

PAPER • OPEN ACCESS

## Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy

To cite this article: P N Sagala and A Andriani 2019 *J. Phys.: Conf. Ser.* **1188** 012025

View the [article online](#) for updates and enhancements.

### You may also like

- [Pre-Service Mathematics Teachers' Knowledge about Higher-Order Thinking Skills](#)  
Veronika Fitri Rianasari and Maria Suci Apriani
- [Content Analysis of Higher Order Thinking Skills \(HOTS\) in Mathematics National Exam](#)  
A M Ratna and H Retnawati
- [Mathematical disposition of strategic thinking ability in working on HOTS questions](#)  
E Firmansyah, M P Mubarika and L Saniah



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy

P N Sagala<sup>1</sup> and A Andriani<sup>1</sup>

<sup>1</sup> Mathematics Education Department, State University of Medan, North Sumatera, Indonesia, 20221

E-mail: sabikhah\_81@yahoo.com

**Abstract.** This research aims at producing assessment instruments of Mathematics in the form of HOTS questions of Probability Theory subject based on Bloom's Taxonomy. Research Design used is Research and Development. This research refers to the Tessmer progressing model which consists of two phases: a preliminary and formative evaluation. The questions are improved based on their validity, practicality, and effectiveness. The prototype I question as the beginning design is validated by 3 experts (expert review) and stated that 20 questions are valid with score 4.3. One-to-one check obtains that questions no. 10 and no. 13 need to be revised. The result of the Prototyping phase revision is called Prototype II. In a small group phase, Prototype II is tested on 5 students. The result shows that the readability level of the questions has met the element of readability very well about 88%. Practicality test of the questions obtained from the correct answers of the students is 88.3%. Considering the range of practicality criteria, the questions can be applied without revision. Effectiveness test of HOTS questions is implemented on a field test. The average score on the higher-order thinking skill of the two classes is 86.22. Counted from the completeness of the learning process of the student's proportion, the score above 70 obtained is 82.25. There are 51 of 62 students obtain score higher than 70. It means that the HOTS score of the students is at a good level. The 20 HOTS questions of Prototype III is rated Good and become the final questions without revision developed based on Tessmer.

## 1. Introduction

Educational successfulness which its primary goal is to increase human resource is influenced by many factors. In the world of education, one of the competencies that must be mastered by the lecture is the evaluation of learning [1]. One of the factors is lecturer ability in performing and utilizing the assessment, evaluation process, and learning outcome. The ability is very important to know that the aims of learning process determined in the curriculum reached. Besides, the ability can also be used to revise or to increase the learning process performed by lecturers. Educational assessment based on the Decree of the Ministry of National Education No. 20 the year 2007 about The Standard of Educational Assessment is the process of collecting and processing information to determine the students' achievement of learning outcome. The assessment principle and standard emphasizes two main ideas which are assessment must improve the learning process of the student and assessment is a valuable tool for teaching decision making [2]. Assessment Questions is not only data collection of the students but also data processing to obtain an overview of the learning process and learning outcome of the



students [3, 4]. Assessment is not only about questioning the students then finished but also about following up for the learning significance [5]. To perform an assessment, the lecture needs assessment instruments in the form of good questions for testing the cognitive, affective, and psychomotor abilities of the students. For that reason, the questions need to be analyzed and then the questions are developed based on HOTS with Bloom's Taxonomy based which refers to Tesser's development model which consists of two phases: a preliminary and formative evaluation. A learning outcome is a statement of learning accomplishment which may be the acquisition of knowledge, understanding, or an intellectual/practical skill [6]. Well, articulated statements of intended learning outcomes help both tutors and students, as they provide a clear explanation of what is required to successfully complete a module provided there are strong links between the learning outcomes and the assessment methods. As long as we can define appropriate Los (learning outcome), we should see students being motivated to focus on the skills and knowledge that a module is expected to deliver [1].

