JTK
— JURNAL —
TADRIS KIMIYA

Development of High Order Thinking Skills (HOTS) Test Instruments on Thermochemistry Topics

Dwi Wida Novatania[1*] and Agus Kamaludin[1]
[1]Department of Chemistry Education, State Islamic University Sunan Kalijaga, Laksda Adisucipto St, Papringan, Caturtunggal, Depok, Sleman, DIY, 55281, Indonesia
*E-mail: widanovatania@gmail.com

_____

Abstract

This research aims to develop Higher Order Thinking Skill (HOTS)-based test instruments on thermochemistry in the form of a valid and reliable multiple-choice question. This research applied Research and Development (R&D) design by using the development model of Borg & Gall. The respondents of this research were 55 high school students which came from nine different schools. The research objects were multiple-choice HOTS questions numbering 20 questions. The research data was obtained by validating the thermochemistry HOTS questions by material and media expert lecturers. High school chemistry teacher as a reviewer who assesses the questions and tested on students. The results showed that the product quality assessment by material expert 92.00%; product quality assessment by media expert 92.00%; and quality assessment by reviewers (high school chemistry teachers) 94.89% in the Very Good category. The trial result data from the students were processed by using IBM SPSS Statistics 24 and it showed that nine questions were valid but not reliable.

_____

1. Introduction

Scientific literacy is one of the people's keys to becoming successful in the 21ˢᵗ century (Nofiana & Julianto, 2018). It can make people have the ability to understand what science is and its application; the ability to solve the problem by using scientific knowledge; and the ability to think critically (Holbrook & Rannikmae, 2009). Therefore, It is used as one of the international assessments in PISA (Program for International Students Assessment) (Purnomo & Dafik, 2015). PISA is an assessment institution that examines students' literacy skills in solving problems (Agustin, Wardono, & Kartono, 2013). Indonesia has been a member of PISA since 2001 (Pratiwi, 2019). However, until 2018, Indonesia's PISA results have not indicated any significant increase in study results. It even tends to decrease when it was compared to the PISA 2015. Indonesia's PISA 2015 for

reading, math, and science skills respectively shows an average score of 397, 386, and 403 while the PISA 2018 shows an average score of 371, 379, and 396 (OECD, 2019).

The government's effort to catch up on students' literacy skills is carried out by leveling up the National Examination (UN) which is directed to the Higher Order Thinking Skill (HOTS) model around 10-15% (BSNP, 2018). The HOTS question model is an instrument to measure students' higher-order thinking skills which use multiple concepts, link one information to another, and think critically (Widana, 2017: 3). The HOTS model is developed based on the levels of Bloom's taxonomy. The levels start from the C4 level (analyze); C5 (evaluate), and C6 (create) (Sofyan, 2019). However, according to an analysis of the UN results by using the PAMER UN application (Operation of Application for the Utilization Report of National Examination

Results), it shows that the average score of UN for chemistry subjects in 2019 was 50.29. This score belongs to a low category.

With the low average score of UN for chemistry subjects, it is necessary to increase the ability or competence of teachers in the teaching and learning process in the classroom (BSNP, 2018). Following Law No. 14 the Year 2005 in article 10, Teacher and Lecturer Law states that teachers must have four competencies, one of which is pedagogical competence. Pedagogical competence is the teachers' ability to be able to design the appropriate learning models, prepare learning materials, and make their questions (Zulkardi & Ilma, 2006). Therefore, teachers need to have the competence to make questions, especially context-based questions or questions related to real-life to improve students' abilities to work on HOTS questions (Mitari & Zulkardi, 2019). However, in reality, many teachers still use non-contextual assessment instruments in the learning process so that students are less able to develop their thinking processes and their arguments, as well as they, are less able to work on logical reasoning questions (Wardhani & Rumiyati, 2011).

The quality of the assessment instrument is an important thing because it can be a reference for teachers and schools in evaluating the achievement of student learning outcomes, especially higher-order thinking skills (Budiman & Jailani, 2014). A good assessment instrument is an instrument that meets the criteria of validity and reliability, as well as has practical value so that it can be used as a reference for the next learning process (Aji & Winarno, 2016).

