

## Research Article

## Developing interactive questions to measure the higher-order thinking skills of senior high schools' students

Afandi <sup>a,1,\*</sup>, Saleh Hidayat <sup>a,2</sup>, Indawan Syahri <sup>b,3</sup><sup>a</sup> Biology Education Study Program, Postgraduate Program, Universitas Muhammadiyah Palembang, Jl. Jenderal Ahmad Yani 13 Ulu, Palembang, South Sumatera 30263, Indonesia<sup>b</sup> English Language Education Study Program, Faculty of Teacher Training and Education, Universitas Muhammadiyah Palembang, Jl. Jenderal Ahmad Yani 13 Ulu, Palembang, South Sumatera 30263, Indonesia<sup>1</sup> afandimoved95@gmail.com\*; <sup>2</sup> salehhidayat29@gmail.com; <sup>3</sup> indawansyahri\_ump@yahoo.co.id

\* Corresponding author

## ARTICLE INFO

## ABSTRACT

## Article history

Received February 27, 2019

Revised June 11, 2019

Accepted June 20, 2019

Published July 12, 2019

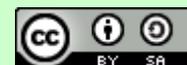
## Keywords

Higher-order thinking skills

Information technology

Interactive questions

The traditional assessments which only covered lower-order thinking skills are still commonly used by teachers in Palembang. Yet the 21<sup>st</sup>-Century education demands the empowerment of higher-order thinking skills (HOTS). The objective of this study was to find out the validity, practicality, and effectiveness of interactive questions in measuring students' HOTS. The research and development applied was 4-D model which included four stages (define, design, develop and disseminate). The study was conducted on the tenth graders of SMAN (*Sekolah Menengah Atas Negeri*-State Senior High School) 4 Palembang, and SMAN 9 Palembang, in the even semester of academic year 2018/2019. The instruments of data collection were validation sheets and interactive questions. The data analysis techniques used were Microsoft Excel 2010 and WINSTEPS version 3.73.0. In conclusion, the study showed that: 1) the expert validation results were very high category; 2) the instrument was considered as good by teacher responses; and 3) the interactive questions were effective in measuring students' HOTS.

Copyright © 2019, Afandi et al  
This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

*How to cite:* Afandi, A., Hidayat, S., & Syahri, I. (2019). Developing interactive questions to measure the higher-order thinking skills of senior high schools' students. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 5(2), 313-324. doi: <https://doi.org/10.22219/jpbi.v5i2.7747>

## INTRODUCTION

The objective of education in the 2013 Curriculum refers to the Ministerial Regulation of Education and Culture of Indonesia, Number 69 Year 2013 about the Basic Framework and Curriculum Structure of Senior High Schools and Islamic Senior High School, is to prepare Indonesian people to have life skills as individuals, citizens who are faithful, productive, creative, innovative and effective, as well as being able to contribute to the life of society, nation, state and world civilization (BSNP, 2013). However, the 2013 Curriculum in schools which have been determined by the government has not run properly (Gunawan, 2017; Kusaeri, 2014; Kustijono & Wiwin HM, 2017). Because the role of the teacher in the school is still too dominant as a disseminator of knowledge or a source of knowledge has not been based on students, so that it

does not train the Higher-Order Thinking Skills (HOTS) of students (Istiyono, 2017; Istiyono, Mardapi, & Suparno, 2014).

HOTS is an important part of educational goals and becomes a necessity for students to face real life (Almeida & Franco, 2011; A. J. Khoiriyah & Husamah, 2018; Papatthanasidou, Kleisiaris, Fradelos, Kakou, & Kourkouta, 2014; Pratama & Retnawati, 2018). The ability to think that develops in individuals as expected in the 2013 Curriculum and Ministerial Regulation of Education and Culture Number 22 Year 2006, cannot happen suddenly. Educational institutions as institutions that are responsible for managing and organizing education, play a role to equip students with abilities that are useful for facing their future lives.

In order to realize this need, it becomes a must to familiarize students to develop HOTS that can be seen from several aspects such as critical thinking, creative, problem solving, and high level decision making that can be developed in the learning process (Husamah, Fatmawati, & Setyawan, 2018; Pramesti, Sajidan, Dwiastuti, & Setyaningsih, 2019; Ramdiah, Abidinsyah, Royani, & Husamah, 2019; Ramdiah et al., 2018), both in the classroom and in the laboratory (Ghani, Ibrahim, Yahaya, & Surif, 2017; Hugerat & Kortam, 2014; Madhuri, Kantamreddi, & Goteti, 2012; Ramesh & Rao, 2015; Setiawan, Malik, Suhandi, & Permasari, 2018; Yonata & Nasrudin, 2018), and in assessment activities (Budiman & Jailani, 2014; Ghani et al., 2017; Razmawaty & Othman, 2017; Retnawati, Djidu, Apino, & Anazifa, 2017). The ability to think high according to Brookhart (2010) is at the top of Bloom's cognitive taxonomy. Teaching purpose underlying cognitive taxonomy is to be able to equip students to carry out knowledge transfer, being able to think means that students are able to apply the knowledge and skills they develop during learning in new contexts, including logic and reasoning, analysis, evaluation, and creation, problem solving and decision making.

The 2015 Program for International Student Assessment (PISA) study showed that Indonesian students were only able to work on questions with type C1-C3 which was a low level of thinking ability (BKLM-Kemendikbud, 2016). The low PISA score was certainly caused by many factors, including the current learning system that familiarizes students with only receive the information, so students are only able to solve procedural problems. The lack of habituation of students completing questions that measure HOTS based on information technology is also one of the factors causing low PISA assessment results. These problems could be solved, if the teacher was able to make questions optimizing HOTS and in the application utilizing information technology.

The results of the preliminary survey through Biology subject questions sampling of the tenth grade in 2018/2019 academic year of several Senior High Schools in Palembang could be seen that the questions used by the teacher in learning evaluation activities were still limited to the types of questions C1, C2, C3 and paper-based applications. This was due to difficulties in finding references to questions that were HOTS and the teacher's low ability to use information technology in the learning process.

