Journal of Biological Science and Education
~JBSE~
Website: http://usnsj.com/index.php/biology
Email: biologi_jbse@usn.ac.id

Developing Higher Order Thingking Skills Assesment Instruments On Invironmental Pollution Material For Class VII

AUTHORS INFO

Asra Indah
Universitas Bung Hatta
indahasra@gmail.com
+6282283178967

Azrita
Universitas Bung Hatta
azrita31@bunghatta.ac.id
+628116624222

Rona Taula Sari
Universitas Bung Hatta
ronataulasari21@gmail.com
+6281363353115

ARTICLE INFO

E-ISSN: 2721-0804
P-ISSN: 2723-6838
Vol. 3, No. 1, June 2021
URL: http://usnsj.com/index.php/biology

Suggestion for the Citation and Bibliography
Citation in Text:
Indah, A., Azrita, & Sari, R.T. (2021)
Bibliography:
Indah, A., Azrita, and Sari, R.T. (2021). Developing higher order thinking skills assesment instruments on invironmental pollution material for class VII. *Journal of Biological Science and Education, 3*(1), 1-13.

Abstract

The seventh grade science teacher at MTsN 3 Kota Pariaman has problems in developing an instrument for assessing higher order thinking skills. In the 2013 curriculum, teachers are required to be able to provide questions in the C4-C6 domain, namely High Order Thinking Skills (HOTS). This study aim to produce a valid and practical assessment instrument for higher order thinking skills on environmental pollution material at the C4-C6 cognitive level. The background of this research is that the questions made by the teacher are still at the C1-C3 level and there are still obstacles for teachers in developing higher order thinking skills assessment instruments, while in the 2013 curriculum teachers are required to be able to provide questions in the C4-C6 domain. This type of research is research development (Research and Development) which is a research method used to produce certain products. This type of research uses 3 stages of the 4-D model, namely the define, design and develop stages.Based on the research conducted, a very valid assessment instrument of 3,32 was produced by the validator, the empirical validity there were 30 valid questions and 10 invalid questions, the reliability in this study was 0,77 with the category of reliability, 94,79% teacher practicality and practicality. students amounted to 80,80%, the difficulty level of the questions was between 0,31 to 0,70 with moderate criteria and the question difference had sufficient criteria. This research can be used as a reference for other researchers related to the development of higher order thinking skills assessment instruments.

Keywords : High Order Thinking Skills, Validity

A. Introduction

The curriculum 2013 focuses on students to be able to observe, ask, reason, and communicate what they have gained after receiving lessons (Budiani et al, 2017). Furthermore, according to Kunandar (2015), the 2013 curriculum aims to prepare Indonesian people to have the ability to live as individuals and citizens who are faithful, productive, innovative and

affective and able to contribute to the life of society, nation and state in world civilization. The 2013 curriculum revision emphasizes high order thinking skills (HOTS) in learning. This shows that learning must provide training not only for basic learning for students to understand conceptually, but also for higher-order thinking skills.

Thinking skills are divided into three, namely low-level thinking skills (Lower Order Thinking Skills, LOTS), middle-level thinking skills (Middle Order Thinking Skills, MOTS), and high-order thinking skills (High Order Thinking Skills, HOTS) (Anderson & Krathwohl, 2001). High-level thinking includes three ability criteria that must be mastered, namely analyze, evaluate, and create. Higher order thinking skills are defined as the use of the mind broadly to find new challenges (Heong et al, 2011). Critical thinking or high-order thinking skills in science and technology also play an important role in instilling scientific attitudes in students. High-level thinking is not only developed in learning, but must also be supported by assessment instruments that reflect higher-order thinking (Rosidah, 2018).

Knowing the development of higher order thinking skills requires an assessment in the aspect of knowledge. Based on Permendikbud number 53 of 2015, the assessment of learning outcomes by educators is the process of collecting information or data about the learning outcomes of students. Questions about higher-order thinking skills can encourage students to think deeply about learning material, so it can be said that the higher order thinking skills assessment instrument can stimulate students to develop higher-order thinking skills (Barnett & Francis, 2012). The improvement of students' critical thinking skills can be evaluated in the presence of measuring tools or relevant instruments. The instrument is said to be good if it is able to evaluate or assess something with results such as the condition being evaluated, to get a good test instrument, an analysis of the instrument must be carried out (Rosidah, 2018).

