JPPIPA 9(11) (2023)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of HOTS (High Order Thinking Skill) Question Instrument on Thermochemistry Material

Mizzan Ayubi[1*], Vegha Dwi Arthamena[1], Nurfina Aznam[2]

[1] Magister of Chemistry Education, Faculty of Mathematics and Natural Science, Yogyakarta State University, Yogyakarta, Indonesia.
[2] Department of Chemistry Education, Yogyakarta State University, Yogyakarta, Indonesia.

Received: September 1, 2023
Revised: October 7, 2023
Accepted: November 25, 2023
Published: November 30, 2023

Corresponding Author:
Mizzan Ayubi
mizzan0019fmipa.2022@student.uny.ac.id

DOI: 10.29303/jppipa.v9i11.5162

Abstract: This study aims to make High Order Thinking Skill (HOTS) items in the form of multiple-choice types of thermochemistry material for SMA/MA. The development of the HOTS question instrument was carried out to produce valid and reliable question items. The HOTS question instrument development method uses the Mc development model. Intire. Data results in development research can be carried out using rater validation and test results. Based on the results of the rater, it can be stated that the items are feasible to test by looking at the average value of validity, which is 0.727 so that they are included in the high category by looking at aspects, namely material, construction and language aspects. Based on the calculation of the results by producing 15 questions with thermochemical material for SMA/MA which are valid with the criteria for the average level of difficulty, namely (0.584), the criteria for discriminating power of the average questions, namely (0.576) with the criteria for the average validity test, namely (0.493) reliability criteria with an average of (0.686). The conclusion obtained from the data is that the HOTS question instrument can be declared suitable for student use.

Keywords: HOTS; Instrument assessment; Mc.Intire model

Introduction

PISA (Program for International Student Assessment) in 2018 released their latest report data regarding the literacy level of each member country, one of which is Indonesia. Based on the report, the literacy level of Indonesian students is ranked 74th out of 79 member countries (OECD, 2019; Sidiq et al., 2021). 21st century skills are the main target in the curricula of educational institutions around the world because good skills in the 21st century are one of the solutions to answer the challenges of the industrial revolution era 4.0 (Hujjatusnaini et al., 2022).

According to sani (2019), individuals who have the ability to think High level or HOTS is able to process and implement new information or knowledge with the aim of finding useful solutions to new problems. When faced with a problem that cannot be solved using a conventional approach or a complex problem, the individual is expected to be able to solve it by using higher-order thinking skills. very individual has different problems, so the solutions or solutions found will also vary. Sani (2019) also stated that HOTS ability to solve problems is important for students to face challenges in everyday life that require a high level of thinking.

Budiarta et al. (2018) explains that HOTS or higher order thinking skills are abilities that include decomposing material, being critical and making solutions to these problems so that it can be said to be complex thinking skills. Similar to the opinion, Thomas et al. (2009) that HOTS or high-level thinking ability is a thinking ability that has ties between facts and problems that occur. the solution to the problems that occur are not only in the form of memorizing, counting and remembering. However, someone is required to create relationships and conclusions on a problem. The same opinion was expressed by Annuuru et al. (2017) providing an explanation that HOTS is a combination of existing facts and an idea from the analyzing stage to the

How to Cite:
Ayubi, M., Arthamena, V. D., & Aznam, N. (2023). Development of HOTS (High Order Thinking Skill) Question Instrument on Thermochemistry Material. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 9945–9954. https://doi.org/10.29303/jppipa.v9i11.5162

creating stage by assessing a fact that has been learned or the creating stage that has been studied

In the law of the republic of Indonesia number 20 of 2003 concerning the National education system, learning is a term that is often used in society. Article 1 point 20 in the law explains that learning is a process of interaction between students, educators, and learning resources in a learning environment. this interaction involves planning that begins with a lesson plan and ends with an evaluation. these three components become important in learning. learning objectives have three aspects, namely knowledge, attitudes, and skills. The success of learning is determined through a process of evaluation or assessment of student learning outcomes.

Critical thinking ability or HOTS is a much-needed basic ability that should be mastered by a learner such as the ability to read, memorize and count (Rositawati, 2019). As time goes on, students are required to have higher order thinking skills or HOTS and must accept technology that is growing rapidly (Maskur et al., 2020). High-level thinking ability or HOTS is an ability that makes students think logically and rationally. Then from this ability students are able to perform several skills such as analyzing, identifying questions, evaluating ideas and being able to make conclusions (R. Dewi et al., 2019). The development of higher order thinking skills or HOTS in chemistry learning has not been fully supported under conditions in the field. Supporting devices or facilities during chemistry learning have not fully facilitated students (Lestari et al., 2019).