There were many research about the development of instruments in measuring HOTS and showed positive results. It was easier for student to understand and work on the questions well (Budiman & Jailani, 2014; Istiyono et al., 2014; Nofiana, Sajidan, & Pugu, 2014; Rofiah, Aminah, & Ekawati, 2013; Sa'adah, Sugianto, & Sutarnan, 2014). But the weaknesses of some of the studies are that in its implementation. It has not fully utilized the development of information technology today (Gardner, 2016; Istiyono, 2017; Nursalam, Angriani, Darmawati, Baharuddin, & Aminuddin, 2018; Razmawaty & Othman, 2017). According to the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2011), the development of information technology is expected to provide success in universal education throughout the world, one of which is to improve students' skills.

After looking at a number of previous studies which focused on developing instruments in measuring HOTS, several weaknesses were found, including the application of the information technology to the maximum extent of the development of information technology, therefore this study intends to determine the validity, practicality, and effectiveness of interactive questions in measuring students' HOTS.

## METHOD

The model used in this study was a 4-D model according to Thiagarajan, Semmel, and Semmel (1974), which consists of defining, designing, developing and distributing. The subjects of this study were students from two SMAN (*Sekolah Menengah Atas Negeri*-State Senior High School) i.e. SMAN 4 Palembang, and SMAN 9 Palembang, with totaling 100 students. The instruments used in this study were questionnaires, interviews, and interactive questions. Analysis of the data in this study was conducted in two stages, the first stage was carried out by analyzing the validity of the questions assessed by the validator before testing and

the practicality based on the analysis of the use of the questionnaire response used by Biology teachers and students with the help of the Microsoft Excel 2010 application. The validity assessment of expert lecturers was analyzed by Aiken's V statistics (Aiken, 1985), using the Microsoft Excel 2010 application which was formulated (Azwar, 2012), as in Formula 1:

$$V = \frac{\sum s}{[n (c-1)]} \quad (1)$$

Where: V = expert agreement index regarding item validity, s = r – l0, r = Number given by expert, l0 = lowest validity assessment number, c = highest validity assessment number, n = number of experts / validator.

To interpret the score of expert validity obtained from the calculation above, the classification of validity was used as shown in Table 1.

**Table 1.** The criteria for expert lecturers validity

No	Validity Result	Validity Criteria
1	0,80 < V ≤ 1,00	Very high
2	0,60 < V ≤ 0,80	High
3	0,40 < V ≤ 0,60	Enough
4	0,20 < V ≤ 0,40	Low
5	0,00 < V ≤ 0,20	Very low

(Source: Aiken, 1985).

Furthermore, in order to analyze the data filling in the user responses questionnaire by Biology teachers and students using Microsoft Excel 2010 application as in Formula 2 (Wicaksono, Kusmayadi, & Usodo, 2014):

$$\% \text{ NRP} = \frac{\sum \text{NRP}}{\text{NRP Maximum}} \times 100 \quad (2)$$

where: % NRP = Percentage of User Response Score,  $\sum$  NRP = User Response Total Score (SB NRP + NRP B + NRP C + NRP K + NRP SK), and Maximum NRP =  $\sum$  R x Best choice score (5).

The results of the analysis interpreted based on the percentage criteria of the user response score per item statement as shown in Table 2.

**Table 2.** NRP percentage criteria

No	Validity result (%)	Validity criteria
1	0 ≤ NRS < 20	Very weak
2	20 ≤ NRS < 40	Weak
3	40 ≤ NRS < 60	Enough
4	60 ≤ NRS < 80	Strong
5	80 ≤ NRS < 100	Very strong

Meanwhile, the analysis in the second stage was carried out on the students' answers after answering interactive questions on Kingdom Plantae material in the tenth grade of Biology subjects with the help of WINSTEPS application 3.73.0 version to determine the effectiveness of interactive questions to measure students' HOTS based on validity, reliability, the level of difficulty of the questions, the differences of the questions, the bias item, the level of student ability, the pattern of student responses, the relevance of the quality of the response to the instrument, and measurement information (Sumintono & Widhiarso, 2015).

Next, to analyze the percentage of high order thinking skills by looking at the total score column. The scores obtained by students will be presented using the formula from Purwanto (2009), as in Formula 3:

$$NP = \frac{R}{SM} \times 100\% \quad (3)$$

where: NP = the expected percentage score, R = Raw score obtained by students, SM = Ideal maximum score of the test in question, and 100 = Fixed number.

The interpretation of the percentage of HOTS of students based on categories according to the Prasetyani, Hartono, and Susanti (2016), as seen in Table 3.

**Table 3.** The categories of students' HOTS

No	Percentage (%)	Predicate
1	81 – 100	Very Good
2	61 – 80	Good
3	41 – 60	Sufficient
4	21 – 40	Poor
5	0 – 20	Very Poor

## RESULTS AND DISCUSSION

### The validity of interactive questions in measuring students' HOTS

The validity of interactive questions in measuring students' HOTS can be known based on the results of the validity test by the expert appraisal, namely the evaluation expert lecturer, material expert lecturer, and linguist lecturer. The validation of the evaluation expert lecturer covers aspects of the formulation of questions, choice of answers, language, and layout. The validation results of the evaluation expert lecturers can be seen in [Table 4](#).

**Table 4.** The results of the validation by the expert evaluation lecturer

Question number		
Valid without revision	Valid with revision	Not Valid
1, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 26, 27, 29, 30.	2, 3, 8, 9, 17, 19, 24, 25, 28.	-
<b>21 items</b>	<b>9 items</b>	<b>-</b>

Based on [Table 4](#), it appears that the number of items declared valid without revision is 21 items, valid with revisions totaling 9 items, and there are no invalid items. Suggestions are given for problems that are declared valid with a revision that is related to the question formulation, language, and pictures/graphs/tables/diagrams.