In practice, an assessment requires an assessment technique. These techniques consist of test and non-test assessment techniques. Hamzah (2013) states that the test is a tool and has a systematic procedure that is used to measure and assess a knowledge or control of a measuring object of a certain set of content and material. Mardapi (2012) states that the test is a form of instrument used to take measurements consisting of a number of questions that have a right or wrong answer, or are all true or partially correct with the aim of knowing the learning achievements or competencies that students have achieved. for a particular field. Meanwhile, non-test techniques according to Hamzah (2013), that non-test is one of the evaluation instruments at the SD education unit level is called an assessment technique to obtain a description of characteristics, attitudes, or personalities. Non-test evaluation instruments include: questionnaires, interviews, observations, portfolios and journal rubrics. So far, the non-test technique is less popular than the test technique.

In the learning process, in general, the assessment activities prioritize test techniques. This is because the aspects of knowledge and skills play a greater role in the decisions made by the teacher when determining the achievement of learning outcomes. Teachers as learning managers are required to be able to prepare and carry out assessments with correct procedures so that the learning objectives set are achieved. Along with the enactment of the education unit level curriculum which is based on competency standards, basic competencies, the assessment technique must be adjusted to the following matters, namely the competence to be measured, the aspects to be measured (knowledge, skills, or attitudes), the measured student abilities, and existing infrastructure.

However, the problems that occur in the field are the implementation of High Order Thinking Skills (HOTS) learning is not easy for teachers to do. The teacher must really master the material and learning strategies and the teacher is also faced with challenges with the students' environment. Learning will be intertwined if students can be invited to think at higher levels. The success of mastery of a concept will be obtained when students are able to think at high levels, where students can not only remember and understand a concept, but students can analyze and synthesize, evaluate, and create a concept well. Another problem is that there are obstacles for teachers in developing higher order thinking skills assessment instruments for students, the questions given to the medium level (C3) are analyzed according to the bloom taxonomy. Whereas in the 2013 Curriculum teachers are required to be able to provide questions in the C4-C6 domain, namely high-order thinking skills (HOTS). Apart from not being used to using HOTS, other factors that cause students to be in moderate or sufficient criteria are culture and character. According to Thomas (2012), culture is generally passed down from parents and children, so that what parents experience will shape the child's personality, so that for years the habits and culture will be attached to the child. From that, the good and bad things done by children as individuals are influenced by the prevailing culture in the environment.

Putri's Research (2018) entitled "Development of Instruments for Assessment of Higher Order Thinking Skills on Biodiversity Materials for Class X SMA/MA Students". This research produces a higher order thinking assessment instrument for viral material that is valid and practical. Furthermore, Safitri (2017) entitled "Development of an Instrument for Assessment of Higher Order Thinking Skills on Virus Materials for Class X SMA/MA Students". This research produces a higher order thinking assessment instrument for valid and practical viral material.

B. Literature Review

1. Assessment Instruments

The learning outcome assessment instrument is a tool (measuring) used in the context of collecting and processing information to determine the achievement of student learning outcomes (Hamzah, 2013). The instrument is a measuring tool used to collect data for student assessment. This instrument will provide information to the teacher about the circumstances and achievements achieved by their students. This assessment can be in the form of test assessments, non-tests, class-based assessments, performance assessments, and also portfolio assessments (Wati, 2016).

The increase in students' critical thinking skills can be evaluated with the presence of measuring instruments or relevant instruments. This instrument is said to be good if it is able to evaluate or assess something with results such as the condition being evaluated, to get a good test instrument, an analysis of the instrument must be carried out (Rosidah, 2018). In practice, an assessment requires an assessment technique. This technique consists of test and non-test assessment techniques.

2. High Order Thinking Skills

Higher order thinking skills are defined as the use of the mind broadly to find new challenges (Heong et al, 2011). Critical thinking or higher-order thinking skills in science and technology also play an important role in instilling scientific attitudes in students. High-level thinking is not only developed in learning, but must also be supported by assessment instruments that reflect higher-order thinking (Rosidah, 2018).