Program developed on the basis of learning and by the Ministry of Education and Culture (Kemdikbud). The program was first registered in 2018 and its focus is to teach students the use of higher order thinking skills, also known as HOTS Higher Order Thinking Skills (Dwijayanti, 2021). This concept of higher order thinking is divided into two levels. The first level is low-level thinking or LOT (Lower Order Thinking), and the second level is higher-level thinking capacity (Kurniawan et al., 2022). In the learning process, HOTs assessment can be used as a guide so that students can change or develop the knowledge they already have, or even create something new (Rozi et al., 2019). The main purpose of training students using the HOTs approach is so that they can develop better analytical and critical thinking skills (Inayati, 2020). Based on previous research, HOTs are an effective tool to promote higher-level thinking through the stages of analysis (c-4), evaluation (c-5), and creativity (c-6) (Lamhatin et al., 2022). By encouraging students to think critically, logically, reflectively, constructively, and creatively, the use of HOTs is expected to increase student productivity in the learning process (Agustina et al., 2019). The HOTS approach also emphasizes the importance of education that can motivate learners to be creative and use various active student-centered learning strategies. This approach gives them the motivation and ability to participate in discussions, ask questions, and communicate more effectively (Fanani, 2018).

This research was carried out because many students or students did not fully have critical thinking skills and chemical literacy in the current era. There are still a lot of them and this is proven in several data such as PISA and OECD at the world level. The instrument function itself makes students have these skills so that students in villages have the same abilities as students in cities. From some of the research that has been done, we can draw the conclusion that the development of HOTS instruments is able to make students' ability to understand to perform tasks based on the reality that exists in the current curriculum by connecting existing facts. In this regard, the researcher is developing HOTS-based question instruments. The development was carried out with the aim of making HOTS instruments with thermochemical material seen valid and reliable then knowing the differentiating power of questions and the level of difficulty of chemistry questions with HOTS-based that have been developed on thermochemical material. Therefore, the researchers developed HOTS instruments which became the basis of this study "Development of HOTS (High Order Thinking Skill) Instruments on Thermochemistry".

Method

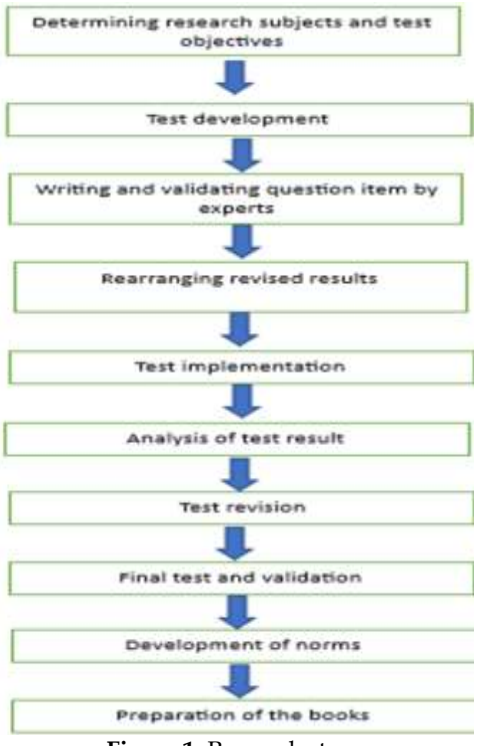

Figure 1. Research stage

The development model applied in this research is the Mc development model Intire. This model consists of 10 stages which include 1) determining research subjects and test objectives, 2) test development, 3) writing and validating question items by experts, 4) rearranging revised results, 5) test implementation, 6) analysis of test results, 7) test revision, 8) final test and validation, 9) development of norms, and 10) preparation of test books.

Determining the Research Subject and Test Objectives

The research was conducted using the type of development research that aims to develop existing question products or improve question items. The research subjects were class XII students at UPT SMAN 8 Musi Rawas with a total of 67 people and class XII UPT SMAN 1 Musi Rawas with a total of 35 people. The research was conducted during the even school year for high school / MA level students in the 2021/2022 school year. To determine the validity of the question product, validation was carried out by 6 experts by assessing in terms of the aspects of the question construct, the material on the question and the language used. Validation was carried out 2 times before revision and after revision. In expert validation in the form of qualitative data with the results of suggestions or comments on the instruments that have been made.