The teacher as the subject supervisor needs to be able to make assessment instruments under procedures so that learning objectives can be achieved (Hanifah, 2019). However, due to the lack of information regarding HOTS questions, the teacher finds it difficult to develop HOTS-based assessment instruments. Eventually, students only receive Low Order Thinking Skills (LOTS) questions (Netri et al., 2018). This reality was strengthened based on interviews with chemistry teachers in Yogyakarta. The interviews showed that the availability of references of HOTS-based assessment instruments in chemistry learning was still very low.

Chemistry is a subject whose context is close to everyday life (contextual). However, many students have difficulties learning it. This is due to the lack of firm conceptualization and the lack of relevant exercises (Yakina et al., 2017). The difficulties of chemistry subjects experienced by students are also due to the abstract properties of chemistry, concerning reactions and calculations. Furthermore, it is also a relatively new subject (Ristiyani & Bahriah, 2016). Therefore, students' accuracy and persistence are required to learn and understand these chemical concepts. However, in reality, most students only apply the memorization method in studying chemistry, so that the memorized concepts are only temporary and have an impact on misunderstanding in developing basic concepts that are mastered to solve various kinds of problems (Marsita et al., 2010). One of the study materials of chemistry that requires concept analysis comprehension and problem-solving skills is thermochemistry (Siswanti et al., 2016). Thermochemistry is one of the 11th grade chemistry materials that studies the concept of enthalpy change in chemical reactions, exothermic reactions, endothermic reactions, enthalpy diagrams, thermochemical equations, types of enthalpy changes, calculation of enthalpy change of a reaction based on experiment, Hess's law, the data of enthalpy change of formation reaction, and the bond energy data (Zakiyah et al., 2018).

The development of the HOTS-based assessment instrument on thermochemistry is hopefully able to help students improve their scientific literacies, sharpen their higher-order thinking skills, and be able to solve logical reasoning questions on the exam. In addition, it is expected that the HOTS-based assessment instrument on thermochemistry can help the teacher develop good and correct HOTS-based questions.

2. Research Method

This research used a Research and Development (R&D) approach by applying Borg & Gall (1983) development model. Borg & Gall's development model used was only up to the sixth stage because to make a good instrument for use only need to sixth stage. The six stages are research and data or information collecting, planning, developing a preliminary form of product, preliminary field testing, main product revision, and main field testing. The resulted product is Higher Order Thinking Skill (HOTS)-based assessment instrument on thermochemistry.

The instrument, HOTS questions on thermochemistry were tested on respondents, 55 students from nine high schools in Yogyakarta. The trial process was carried out online via Google Forms whose link was sent to the students via WhatsApp. This procedure was applied because students were still accessing online learning by government regulation during the Covid-19 pandemic.

The HOTS-based assessment instrument on the developed study materials of thermochemistry was reviewed by four peer reviewers and validated by one material expert lecturer, one media expert lecturer, and six reviewers (high school chemistry teachers). The validated assessment instrument was tested on students for validity and reliability analysis by using the IBM SPSS Statistics 24 application.

2.1. Product Validation Data

The product validation data are product quality data from the material expert lecturer, media expert lecturer, and reviewers (high school chemistry teachers). The data obtained were the conversion of letters to numbers by using a Likert scale whose score is from 1 to 5. The Likert scale in Setyawati, Hastuti, and Widowati (2016) is presented in Table 1.

Table 1. Likert scale for Validation of Material Expert Lecturer, Media Expert Lecturer, and Reviewers (High School Chemistry Teachers)

Score	Value
1	Very Poor
2	Poor
3	Fair
4	Good
5	Very Good

Interpretation of validation value obtained used Table 1. Furthermore, the data obtained were calculated by using the formula for the total average score of each aspect by using the equation (Widoyoko, 2009):

$$\bar{x} = \frac{\sum X}{n} \quad (1)$$

Information:
X   = average score
$\sum X$ = total score
n   = number of evaluator

The average score of all product assessment aspects was converted to qualitative data based on ideal assessment criteria. It is shown in Table 2.

