1, 2, 3, 6, 7, 10 (multiple choices) 2, 3, 5 (matching)	60%	4, 5 (multiple choices) 1 (matching)	20%

Tabel 13. Result of the 2nd Difficulties Index Test

Group	Criteria					
	Difficult	%	Medium	%	Easy	%
A	10 (multiple choices) 5 (matching)	13,33%	2, 4, 5, 6, 7, 8, 9 (multiple choices) 1, 2, 3, 4 (matching)	73,33%	1, 3 (multiple choices)	13,33%
B	7 (multiple choices)	6,67%	1, 2, 4, 6, 8, 9, 10 (multiple choices) 1, 2, 3, 4, 5 (matching)	80%	3, 5 (multiple choices)	13,33%
C	9, 10 (multiple choices) 5 (matching)	20%	2, 3, 4, 5, 6, 7, 8 (multiple choices) 1, 2, 3, 4 (matching)	73,33%	1 (multiple choices)	6,67

Based on the analysis result, it shows that the developed HOTS questions have simple to difficult criteria. If it is correlated to the percentage of first difficulty test level, group A had 33% with difficult criteria, 60% with medium criteria, and 6,67% with easy criteria. Group B had 20% with difficult criteria, 66,67% with medium criteria, and 13,33% with easy criteria. Also, group C had 20% with difficult criteria, 60% with medium criteria, and 20% with easy criteria. Meanwhile, the second difficulty test level in group A with difficult criteria was 13,33%, 73,33% of medium criteria, and 13,33% of easy criteria. In group B, the percentage of difficult criteria was 6,67%, 80% of medium criteria, and 13,33% of easy criteria. Group C had a Difficulty index of 20% with difficult criteria, 73,33% with medium criteria, and 6,67% with easy criteria.

A good instrument is neither too easy nor difficult (Arikunto, 2016). It shows that too easy instrument cannot develop student's thinking ability and the too difficult instrument also causes students too lazy to do it. The aim of developing an instrument affects the Difficulty index (Azwar, 2015). As a result, the type of questions must be adjusted to the achieved purposes. Therefore, the questions developed must be aimed to train the students' higher-order thinking skills that must be distinguished to the Middle examination and Final examination questions. Higher-order thinking skills differ from the question Difficulty index (Putri M. et al., 2019). The questions that are developed in a high level of difficulty are not necessarily HOTS questions. Because of the way of solving the problem, HOTS aimed to train higher-order thinking.

Discriminant index

The discriminant index is to differ students' intelligence (smart and less smart) (Uno & Koni, 2012). A discriminant index is defined as a measurement tool that can describe the test result with the acquisition of high and low scores (Kocdar et al., 2016). The analysis of the discriminant index of multiple choice and matching questions uses the same formula. The analysis result of the first test's discriminant index can be seen in Table 14 and the second test in Table 15.

Table 14. The Result of The 1st Discriminant Index Test

Group	Criteria				
	Very Poor	Poor	Sufficient	Good	Very Good
A	3, 6 (multiple choices)	1, 2, 4, 5, 7, 8, 9, 10 (multiple choices) 1, 2, 3, 4, 5 (matching)	-	-	-
B	8 (multiple choices) 2 (matching)	3, 5, 7 (multiple choices) 3, 4 (matching)	1, 4, 6, 9, 10 (multiple choices) 1, 5 (matching)	2 (multiple choices)	-
C	6 (multiple choices)	3, 4, 5, 8, 9, 10 (multiple choices) 3, 4 (matching)	1, 2, 7 (multiple choices) 1, 2, 5 (matching)	-	-

Tabel 15. The Result of The 2nd Discriminant Index Test

Group	Criteria			
	Poor	Sufficient	Good	Very Good
A	1, 3, 4, 5, 6, 7, 9, 10 (multiple choices) 5 (matching)	8 (multiple choices) 1 (matching)	2 (multiple choices) 2, 3, 4 (matching)	-
B	2 (matching)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (multiple choices) 1, 3, 4, 5 (matching)	-	-
C	7 (multiple choices) 2, 5 (matching)	1, 2, 3, 4, 5, 8, 9, 10 (multiple choices) 1, 3, 4 (matching)	6 (multiple choices)	-

The analysis as mentioned above result shows that the second test increases. In the first test discriminant index in group A, the discriminant index consists of 13 questions which 2 questions do not have a discriminant index. Group B has 1 question with good criteria, 7 questions with sufficient criteria, 5 questions with insufficient criteria, and 2 questions that do not have a discriminant index. Group C has 6 questions with sufficient criteria, 8 questions with insufficient criteria, and 1 question that does not have a discriminant index. Meanwhile, in the second test discriminant index, group A consists of 4 questions with good criteria, 2 questions with sufficient criteria, and 9 questions with insufficient criteria. Group B has a discriminant index with sufficient criteria of 12 questions and insufficient criteria of 1 question. Group C has the discriminant index of 1 question with good criteria, 11 questions with sufficient criteria, and 3 questions with insufficient criteria.