Anderson & Krathwohl conducted research in 2001 and resulted in improvements to Bloom's taxonomy. The improvements made were to change Bloom's taxonomy from a noun to a verb. Anderson & Krathwohl (2001) state that the cognitive domain according to Bloom's taxonomy has six levels of thought processes, starting from the lowest level to the highest level.

3. Quality of Development Result Based on Validity, Practicality, Reliability, Difficulty Level and Difference

Validity

Validity is evidence and theory support for the interpretation of test scores in accordance with the purpose of using the test so that validity is the most basic foundation in developing and evaluating a test (Mardapi, 2012). Meanwhile Siskandar & Basrowi (2012), state that to be valid an instrument is not only consistent in its use, but what is important is that it must be able to measure its target size. A test can have multiple levels of validity: high, medium, low depending on its purpose.

Reliability

A test is said to be reliable if the test results show consistency. This means that if students are given the same test at different times, each student will remain in the same order (rank) in their group (Arikunto, 2012). In line with this Supardi (2015) an item of assessment instrument is said to be reliable if it is used to measure at different times the results will be the same, thus reliability can also be interpreted as stability.

Practicality

Practicality means the ease of a test, both in preparing, using, processing, and interpreting, as well as administering it (Arifin, 2012). Factors that influence the practicality of the evaluation instrument include ease of administration, time provided for smooth evaluation, ease of scoring, ease of interpretation and application, availability of an equivalent or comparable form of evaluation instrument.

Arikunto, (2012) a test is said to be practical if it has the following characteristics:

a) Easy to implement, for example it does not require a lot of equipment and gives students the freedom to do the parts that are considered easy by the students first.
b) Easy to check, meaning that the test is equipped with both an answer key and a scoring guide.
c) Equipped with clear instructions.

Difficulty Level

A good question is a question that is not too easy or not too difficult (Arikunto, 2012). Meanwhile Surapranata (2005), also states that, there are two characteristics of the level of difficulty, namely:

a) The level of difficulty is a measure of the question which does not indicate the characteristics of the question
b) The level of difficulty is a characteristic of the item itself and the taking of the test.

Based on the above, the difficulty level is the level of ease of the question. The higher the difficulty index value, the easier the questions are given and conversely the lower the difficulty index value, the more difficult the questions are. Good questions are moderate questions that have a moderate difficulty index, namely 0.30 to 0.70 (Arikunto, 2012).

Difference

The distinguishing power of questions is the ability of questions to distinguish between students who have mastered the material and students who have not mastered the material (Kunandar, 2015). In line with this, Arikunto, (2012), states that the distinguishing power of a question is the ability of a question to distinguish between students who are smart (high ability) and students who are less intelligent (low ability).

C. Methodology

1. Research Design

This high-order thinking ability assessment instrument was developed using the Four-D-Models learning tool development model suggested by Thiagarajan, Semmel et al (1974). This model consists of four stages, namely define, design, develop, and disseminate. Due to time constraints, this research was only carried out until the develop stage.