Test Development Stage

At the test development stage is to design the overall test process. Test development can be taken from chemistry subject materials that have been studied by students based on the curriculum used, namely the 2013 Curriculum. From the materials that have been studied, identify the material and the desired Basic Competencies (KD). Determination of Higher Order Thinking Skill (HOTS) items must use at least KKO at the C4 level with one of the words "analyze". After obtaining the appropriate KD, then the KD is identified or analyzed so that it can be developed again so as to get the appropriate GPA and question indicators. Determining the GPA of the selected KD determines the selection of material so that the material can be used in the form of questions.

Writing Test Items

After the test development stage, then writing question items can be done in accordance with the question indicators and GPA that have been made previously. The question items that have been made are equipped with appropriate answers and scoring guidelines contained in the questions. Then at this stage prototype 1 is produced. After completion in making the question items, the question is validated by the expert (Rater). Before validation by raters, namely doing self-evaluation of their own assessment of the question items. Each rater conducts an assessment in accordance with the scoring guidelines that have been prepared. Raters assess several things such as items that have been made, GPA and indicators that have been developed. There are 6 raters for item development assessment who assess the material, construction and language aspects. After the assessment by the raters, revisions are made according to comments or suggestions.

Reconstruction of the Revised Results

The results of the comments from the 6 raters were carried out to qualitatively improve the question items by looking at the comments and suggestions by the raters. Then the results of the revision of the question items that have been revised can be used as prototype 2 which can then be field tested on research subjects.

Test Implementation Stage

In the test implementation stage, HOTS questions with Thermochemistry material that has been developed are tested on the research subjects. The stage carried out was to carry out the implementation from February to March 2022 which was attended by 35 students. In the first stage of the test, it was conducted at UPT SMAN 8 Musi Rawas from class XII MIA 3. The test conducted at the school was conducted for 90 minutes. At the time the test conducted at UPT SMAN 10 Musi Rawas was conducted in one day. Then in working on HOTS items, students are allowed to open books to make it easier to work on the questions.

Data collection using the validity test of Hots Instruments on Students of Class XII UPT SMAN 8 Musi Rawas and Xii UPT SMAN 1 Musi Rawas. The field test was conducted to see whether the instrument was categorized as suitable for use or not suitable for use with students. There are data analysis techniques used.

Analysis of Test Results

The results that have been obtained from the answers of students to make quantitative calculations. Some of the things that are analyzed are validity, reliability, level of difficulty and discriminatory power.

Test Revision

After carrying out quantitative calculations, it can be seen from the analysis of the questions that fall into the good and bad categories as well as the questions that have been revised. Criteria for decision making on questions can be seen from the aspects of validity and reliability as well as the level of difficulty of the questions and distinguishing power and can be seen from several comments from students.

Final Test and Validation

Then carry out the test again with a different sample and population over a different time period in order to obtain valid question items. This validity stage is carried out by correlating the current test results with the results that were previously carried out. Test validity is used to prove that the question items are reliable. Test questions are declared reliable if the scores obtained are consistent after carrying out several tests on the same subject but in different time periods

Development of Norms

At this stage, namely the development stage of reference norms. At this stage, valid item questions have been obtained and the determination of test scores can be based on the minimum completeness criteria (KKM). Determining the score for the question items is the limit score that student must obtain. The score limit can be used to find out which students can be categorized as passing and not passing. The KKM for the questions set is 70.

Preparing Test Bookkeeping

This stage is the last stage to be carried out. The process at this stage completes the test manually by compiling a test use book by producing items based on High Order Thinking Skills (HOTS). The HOTS questions collection book contains an explanation of the background of making the test, history of the test development process, results of validation studies, instructions for administering the test, how to score the test, and can provide information about interpreting individual student scores.

Content and Construct Validity

Testing of items usually uses a test validity test to determine the quality of the items (Surapranata, 2009). According to Arikunto (2013), if it is said to be valid if a test that is carried out can measure something that is measured. The word valid itself can be said to be valid which means that the validity of an instrument so that it cannot be doubted. This study uses content validity and construct validity. The validity of a test can be found using a formula, one of which is the product moment formula (Surapranata, 2009), namely as equation 1.

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (1)$$

Description:
rxy = validity
∑X = the total score of all students on the question
∑Y = total score of all students on the test
X = the score of each student on the question
Y = total score of each student
N = number of students

The criteria for product moment correlation analysis results can be seen as table 1.

Table 1. Criteria for Validity Analysis Results

Correlation Coefficient	Description
0.80 – 1.00	Very High
0.60 – 0.80	High
0.40 – 0.60	Moderate
0.20 – 0.40	Low
0.00 – 0.20	Very Low

Criterion Validity

Testing of items usually uses a test validity test to determine the quality of the items (Surapranata, 2009). According to Arikunto (2013), if it is said to be valid if a test that is carried out can measure something that is measured. The word valid itself can be said to be valid which means that the validity of an instrument so that it cannot be doubted. This study uses content validity and construct validity. The validity of a test can be found using a formula, one of which is the product moment formula (Surapranata, 2009), namely as equation 2.