The results of the validation by an evaluation expert are interpreted based on the score category and analyzed by Aiken's V statistics using Microsoft Excel 2010. The results of the analysis can be seen in [Table 5](#).

**Table 5.** Statistical analysis of the validation results of expert evaluation lecturers

No	Validity result	Validity criteria	Question number
1	$0,80 < V \leq 1,00$	Very high	1, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 26, 27, 29, and 30.
2	$0,60 < V \leq 0,80$	High	-
3	$0,40 < V \leq 0,60$	Enough	2, 3, 8, 9, 17, 19, 24, 25, and 28
4	$0,20 < V \leq 0,40$	Low	-
5	$0,00 < V \leq 0,20$	Very low	-

[Table 5](#) shows that as many as 21 items had very high levels of validation and as many as 9 items had sufficient validation. The conclusion of the validation results by expert lecturers evaluating interactive questions is that interactive questions have a validity with a very high category in measuring HOTS. From the average statistical value of Aiken's V, the validation of the evaluation expert lecturer is 0.85 and the percentage of valid questions is 70.00%.

Validation of material expert lecturers includes aspects of indicators, concepts, languages, layouts, and construction. The results of the validation of the material expert lecturer can be seen in [Table 6](#).

**Table 6.** The results of the validation by the material expert lecturer

Question number		
Valid without revision	Valid with revision	Not Valid
1, 3, 4, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 29, 30.	2, 5, 12, 13, 14, 25, 28.	-
<b>23 items</b>	<b>7 items</b>	<b>-</b>

Based on [Table 6](#), it can be seen that the number of questions declared valid without revision is 23 items, valid with revisions are 7 items, and there are no invalid items. Suggestions are given for problems that are declared valid with a revision that is related to conception, language, and choice of answers.

The results of the validation of the material experts were interpreted based on the score category and analyzed Aiken's V statistics using Microsoft Excel 2010. The results of the analysis can be seen in [Table 7](#).

**Table 7.** Statistical analysis of the validation results of the material expert lecturer

No	Validity result	Validity criteria	Question number
1	$0,80 < V \leq 1,00$	Very high	1, 3, 4, 6, 7, 8, 9, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 29, and 30
2	$0,60 < V \leq 0,80$	High	-
3	$0,40 < V \leq 0,60$	Enough	2, 5, 12, 13, 14, 25, and 28
4	$0,20 < V \leq 0,40$	Low	-
5	$0,00 < V \leq 0,20$	Very low	-

Based on [Table 7](#), the results show that as many as 23 items have a very high level of validation and as many as 7 items have sufficient validation. The conclusion from the results of the validation by the material expert lecturers is that interactive questions have a very high validity in measuring HOTS. This is seen from the average value of Aiken's V statistics that the validation of material expert lecturers is 0.88 and the percentage of valid questions is 76.60%.

Validation of linguist lecturers covers aspects of the use of Indonesian language rules, terminology, language clarity, and language suitability. The results of the validation of language professors can be seen in [Table 8](#).

**Table 8.** The results of the validation by the linguists

Question number		
Valid without revision	Valid with revision	Not Valid
1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 27, 29, 30.	3, 9, 17, 26, 28.	-
<b>25 items</b>	<b>5 items</b>	-

Based on [Table 8](#), it can be seen that the number of questions declared valid without revision is 25 items, valid with revisions are 5 items, and there are no invalid items. Suggestions are given for problems that are declared valid with revisions, namely relating to Indonesian language rules, language clarity, and language suitability.

The results of the validation of linguists are interpreted based on the score category and analyzed by Aiken's V statistics using Microsoft Excel 2010. The results of the analysis can be seen in [Table 9](#).

**Table 9.** Statistical analysis of the validation results of the linguist

No	Validity result	Validity criteria	Question number
1	$0,80 < V \leq 1,00$	Very high	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 27, 29, and 30
2	$0,60 < V \leq 0,80$	High	-
3	$0,40 < V \leq 0,60$	Enough	3, 9, 17, 26, and 28
4	$0,20 < V \leq 0,40$	Low	-
5	$0,00 < V \leq 0,20$	Very low	-

Based on [Table 9](#), the results show that as many as 25 items have a very high level of validation and as many as 5 items have sufficient validation. The conclusion from the results of the validation of linguists by interactive linguists is that interactive questions have a very high validity in measuring HOTS, seen from the average value of Aiken's V statistics for linguists 'lecturers' validation by 0.92 and the percentage of valid questions by 83.30%.

The validity of interactive questions in measuring HOTS was tested through the expert validity stage and trial analysis. Based on the results of the expert validity, it was found that; a) the average value of Aiken's V was 0.85 with a very high category and the percentage of valid questions was 70.00%, but improvements must be made in the formulation of questions, languages, and images/charts/tables/diagrams; b) material expert validation obtained an average value of Aiken's V statistics was 0.88 with a very high category and the percentage of valid questions amounted to 76.60%, but improvements must be made to concepts, languages, and answer choices; c) the validation of linguists obtained an average value of Aiken's V statistics was 0.92 with a very high category and the percentage of valid questions amounted to 83.30%, but improvements must be made to the rules of Indonesian language, language clarity, and language suitability.

The validity of interactive questions for measuring HOTS could also be seen from the results of the analysis development trials using the WINSTEPS-assisted Rasch Model 3.73.0 version. The results of empirical validity test could be explained as follows: a) the analysis results of the validity/level of item fit using the Winsteps application 3.73.0 version could be concluded that all questions were normal fit/valid and there was no questions that needed to be repaired; b) the results of reliability analysis using the Winsteps

application 3.73.0 version found that the developed questions have Alpha Cronbach reliability was 0.72, which means that the interaction between students and items in the overall question was good. For the score of person reliability and item reliability, in order was 0.84 and 0.82, it could be interpreted that the consistency of the students answers was good and the quality of the items in the instrument developed by the reliability aspects were good; c) the questions that have the highest level of difficulty was item number 7 with item measure logit score was 1.64 and questions that have the lowest difficulty level was item number 30 with item measure logit score was -1.22; d) the questions that have very good different power criteria were 7 items, the criteria for different power problems were 19 items, and the quite good different power criteria were 4 items; e) there was no question that was biased and must be fixed because it had a probability score greater than 5% (0.05).