Questions are an essential element of effective teaching. Daily lectures use questions to stimulate student thinking and reasoning, while final examination papers assess the retention and application skills. The assumption exists that questions relating to application skills should start to dominate the higher academic levels in education, with a corresponding reduction in questions requiring retention skills. Effective questions should help raise issues that need feedback or about which students need to think, should include informational or problem-solving questions, and significantly more complex thinking questions that stimulate a student's mental activities [7]. Questions cannot be unclear and ambiguous by nature, and should not contain difficult vocabulary, complex syntax or unintentional clues [8, 9]. Biggs [10] talks about '*constructive alignment*' where academics support students by aligning teaching methods, assessment, and classroom environment to attain the skills and understanding required of them. When assessing the acquired skills of final-year students academics cannot create an examination using numerous LOCQ (simple recall of information). Similarly, first-year students cannot be expected to answer many HOCQ (evaluation of complex problems), as they are still assimilating new information. Therefore examination papers must be given appropriate attention in terms of maintaining the correct balance between lower, intermediate and higher order cognitive questions. This work attempts to distinguish between three different types of questions, namely LOCQ, IOCQ and HOCQ in light of Bloom's taxonomy. The aim is to ascertain whether academics are assessing critical-thinking and problem-solving skills by using effective questions. Questions analysis generally aims to know whether every question item is actually correct. It is a study of the test questions to obtain question devices that have adequate quality. Questions analysis is an activity performed by teachers to improve the quality of questions had been written. Test quality analysis is a phase that should be performed to know the quality level of a test both the overall test and the items that are part of the test. Question item analysis is an analysis performed to identify the good, poor, or bad questions. The result obtained is information about the quality of questions created to be revised as needed. Question item revision should be done by the teacher or by the school itself [9].

Based on the triennial test and evaluation result of PISA (*Programme for International Student Assessment*) performed in 2015 entitled "PISA 2015 Result in Focus", Indonesia has performance that was still far from expectations. Table 1 below shows the result of PISA 2015.

**Table 1.** Indonesia Score of PISA 2015.

	Score	Level
Mathematics	386	63 of 72 countries
Science	403	62 of 72 countries
Reading	397	64 of 72 countries

(Source Processed from OECD, PISA 2015 Result in Focus [11])

Table 1 above shows that Indonesia is at the level the 60s of 72 countries followed the programme. This issue also appears on *Trends in International Mathematics and Science Study* (TIMSS), four-yearly research which measures the student's ability of Class VIII Junior High School. State that the achievements of TIMSS in 2017 and 2011 showed the learning achievement scores of eighth-grade students of Junior High School successively were 397 and 386 (scale 0 to 800) with an average score of 500. It means that the ability of the students was below average. The result is not much different for each its participation. The low TIMSS achievements are indeed caused by several factors. One of the causing factors was that Indonesia students were poorly trained in solving contextual questions which need reasoning power, argumentation, and creativity. Such questions are the characteristics of TIMSS. The research result of TIMSS shows that Indonesia student was at the rank 36 of 49 countries in terms of natural procedure. In accordance with the result, Indonesia was left behind compared with many other countries, where Indonesia students were less competent in answering the questions that measure the higher order thinking skills. So, try out in answering non-routine questions items such as HOTS is needed to improve the student thinking ability. In line with the Zoller [12] theory that Higher Order Cognitive Skills (HOCS) items as *“quantitative problems or qualitative conceptual questions, unfamiliar to the students, that require for their solution more than knowledge and application of known algorithms. Such an application may further require (partially or fully) the abilities of reasoning, decision making, analysis, synthesis, and critical thinking”*

The research results of Agus Budiman and Jailani [13] in their article entitles Developing an Assessment Instruments of Higher Order Thinking Skill (HOTS) in Mathematics For Junior High School Grade VIII Semester 1 show that assessment instruments of HOTS in the form of HOTS test questions consists of 24 multiple choices and 19 essay questions are valid and feasible to be used seeing from the material, construction, and language aspects. Those instruments have a coefficient of reliability of 0.713 (for multiple choice) and 0.920 (for the essay). Multiple choice questions have an average of difficulty level 0.406 (medium), an average of distinguishing power 0.330 (good), and all the spieler work on the track. Essay questions have an average of difficulty level 0.373 (medium) and an average of distinguishing power 0.508 (good). First step assessment is performed by educational mathematicians to assess the validity of assessment instruments. Second step assessment is a field test which involves 178 students from three schools. The assessment is focussed on the characteristics of the HOTS question item test. Many students involved aims at adapting the students on HOTS questions.

Based on the experience of the writer as the lecturer of Probability Theory, the questions given to the students on the examinations were not analyzed yet. The file of questions saved will be a recommendation for the lecturer of the subject matter. So, the goal of this research is to describe the development of the questions of Mid-Term and Final Term examinations of the subject matter based on HOTS in Mathematics Education Department, State University of Medan. Questions developed are based on its validity, practicality, and effectiveness. Besides, the understanding of taxonomies and student understanding rating to master cognitive level in solving problems becomes one of the attempts to improve learning quality. To help the student in developing the ability, practices of HOTS are needed. So, the lecturer can give questions or exercise concerning to HOTS in the teaching-learning process in the class such as at the time of the daily test, Mid-Term Test or Final Term Test.