In making the answer key cause poor discriminant index, errors existed more than one correct answer, unworking effectiveness of the distractor, too difficult material, and unmatching learning objectives (Syarif & Syamsurizal, 2019). Therefore, it is necessary to re-analyze by evaluation experts when they have not met the standard criteria. The discriminant index will function when determining the level of student ability (Hanifah & Hanifah, 2017). The questions conveying to the students can cause the discrimination of upper and lower intelligence level groups of the students. The discriminant index must distinguish between smart and less smart students (Uno & Koni, 2012). It shows that the questions with discriminant index can be answered by smart students but cannot be answered by less smart students.

Distractor efficiency

The distractor's effectiveness will function when the respondents do not understand the theory yet are difficult and tricked into choosing the incorrect answer (Sudijono, 2015). The effectiveness of a

distractor has a target to determine how successful a distractor is so that the students tend to choose the distractor as the answer (Kurniawan, 2015). It shows that the distractor will be useful when the respondent is interested in choosing the distractor. The analysis of the distractor's effectiveness in this study is carried out by measuring the distractor index. The analysis result of the first test distractor effectiveness in group A shows that 9 distractors are not useful and 6 questions also are not useful, while group B 5 distractors are not useful, and 10 questions are also not useful. Then 9 distractors are not useful, and 6 questions also are not useful in group C. Distractor efficiency of HOTS questions of the second test in group A has 4 useful questions and 11 unuseful questions, whereas 3 unuseful distractors and 12 useful questions in group B, and 4 unuseful distractors and 11 useful questions in group C.

The distractor's effectiveness is a measurement tool that determines the quality of question items (Burud et al., 2019). A good instrument is influenced by discriminant index and question Difficulty index and distractor efficiency (Azwar, 2015). It shows that the distractor's effectiveness that is unuseful will affect the discriminant index and question Difficulty index. The effectiveness will be useful if the lower group chooses the distractor than the upper group (Arifin, 2016). The effectiveness of distractors is used to track the students who do not study. If the upper group chose the distractors, it would be unuseful well. Therefore, the distractors need to be revised adjusted to the analysis of the question items. The example of the revision of distractor effectiveness in this study is shown in Table 16.

Table 16. Distractor Effectiveness Revision

No. Item	Option	Distractor efficiency		The Function of the Revision
		Before Revision	After Revision	
Group A				
Multiple Choice Questions				
3.	D	activity	weather	attached with the answer key
4.	A	influenza	tuberculosis	attached with the answer key
5.	D	5 dan 6	1 dan 2	attached with the answer key
7.	D	expand and remain	expand and thinning	attached with the answer key
Group B				
Multiple Choice Questions				
3.	C	environment	physical disorder	attached with the answer key
5.	D	making physical contact without any distance	making physical contact with the distance of a meter	attached with the answer key
6.	D	maintaining the health of the surrounding environment	maintaining body health	attached with the answer key
Group C				
Multiple Choice Questions				
2.	A	desert	valley	attached with the answer key
4.	A	sinusitis (Sinus Infection)	pneumonia	attached with the answer key
5.	D	asking the smoker to sit at the back	making a distance with the smoker	attached with the answer key
10.	B	spiracles-tracheoles-trachea-air sacs-body cells	spiracles-tracheoles-trachea-air sacs-body cells	attached with the answer key

Based on the experts and empirical test results, it is shown that HOTS questions of human and animal respiratory organs have been fulfilled as good questions. The questions' quality is seen from the analysis of validity and reliability, while every item's quality can be seen from the analysis of the Difficulty index, discriminant index, and distractor's effectiveness (Rahayu & Djazari, 2016). It shows that analysis of validity and reliability is good, the quality of the questions is also considered well as a whole. Otherwise, if the Difficulty index analysis, discriminant index, and distractor's effectiveness are

in good criteria, the validity and reliability are not in good criteria, the developed questions do not have good quality. Good quality instruments include validity and reliability (Sudjana, 2010). Having basic usability, referring to procedures, being actual, valid, and reliable is a good criterion of the questions (Kadir, 2015). These various opinions indicate that the HOTS problem of human and animal respiratory organ material is of good quality so that it is feasible to be implemented. Therefore, the teacher of grade V of elementary school is expected to develop the HOTS category instruments.