A valid instrument, will be able to measure aspects of HOTS student accurately (Erfianti, Istiyono, & Kuswanto, 2019; K. Khoiriyah, Jalmo, & Abdurrahman, 2018; Puteh, Aziz, Tajudin, & Adnan, 2018; Sumarni, Supardi, & Widiarti, 2018). Instruments must have accuracy when used (R. Williams, 2003). Consistent and stable, in the sense of not changing from one measurement time to another measurement. Data that lack validity, will produce biased conclusions, are not in accordance with what they should, and may even conflict with custom. To make an instrument to measure the instrument, it is necessary to study theories, expert opinions, and experiences that are sometimes needed if the operational definitions of the variables are not found in theory (Cadorn, Bagnasco, Tolotti, Pagnucci, & Sasso, 2016; Yeager & Lee Duckworth, 2015).

### The practicality of interactive questions in measuring students' HOTS

According to Nieveen (1999), that aspect of practicality is seen in terms of users: a) do experts and practitioners think that something developed can be used in normal conditions?; and b) does reality show that something developed can be applied by teachers and students?. Based on these, the practicality of interactive questions for measuring HOTS was obtained from the results of the responsiveness analysis of HOTS.

In the implementation of extensive trials carried out the distribution of user response questionnaires to biology teacher of SMAN 4 Palembang and biology teacher of SMAN 9 Palembang. The results of the questionnaire were analyzed using Microsoft Excel 2010 application, to find out the response criteria for using interactive questions in measuring HOTS. Analysis of the data to determine the response criteria for using interactive questions is to determine the percentage of response values from instrument users, the results of which are shown in Table 10.

Table 10. The results of the analysis of the response to the use of interactive questions

No	User response	P	Criteria	NRP	
1	Very good	5	5	25	
2	Good	84	4	336	
3	Enough	0	3	0	
4	Less	0	2	0	
5	Ver less	0	1	0	
				$\Sigma$ NRS	361
				$\Sigma$ Max	550
				% NRS	65.60
				Criteria	Strong

Based on Table 10, it can be seen that the percentage of responses using interactive questions is equal to 65.60%. This shows the teacher's response to interactive questions to measure HOTS included in the criteria of strong and positive, because the value of user responses above 50%. The data is supported by comments from the teacher. Based on the results of the responsiveness analysis for the use of interactive questions to measure HOTS, it can be concluded that interactive question was practical to measure students' HOTS. It was seen from the analysis result of questionnaire by the teacher. Biology teacher stating that the suitability of the content, information clarity, language usage, clarity of purpose, ease of use, systematic presentation, and display of interactive questions included in strong and positive criteria.

Consequently, we need methods and criteria in order to assess the practicality of an instrument before it is implemented in their practice (Gamel, 2007). Compliance with completing education modules or teaching sets/instrument and the perceived usefulness are important factors in minimising bias and enhancing reliability in how raters complete the measure (R. G. Williams, Klamen, & McGaghie, 2003). Criteria that can be seen for example suitability of the content (Miles, Fulbrook, & Mainwaring-Mägi, 2016), information clarity (Silveira et al., 2018), language (Jones, 1976), and others .

### The effectivity of interactive questions in measuring students' HOTS

The effectiveness of interactive questions to measure HOTS was seen using Duncan's measure of effectiveness theory in Steers (1985), one of which is the achievement of goals. The purpose of developing interactive questions was as an instrument for assessing students' HOTS. The assessment of students' HOTS was seen based on the results of the analysis of large-scale trials using the WINSTEPS version 3.73.0. Large scale trials were conducted at SMAN 4 Palembang and SMAN 9 Palembang with a total of 70 students. This stage produces an overview of the effectiveness of interactive questions in measuring HOTS. The results of wide-scale trials were analyzed to find out the categories of students' high-level thinking skills, student response patterns, the relationship of the quality of student responses to instruments, and the measurement information function.

Based on the analysis of students' answers using the WINSTEPS application, students who have the highest level of ability in answering questions are 34P code students (student of SMAN 4 Palembang; female) with a logit item measure value of 1.70 and students who have the lowest ability level in answering questions namely 67L code students (student of SMAN 9 Palembang, male) with a logit item measure value of -0.43. The average value of the person measure obtained 0.29, which means that the ability of students above the level of difficulty of the problem, or most students are able to do the problems correctly.

The percentage of HOTS is obtained by looking at the total score column. The results of the HOTS category analysis of students can be seen in Table 11 and Table 12.

**Table 11.** HOTS category of students of SSS 4 of Palembang

No	Percentage (%)	Predicat	Student	Amount
1	81 – 100	Very good	34P	1
2	61 – 80	Good	07P, 15P, 16P, 17L, 18L, 19P, 21L, 22L, 23P, 25L, 26L, 27L, 28L, 30L, 31L, dan 32L	16
3	41 – 60	Enough	01P, 02P, 03P, 04P, 05P, 06P, 08P, 09P, 10P, 11P, 12P, 13L, 14L, 20P, 24P, 29L, 33L, dan 35P	18

**Table 12.** HOTS category of students of SSS 9 of Palembang

No	Percentage (%)	Predicat	Student	Amount
1	61 – 80	Very good	50P, 54P, 56L, 58P, dan 69L	5
2	41 – 60	Good	36P, 37P, 38L, 39P, 40L, 41P, 42P, 43P, 44P, 45P, 46P, 48P, 49P, 51P, 52L, 53L, 55L, 57L, 59P, 60P, 61L, 62P, 63L, 65P, 66P, 68P, dan 70L	27
3	21 – 40	Enough	47L, 64P, dan 67L	3

Based on Table 11 and Table 12, it can be seen that students of SMAN 4 Palembang have HOTS with very good and good categories; more compared to students of SMAN 9 Palembang. This is directly proportional to the results of the computer-based national exam in 2017, where the scores of students of SMAN 4 Palembang are higher than the grades of students of SMAN 9 Palembang.