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (2)$$

Description:
rxy = validity
∑X = the total score of all students on the question
∑Y = total score of all students on the test
X = the score of each student on the question
Y = total score of each student
N = number of students

The criteria for product moment correlation analysis results can be seen as table 2.

Table 2. Criteria for Validity Analysis Results

Correlation Coefficient	Description
0.80 – 1.00	Very High
0.60 – 0.80	High
0.40 – 0.60	Moderate
0.20 – 0.40	Low
0.00 – 0.20	Very Low

Reliability

Reliability is the level of consistency of an instrument. If a test has a high or consistent reliability value when measuring concepts on a material to be measured then the research instrument has reliability in a good category (Suwarto, 2013). Reliability can be interpreted as trustworthiness. Reliability is related to fixity and consistency. To calculate the reliability of the description form test can be done by using the Cronbach-Alpha formula, namely as equation 3.

$$r_{11} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} s^2 i}{s_t^2}\right) \quad (3)$$

Description:
r11 = reliability coefficient
n = number of items
s2i = variance of i-th question score
s2t = total score variance
The correlation value (r) obtained is then consulted with the correlation coefficient table value so that it is like table 3.

Table 3. Reliability Qualifications

Correlation Coefficient	Qualification
0.91 – 1.00	Very High
0.71 – 0.90	High
0.41 – 0.70	Fair
0.21 – 0.40	Low
Negative – 0.20	Very Low

Difficulty Level

The item has a level of difficulty which is one of the indicators so that the item is included in the difficult, medium and easy categories. The items developed can be categorized as good items if the items are not too difficult or not too easy. In other words, the level of difficulty of the test is moderate or sufficient. The number that shows the difficulty and ease of a problem is called the difficulty index. The higher the difficulty index of the question, the easier the question. The equation used to determine the level of difficulty with the proportion of correct answers is as equation 4.

$$p = \frac{\sum x}{s_m N} \quad (4)$$

Description:
p = proportion of correct answers or difficulty level
∑x = number of test takers who answered correctly
Sm = maximum score
N = number of test takers
The classification value of the level of difficulty on HOTS items is as shown in table 4

Table 4. Classification of Level of Difficulty

Score range	Category
0.00 – 0.30	Difficult
0.31 – 0.70	Medium
0.71 – 1.00	Easy

According to Kunandar (2013) a good proportion for the level of difficulty of the item is, the question consists of 30% of items in the easy category, 50% in the medium category and 20% in the difficult category, but in this study it will not focus on these proportions because the questions made are based on higher-level thinking skills so that questions in the easy category will not be used.

Result and Discussion

Result

The development of items with HOTS (High Order Thinking Skill) resulted in 10 items with descriptions that had passed the 15 stages of the MC development model Intire. The results of the development of these items have been tested by paying attention to 4 aspects, namely the reliability of the questions, the validity of the questions, the difficulty level of the questions and the differentiating power of the questions. The source of question instruments that have been developed is found in SMA/MA chemistry books, chemistry question bank books and National Exam questions. The question items that have been selected from various sources are then modified and re-examined so as to produce question items in the form of descriptions. The question instrument developed has a stimulus and has questions that contain higher-level cognitive.

Learners in applying 4C skills (creative, critical thinking, communicative, collaborative) really need these skills in this modern era (Trisnawati et al., 2019). Skills in critical thinking abilities possessed by students so that they can solve a problem, quickly and easily adapt to social activities, can make a question critically, have good analogy skills in terms of science such as chemistry, physics and mathematics (Agustiana, 2019; Dilley, 2005; Paul et al., 1997).

Subjective tests using description-type items aim for students to have more complex abilities, to implement students' thinking skills in terms of solutions to problems and to make students more motivated in teaching and learning activities so that they are suitable for use for items based on higher-level thinking.