### *1.1. Higher order thinking skill (HOTS)*

Higher Order Thinking Skill (HOTS) is divided into 4 groups, those are problem-solving, decision making, critical thinking, and creative thinking [14, 15]. Educational researchers explain that to learn critical thinking is indirect as to learn about materials, but to learn how to relate the critical thinking inside self effectively. It means that each critical thinking skill to solve a problem relates to each other in its use.

The indicators of critical thinking skill are divided into five groups, those are providing a simple explanation, building basic skill, concluding, explaining further, and also managing strategy and tactics. The skill of 5 groups of critical thinking in detail are a) providing simple explanation consists of focusing question, analyzing argument, asking and answering question skills, b) building basic skill consists of adjusting to the sources, observing and reporting the result, c) concluding consists of considering conclusion, doing generalization and performing evaluation skills, d) explaining further consists of defining terms and creating definition as examples, and e) managing strategy and tactics consists of defining an action and interacting and communicating with the other people as examples. Critical thinking skills of the students can also be trained through giving problems in vary questions.

Techniques of writing question item based on HOTS are a) pay attention to the material coverage required for education level, b) pay attention to the several competencies required for every education level which is then lowered to be several indicators and goals of learning based on the recommendation contained in the curriculum, c) pay attention on the use of basic knowledge of a material coverage which is very possible different in accordance with the education level, use its basic knowledge or skill to solve the problems exist, d) in Bloom's Taxonomy, the lowest level can be basic knowledge to answer the question to the next level, e) providing vary data (statement, table, graph, result of the experiment is done, report, reading materials, observation result, etc.) as a stimulus to answer HOTS based questions is recommended, f) vary data provided should give information to the students which refer to basic knowledge or skill to be able to be processed further, and g) data proposed as stimulus to the student is as possible as made relevant to the authentic or real situation [8,16].

Resnick defines higher order thinking in a journal of Lestari [17] as follow (1) Higher order thinking is nonalgorithmic. It means that the action order can be completely stated first. (2) Higher order thinking tends to be complex. The overall orders or steps cannot be 'seen' only from one certain point of view. (3) Higher order thinking often results in a multi-solution. Every solution has its own weakness and strength. (4) Higher order thinking involves careful consideration and interpretation. (5) Higher order thinking involves the implementation of multi-criteria which sometimes creates criteria conflict with one another. (6) Higher order thinking often involves uncertainty. Not all things related to the task being handled can be fully understood.

Higher order thinking involves self-control in the thinking process. An individual can not be considered as a higher order thinking skill if someone else helps in every phase. Why is it that so many faculty want their students to think critically but are hard-pressed to provide evidence that they understand critical thinking or that their students have learned how to do it? We identified two major impediments to the assimilation of pedagogical techniques that enhance critical-thinking abilities. First, there is the problem of defining "critical thinking." Different definitions of the term abound [14,18]. Not surprisingly, many college instructors and researchers report that this variability greatly impedes progress on all fronts. However, there is also widespread agreement that most of the definitions share some basic features, and that they all probably address some component of critical thinking [15] Thus, we decided that generating a consensus definition is less important than simply choosing a definition that meets our needs and consistently applying it. We chose Bloom's taxonomy of educational objectives[4,8], which is a well-accepted explanation for different types of learning and is widely applied in the development of learning objectives for teaching and assessment [2].

### *1.2.Bloom's taxonomy*

Bloom's taxonomy delineates six categories of learning: basic knowledge, secondary comprehension, application, analysis, synthesis, and evaluation [6,19]. The first two categories, basic knowledge, and secondary comprehension do not require critical-thinking skills, but the last four-application, analysis, synthesis, and evaluation-all require the higher-order thinking that characterizes critical thought. The definitions for these categories provide a smooth transition from educational theory to practice by

suggesting specific assessment designs that researchers and instructors can use to evaluate student skills in any given category. Other researchers and even entire departments have investigated how to apply Bloom's taxonomy to refine questions and drive teaching strategies [19]. Nonetheless, the assessments developed as part of these efforts cannot be used to measure critical thinking independent of content. The difference between the new version of Bloom's Taxonomy and the old one can be seen in Table 2 [17].