Table 2. Conversion of the Ideal Score into a Scale Value 5

Score Interval	Criteria
$X > \bar{X}_i + 1.8\ SB_i$	Very Good
$\bar{X}_i + 0.60\ SB_i < X \le \bar{X}_i + 1.8\ SB_i$	Good
$\bar{X}_i - 0.60\ SB_i < X \le \bar{X}_i + 0.60\ SB_i$	Fair
$\bar{X}_i - 1.8\ SB_i < X \le \bar{X}_i - 0.60\ SB_i$	Poor
$X \le \bar{X}_i - 1.8\ SB_i$	Very Poor

To find out the validity of the HOTS-based assessment instrument on thermochemistry, the ideal percentage of product quality for each aspect was calculated by using the equation:

$$\% = \frac{\text{The average score of each aspect}}{\text{The highest ideal score of each aspect}} \times 100\% \quad (2)$$

Meanwhile, the ideal percentage of product for all assessment aspects used the equation:

$$\% = \frac{\text{The average score of all aspect}}{\text{The highest ideal score of all aspect}} \times 100\% \qquad (3)$$

2.2. Validity Test

The validity test is the validity or accuracy level of a test that measures what it is supposed to measure. In another word, the validity test is conducted to show the level of accuracy of a test instrument in measuring the target to be measured (Arikunto, 2013).

The validity test was conducted by using the IBM SPSS Statistics 24 application. The test results were compared with the r_{table} of product-moment at the 95% confidence level for n = 55 (dF = 53), which is 0.2656. If $r_{count} \geq r_{table}$, the instrument or the items of the question have a significant correlation with the total score (to be declared valid). If $r_{count} < r_{table}$, the instrument does not have a significant correlation with the total score (to be declared invalid). The interpretation for the magnitude of the correlation coefficient can be seen in Table 3 as follows:

Table 3. Interpretation of Validity Value for Assessment Instrumen

Limitation	Category
0.80 < Rxy ≤ 1.00	Very High
0.60 < Rxy ≤ 0.80	High
0.40 < Rxy ≤ 0.60	Fair
0.20 < Rxy ≤ 0.40	Low
0.00 < Rxy ≤ 0.20	Very Low

The instrument used must be valid in research because a suitable device describes the data according to the actual situation. Therefore, the tool used must be accurate.

2.3. Reliability Test

According to Sugiyono (2010), the assessment instrument can be considered to be reliable (trustworthy) if the instrument is used repeatedly to measure the same object and it will produce the same data. The analysis of the reliability test is if the Cronbach's Alpha value is > 0.60, the assessment instrument has a high or reliable correlation. Whereas, if the Cronbach's Alpha value is < 0.60, the assessment instrument has a low or unreliable

correlation (Basuki & Hariyanto, 2014). The interpretation of the reliability value which can be seen from Cronbach's Alpha value refers to Guildford's opinion in Jihad & Haris (2009) which can be seen in Table 4 below:

Table 4. Interpretation of Reliability Value for Assessment Instrument

r_{11} Value	Criteria
0.00 – 0.199	Very Weak
0.20 – 0.399	Weak
0.40 – 0.599	Fair
0.60 – 0.799	Strong
0.80 – 1.000	Very Strong

In addition to the test, an instrument must be valid, then another criterion that an instrument must meet is reliability. Reliable shows the data obtained can be trusted. Therefore, a good tool must be valid and reliable.

3. Result and Discussion

The developed product from the development research that has been carried out is a HOTS-based assessment instrument on thermochemistry for 11th grade. It has been validated by the material expert lecturer, media expert lecturer, and reviewers (high school chemistry teachers), as well as it has been tested on students to determine the validity and reliability. The development model used is the Borg and Gall model which is up to the sixth stage. The description of the research results at each of the Borg and Gall development stages that have been carried out, the validation results from experts, and the results of the validity and reliability tests are presented as follows.

3.1. Research and Data or Information Collection

This stage aims to find out an overview of the learning process in schools and the problems which emerge in it. This stage was carried out through needs analysis and surveys in schools. The results obtained at the stage of needs analysis were that students rarely worked on HOTS-based questions. Based on the literature studies, the references of HOTS-based questions in schools were still limited and the

research that developed HOTS-based questions was still little. According to the interviews and surveys in schools, teachers were still using questions with the Low Order Thinking Skill (LOTS) level to measure students' abilities.