Conclusion

Based on the research result, it can be concluded that (1) there are three groups of HOTS questions of the human and animal respiratory organ for students grade V of elementary students as the result of the development of HOTS questions in this study. Every group consists of 10 multiple choice questions and 5 matching questions; (2) the quality of question development reviewed by the experts shows that the assessment of HOTS questions from material and evaluation experts has a good average. The empirical test and validity test assessment stated that the first validity test is categorized as valid by 13 questions and invalid by 32 questions, while the second validity test is categorized as valid by 44 questions and invalid by 1 question. The reliability test shows that the first reliability test has sufficient and low criteria in intervals 0,21-0,70, whereas the second reliability test has high and sufficient criteria in intervals 0,41-0,90. The first test Difficulty index has difficult criteria by 24,33%, medium criteria by 62,22%, and easy criteria by 13,33%, while the second test Difficulty index has difficult criteria by 13,33%, medium criteria by 75,55%, and easy criteria by 11,11%. The first test discriminant index has the criteria of good, sufficient, and insufficient, and the second test discriminant index has the criteria of good, sufficient, and insufficient. The first test of the distractor's effectiveness has functional distractors of 48,89% and non-functional distractors of 51,11%, whereas the second test has functional distractors of 75,55% and non-functional distractors of 24,44%. Non-functional distractor effectiveness is further revised according to the analysis of the question items. According to the experts and empirical test analysis result, it is stated that HOTS questions of human and animal respiratory organs can be applied in the learning. Therefore, the teacher of grade V of elementary school is expected to develop the instruments that lead to the HOTS category.

References

Aini, D. F. N., & Sulistyani, N. (2019). Pengembangan instrumen penilaian e-quiz (electronic quiz) matematika berbasis HOTS (higher of order thinking skills) untuk kelas V sekolah dasar. *Edumaspul: Jurnal Pendidikan*, *3*(2), 1–10. https://doi.org/10.33487/edumaspul.v3i2.137

Alwi, I. (2015). Pengaruh jumlah alternatif jawaban tes obyektif bentuk pilihan ganda terhadap reliabilitas, tingkat kesukaran dan daya pembeda. *Faktor Exacta*, *3*(2), 184–193. https://doi.org/10.30998/faktorexacta.v3i2.19

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Meyer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives*. Longman.

Arifin, Z. (2016). *Evaluasi pembelajaran: Prinsip, teknik, prosedur*. Rosdakarya.

Arikunto, S. (2005). *Manajemen penelitian*. Rineka Cipta.

Arikunto, S. (2010). *Prosedur penelitian: Suatu pendekatan praktik*. Rineka Cipta.

Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (5th ed.). Bumi Aksara.

Azis, A. (2016). Analisis tes buatan guru bidang studi matematika kelas V SD 1 Katobengke. *Edumatica : Jurnal Pendidikan Matematika*, *6*(01), 15–24. https://doi.org/10.22437/edumatica.v6i01.2998

Azwar, S. (2015). *Tes Prestasi: fungsi dan pengembangan pengukuran prestasi belajar*. Pustaka Pelajar.

Bajpai, R., & Bajpai, S. (2014). Goodness of measurement: Reliability and validity. *International Journal of Medical Science and Public Health*, *3*(2), 112. https://doi.org/10.5455/ijmsph.2013.191120133

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.

Burud, I., Nagandla, K., & Agarwal, P. (2019). Impact of distractors in item analysis of multiple choice questions. *International Journal of Research in Medical Sciences*, *7*(4), 1136. https://doi.org/10.18203/2320-6012.ijrms20191313

Chauhan, P., Chauhan, G. R., Chauhan, B. R., Vaza, J. V., & Rathod, S. P. (2015). Relationship between difficulty index and distracter effectiveness in single best-answer STEM type multiple choice questions. *International Journal of Anatomy and Research*, *3*(4), 1607–1610. https://doi.org/10.16965/ijar.2015.299

Desstya, A. (2016). Kedudukan dan aplikasi pendidikan sains di sekolah dasar. *Profesi Pendidikan Dasar*, *1*(2), 193–200. https://journals.ums.ac.id/index.php/ppd/article/view/1002

Hanifah, N., & Hanifah, N. (2017). Perbandingan tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi mata pelajaran ekonomi. *Sosio E-Kons*, *6*(1), 41–55. https://doi.org/10.30998/sosioekons.v6i1.1715

Herwin, & Nurhayati, R. (2021). Measuring students' curiosity character using confirmatory factor analysis. *European Journal of Educational Research*, *10*(2), 773–783. https://doi.org/10.12973/EU-JER.10.2.773