**Table 2.** The old and new versions of Bloom's taxonomy.

The old version of Bloom's Taxonomy	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
New version of Bloom's Taxonomy	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating

(in article journal of the progress of HOTS questions, Poppy Kamalia Devi [20])

The difference of old version of Bloom's Taxonomy and the new one lies on synthesis aspect, wherein revised taxonomy synthesis aspect is no longer, but actually, be mixed to analysis. The addition is creating which comes from creating. The order of evaluation is now at the fifth meanwhile creating at the sixth, so creating becomes the highest aspect. The second difference is on the lowest cognitive aspect, knowledge. Knowledge is changed to be remembering. There is an improvement in the cognitive process, for example, the students are no more asked for knowing only a concept but must remembering the concept learned [21]. A thinking level which is accordance with HOTS seen from the cognitive aspect of the old version of Bloom's Taxonomy is at analysis, synthesis, and evaluation levels which means that seeing at the new version of Taxonomy the level is until creating a level.

HOTS questions based on Bloom's Taxonomy revised are questions type C4 (analyzing questions), C5 (evaluation questions), C6 (creating questions). [6, 19] A Revision of Bloom's Taxonomy: An Overview – Theory into Practice states that indicator to measure higher order thinking skill involves:

1) Analyzing

- analyze entered information and dividing or structuring the information into smaller parts to identify its formula or relationship.
- be able to identify and to differentiate the causing factors and effects of a complex scenario.
- identify or formulate questions.

2). Evaluating

- assess on the solution, idea, and methodology by using appropriate criteria or standard exists to make sure its effectiveness and utility score.
- hypothesize, criticize, and examine.
- accept or reject a statement based on the criteria stated.

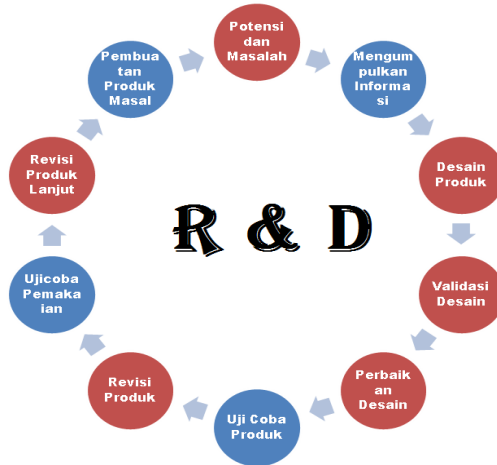
3). Creating

- generalize the idea or perspective of something
- design a way to solve a problem
- organize the elements or the parts to be a new structure which does not exist before.

## 2. Method

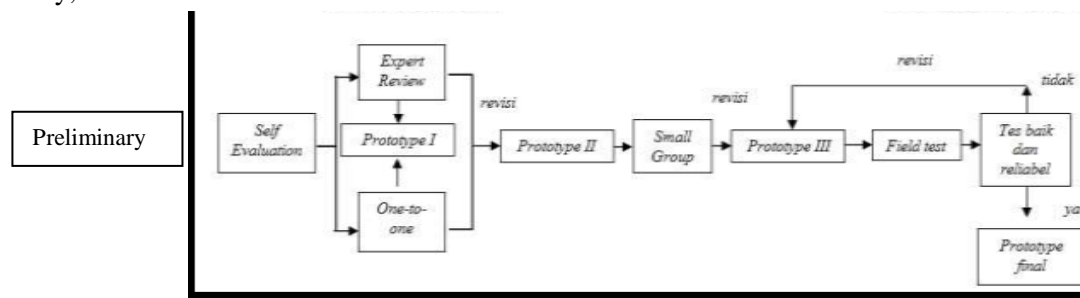
This research locates at Mathematics Education Department, State University of Medan, Willem Iskandar Street, Pasar V, Medan, North Sumatra, Indonesia in Odd Semester of 2018/2019 academic year. The subjects of this research are 60 students of Class E 2017 and Class Bilingual 2017. This

research uses the Research and Development (R&D) model [22, 23] which researcher improves the questions of Probability Theory with HOTS based on Bloom's Taxonomy.



**Figure 1.** The plot of research and development (R&D) model

This research refers to Tessmer's Development Design which consists of 2 phases: preliminary and formative evaluation that can be seen in Figure 2. Questions are improved based on its validity, practicality, and effectiveness.



**Figure 2.** The plot of Formative Evaluation Design (Tessmer [24])

#### *a. Preliminary Phase*

At this phase, researcher defines the location and the subject of research by choosing the Class to be researched and also prepares the materials needed such as textbooks and Syllabus of Probability Theory used and manages the research schedule.

#### *b. Self Evaluation Phase*

**Analysis.** At this phase, the researcher analyses the material. The material analysis uses Ronald E Wlapole and Raymond H Myers' book entitles Probability and Statistics for Engineers and Scientists as a compulsory reference in the lecture. It is aimed to know the material taught in line with the Syllabus of Probability Theory, questions designed is in line with the indicators of material competence as a comparison with the material of improvement questions which is in line with the indicator of higher order thinking skill for the questions to be improved.