3.2. Planning
At this stage, the research and development plan was drawn up. From the stage of planning, it is obtained analysis results as follows:

3.2.1. The Required Ability
The required ability in this research was the ability to make a HOTS-based assessment instrument in the form of a multiple-choice question. An analysis of the module of HOTS question development was conducted. This analysis aims to find out the characteristics of the HOTS-based question.

3.2.2. Analysis of 11th Grade Chemistry Syllabus on Thermochemistry
This analysis was conducted to determine the chemistry's basic competence for 11th grade. The basic competencies and the thermochemistry syllabus can be seen in Table 5.

Table 5. Basic Competence of Thermochemistry

No.	Basic Competence
3.4	Distinguish exothermic and endothermic reactions based on experimental results and energy level diagram
3.5	Determine the ΔH of a reaction based on Hess's law, the standard enthalpy change of formation data, and bond energy data

Based on the basic competencies above, it is known that thermochemistry is one of the study materials which requires students' abilities to be able to understand and analyze the phenomena and events in everyday life as well as to solve the problems. Those are used as the research basis to develop the product, in the form of a HOTS-based assessment instrument. Therefore, the learning material used in the development of the HOTS-based assessment instrument is thermochemistry.

3.2.3. Determination of Indicators and Question Guidelines
This stage was performed by analyzing the basic competence. Two basic competencies, mentioned in Table 5 were elaborated into indicators that were adjusted to the HOTS question criteria, these were C4, C5, and C6 from Bloom's Taxonomy at the cognitive level. The indicators were used as a reference for the preliminary design of the HOTS assessment instrument on thermochemistry.

3.2.4. Design of Research Steps
This stage began by determining the research and development model to be used. This research used the Borg and Gall development model which consisted of six steps. In the research steps, there was a product validity test by the material expert lecturer, media expert lecturer, and reviewers (high school chemistry teachers), so that research instrument was required. The assessment instrument produced was then validated by an instrument expert lecturer.

3.3. Developing Preliminary Form of Product
At this stage, the preliminary product in the form of a multiple-choice question with a total of 20 questions that are equipped with the discussion was made up. The initial section of the product also included basic competencies and question guidelines. These questions are self-made and modified questions from relevant sources. The result of the preliminary product was then reviewed by four peer reviewers and validated by a material expert lecturer and media expert lecturer by using a five-point scale Likert questionnaire.

3.4. Preliminary Field Testing
This stage was performed by the research subjects, chemistry teachers of 11th grade. It was under the condition that it came from one to three schools and three to 12 trial subjects (teachers). In this research, a preliminary test was carried out by testing the quality of the assessment instrument or the developed product on six high school chemistry teachers by using a five-point scale Likert questionnaire. In addition to assessing the quality of HOTS questions, the teachers also provided inputs.

3.5. Main Product Revision

The revision result of this stage was the final result of the developed HOTS-based assessment instrument. The following is an example of the developed HOTS question which is equipped with the discussion and elaboration of the HOTS question characteristics.

3.5.1. Sample of HOTS Question

A cold pack is an item that is often used to relieve pain from injury. The inside of the pack contains ammonium nitrate powder and the outer part that is fragile contains water. The mass of ammonium nitrate in the pack is 120 grams and the 500 mL of water is on the outside. If the pack is massaged, the plastic pack filled with water will burst and by a little shock, the ammonium nitrate will dissolve in the water. When ammonium nitrate mixes with water, the temperature of the solution drops. The coldness is what is used to wipe the injured part of the body. If the heat absorbed per mole is 26 kJ, then how much does the solution temperature drop according to your opinion? (Ar N = 14, Ar H = 1, Ar O = 16, specific heat = 4.2 $Jg^{-1}K^{-1}$).

A. 1.24 °C   D. 18.6 °C
B. 1.86 °C   E. 23.2 °C

The answer for the question is:

Mass NH_4NO_3 = 120 grams

Mole NH_4NO_3 = $\dfrac{Mass}{Molecular\ mass}$

Mole NH_4NO_3 = $\dfrac{120\ grams}{80\ grams/mole}$

Mole NH_4NO_3 = 1.5 mole

The total of heat absorbed = 1,5 mole x 26 kJ/mole = 39 kJ = 39000 J

Q   = m x c x ΔT
39000 J   = 500 grams x 4.2 J/gramK x ΔT
ΔT   = 18.6 °C

Thus, the temperature change that occurs is 18.6 °C.