Expert Validation Results

Expert validation aims to see the validity of the question items that have been developed empirically by paying attention to the question criteria by translating the product moment correlation number. Item validation was carried out by 6 experts who can also be called raters. 6 raters validated the question items developed by paying attention to 3 aspects, namely material, construction and language with 15 indicator descriptions. The 6 raters assessed the chemistry question grids, chemistry items that had been developed with a total of 15 multiple choice type questions on Thermochemistry material and scoring guidelines. The study conducted by Jannah et al. (2020) showed that the results of the assessment by the validators were very high, with a score of 98% from the material expert, 87%

from the evaluation expert, and 92% from the linguist, all of which fell into the "very strong" criteria. With an average validation score of 92%, this study also met the "very strong" criteria. Similar research conducted by Khaldun et al. (2020) in evaluating computer-based HOTS category chemistry questions using WQC also received good results, with a score of 85% which met the criteria of "very valid". The rater provides suggestions or comments on the items until the items can be declared suitable for testing on students. By considering all aspects measured, it can be concluded that this WQC media is very suitable for use. According to research conducted by Adah et al. (2019), the overall media validation results reached a percentage of 95.48, which exceeded 80, so it can be categorized as very good.

Test Results

The result of the test is quantitative data which is conducted twice with the subject of students from class XII. The subject was class XII UPT SMAN 1 Musi Rawas. At the first field test stage, it was carried out with a small number of students as a trial. Data from the results obtained then made improvements by taking into account the comments or suggestions of students on the items developed. The second stage of the field test was carried out with more students to get more valid, reliable, differentiating power and level of difficulty of the items. The results of the data obtained are then analyzed as follows:

Differentiating Power

Quantitative data analysis of distinguishing power uses the help of Anates ver 4.0.5 to make it easier so that the results are obtained as in table 5.

Table 5. Analysis of Instrument Distinguishing Data

Question number	Differentiating Power	Description
1	0.224	Enough
2	0.224	Enough
3	0.884	Enough
4	0.448	Good
5	0.884	Very Good
6	0.777	Very Good
7	0.113	Less
8	0.334	Enough
9	0.448	Good
10	0.448	Good
11	0.779	Very Good
12	0.880	Very Good
13	0.667	Good
14	0.884	Very Good
15	0.779	Very Good

According to Dewi et al. (2019) it is important to conduct a power difference analysis to evaluate the extent to which a question can measure the true ability of students. If a question has a low power difference, then it means that the question is not effective in measuring students' abilities. On the other hand, if a question has a high-power difference, then it is more efficient in measuring students' ability better. Furthermore, calculations were carried out to test the distinguishing power of the items with an average result of 0.584. Based on table 1, of the 15 multiple choice type items tested on students, 6 items scored in the excellent category. Then 4 items scored in the good category, so they can distinguish between students with high and low abilities. In addition, there are 4 items that score in the sufficient category, which can also differentiate the abilities of high and low learners and 1 item scores in the poor category. Item number 7 can be influenced by questions that are too easy so that students can answer correctly. According to Arikunto (2007), the differentiating power of items to prove the difference between students with high and low abilities can be taken as 50% with high abilities and 50% with low abilities.

Level of Difficulty

Analysis of the level of difficulty of the items can be done with the help of ANATES version 4.0.5 so that the results are obtained as in table 6.

Table 6. Data Results of Instrument Difficulty Level

Question number	Level of difficulty	Description
1	0.941	Easy
2	0.911	Easy
3	0.470	Medium
4	0.617	Medium
5	0.500	Medium
6	0.617	Medium
7	0.911	Easy
8	0.588	Medium
9	0.411	Medium
10	0.176	Hard
11	0.588	Medium
12	0.470	Medium
13	0.676	Medium
14	0.323	Medium
15	0.441	Medium

After doing the calculations in the second field test, the average level of difficulty of the items was 0.576. Based on Table 2, there are several items that still get scores in the difficult or difficult category. This problem is caused by limited time constraints when conducting field tests, where students have not been maximized in working on items and the stimulus used is still difficult to understand. During the chemistry learning process in class, students only get 30 minutes in one lesson hour, so the material that can be delivered by the teacher is very limited. In addition, there are 11 items that score in the

medium category, which indicates that the items have a balanced level of difficulty and are not too easy or too difficult for students. There are 3 items with scores in the easy category, which means that almost all students can answer these items correctly. Items in the difficult category indicate that few learners can work on these items.

Validity

Testing the validity of the items can be done which is useful to prove that the items are empirically valid by fulfilling the criteria of the question by using the product moment correlation number translator. Based on research conducted by Riyani et al. (2017) the concept of "empirical validity" contains the term "empirical," which refers to experience. Therefore, an instrument is considered suitable for use when empirical validity has been tested through empirical experience. The results of validation data in the second field test with product moment correlation can be seen in table 7.