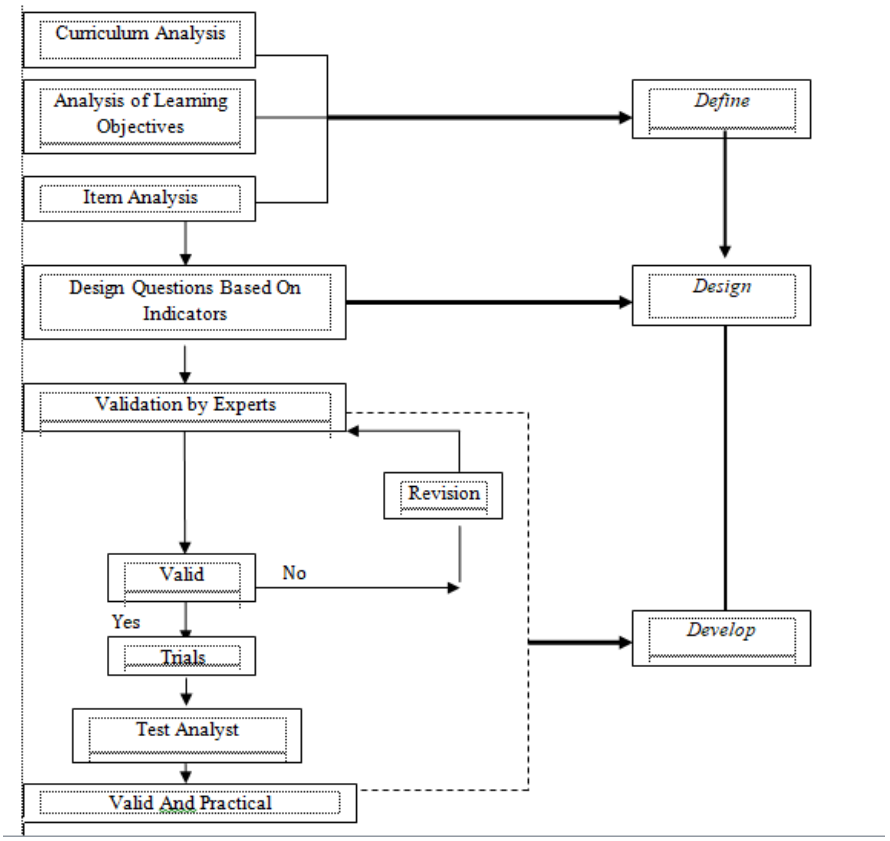

Figure 1. Research Procedure for the Development of Higher-Level Thinking Ability Assessment Instruments

2. Instruments

The instruments used to collect data in this study were a validity test questionnaire and a practicality test questionnaire.

Validity Questionnaire

The validity questionnaire is filled in by the validator, namely the lecturer. The purpose of the validity questionnaire is to find out data about the validity of the higher order thinking skills instrument that will be developed.

Practicality Questionnaire

The questionnaire for the practicality test of the higher order thinking skills instrument was filled in by teachers and students. This questionnaire contains questions related to the ease of implementation, examination, and instructions for the assessment of higher order thinking skills.

The validity test questionnaire and practicality test questionnaire were arranged according to a modified Likert scale from Riduwan (2012) with a scale of 4 alternative answers, namely:

Table 1. *The questionnaire criteria for validity and practicality were compiled by A Likert Scale*

Symbol	Criteria	Skor
SA	Strongly Agree	4
A	Agree	3
D	Disagree	2
SD	Strongly Disagree	1

3. Technique of Data Analysis

The research data were analyzed using descriptive statistics. This technique describes the results of the validity test, practicality test, and item quality test of the higher order thinking skills assessment instrument. This analysis includes the following matters.

Analysis of the Validity of the High-Level Thinking Ability Assessment Instrument

Validity comes from the word validity which means the extent to which the accuracy and accuracy of a measuring instrument in performing its measuring function. A test or non-test of a measuring instrument or measuring instrument is said to have high validity if the tool performs its measuring function or provides measurement results that are in accordance with the intended purpose. take that measurement. In this study, the validity used in determining the assessment instrument was twofold, namely

Logical Validity Analysis

From the results of the media validity obtained, it was analyzed on all aspects presented in tabular form using a Likers scale, then the mean value was searched using the following formula*:*

$$R = \frac{\sum_{i=1}^{n} Vij}{nm}$$ (Muliyardi, 2006)

Information :

R = Average research results from the research results of experts / practitioners
Vij = Score of research results of experts / practitioners to-j criteria i
n = The number of experts / practitioners who judge
m = Number of criteria

Table 2. Criteria for the assessment of validity

Category	Range
3,25 - 4,00	Very valid
2,50 - 3,24	Valid
1.75 - 2.49	Invalid
1.00 - 1.74	Not valid

Empirical Validity Analysis

An instrument can be said to have empirical validity if it has been tested from experience. The internal quantitative characteristics are intended to include the parameters of the difficulty level, the distinguishing power and the reliability. Especially for multiple choice questions, two additional parameters are seen from the opportunity to guess or answer the questions correctly and the function of the answer choices, namely the distribution of all alternative answers from the tested subjects.