Figure 1. Sample of HOTS Question

3.5.2. The Elaboration of HOTS Question Characteristics

The question in 3.5.1 can be considered as a HOTS-based question because it has a stimulus in the form of images and contextual reading text. The question is at the C4 cognitive level (analysis) because to work on the question, students need to be able to analyze how the cold pack works. Furthermore, multi-concept is applied to work on the question because students need to be able to determine the compound formula described in the question to determine the Mr value of the compound. Compound

nomenclature and Mr determination are part of study materials for 10th grade, namely compound nomenclature and stoichiometry.

3.6. Main Field Testing

This stage was carried out by the students who had received thermochemistry materials at school. The questions which were revised in the previous stage were distributed via Google Form.

The collected data from the main field testing via Google Form were collected from 55 students. These students came from nine different schools.

3.7. Validity Result, Validity Test, and Reliability Test of HOTS-Based Assessment Instrument

The validation of the HOTS-based assessment instrument on thermochemistry was carried out by one media expert lecturer, one material expert lecturer, and six reviewers (high school chemistry teacher) by using a five-point scale Likert questionnaire. The validity and reliability test of a HOTS-based assessment instrument was carried out by using the IBM SPSS Statistics 24 application.

3.7.1. The Validation of Hots-Based Assessment Instrument

The validation stage of a HOTS-based assessment instrument contained inputs and assessments of the product in terms of quality. The inputs obtained from the material expert lecturer were the writing improvement of chemical reaction in glucose fermentation and the lack of writing (+) or (-) signs on the questions.

The quality assessment of HOTS-based assessment instrument on thermochemistry in terms of study materials consists of four aspects including content feasibility, characteristics of HOTS question, the role of HOTS question, and language which are elaborated into ten indicators. Data of quality assessment results by material expert lecturer can be seen in Table 6.

Table 6. Quality Assessment Data of HOTS Based Assessment Instrument on Thermochemistry by Material Expert

Assess-ment Aspects	Σ Score	Σ Ideal Max. Score	% Ideal	Category
Content	14	15	93.33%	VG
Character-istics of HOTS Question	9	10	90%	VG
Role of HOTS Question	5	5	100%	VG
Language	18	20	90%	VG
Total	46	50	92 %	VG

The assessment result by the material expert lecturer overall has an average score of 46 with a maximum score of 50 and an ideal percentage of 92.00%. Based on the assessment criteria by the material expert lecturer, the average score of the assessment result is in the range of X > 42, so that the HOTS-based assessment instrument on thermochemistry is categorized as the Very Good (VG) quality.

The quality assessment of HOTS-based assessment instrument on thermochemistry in terms of media consists of two aspects including graphics and construction which are elaborated into five indicators. Data of quality assessment results by media expert lecturers can be seen in Table 7.

Table 7. Quality Assessment Data of HOTS-Based Assessment Instrument on Thermochemistry by Media Expert

Assess-ment Aspects	Σ Score	Σ Ideal Max. Score	% Ideal	Category
Graphics	9	10	90%	VG
Constru-ction	14	15	93.3 %	VG
Total	23	25	92%	VG

The assessment result by the media expert lecturer relatively has an average score of 23 with an ideal maximum score of 25 and an ideal percentage of 92.00%. Based on the

assessment criteria by the media expert lecturer, the average score is in the range of X > 21, so the HOTS-based assessment instrument on thermochemistry is categorized as the Very Good (VG).

The quality assessment of the HOTS-based assessment instrument on thermochemistry from six reviewers (high school chemistry teachers) consists of six aspects which are elaborated into 15 indicators relatively. The quality assessment data by six reviewers can be seen in Table 8.