**Design.** At this phase, the researcher designs HOTS questions based on higher order thinking indicators in accordance with Bloom's Taxonomy which consists of Analysis (C3), Evaluation (C4), and Creation (C6).

#### *c. Prototyping (Validation, Evaluation, and Revision)*

At this phase, the prototype will be tested on the below groups.

- Expert Review

After doing self-evaluation, a draft of HOTS questions submitted to the expert to be validated. At this phase, the first prototype will be carefully observed, assessed, and evaluated by the experts. It is often called a validity test. The experts are asked for giving suggestion and perception on the validity sheet as the material to revise the first prototype and state that the first prototype is valid. This will be a material to revise draft I to result in draft II. Revision from three experts is used as the material to test one to one.

- One to One

At this phase, the researcher tests the first prototype to a student as a tester. The comments of the student will be used to revise the items. Researcher communicates the aim of HOTS question item test to the student which are to know his ability in understanding the language used in HOTS questions, and to know whether the questions are clear or not. The weaknesses of the items then are revised so it will result prototype II to be tested to a small group.

- Small Group

The result of the revised decision on the prototype I will result in prototype II. Then, prototype II will be tested to 5 students who are not the subject of the research. At this phase, the five students are asked for answering the question. The result of the test and the comment of the students will be used to revise the items. Suggestion and comment from a small group become the background to revise prototype II. Revised result of prototype II becomes prototype III that will be used to assess the practicality. Practicality is obtained from teaching and learning process implementation in the class which is used to observe the ease in the implementation of Probability Theory examination.

- Field Test

The third prototype is tested at the subject of research those are 62 students of DikMat E 2017 and Dikmat Bilingual 2017. The effectiveness of questions obtained by the test result of students' higher order thinking skill. HOTS questions are effective if the effectiveness reaches the Classical completeness level.

## 2.1 Data Analysis Technique

The technique of data analysis uses descriptive quantitative in analyzing the questions of Probability Theory subject matter in Odd Semester in academic year 2018/2019. The analysis is performed to find out the validity, practicality, effectiveness of the questions and also the level of higher order thinking the skill of the students.

### 2.1.1 Validity Analysis

Research result of the experts on the validation sheet is found by using the following ways.

- scoring every item with an answer in very valid (5), valid (4), valid enough (3), less valid (2), and invalid (1) categories.
- calculating the total score of every validator for every aspect.
- averaging every aspect of all validators.
- assessing the validity using  $V_a = \frac{\sum_{i=1}^n V_{Li}}{n}$

Note:

$V_a$  = total average score of every aspect

$L_i$  = the average score of aspect no. I

n = number of aspect

- e. matching the average validity (V) to the criteria of question item validity.



**Table 3.** Criteria of the level of question item validity.

Score $V_a$	Validity Level
$V_a = 5$	Very Valid
$4 \leq V_a < 5$	Valid
$3 \leq V_a < 4$	Valid Enough
$2 \leq V_a < 3$	Less Valid
$1 \leq V_a < 2$	Not Valid

The criteria above states that learning devices in the form of examination question developed using Bloom's Taxonomy approach have a good level of validity if the minimum level of validity obtained is a valid level. If the level of validity obtained is below valid, revision based on the correction from the experts needs to be performed. Moreover, further validation action is done and so on until ideal learning devices in its content and construction validity obtained.

## 2.2 Practicality of Questions

The practicality of an evaluation tool emphasizes more on the efficiency and effectiveness level of the evaluation tool. The practicality of evaluation tool will result in a huge benefit for its implementation and for the students because of systematically designed especially the instrument material. Practicality is obtained by calculating the percentage of correct answers of every student using the formula is the percentage of correct answers of the student no divided to all answers.  $i$  is the student. The percentage of correct answers is as the following category.

**Table 4.** Criteria of the practicality of the questions of student's answer analysis.

Percentage (%)	Criteria of Practicality
$75\% \leq JB \leq 100\%$	can be used without revision
$50\% \leq JB < 75\%$	can be used with minor revisions
$25\% \leq JB < 50\%$	can be used with major revisions
$0\% \leq JB < 25\%$	cannot be used

(Adapted from Siti Khabibah [25])

Readability score is

$$k = \frac{\text{score obtained}}{\text{maximum score}} \times 100\%$$

$k$  is readable questions

**Table 5.** Assessment Criteria of the percentage of readable HOTS questions

No.	Percentage	Criteria
1.	81% - 100%	very good
2.	61% - 80%	good
3.	41% - 60%	good enough
4.	21% - 40%	bad enough
5.	0% - 20%	bad

(adapted from Akbar [26])

## 2.3 Effectiveness of questions

The effectiveness of the question can be seen from the potential effect in the form of quality of learning outcome, attitude, and motivation of the students. There is two aspects of effectiveness to be fulfilled by HOTS questions, those are:

- Experts and Practitioners based on their experience state that the questions are effective.
- Operationally the questions give the result as expected.