If the dichotomy score is (0.1) then to calculate the correlation coefficient between the item score and the total instrument score, the biserial point correlation coefficient (rpbis) is used which uses the formula:

$$r_{pbis} = \frac{Mp - Mt}{St}\sqrt{\frac{p}{q}}$$

Information :
rpbis = Point biserial correlation coefficient
Mp =The mean score of the subjects who answered correctly the item being sought was correlated
Mt = Mean total score
St = Standard deviation
p = The proportion of subjects who answered the question correctly
q = 1- p

Reliability Analysis of High-Level Thinking Ability Assessment Instruments

Reliability is the determination of the results obtained from a measurement result. The reliability used to measure the learning outcome test is to use the Alpha Crownbach formula, namely:

$$r_{11} = \left[\frac{k}{(k-1)}\right]\left[\frac{\sum\sigma^2}{\sigma^2 t}\right]$$

Information:
r_{11} = reliability coefficient alpha
k = Number of question items
$\sum\sigma^2$ b = The number of score variants for each item
σ^2 t = total variant

The test criteria, if Rcount > Rtable with a significant level (α) = 0.05, the instrument meets the reliability requirements. Likewise, if Rhcount < Rtable with a significant level of 0.05, the instrument does not meet the reliability requirements. A test is said to be reliable (high reliability) if r_{11} is equal to or greater than 0.70 (Supardi, 2015).

Table 3. *The terms of the instrument reliability coefficient*

Cronbach's Alpha Value Reliability	Criteria
0.81-1.00	Very High
0.61-0.80	High
0.41-0.60	Moderate
0.21-0.40	Low
0.00-0.20	Very Low

Analysis of Practicality Test Instruments for High-Level Thinking Ability Assessment

The practicality test data for the assessment of higher order thinking skills were analyzed by percentage (%) using the following formula:

$$\text{Practicality value} = \frac{\text{The sum of all scores}}{\text{Maximum number of scores}} \times 100\%$$

After the percentage is obtained, then grouping is carried out according to the modified criteria by Purwanto (2009) which is modified as follows:

Table 4. Criteria for practicality assessment

Practicality Value (%)	Practicality Criteria
90-100	Very practical
80-89	Practical
60-79	Enough practical
0-59	It's not practical

Difficulty Level Analysis

The level of difficulty of the questions is searched using the following formula:

$$P = \frac{B}{J}$$

Information :
P = level of difficulty
B = Many subjects answered correctly
J = Many subjects took the test

Table 5. *Classification of difficulty level*

Value of Difficulty Level	Criteria for Level of Difficulty
0.00 - 0.30	Hard
0.31 - 0.70	Moderate
0.71 - 1.00	Easy

According to Arikunto (2012), good questions are questions with a difficulty index of 0.31 to 0.70, namely questions in the medium category.

Difference

The formula used to find the difference power is as follows:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Information :
D = Distinguishing power of the item
BA = The number of upper groups who answered the question correctly
BB = The number of lower groups who answered the question correctly
JA = The number of subjects in the top group
JB = The number of subjects in the lower group

The classification of distinguishing power according to Arikunto (2012) is as follows.

Table 6. *Criteria for differentiation*

Value	Criteria
0.00-0.20	Poor
0.21-0.40	Enough
0.41-0.70	Good
0.71-1.00	Very Good

The instrument for assessing high-order thinking skills that is good is an instrument with most of the discriminating power which is categorized as sufficient, good and excellent. Assessment instruments that have sufficient, good, and excellent discriminating power can differentiate between low and high ability students.

D. Findings and Discussion

1. Findings

The development stage (develope) in this research includes logical validity test, empirical validity, reliability test, practicality test, difficulty level, and distinguishing power.

1.1 Analysis of the Validity of the High-Level Thinking Ability Assessment Instrument
The logical validity of the higher-order thinking skills assessment instrument

The logical validity of this higher order thinking ability assessment instrument was carried out by two validators consisting of FKIP Bung Hatta University lecturers using a validation questionnaire. During the validation phase, there were various suggestions and criticisms received from the validator so that it became the basis for consideration for making revisions to the higher order thinking skills assessment instrument made. According to the validator, of the questions that have been validated, nothing should be discarded, it just has to be a question. revised so that it can be a valid question according to the criteria for logical validity. Revision of the questions carried out can be based on the validator's suggestions in a nutshell can be seen in Table 7.

Table 7. *Validators' suggestions for the assessment instrument*

Validator	Suggestion	Corrective
1	Fix questions that have too long sentences	The questions have been corrected with more efficient sentences
2	Look again at the cognitive domain of each question or question instrument	The problem has been fixed based on the cognitive domain of the bloom taxonomy

Based on suggestions and input from the validator of the higher order thinking skills assessment instrument, this instrument can be tested on students. The results of the validity test of the validator can be seen in Table 8.