Table 8. Quality Assessment Data of HOTS-Based Assessment Instrument on Thermochemistry by Reviewers

Assessment Aspects	Σ Score	Σ Ideal Max. Score	Average Score	% Ideal Category
Content	89	90	14.83	98.87% (VG)
Characteristics of HOTS Question	59	60	9.83	98.30% (VG)
Role of HOTS Question	29	30	4.83	96.60% (VG)
Language	112	120	18.67	93.35% (VG)
Graphics	56	60	9.34	93.40% (VG)
Construction	82	90	13.67	91.13% (VG)
Total	427	450	71.17	94.89% (VG)

The assessment result by the six reviewers overall has an average score of 71.17 with a maximum score of 75 and an ideal percentage of 94.89%. Based on the assessment criteria by six reviewers, the average score is in the range of X > 63, so that the HOTS-based assessment instrument on thermochemistry gets the Very Good (VG) category.

3.7.2. Validity Test of HOTS-Based Assessment Instrumen

The stage of validity test used the data of student's work results. The students' answers were corrected by using Microsoft Excel and the assessment guidelines are if each number's answer is true, then gets 1 point; and if it is false, then gets 0. Furthermore, the total score of each student was calculated relatively. The score data of each number and the total score will be used to test the validity by using the IBM SPSS Statistics 24 application. The data obtained were 55 students.

Based on the results of the analysis on the developed HOTS questions in the form of multiple choices, there were nine valid questions and eleven invalid questions. The validity test on HOTS questions about thermochemistry can be seen in Table 9.

Table 9. Validity Test of HOTS Question

Validity Index	Question Number	Number of Questions	Percent-age
$r_{count} > 0,2656$ (valid)	1, 2, 3, 4, 5, 7, 11, 16, 20	9	45%
$r_{count} < 0,2656$ (invalid)	6, 8, 9, 10, 12, 13, 14, 15, 17, 18, 19	11	55%

Based on the data in Table 9, it can be concluded that nine questions consisting of numbers 1,2,3,4,5,7,11,16, and 20 are considered valid because they have r_{count} values that are greater than the r_{table} ($r_{count} > 0.2656$). Meanwhile, the other numbers, namely numbers 6, 8, 9, 10, 12, 13, 14, 15, 17, 18, and 19 are considered invalid because the r_{count} values are less than the r_{table} ($r_{count} < 0.2656$).

The following is a diagram of the percentage distribution of HOTS questions based on the validity in Figure 2.
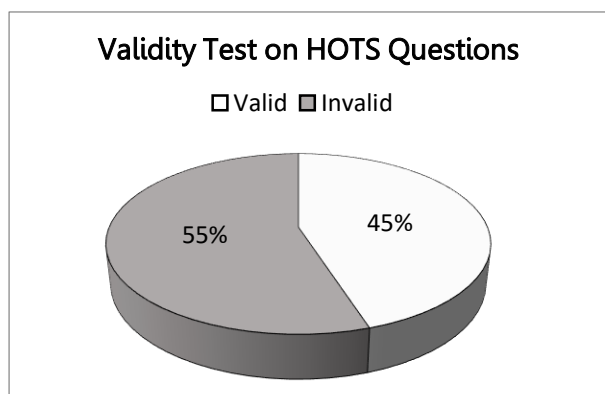
Figure 2. Diagram of Percentage Distribution of HOTS Questions Based On Validity

3.7.3. Reliability Test on HOTS-Based Assessment Instrument

The reliability test was carried out only on valid questions so that only nine questions were tested for reliability, those are numbers 1, 2, 3, 4, 5, 7, 11, 16 and 20. Reliability test of HOTS-based assessment instrument on Thermochemistry was conducted by using the IBM SPSS Statistics 24 application. The results of the SPSS output showed that Cronbach's Alpha value was 0.453. Therefore, the questions were considered unreliable or had a low correlation when they were viewed from the Cronbach's Alpha value because they were less than 0.6. The results of the SPSS output for the reliability test can be seen in table 10.

Table 10. The Result of Reliability Test

Reliability Statistics	
Cronbach's Alpha	N of Items
0.453	9

If the interpretation of the reliability value in Table 4 is viewed, the category of reliability value has a medium reliability category or it can be considered unreliable. This is due to students get difficulties working on the HOTS-based questions. Based on the results of interviews with several students, they revealed that the learning process at school did not develop HOTS-based learning. The teachers only provided the LOTS exercises. In addition, the shift of students' learning methods that was initially offline (at school) to online (at home) due to the Covid-19 pandemic also made students' difficult to understand the subject matters, especially thermochemistry.