If seen from the number of students who are able to answer correctly based on the cognitive level of the questions can be seen in Table 13.

**Table 13.** Results of analysis of student answers based on cognitive levels

No	Cognitive level	Question number	Percentage of students who answered right (%)
1	C4 (Analyze)	1, 3, 4, 11, 12, 14, 15, 16, 14, 22, 24, 26, 30	67.03
2	C5 (Evaluate)	2, 5, 6, 9, 10, 13, 18, 19, 21, 23, 28, 29	54.81
3	C6 (Create)	7, 8, 20, 25, 27	36.57

Based on Table 13, it can be seen that students who are able to answer questions with a cognitive level C4 (Analyze) are 67.03%. Students who are able to answer questions with a cognitive level of C5 (Evaluate) are 54.81%. Students who are able to answer questions with a cognitive level of C5 (Create) are 36.57%.

The results of the analysis of student answers using the WINSTEPS application were also carried out to determine the pattern of student responses. The complete response patterns of SMAN 4 Palembang students can be seen from the Guttman Scalogram in Figure 1.

```

GUTTMAN SCALOGRAM OF RESPONSES:
Person | Item
      | 21231 1 11 11221 2112 222 2
      | 254071134842636999031256805877
      | -----
22L +011110101111110010010111111101 22L
8  +100111011011000011111010000001 08P
    
```

Figure 1. Guttman Scalogram SSSH 4 of Palembang

Based on Figure 1, it appears that the 22L code students are categorized as careless students. This is because the first easiest item, question number 22, cannot be done correctly, while the most difficult question, problem number 27, can be done correctly. Furthermore, there is no similar pattern of student responses in answering questions, this means that no students are indicated to cooperate in answering questions. But it is indicated that there are students who answer the questions by guessing (lucky guess) the students are 08P code. this is because the hardest problem, problem number 27, can be done correctly, while the second easiest problem, problem number 15, cannot be done correctly.

The response patterns of SMAN 9 Palembang students can be seen from the Guttman Scalogram in Figure 2.

```

GUTTMAN SCALOGRAM OF RESPONSES:
Person | Item
      | 21231 1 11 11221 2112 222 2
      | 254071134842636999031256805877
      | -----
69L +011110101101110100111011100110 69L
49P +101010101010110100010010010100 49P
    
```

Figure 2. Guttman Scalogram SSSH 9 of Palembang

Based on Figure 2, it appears that 69L code students are included in the category of careless students. This is because the easiest question item, question number 25, cannot be done correctly, while the second most difficult problem, question number 7, can be done correctly. Furthermore, there is no pattern of student responses that are the same in answering questions, this means that no students are indicated to cooperate in answering questions. But it is indicated that there are students who answer the questions by guessing (lucky guess), namely students 49P code. The most difficult item, problem number 8, can be done correctly, while the second easiest problem, problem number 15, cannot be done correctly.

The linkage of the quality of student responses to instruments can be assessed by looking at the mean values of MNSQ and ZSTD, as can be shown in Figure 3.

```

-----
Person: REAL SEP.: .00 REL.: .00 ... Item: REAL SEP.: .00 REL.: .00

Item STATISTICS: MEASURE ORDER

-----
|ENTRY  TOTAL  TOTAL  MODEL|  INFIT |  OUTFIT |PT-MEASURE|EXACT MATCH|
|NUMBER SCORE  COUNT MEASURE S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP. |OBS%  EXP%| Item |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN   39.4  70.0   .00   .26|1.00   .0|1.00   .1|         | 64.3  64.1|
| S.D.   8.6    .0     .55   .01|.05   .8|.07   .8|         | 7.7   5.4|
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
    
```

Figure 3. MNSQ and ZSTD mean value

Based on Figure 3, it can be explained that the mean INFIT and OUTFIT MNSQ values obtained are 1.00, and the mean INFIT and OUTFIT ZSTD values obtained respectively are 0 and 0.1. According to Sumintono and Widhiarso (2015) the mean value of INFIT and OUTFIT MNSQ approached the ideal value of 1.00 and the mean value of INFIT and OUTFIT ZSTD approached the ideal value of 0.0 then the quality of the

instrument in terms of overall student response and items were getting better. So it can be concluded that the quality of the instrument seen based on student responses and items included in both categories.

The next stage is knowing the function of measurement information. The results of the analysis of student answers using the WISTEPS application to determine the function of measurement information, can be seen in [Figure 4](#).