Table 8. *Test results of the logical validity of the assessment instrument*

No	Assessment Component	Validator		Total	Validity Value	Criteria
		1	2			
1	Theory	33	27	60	3,33	Valid
2	Construction	28	24	52	3,25	Valid
3	Language	12	9	21	3,50	Very valid
4	Higher Order Thinking	17	15	32	3,20	Valid
Total Value of Validity					13,28	
Average Value of Validity					3,32	Very valid

The final result of the logical validity of the validator gets an average of 3.33 with very valid criteria. This shows that the higher order thinking skills assessment instrument that has been made is very valid, both in terms of material, construction, language and higher order thinking skills so that it can be used in research. The instrument for assessing higher order thinking skills is then given to science teachers at MTsN 3 Kota Pariaman for practicality test.

Empirical Validity of Higher-Level Thinking Ability Assessment Instruments

The empirical validity aims to determine the level of reliability of the questions. The results can be seen in Table 9.

Table 9. *The results of the test of the empirical validity of the assessment instrument*

No	Empirical Validity	Total
1	Valid Question	30 Questions
2	Invalid Question	10 Questions
The Total Number Of Questions		40 Questions

Based on Table 9, the empirical validity of the instrument for assessing high-level thinking skills in environmental pollution material is 30 valid questions and 10 invalid questions with a total of 40 questions. 10 Invalid questions can not be defended or in other words, they are not used.

1.2 Instrument Reliability Analysis of High Level Thinking Ability Assessment

The test criteria, if Rhitung>Rtabel with a significant level (α) = 0.05, then the instrument meets the reliability requirements. Likewise, if Rcount < Rtable with a significant level of 0.05, the instrument does not meet the reliability requirements. A test is said to be reliable (high reliability) if r_{11} is equal to or greater than 0.70 (Supardi, 2015).

Table 10. *Results of the reliability of the assessment instrument*

Reliability	Criteria
0,77	High reliability

Based on Table 10, the reliability of the instrument for assessing high-level thinking skills in environmental pollution material is reliable because the results obtained are 0.77. Based on this, the resulting instrument has high reliability. This means that reliability can be used and provides consistent results for the same measurement.

a. Practicality Analysis of Higher-Order Thinking Ability Assessment Instruments

An instrument for assessing high-order thinking skills that is valid and ready to be tested, then a practicality test is carried out which aims to determine the practicality level of the

instrument for assessing higher-order thinking skills. The practicality test was carried out by the science teacher at MTsN 3 Kota Pariaman by filling out a practicality test questionnaire. The results can be seen in Table 11.

Table 11. *The results of the practicality of the teacher's assessment instrument*

No	The assessment aspect	Persentase (%)	Criteria
1	Implementation	87,50	Practical
2	Examination	100	Very practical
3	Instructions for questions	100	Very practical
4	Theory	91,67	Very practical
	Average	94,79	Very practical

In addition to the practicality test by the teacher, the practicality of the instrument for assessing the ability to think highly of environmental pollution material was also carried out by students. The practicality data of students were obtained through the results of a practicality questionnaire. A total of 36 students of class VII MTsN 3 in Pariaman City conducted a practicality test by filling out the practicality questionnaire that the researcher had given. Analysis of practicality questionnaire results by students can be seen in Table 12.

Table 12. The results of the practicality of the assessment instruments by students

No	The assessment aspect	Persentase (%)	Criteria
1	Implementation	81,08	Practical
2	Examination	80,06	Practical
3	Instructions for questions	81,25	Practical
4	Theory	80,79	Practical
	Average	80,80	Practical

Based on Table 12, it is known that the practical value of the instrument for assessing the ability to think highly of environmental pollution material filled by students, in terms of implementation it is obtained 87.50%, examination is 100%, question instructions are 100% and in terms of material 91.67%. This shows that, the instrument for assessing the ability to think highly of environmental pollution material that has been developed is practical for use by students with an average of 80.80%.

b. Analysis of Difficulty Levels of High-Level Thinking Ability Assessment Instruments

Item analysis is the assessment of test questions in order to obtain a set of questions of adequate quality. Test items must be known about the level of difficulty, because each test maker needs to know whether the questions are difficult, medium or easy. The level of difficulty can be seen from the students' answers. The fewer the number of students who can answer the question correctly, it means that the question is considered difficult and conversely the more students can answer the question correctly, meaning that it indicates the question is not difficult or the question is easy. The results of research conducted for the difficulty level of problems on environmental pollution material can be seen in Table 13.