This is due to the online learning makes teachers less effective in teaching thermochemistry.

4. Conclusion

Based on the research results and discussion that has been carried out, the HOTS-based assessment instrument on thermochemistry for 11th grade is found. Based on the validity test on the material expert lecturer's assessment, the product gains an ideal percentage of 92.00% in the Very Good (VG) category. According to the assessment by the material expert lecturer, the product obtains the ideal percentage of 92.00% in the Very Good (VG) category. In accordance with the assessment of the reviewers (high school chemistry teacher), the product gets the ideal percentage of 94.89% in the Very Good (VG) category. The results of the validity test shows that nine questions are considered valid, namely numbers 1, 2, 3, 4, 5, 7, 11, 16, and 20 with sufficient validity categories; numbers 1, 7, and 16 with sufficient validity categories; and number 2, 3, 4, 5, 11, and 20 with the low validity categories. Furthermore, the rest, namely numbers 6, 8, 9, 10, 12, 13, 14, 15, 17, 18, and 20 are invalid. The reliability test is performed by using only valid question data and it shows that the questions are not reliable or they are in the medium reliability category. This is due to the questions having a Cronbach's Alpha value < 0.6.

References

Agustin, E. N., Wardono, & Kartono. (2013). Pembelajaran matematika realistik berorientasi *programme for international student assessmen* (PISA) untuk meningkatkan kemampuan pemecahan masalah siswa SMA. *Proceeding of Seminar Nasional Evaluasi Pendidikan Tahun 2013, Semarang, 125-137.*

Aji, B. S., & Winarno, M.E. (2016). Pengembangan instrumen penilaian pengetahuan mata pelajaran pendidikan jasmani olahraga dan kesehatan (PJOK) kelas VIII semester

gasal. *Jurnal Pendidikan*, *1*(7), 1449-1463. Diambil dari http://dx.doi.org/10.17977/jp.v1i7.6594

Arikunto, Suharsimi. (2013). *Dasar-dasar evaluasi pendidikan (2nd ed.)*. Jakarta: Bumi Aksara.

Basuki & Hariyanto. (2014). *Asesmen pembelajaran*. Bandung: PT Remaja Rosdakarya.

Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction*. New York: Longman Publishing.

BSNP. (2018). Penerapan soal model penalaran dalam ujian nasional: Apa dan mengapa?. *Buletin BSNP Media Komunikasi dan Dialog Standar Pendidikan*, *XIII*(2), 1-27. Retrieved from https://bsnp-indonesia.org/wp-content/uploads/2018/09/Buletin-Edisi-2-2018-Ok.pdf

Budiman, A., & Jailani. (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. https://doi.org/10.21831/jrpm.v1i2.2671

Hanifah, Nurdinah. (2019). Pengembangan instrumen penilaian *higher order thinking skill* (HOTS) di sekolah dasar. *Conference Series Journal*, *1*(1), 1-8. Retrieved from https://ejournal.upi.edu/index.php/crecs/article/view/14286/pdf

Holbrook, J., & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental & Science Education*, *4*(3), 275-288. Retrieved from https://eric.ed.gov/?id=EJ884397

Marsita, R. A., Priatmoko, S., & Kusuma, E. (2010). Analisis kesulitan belajar kimia siswa SMA dalam memahami materi larutan penyangga dengan menggunakan *two-tier multiple choice diagnostic instrument*. *Jurnal Inovasi Pendidikan Kimia*, *4*(1), 512-520. https://doi.org/10.15294/jipk.v4i1.1308

Mitari, Osi & Zulkardi. (2019). Pengembangan soal matematika model PISA dengan konteks wisata Jakabaring Sport City. In *Seminar Nasional Pendidikan Matematika Ahmad Dahlan*, *Palembang*, *172-178*.