Table 7. Results of Problem Validity Data

Question number	Validity	Description
1	0.316	Low Validity
2	0.367	Low Validity
3	0.667	High validity
4	0.308	Low Validity
5	0.606	High validity
6	0.586	Fair validity
7	0.231	Low Validity
8	0.311	Low Validity
9	0.406	Fair validity
10	0.521	Fair validity
11	0.639	High validity
12	0.595	Fair validity
13	0.564	Fair validity
14	0.667	High validity
15	0.622	High validity

After students work on the items, calculations are made using the SPSS application to determine the validity of the items. The average calculation of R-count with the help of SPSS is 0.493. The R-table value obtained is 0.244 with a total of 67 students and the significance level used is 5%. After doing the calculation, it can be concluded that the 10 items are valid and suitable for testing. Question items with low validity categories can be tested on condition that improvements are made so that they become a sufficient category. This is because the R-count value obtained is greater than R-table. Based on the table, there are 5 items that get a validity value in the sufficient category. Meanwhile, there are 5 items that obtain question validity in the high category and 5 items are in the low validity category.

Reliability

The reliability test is carried out after the question validity test. Reliability test can be done to see the consistency of the items after being done by different or the same students at different times. The reliability test was carried out with the help of an application, namely SPSS, so that the data obtained was as shown in Table 8.

Table 8. Results of Problem Reliability Data

Question number	Cronbach's alpha if item deleted	Description
1	0.685	Fair Correlation
2	0.692	Fair Correlation
3	0.668	Fair Correlation
4	0.705	Fair Correlation
5	0.667	Fair Correlation
6	0.666	Fair Correlation
7	0.698	Fair Correlation
8	0.692	Fair Correlation
9	0.701	Fair Correlation
10	0.688	Fair Correlation
11	0.725	High correlation
12	0.712	High correlation
13	0.683	Fair Correlation
14	0.648	Fair Correlation
15	0.666	Fair Correlation

Figure 2. Students work on instruments

Based on table 8, the average reliability obtained from the reliability test is 0.686. The reliability test is conducted to evaluate the extent to which the items are consistent when done by different students or at different times. The reliability test results show that there are 13 items with sufficient correlation categories and 2 items with high correlation categories. Thus, it can be concluded that the 15 items that have been developed can be used for students because the correlation is classified as good and can be used both for the same and different samples of students at different times.

After calculating the results of the question's distinguishing power, difficulty level, validity test and reliability test so that these 4 aspects can be concluded based on the table Based on the table, 15 items that have

been developed are accepted and corrected according to comments and suggestions and can improve the items better. Based on the table, question items number 10 and 7 are not included in the criteria for appropriate question items because the results of the analysis on the level of difficulty are included in the difficult and easy categories, sufficient and less differentiating power, high validity and high correlation. Although item number 8 can distinguish students with high abilities and students with low abilities. However, only a few students can work on these items even though the differentiating power is in the sufficient category.

Based on the results of the analysis in the table, the items included in the good criteria are for the level of difficulty included in the medium category because the items are not too difficult to understand for students and are not included in the easy category because the items are not included in the Lower Order Thinking Skill (LOTS). Then for distinguishing questions included in the good or excellent category so that the items can distinguish students with high abilities and students with low abilities. Then for the validity test included in the category of sufficient validity, high validity and very high. For the reliability test, it is included in the sufficient, high and very high categories so that the items can be used to work on students at different times with the same or different samples.

Conclusion

Based on the research conducted with the title, namely the development of HOTS (High Order Thinking Skill) questions on Volta cell material for class XII SMA/MA, the results of the research can be concluded as follows Development of HOTS (High Order Thinking Skill) instruments by producing 15 items in the form of multiple choice and paying attention to aspects, namely the differentiation of questions, the level of difficulty of the questions, the validity of the questions and the reliability of the questions. Testing is also carried out by expert validation to pay attention to aspects of question construction, question material and language used. Then carried out for quantitative data so that it is obtained from 15 items that have been developed included in the category of valid items and are suitable for testing.

Acknowledgment
Thank you to the school for giving me permission to conduct the research. Thank you also to the local government for helping the research run smoothly.

Authors Contributions
Author one is the main author who came up with the research idea and designed the study. The second author is in charge of translating Indonesian into English. While the third author is a lecturer in charge of revising the article before submitting it.

Funding
This research was conducted with funding independently. There is no funding from any agency or institution, everything is done independently.

Conflicts of Interests
This research was conducted for the purposes of obtaining a Master's degree at graduation. Then for the benefit of a hobby because I have a hobby of writing in any form.