Julia, J., Isrok'atun, I., & Safari, I. (2018). Membangun generasi emas 2045 yang berkarakter dan melek IT dan pelatihan berpikir suprarasional. *Prosiding Seminar Nasional*, 1–108. http://repository.pelitabangsa.ac.id/xmlui/handle/123456789/6157

Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *AL-TA'DIB : Jurnal Kajian Ilmu Kependidikan*, *8*(2), 70–81. https://doi.org/10.31332/atdb.v8i2.411

Kocdar, S., Karadağ, N., & Doğan Şahin, M. (2016). Analysis of the difficulty and discrimination indices of multiple-choice questions according to cognitive levels in an open and distance learning context. *TOJET: The Turkish Online Journal of Educational Technology*, *15*(4), 16–24.

Kurniawan, T. K. (2015). Analisis butir soal ulangan akhir semester gasal mata pelajaran IPS Sekolah Dasar. *Journal of Elementary Education*, *4*(1), 1–6. https://journal.unnes.ac.id/sju/index.php/jee/article/view/7488

Kusumawati, N., & Rulviana, V. (2017). *Pengembangan kurikulum di sekolah dasar*. CV AE Media Grafika.

Marzano, R. J., & Kendall, J. S. (2010). Praise for the second edition of the new taxonomy of educational objectives. In *Coaching* (pp. i–ii). Elsevier. https://doi.org/10.1016/B978-1-85617-816-7.10013-X

Muhammedi, M. (2016). Perubahan kurikulum di Indonesia: Studi kritis tentang upaya menemukan kurikulum pendidikan Islam yang ideal. *Jurnal Raudhah*, *4*(1), 49–70. https://doi.org/10.30829/raudhah.v4i1.61

Putra, G. Y. M. A., & Agustiana, I. G. A. T. (2021). E-LKPD materi pecahan dalam pembelajaran daring. *MIMBAR PGSD Undiksha*, *9*(2). https://doi.org/10.23887/JJPGSD.V9I2.35813

Putri M., D. A. A., Maryani, M., & Putra, P. D. A. (2019). Development of higher order thinking skill (HOTS) test instruments based on wondershare quiz creator. *Jurnal Pembelajaran Fisika*, *9*(1), 33–45. https://doi.org/10.23960/jpf.v9.n1.202104

Rahayu, R., & Djazari, M. (2016). Analisis kualitas soal pra ujian nasional mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, *14*(1), 85–94. https://doi.org/10.21831/jpai.v14i1.11370

Siregar, S. (2013). *Metode penelitian kuantitatif: dilengkapi perbandingan perhitungan manual & SPSS*. Kencana Prenada Media Group.

Sudijono, A. (2015). *Pengantar evaluasi pendidikan*. PT Raja Grafindo Persada.

Sudjana, N. (2010). *Penilaian hasil proses belajar mengajar*. PT Remaja Rosdakarya.

Sugiyono, S. (2014). *Statistik untuk penelitian*. Alfabeta.

Sugiyono, S. (2016). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R & D*. Alfabeta.

Syarif, E. A., & Syamsurizal, S. (2019). Analyzed quality of senior high school biology olympiad

questions at West Sumatera, Riau, Jambi, and Bengkulu in 2018. *Bioeducation Journal*, *3*(2), 142–150. https://doi.org/10.24036/bioedu.v3i2.194

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

Triyuni, N. N. E., Kusmariyatni, N. N., & Margunayasa, I. G. (2019). Pengembangan perangkat pembelajaran inkuiri terbimbing berbasis aktivitas higher order thinking (HOT) pada tema 8 subtema 1 kelas V SD. *Journal of Education Technology*, *3*(1), 22. https://doi.org/10.23887/jet.v3i1.17960

Uno, H. B., & Koni, S. (2012). *Assesment pembelajaran* (D. Ispurwanti (ed.)). PT Bumi Aksara.

Utami, N. D., Widowati, A., & Wibowo, W. S. (2017). Pengembangan media virtual laboratorium IPA materi global warming berpendekatan inkuri untuk meningkatkan kemampuan analisis siswa kelas VII. *Jurnal Pendidikan Matematika Dan Sains*, *6*(1), 62–67.

Yuniar, M., Rakhmat, C. R., & Saepulrohman, A. (2015). Analisis HOTS (higher-order thinking skill) pada soal objektif dalam mata pelajaran ilmu pengetahuan sosial (IPS) kelas V SD Negeri 7 Ciamis. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *2*(2), 187–195. https://ejournal.upi.edu/index.php/pedadidaktika/article/view/5845