Netri, N., Holiwarni, B., & Abdullah. (2018). Development of test instruments based higher order thinking skills (HOTS) on chemical equilibrium at second grade in senior high school. *JOM FKIP*, *5*(2), 1-11. Retrieved from https://jom.unri.ac.id/index.php/JOMFKIP/article/view/20515/19843

Nofiana, M., & Julianto, Teguh. (2018). Upaya peningkatan literasi sains siswa melalui pembelajaran berbasis keunggulan lokal. *BIOSFER Jurnal Tadris Pendidikan Biologi*, *9*(1), 24-35. https://doi.org/10.24042/biosf.v9i1.2876

OECD. (2019). *Programme for international student assessment (PISA) results from PISA 2018*. Retrieved from https://www.oecd.org/pisa/publications/PISA2018_CN_IDN.pdf

Pratiwi, Indah. (2019). Efek program PISA terhadap kurikulum di Indonesia. *Jurnal Pendidikan dan Kebudayaan*, *4*(1), 51-71. https://doi.org/10.24832/jpnk.v4i1.1157

Purnomo, S., & Dafik. (2015). Analisis respon siswa terhadap soal PISA konten *shape and space* dengan rasch model. In *Seminar Nasional Matematika dan Pendidikan Matematika, Jember, 1155-1160.*

Ristiyani, Erika & Bahriah, E. S. (2016). Analisis kesulitan belajar kimia siswa di SMAN X kota Tangerang Selatan. *Jurnal Penelitian dan Pembelajaran, 2*(1), 18-29. https://doi.org/10.30870/jppi.v2i1.431

Setyawati, T. S., Hastuti, P. W., & Widowati, Asri. (2016). Pengembangan komik IPA berbasis paikem untuk menanamkan sikap peduli lingkungan dan mengembnagkan kreativitas siswa. *Jurnal Pendidikan Ilmu Pengetahuan Alam, 5*(9), 24-31. 10.21831/jpms.v6i2.23943

Siswanti, S., Saputro, S., & Utomo, Suryadi Budi. (2016). Pengembangan modul termokimia berbasis *problem solving* untuk siswa SMA/MA kelas XI semester 1 kurikulum 2013. *Jurnal Inkuiri, 5*(1), 28-36. https://doi.org/10.20961/inkuiri.v5i1.9500

Sofyan, Fuaddilah Ali. (2019). Implementasi HOTS pada kurikulum 2013. *Jurnal Inventa, III*(1), 1-17. https://doi.org/10.36456/inventa.3.1.a1803

Sugiyono. (2010). *Metode penelitian kuantitatif kualitatif dan R&D.* Bandung: Alfabeta.

Wardhani, S., & Rumiati. (2011). *Instrumen penilaian hasil belajar matematika SMP: Belajar dari PISA dan TIMSS.* Yogyakarta: Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan (PPPPTK) Matematika. Retrieved from http://repositori.kemdikbud.go.id/15137/

Widana, I Wayan. (2017). *Modul penyusunan soal higher order thinking skills (HOTS).* Jakarta: Direktorat Pembinaan SMA, Direktorat Jenderal Pendidikan Dasar dan Menengah, Departemen Pendidikan dan Kebudayaan. Retrieved from http://suwatno.staf.upi.edu/files/2017/11/10.-Modul-Penyusunan-Soal-HOTS-Tahun-2017.-Modul-Penyusunan-Soal-HOTS-Tahun-2017.-Modul-Penyusunan-Soal-HOTS-Tahun-2017_1.pdf

Widoyoko, S. E. P. (2009). *Evaluasi program pembelajaran.* Yogyakarta: Pustaka Pelajar.

Yakina, Kurniati, T., & Fadhilah, Raudhatul. (2017). Analisis kesulitan belajar siswa pada mata pelajaran kimia kelas X di SMA Negeri 1 Sungai Ambawang. *Ar-Razi Jurnal Ilmiah, 5*(2), 287-297. http://dx.doi.org/10.29406/arz.v5i2.641

Zakiyah, Ibnu, S., & Subandi. (2018). Analisis dampak kesulitan siswa pada materi stoikiometri terhadap hasil belajar termokimia. *EduChemia, 3*(1), 119-134. https://doi.org/10.30870/educhemia.v3i1.1784

Zulkardi, & Ilma, Ratu. (2006). Mendesain sendiri soal kontekstual matematika. *Proceeding of KNMI13, Semarang, 1-7.* Retrieved from http://repository.unsri.ac.id/6350/1/mendesain_sendiri_soal_kontekstual.pdf