Effectiveness measurement will be seen refers to point 2 above. Its category is can be seen in Table 6 below. The percentage calculation using Learning Completeness formula is performed to calculate the percentage of the result of higher order thinking skill by dividing the students with score  $> 70$  to the total students [6]. All results obtained are converted to some categories to make the achievement level of the learning process obtained easy to be seen.

**Table 6.** Completeness criteria of students' learning outcome.

Assessment criteria of students' learning outcome	Criteria score
Excellent	$89 < A \leq 100$
Good	$79 < B \leq 89$
Enough	$70 < C \leq 79$
need guidance	$D < 70$

### 3. Result and Discussion

#### 3.1 Prototyping (Validation, Evaluation, and Revision) Phase

##### 3.1.1 Expert Review

Researchers give prototype I of HOTS questions to 3 experts as validator which their perception and suggestion become material to perform a revision. At one-to-one phase, researchers give the prototype of Hots questions to 3 non-subject students which have high, medium, and low level of skills. Researchers observe and find the difficulties experienced by the students when answering the questions. This also will be used as consideration in the development of assessment instruments of HOTS in the form of examination question of Probability Theory subject matter. After analyzing validation sheet given to 3 experts, the calculation of validity obtained is as follows.

**Table 7.** Final result data of expert validation.

No	Name of Validator	Score	Criteria
1	Expert validator of development	4,6	Valid
2	Expert validator of material	4,3	Valid
3	Expert validator of material	4,0	Valid
	Average	4,3	Valid
	Criteria: Valid		

Based on table 7 above, it can be concluded that questions developed is valid and further testing process to the small group can be performed. Then based on Table 8 and 9, The three validators gives a valid score for the components of aspects of content, construction, the language used, and time allocation in the test. This shows that the questions can be used at the field test after revising as suggested by validators. Meanwhile, comments in the form of suggestion and critics are used to revise the developed questions.

**Table 8.** Test validation result based on Bloom's taxonomy.

No.	Aspect observed	Average		
		1	2	3
1.	Content Validation			
	• The aim of the questions is formulated simply and clearly	5	5	4
	• The validity of content/material	5	5	3
	• Compatibility questions to an indicator of Competency	5	5	4
	• Clearness of question and answer expected	4	4	4
	• Compatibility with Bloom's Taxonomy approach	5	5	4
	• Clearness of Scoring Guidance			
2.	Construction Validity			
	• Problems served are questions in accordance with Bloom's Taxonomy	5	4	4
	- <i>analysis type</i>			
	- <i>evaluation type</i>			
	- <i>Creation type</i>			
	• Problems served to have solving solution or strategy which need HOTS	3	4	4
	• Problems served is in line with the level of the students.	5	5	3
3.	Language used			
	• Correct Grammatical order	4	5	4
	• Compatibility of the sentence to the thinking level, reading ability, and the age of the students.	4	4	4
	• Encourage students' interest to think deep	4	3	4
	• The simplicity of sentence structure of the questions	5	4	4
	• Sentence of the questions are not ambiguous.	5	5	4
	• Clarity of How to answer the questions	5	5	5
4.	Time allocation: in line with the questions given	5	5	4
	Overall average	4,6	4,3	4,0
	Validation result			VALID

**Table 9.** Revised suggestion from the validator.

No.	Validator	Instrument	Revised suggestion
1.	Validator 1	HOTS question grid HOTS questions	Revise the writing indicator of the questions Revise the sentence of the questions Give two HOTS questions for every indicator
No.	Validator	Instrument	Revised suggestion
2.	Validator 2	HOTS questions	Notice punctuation marks
	Validator 3	Answer Key	Notice the answer calculation

### 3.1.2 One to One

Besides the questions of test instrument of Mathematics problem solving is validated by an expert, the questions must be tested one to one to a nonsubject student. This test aims at knowing their ability in

understanding the aim of the language of HOTS questions and at knowing the questions are clear or not. The revision of Prototype I obtains decisions of the error of sentence writing of the questions to be revised and also some questions such as question no. 10 and 13 will be cleared more. The revision result of this phase obtains prototype II.