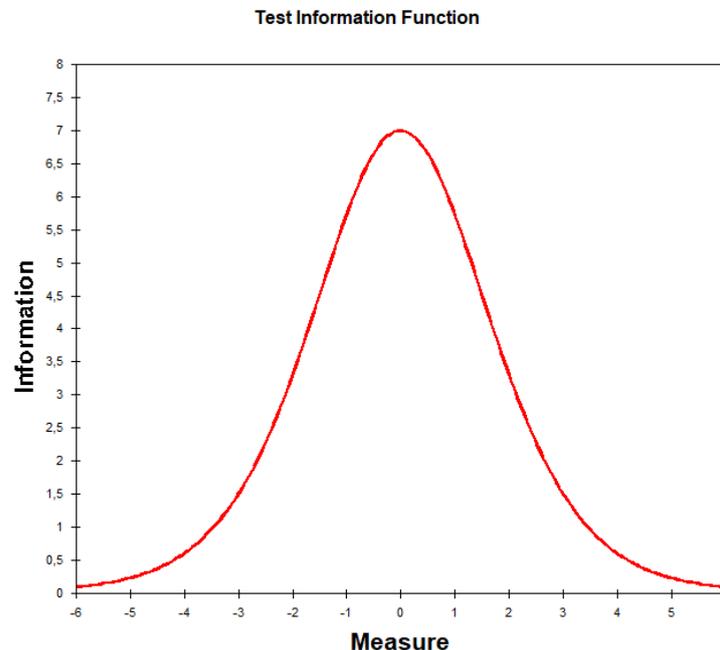


Figure 4. The graphic of test information function

Based on [Figure 4](#), it can be seen that the X axis shows about the ability of students, while the Y axis shows the magnitude of the information function obtained. It can be concluded that 30 items given to 70 students in class X of SMAN 4 Palembang and SMAN 9 Palembang are suitable for knowing the level of ability of students with moderate ability. This means that HOTS interactive question instruments produce optimal information when given to students who have HOTS at a moderate level.

The problem, which occurs at school, is that questions tend to test more cognitive low level aspects which do not train the HOTS aspects of students ([Hugerat & Kortam, 2014](#); [K. Khoiriyah et al., 2018](#); [Ramesh & Rao, 2015](#); [Razmawaty & Othman, 2017](#)), so it is only natural that the ability to think of Indonesian children is scientifically considered low. This is for example seen from the TIMSS survey results ([Budiman & Jailani, 2014](#)) atau PISA ([BKLM-Kemendikbud, 2016](#)). The existence of HOTS assessment instruments, especially in the form of interactive questions, in addition to being used to determine a student's ability profile, can also be used as a means of training students' HOTS abilities. The questions used as exercise can contain questions that test students, according to the categories in HOTS.

## CONCLUSION

Based on the results and discussion presented above, it could be concluded that: 1) the interactive question was valid to measure the HOTS of the tenth grade students of high school on kingdom plantae material, based on the results of tests through expert validation stages and analysis of trial results development. The results of the validity of expert (evaluation, material, and language) obtaining the average Aiken's V statistics in order was 0.85, 0.88, and 0.92 in the very high category. Meanwhile, the development test results showed that the normal functioning of measuring validity, reliability aspects were good and there was no question that need to be corrected; 2) the interactive questions were stated practically to measure the HOTS of the tenth grade students of high school on kingdom plantae material, response analysis using interactive questions show that the question had a strong and positive response criterion based on the percentage of teacher was 65.60% (Biology teachers of SMAN 4 Palembang and SMAN 9 Palembang); and 3) the interactive question was effective for measuring HOTS of the tenth grade students of high school on

kingdom plantae material, based on the level of students' HOTS, number of correct answers based on cognitive level questions, student response patterns, quality questions, and optimization of using interactive questions.

## REFERENCES

- Aiken, L. R. (1985). *Psychological testing and assessment* (5th ed.). Needham Heights, MA, US: Allyn & Bacon. Retrieved from <https://psycnet.apa.org/record/1985-97221-000>
- Almeida, L. da S., & Franco, A. H. R. (2011). Critical thinking: Its relevance for education in a shifting society. *Revista de Psicologia*, 29(1), 175–195. Retrieved from [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S0254-92472011000100007](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0254-92472011000100007)
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar. Retrieved from <https://scholar.google.com/scholar?cluster=12819496658226871221&hl=en&oi=scholar>
- BKLM-Kemendikbud. (2016). Peringkat dan capaian PISA Indonesia mengalami peningkatan. 4 January 2016. Jakarta: Biro Komunikasi dan Layanan Masyarakat Kementerian Pendidikan dan Kebudayaan. Retrieved from <https://www.kemdikbud.go.id/main/blog/2016/05/rumah-kunci-sukses-pola-asuh-anak>
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria, VA-USA: ASCD. Retrieved from <http://www.ascd.org/publications/books/109111.aspx>
- BSNP. (2013). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 69 Tahun 2013 Tentang Kerangka Dasar dan Struktur Kurikulum Sekolah Menengah Atas/Madrasah Aliyah. Jakarta-Indonesia: BSNP. Retrieved from <http://bsnp-indonesia.org/id/wp-content/uploads/2013/06/Salinan-Permendikbud-No.-69-th-2013-ttg-ttg-KD-dan-Struktur-Kurikulum-SMA-MA.zip>
- Budiman, A., & Jailani, J. (2014). Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1. *Jurnal Riset Pendidikan Matematika*, 1(2), 139–151. doi: <https://doi.org/10.21831/jrpm.v1i2.2671>
- Cadorin, L., Bagnasco, A., Tolotti, A., Pagnucci, N., & Sasso, L. (2016). Instruments for measuring meaningful learning in healthcare students: A systematic psychometric review. *Journal of Advanced Nursing*, 72(9), 1972–1990. doi: <https://doi.org/10.1111/jan.12926>
- Erfianti, L., Istiyono, E., & Kuswanto, H. (2019). Developing lup instrument test to measure higher order thinking skills (HOTS) Bloomian for senior high school students. *International Journal of Educational Research Review*, 4(3), 320–329. doi: <https://doi.org/10.24331/ijere.573863>
- Gamel, C. J. (2007). Establishing the practicality of an instrument. In *The 39th Biennial Convention*. Retrieved from [https://stti.confex.com/stti/bc39/techprogram/paper\\_36004.htm](https://stti.confex.com/stti/bc39/techprogram/paper_36004.htm)
- Gardner, R. M. (2016). Clinical information systems – From yesterday to tomorrow. *IMIA Yearbook*, 62–75. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5171508/pdf/yimi-11-0s62.pdf>
- Ghani, I. B. A., Ibrahim, N. H., Yahaya, N. A., & Surif, J. (2017). Enhancing students' HOTS in laboratory educational activity by using concept map as an alternative assessment tool. *Chemistry Education Research and Practice*. doi: <https://doi.org/10.1039/C7RP00120G>
- Gunawan, I. (2017). Indonesian Curriculum 2013: Instructional management, obstacles faced by teachers in implementation and the way forward. In *Advances in Social Science, Education and Humanities Research* (Vol. 128, pp. 56–63). Atlantis Press. doi: <https://doi.org/10.2991/icet-17.2017.9>
- Hugerat, M., & Kortam, N. (2014). Improving higher order thinking skills among freshmen by teaching science through inquiry. *Eurasia Journal of Mathematics, Science and Technology Education*, 10(5), 447–454. doi: <https://doi.org/10.12973/eurasia.2014.1107a>
- Husamah, H., Fatmawati, D., & Setyawan, D. (2018). OIDDE learning model: Improving higher order thinking skills of biology teacher candidates. *International Journal of Instruction*, 11(2), 249–264. doi: <https://doi.org/10.12973/iji.2018.11217a>
- Istiyono, E. (2017). The analysis of senior high school students' physics HOTS in Bantul District measured using PhysReMChoTHOTS. *AIP Conference Proceedings*, 1868(August). doi: <https://doi.org/10.1063/1.4995184>
- Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi Fisika (PhysTHOTS) untuk peserta didik. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 1–12. doi: <https://doi.org/10.21831/pep.v18i1.2120>
- Jones, J. E. (1976). Instrumentation. In *Criteria for evaluating instruments* (1st ed., pp. 502–506). University Associates, Inc. doi: <https://doi.org/10.1177/105960117600100417>