References

Adah, L. S., & Qosyim, A. (2019). Kelayakan tes hots berbasis komputer pada materi sistem peredaran darah manusia kelas viii. *Pensa E-Jurnal: Pendidikan Sains*, *7*(3), 382–387. Retrieved from https://jurnalmahasiswa.unesa.ac.id/index.php/pensa/index

Agustiana, J. (2019). Analisis kemampuan berpikir kritis siswa pada materi koloid. *SPEKTRA : Jurnal Kajian Pendidikan Sains*, *5*(1), 91. https://doi.org/10.32699/spektra.v5i1.80

Agustina, D. R., & Wibawa, R. P. (2019). Peran Pendidikan Berbasis Higher Order Thinking Skills (Hots) Pada Tingkat Sekolah Menengah Pertama di Era Society 5.0 Sebagai Penentu Kemajuan Bangsa Indonesia. *EQUILIBRIUM: Jurnal Ilmiah Ekonomi Dan Pembelajarannya*, *7*(2), 137. https://doi.org/10.25273/equilibrium.v7i2.4779

Annuuru, T. A., Johan, R. C., & Ali, M. (2017). Peningkatan Kemampuan Berpikir Tingkat Tinggi dalam Pelajaran Ilmu Pengetahuan Alam Peserta Didik Sekolah Dasar Melalui Model Pembelajaran Treffinger. *Edutcehnologia*, *3*(2), 136–144. Retrieved from https://ejournal.upi.edu/index.php/edutechnologia/article/view/9144

Arikunto, S. (2007). *Dasar-Dasar Evaluasi Pendidikan Edisi Revisi*. Bumi Aksara.

Arikunto, S. (2013). *Dasar-Dasar Evauasi Pendidikan Edisi 2*. BumiAksara.

Budiarta, K., Pendas, F., & Mailani, E. (2018). Potret Implementasi Pembelajaran Berbasis High Order Thinking Skills (HOTS) di Sekolah Dasar Kota Medan. *Jurnal Pembangunan Perkotaan*, *6*(2), 102–111. Retrieved from https://www.researchgate.net/publication/332212877

Dewi, R., & Azizah, U. (2019). Pengembangan Lembar Kerja Peserta Didik (LKPD) Berorientasi Problem Solving Untuk Melatihkan Keterampilan Berpikir Kritis Peserta Didik Kelas XI Pada Materi Kesetimbangan Kimia. *Unesa Journal of Chemical*

Education, *8*(3), 332– 339. https://doi.org/10.26740/ujced.v8n3.p

Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). Analisis Tingkat Kesukaran Dan Daya Pembeda Soal Olimpiade Matematika (OMI) Tingkat SMP Tahun 2018. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, *3*(1), 15–26. https://doi.org/10.36526/tr.v3i1.388

Dilley, P. (2005). The Power of Critical Theory: Liberating Adult Learning and Teaching (review). *The Review of Higher Education*, *28*(4), 626–627. https://doi.org/10.1353/rhe.2005.0038

Dwijayanti, N. (2021). Pembelajaran Berbasis HOTS sebagai Bekal Generasi Abad 21 di Masa Pandemi. *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, *9*(1). https://doi.org/10.20961/jkc.v9i1.53837

Fanani, K. (2018). Pengembangan Pembelajaran Berbasis HOTS di Sekolah Dasar Kelas V. *Jurnal Pendidikan Dasar*, *1*(9), 1–11. Retrieved from https://journal.unj.ac.id/unj/index.php/jpd/article/view/JPD.91.01/5168

Hujjatusnaini, N., Corebima, A. D., Prawiro, S. R., & Gofur, A. (2022). The Effect of Blended Project-based Learning Integrated with 21st-Century Skills on Pre-Service Biology Teachers' Higher-order Thinking Skills. *Jurnal Pendidikan IPA Indonesia*, *11*(1), 104–118. https://doi.org/10.15294/jpii.v11i1.27148

Inayati, U. (2020). Strategi Guru Dalam Menerapkan Pembelajaran Hots Menggunakan Model Problem Based Learning. *Auladuna : Jurnal Prodi Pendidikan Guru Madrasah Ibtidaiyah*, *2*(2), 27–34. https://doi.org/10.36835/au.v2i2.410

Jannah, K., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Berbasis Higher Order Thinking Skills Berbantuan Aplikasi "Kahoot!" Pada Kompetensi Dasar Menerapkan Penanganan Surat Masuk Dan Surat Keluar Jurusan OTKP Di SMK Negeri 2 Buduran. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, *8*(1), 108–121. https://doi.org/10.26740/jpap.v8n1.p108-121

Khaldun, I., Hanum, L., & Utami, S. D. (2020). Pengembangan Soal Kimia Higher Order Thinking Skills Berbasis Komputer Dengan Wondershare Quiz Creator Materi Hidrolisis Garam dan Larutan Penyangga. *Jurnal Pendidikan Sains Indonesia*, *7*(2), 132–142. https://doi.org/10.24815/jpsi.v7i2.14702

Kunandar. (2013). *Penilaian Autentik*. Rajawali Pers.