The prototype I are:

1. A coin is thrown up 3 times. Suppose that X denotes the number of faces. Is X a random variable?"
2. In the experiment shuffle a playing card, where event A is a red card appears and event B is a red card appears. Do events A and B independent?

The prototype I had been revised to Prototype II:

1. The question is not including the HOTS Test category, so the sentence needs to be added, which is to add the sentence "If so, specify the space of X". Prototype II question no 10 becomes "A coin is thrown up 3 times. Suppose that X denotes the number of faces. Is X a random variable? If so, specify the space of X!"
2. The prototype I no.13 is "In the experiment shuffle a playing card, where the event A is a red card appears and event B is a red card appears. Do events A and B independent?". The question had been revised to "In the experiment shuffle a playing card, where the event A is a red card appears and event B is a black card appears. Do events A and B independent?"

### 3.1.3 Small Group

Prototype II is tested to 5 students. The students will perform a test of question readability. The students are given a questionnaire of question readability. This is to know the readable level of the questions developed. From the questionnaire, it is obtained that the questions have met the elements of readability well about 88%. It can be seen in Table 10.

**Table 10.** The questionnaire result of readability of questions.

No.	Readable Indicator	Average
1.	Readable, the language used easily to be understood	90 %
2.	Simplicity, writing form, typography	95 %
3.	Attractiveness, reader interest, the accuracy of the idea on the questions	82 %
4.	Understanding, word and sentence characteristics.	85 %
Final Average Score		88 %

After the test of readability done, students are asked for answering the questions step by step to adjust the time needed and to comment. The goal is to test the practicality level of the questions. Practicality test of the questions obtained from the correct answer of the students is 88.3%. Adjusted to a range of practicality criteria, the questions can be used without revision. Test of small group results prototype II which is going to be tested at Field Test with 62 students as the subject of research.

### 3.1.4 Field Test Phase

The validated and revised prototype is tested to the students of Class Dikmat 2017 and Class Bilingual 2017 as many as 62 as the subject of this research. Test activity is done for 3 x 50 minutes. Students are asked for doing HOTS test which consists of 20 essay items. In the beginning, researchers divide the question and answer sheets to the student. Before doing the test, researchers give direction to answer the questions first. The data analysis result of HOTS question item test to the students obtains that: 1) for Class Dikmat 2017, the average score of higher order thinking ability based on Bloom's Taxonomy for Class Bilingual 2017, the average score of higher order thinking skill based on Bloom's Taxonomy is 85.00. So, the average score of the two classes is 86.22.

Counted from the completeness of the learning process, the proportion of students with a score more than 70 is 82.25. There are 51 of 62 students obtains score more than 70. From both calculations,

obtained that HOTS score of the students is at a good level in line with the category in Table 6. Prototype III of HOTS questions is considered Good and becomes final questions without revision.

#### 4. Conclusion

Based on the data analysis, the conclusions are as follow. At the prototyping phases (expert review and one-to-one), HOTS questions designed as the prototype I are 50 items validated by 3 experts and be stated that 20 items are valid with score 4.3. Meanwhile, thirty items are not used because cannot be finished for 3 x 50 minutes. Twenty questions chosen are revised based on the experts' note those are revised writing indicator of questions, revise sentences in the questions, give two HOTS questions for every indicator, notice the punctuation marks, and notice the answer calculation. Moreover, at one-to-one phase, students face difficulty for question no. 10 and no. 13. The revision result of prototyping phase is called prototype II.

At the small group phase, prototype II is tested to five students. The result shows that a readable level of the questions has met the elements of good readability about 88%. Practicality test of the questions obtained from the correct answers of the students is 88.3%. Adjusted to the range of practicality criteria, the questions can be used without revision. Test of small group results prototype III that will be tested at Field Test for 62 students.

Effectiveness test of HOTS questions is done at Field test phase. Data analysis result of HOTS questions to the students obtains that 1) for Class Dikmat 2017, the average score of higher order thinking ability based on Bloom's Taxonomy is 82.44, 2) for Class Bilingual 2017, the average score of higher order thinking skill based on Bloom's Taxonomy is 85.00. So, the average score of the two classes is 86.22. Counted from the completeness of the learning process, the proportion of students with a score more than 70 is 82.25. There are 51 of 62 students obtains score more than 70. From both calculations, obtained that HOTS score of the students is at a good level in line with the category in Table 6. Prototype III of HOTS questions is considered "Good" and becomes final questions without revision. So, it can be concluded that there are 20 HOTS questions that have been developed based on Tessmer.