- Khoiriyah, A. J., & Husamah, H. (2018). Problem-based learning: Creative thinking skills, problem-solving skills, and learning outcome of seventh grade students. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 4(2), 151–160. doi: <https://doi.org/10.22219/jpbi.v4i2.5804>
- Khoiriyah, K., Jalmo, T., & Abdurrahman, A. (2018). Development of assessment instrument higher order thinking skills on science subjects for student grade eight junior high school. *The Online Journal of New Horizons in Education*, 8(2), 19–29. Retrieved from <http://repository.lppm.unila.ac.id/3260/25/development-HOTS-instruments.pdf>
- Kusaeri, K. (2014). *Acuan dan teknik penilaian proses dan hasil belajar dalam kurikulum 2013* (I). Yogyakarta: Ar Ruzz Media. Retrieved from <http://digilib.uinsby.ac.id/14615/>
- Kustijono, R., & Wiwin HM, E. (2017). Pandangan guru terhadap pelaksanaan kurikulum 2013 dalam pembelajaran fisika SMK di Kota Surabaya. *Jurnal Penelitian Fisika Dan Aplikasinya (JPFA)*, 4(1), 1. doi: <https://doi.org/10.26740/jpfa.v4n1.p1-14>
- Madhuri, G. V., Kantamreddi, V. S. S. N., & Goteti, L. N. S. P. (2012). Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education*, 37(2), 117–123. doi: <https://doi.org/10.1080/03043797.2012.661701>
- Miles, S., Fulbrook, P., & Mainwaring-Mägi, D. (2016). Evaluation of standardized instruments for use in universal screening of very early school-age children: suitability, technical adequacy, and usability. *Journal of Psychoeducational Assessment*, 36, 1–21. doi: <https://doi.org/10.1177/0734282916669246>
- Nieveen, N. (1999). Prototyping to reach product quality. In J. an van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 125–135). Switzerland: Springer, Dordrecht. doi: [https://doi.org/10.1007/978-94-011-4255-7\\_10](https://doi.org/10.1007/978-94-011-4255-7_10)
- Nofiana, M., Sajidan, S., & Puguh, P. (2014). Pengembangan instrumen evaluasi two-tier multiple choice question untuk mengukur keterampilan berpikir tingkat tinggi pada materi kingdom plantae. *Jurnal Inkuiri*, 3(2), 60–74. doi: <https://doi.org/10.20961/inkuiri.v3i2.9694>
- Nursalam, Angriani, A. D., Darmawati, Baharuddin, & Aminuddin. (2018). Developing test instruments for measurement of students' high-order thinking skill on mathematics in junior high school in Makassar. In *Journal of Physics: Conference Series* (Vol. 1028, pp. 1–5). doi: <https://doi.org/10.1088/1742-6596/1028/1/012169>
- Papathanasiou, I. V., Kleisaris, C. F., Fradelos, E. C., Kakou, K., & Kourkouta, L. (2014). Critical thinking: The development of an essential Skill for nursing students. *Acta Informatica Medica*, 22(4), 283–286. doi: <https://doi.org/10.5455/aim.2014.22.283-286>
- Pramesti, B. N., Sajidan, S., Dwiastuti, S., & Setyaningsih, E. (2019). The feasibility of biology module based on Stim-HOTS models. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 5(1), 101–108. doi: <https://doi.org/10.22219/jpbi.v5i1.7385>
- Prasetyani, E., Hartono, Y., & Susanti, E. (2016). Kemampuan berpikir tingkat tinggi siswa kelas XI dalam pembelajaran trigonometri berbasis masalah di SMA Negeri 18 Palembang. *Jurnal Gantang Pendidikan Matematika FKIP - UMRAH*, 1(1), 31–40. doi: <https://doi.org/10.31629/jg.v1i1.4>
- Pratama, G. S., & Retnawati, H. (2018). Urgency of higher order thinking skills (HOTS) content analysis in mathematics textbook. In *Journal of Physics: Conference Series* (Vol. 1097, pp. 1–8). IOP Publishing. doi: <https://doi.org/10.1088/1742-6596/1097/1/012147>
- Purwanto, M. N. (2009). *Prinsip-prinsip dan teknik evaluasi pengajaran* (10th ed.). Bandung: Tri Remaja Rosdakarya. Retrieved from <https://rosda.co.id/pendidikan-keguruan/418-prinsip-teknik-evaluasi-pengajaran.html>
- Puteh, M., Aziz, A. A. M. A., Tajudin, N. M., & Adnan, M. (2018). Developing a secondary mathematics higher order thinking skills assessment (SMHOTS) instrument. *Turkish Online Journal of Design Art and Communication*, 8(SEPT), 1238–1246. doi: <https://doi.org/10.7456/1080sse/166>
- Ramdiah, S., Abidinsyah, A., Royani, M., & Husamah, H. (2019). Understanding, planning, and implementation of HOTS by senior high school biology teachers in Banjarmasin-Indonesia. *International Journal of Instruction*, 12(1), 425–440. doi: <https://doi.org/10.29333/iji.2019.12128a>
- Ramdiah, S., Abidinsyah, H., & Mayasari, R. (2018). Problem-based learning: Generates higher-order thinking skills of tenth graders in ecosystem concept. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 4(1), 29–34. doi: <https://doi.org/10.22219/jpbi.v4i1.5490>
- Ramesh, R., & Rao, U. R. (2015). Investigating the impact of in-class assignments on higher order thinking skills of students in engineering course. In *Proceedings - 2015 International Conference on Learning and Teaching in Computing and Engineering, LaTiCE 2015* (pp. 95–99). IEEE. doi: <https://doi.org/10.>