Kurniawan, R., Nasution, A. F., Hasibuan, Z., Afrida, A., & Ritonga, W. A. (2022). Pengembangan Instrumen Tes Kompetensi Sains Madrasah Tsanawiyah Berbasis Hight Order Thingking Skill (HOTS) untuk Meningkatkan Kemampuan Pemahaman Konsep Matematis Siswa. *Al-Khawarizmi: Jurnal Pendidikan Matematika*, *3*(1), 16–20. Retrieved from https://ejurnal.univalabuhanbatu.ac.id/index.php/al-khawarizmi/article/download/539/378

Lamhatin, F., Fajarianingtyas, D. A., & Anekawati, A. (2022). Pengembangan Instrumen Penilaian HOTS Memuat Keterampilan 4C Menuju Pembelajaran Abad 21. *EKSAKTA: Jurnal Penelitian Dan Pembelajaran MIPA*, *7*(1), 30–38. https://doi.org/10.31604/eksakta.v7i1.30-38

Lestari, P. A. S., Gunawan, G., & Kosim, K. (2019). Model Pembelajaran Discovery Dengan Pendekatan Konflik Kognitif Berorientasi Pada Kemampuan Berpikir Kritis Peserta Didik. *Jurnal Pendidikan Fisika Dan Teknologi*, *5*(1), 118–123. https://doi.org/10.29303/jpft.v5i1.1161

Maskur, R., Sumarno, Rahmawati, Y., Pradana, K., Syazali, M., Septian, A., & Palupi, E. K. (2020). The effectiveness of problem based learning and aptitude treatment interaction in improving mathematical creative thinking skills on curriculum 2013. *European Journal of Educational Research*, *9*(1), 375–383. https://doi.org/10.12973/eu-jer.9.1.375

OECD. (2019). *PISA 2018 Insights and Interpretations*. OECD Publishing.

Paul, R. W., Elder, L., & Bartell, T. (1997). California teacher preparation for instruction in critical thinking: Research findings and policy recommendations. *Eric*, 194. Retrieved from https://eric.ed.gov/?id=ED437379

Riyani, R., Maizora, S., & Hanifah, H. (2017). Uji Validitas Pengembangan Tes Untuk Mengukur Kemampuan Pemahaman Relasional Pada Materi Persamaan Kuadrat Siswa Kelas VIII SMP. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *1*(1), 60–65. https://doi.org/10.33369/jp2ms.1.1.60-65

Rositawati, D. N. (2019). Kajian Berpikir Kritis Pada Metode Inkuiri. *Prosiding SNFA (Seminar Nasional Fisika Dan Aplikasinya)*, *3*, 74. https://doi.org/10.20961/prosidingsnfa.v3i0.28514

Rozi, F., & Hanum, C., B. (2019). Pembelajaran IPA SD Berbasis HOTS (Higher Order Thinking Skills) Menjawab Tuntutan Pembelajaran di Abad 21. *Seminar Nasional Dasar Universitas Negeri Medan*, *2*(1), 246–311. Retrieved from https://jurnal.unimed.ac.id/2012/index.php/snpu/article/view/16127

Sani, R. A. (2019). *Pembelajaran Berbasis HOTS (Higher Order Thinking Skills)*. Tira Smart.

Sidiq, Y., Ishartono, N., Desstya, A., Prayitno, H. J., Anif, S., & Hidayat, M. L. (2021). Improving elementary school students' critical thinking skill in science

through hots-based science questions: A quasi-experimental study. *Jurnal Pendidikan IPA Indonesia*, *10*(3), 378–386. https://doi.org/10.15294/JPII.V10I3.30891

Surapranata, S. (2009). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. PT. Remaja Rosdyakarya.

Suwarto. (2013). *Pengembangan Tes Diagnostik dalam Pembelajaran*. Pustaka Pelajar.

Thomas, G., & Thorne, A. (2009). *How To Increase Higher Level Thinking*. Center for Development and Learning.

Trisnawati, W. W., & Sari, A. K. (2019). Integrasi Keterampilan Abad 21 Dalam Modul Sociolinguistics: Keterampilan 4C (Collaboration, Communication, Critical Thinking, dan Creativity). *Jurnal Muara Pendidikan*, *4*(2), 455–466. https://doi.org/10.52060/mp.v4i2.179