#### References

- [1] Arifin Z & Retnawati H 2017 *Pengembangan Instrumen Pengukur Higher Order Thinking Skills Matematika Siswa SMA Kelas X PYTHAGORAS* Jurnal Pendidikan Matematika 12(1) 98
- [2] Van de Walle J A 2007 *Elementary and middle school mathematics* 6th Ed (Boston: Pearson)
- [3] Haryanto K 2016 Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap Tahun 2014/2015 Mata Pelajaran IPA Kelas III SD di Kecamatan Depok *Skripsi Yogyakarta Fak. Keguruan dan Ilmu Pendidikan Universitas Sanata Dharma*
- [4] Duhita Raras 2015 *Analisis butir Soal Ujian akhir Semester Gasal Mata Pelajaran Pengantar Akuntansi Kelas X Akuntansi SMKN 1 Godean T.A.2014/2015* Yogyakarta FKIP UNY
- [5] Popham W J 1995 *Classroom Assessment: What Teachers Need to Know* (Allyn and Bacon, Boston) 82
- [6] Anderson LW, Krathwohl D eds 2001 *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (New York: Longman)
- [7] Presseisen 1985 Glossary of Thinking Skill in A L Costa *Developing Mind: A Resource Book for Teaching Thinking* Alexandria ASCD
- [8] Betha K Wardono K 2018 Analisis Soal-Soal Matematika Tipe Higher Order Thinking Skill (HOTS) pada Kurikulum 2013 untuk Mendukung Kemampuan Literasi Siswa *Prosiding Seminar Nasional Matematika 2008* Unnes Semarang
- [9] Leeds D 2000 The seven powers of questions – Secrets to successful communication in life and Work (New York, Perige) *Proceedings of the Int'l Conference InfoTech-2009 (SAER-2009)*
- [10] Biggs J 2001 The reflective institution: Assuring and enhancing the quality of teaching and learning, *The International Journal of Higher Education and Educational Planning* 41(3) 221

- [11] OECD 2015 *PISA 2014 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (Paris: OECD Publishing)
- [12] Zoller U & Tsaparlis G 1997 Higher and lower-order cognitive skills: The case of chemistry *Research in Science Education* 27 117
- [13] Budiman A & Jailani 2014 Pengembangan Instrumen Asesmen HOTS pada Mata Pelajaran Matematika SMP Kelas VIII semester I *Journal of Riset Pendidikan Matematika* 1(2)
- [14] Ahrash N Bissel & Paula P lemons 2006A new method for assessing critical thinking in the classroom *Journal Bioscience Biological science database* 45(3) 66
- [15] Potts B 1994 Strategies for Teaching Critical Thinking in Education *ERIC Clearinghouse Report* ED385606 Washington DC
- [16] Black P C Harrison, C Lee, B Marshall, & D Wiliam 2003 Assessment for learning: Putting it into practice *Buckingham: Open University Press* 42
- [17] Lestari P R 2018 Pengembangan Soal Matematika HOTS (Higher Order Thinking Skills) Kelas X Berdasarkan Triple Theory *Journal EFEKTOR* 5(2) 8
- [18] Dodridge, M. Learning 1999 Outcomes and their assessment in higher education *Engineering Science and Education Journal* 8(4) 161
- [19] Bloom BS, ed 1956 *Taxonomy of Educational Objectives: The Classification of Educational Goals* ( New York: McKay)
- [20] Devi P K 2017 *Pengembangan Soal HOTS dalam Pembelajaran IPA SMP/MTs* Jurnal UNY
- [21] Setyawan F 2015 Conceptual Understanding Profile of LEOV Junior High School Students Based on Kolb's Learning Style *International Conference on Mathematics, Science, and Education (ICSME)* (Semarang: Universitas Negeri Semarang) pp 61–3
- [22] Facione PA 1990 Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction *The Delphi Report. Millbrae (CA): California Academic Press* 1
- [23] Karl O Jones, Jenice H, Juliet M V R, Tom T & Rebecca B 2009 Using Bloom Taxonomy to Compare the Relationship between Examination Questions and Learning Outcomes *Proceedings of International Conference on Information Technologies* 17<sup>th</sup>-20<sup>th</sup> 11
- [24] Tessmer 1993 Formative Evaluation Alternatives *Performance Improvement Quarterly* 7(1) 3
- [25] Khabibah S 2006 *Pengembangan Model Pembelajaran Matematika Dengan Soal Terbuka Untuk Meningkatkan Kreativitas Siswa Sekolah Dasar* (Surabaya: UNESA)
- [26] Akbar S 2013 *Instrumen Perangkat Pembelajaran* (Bandung: Rosdakarya)