1109/LaTiCE.2015.37

- Razmawaty, M., & Othman, L. (2017). Authentic assessment in assessing higher order thinking skills. *International Journal Of Academic Research In Business And Social Sciences*, 7(2), 466–476. doi: <https://doi.org/10.6007/IJARBS/v7-i2/2021>
- Retnawati, H., Djidu, H., Apino, E., & Anazifa, R. D. (2017). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in The 21st Century*, 76(2), 215–230. Retrieved from <http://www.scientiasocialis.lt/pec/node/1121>
- Rofiah, E., Aminah, N. S., & Ekawati, E. Y. (2013). Penyusunan instrumen tes kemampuan berpikir tingkat tinggi Fisika pada siswa SMP. *Jurnal Pendidikan Fisika*, 1(2), 17–22. Retrieved from <https://jurnal.fkip.uns.ac.id/index.php/pfisika/article/view/2797/1913>
- Sa'adah, A., Sugianto, & Sutarman. (2014). *Pengembangan instrumen tes benar-salah untuk menilai kemampuan berpikir tingkat tinggi siswa pada materi dinamika rotasi dan kesetimbangan benda tegar*. Jurusan Fisika - Fakultas MIPA Universitas Negeri Malang. Retrieved from <http://jurnal-online.um.ac.id/data/artikel/artikel662F729E1D404633A7770C86049A5069.pdf>
- Setiawan, A., Malik, A., Suhandi, A., & Permasari, A. (2018). Effect of higher order thinking laboratory on the improvement of critical and creative thinking skills. In *IOP Conference Series: Materials Science and Engineering* (Vol. 306, pp. 1–7). IOP Publishing. doi: <https://doi.org/10.1088/1757-899X/306/1/012008>
- Silveira, M. B., Saldanha, R. P., Leite, J. C. de C., Silva, T. O. F. da, Silva, T., & Filippin, L. I. (2018). Construction and validation of content of one instrument to assess falls in the elderly. *Einstein (São Paulo)*, 16(2), 1–8. doi: <https://doi.org/10.1590/s1679-45082018ao4154>
- Steers, R. M. (1985). *Efektivitas organisasi* (1st ed.). Jakarta: Erlangga. Retrieved from <http://kin.perpusnas.go.id/DisplayData.aspx?pld=128744&pRegionCode=UN11MAR&pClientId=112>
- Sumarni, W., Supardi, K. I., & Widiarti, N. (2018). Development of assessment instruments to measure critical thinking skills. In *IOP Conference Series: Materials Science and Engineering* (Vol. 349, pp. 1–11). IOP Publishing. doi: <https://doi.org/10.1088/1757-899X/349/1/012066>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesment pendidikan*. (B. Trim, Ed.) (1st ed.). Cimahi: Trim Komunikata. Retrieved from <https://www.researchgate.net/publication/282673464>
- Thiagarajan, S., Semmel, D. G., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Bloomington, Indiana: Indiana Univ., Center for Innovation in Teaching the Handicapped. Retrieved from <https://files.eric.ed.gov/fulltext/ED090725.pdf>
- UNESCO. (2011). ICT in education. Retrieved May 23, 2019, from <https://en.unesco.org/themes/ict-education>
- Wicaksono, D. P., Kusmayadi, T. A., & Usodo, B. (2014). Pengembangan perangkat pembelajaran matematika berbahasa inggris berdasarkan teori kecerdasan majemuk (multiple intelligences) pada materi balok dan kubus untuk kelas VIII SMP. *Jurnal Elektronik Pembelajaran Matematika*, 2(5), 534–549. Retrieved from <https://jurnal.fkip.uns.ac.id/index.php/s2math/article/view/4378/3063>
- Williams, R. (2003). Terminology used in instrument accuracy (pp. 279–284). American school of gas measurement technology. Retrieved from <https://asgmt.com/wp-content/uploads/pdf-docs/2003/1/55.pdf>
- Williams, R. G., Klamen, D. A., & McGaghie, W. C. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine*, 15(4), 270–292. doi: [https://doi.org/10.1207/S15328015TLM1504\\_11](https://doi.org/10.1207/S15328015TLM1504_11)
- Yeager, D. S., & Lee Duckworth, A. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. doi: <https://doi.org/10.3102/0013189X15584327>
- Yonata, B., & Nasrudin, H. (2018). Laboratory activity worksheet to train high order thinking skill of student on surface chemistry lecture. In *Journal of Physics: Conference Series* (Vol. 947, pp. 1–7). IOP Publishing. <https://doi.org/10.1088/1742-6596/947/1/012027>