Table 13. *The results of the difficulty level of the assessment instrument questions*

No	Item difficulty level	Total
1	Easy	11 Questions
2	Medium	19 Questions
3	Hard	10 Questions
Total questions		**40Questions**

c. Analysis of the Distinctive Power of Higher-Order Thinking Ability Assessment Instruments

The discriminating power analysis examines the items with the aim of knowing the ability of the questions in distinguishing students who are classified as capable (high achievement) from students who are classified as poor or weak in achievement. That is, if the questions are given to capable children, the results show high achievement and achievement. if given to weak students, the results are low. The results can be seen in Table 14.

Table 14. *The results of the difference of the assessment instrument questions*

No	Distinction of Question Items	Total
1	Bad	13 Questions
2	Enough	16 Questions
3	Good	11 Questions
Total Questions		40 Questions

Based on the results of the study, it can be seen that the daily test questions on environmental pollution material used in MTsN 3 Kota Pariaman class VII are 40 items, there are 13 items with bad criteria, 16 items with sufficient criteria and 11 items with good criteria.

2. Discussion

2.1 Validity and Reliability

The instrument for assessing high-level thinking skills that was developed was very valid based on the four aspects validated by the validator, namely aspects of material, construction, language and high-order thinking skills with an average value of 3.32. In terms of material, the instrument for assessing higher order thinking skills is categorized as very valid with a validation value of 3.33. This means that the questions are in accordance with the 2013 curriculum which has been adjusted to the defined core competencies and basic competencies. These valid results illustrate that the higher order thinking skills assessment instrument developed is suitable for learning so that it can be used in the assessment process.

Viewed from the construction aspect, the instrument for assessing higher order thinking skills that has been made is very valid with 3.25 validation. The construction of the questions is in accordance with the formulation of the questions given clearly and is related to the stated material so that it does not cause confusion for students. This is in line with Widodo's opinion (2010) which states that by knowing the learning objectives students will not deviate from the learning being learned. , due to information about learning objectives.

a validation value of 3.20. The validator suggests that the cognitive level of the question is in accordance with the cognitive level of the students' high-order thinking ability. According to Kurniati (2016), he states that high-order thinking questions stimulate students to interpret, analyze, or even be able to manipulate previous information so that it is not monotonous. The assessment instrument developed can be used as an instrument that can measure and stimulate and train students' higher-order thinking skills.

Based on the results of the test items that have been carried out on the 40 items tested on class VII students of MTsN 3 Kota Pariaman, there were 30 valid questions based on their empirical validity. Meanwhile, 10 questions are invalid based on empirical data because the level of correlation is low. According to Sudarmin (2012), it is stated that a good question has a valid measure for a specific purpose but is not valid for other purposes or even for the same purpose in the group other. Furthermore, according to Rahayuni (2016) states that basically validity is a concept related to the extent to which the test has measured what must be measured.

The resulting high-level thinking ability assessment instrument is reliable, namely the r11 is 0.77. Based on this, the instrument can be said to be reliable with high criteria. The instrument for assessing high-order thinking skills is said to be reliable if this assessment instrument can provide the same results if tested in the same group at different times or occasions. Arikunto (2012) states that a test is said to be reliable if it has a fixed result in the test. Furthermore, Nurwanah (2019) states that questions that already have reliability above 0.70 are said to be reliable and there is no need for revision of the item instrument according to the reliability test.

2.2. Practicality of high-order thinking skills assessment instruments

Based on the results of the practicality test analysis, it was found that the instrument for assessing higher order thinking skills that had been made had very practical criteria. The practicality average value obtained was 94.79% from four aspects in the practicality test, namely aspects of implementation, examination, question instructions and material.

In terms of the implementation aspect, the instrument for assessing high-level thinking skills is considered practical with a practicality value of 87.50%. This means that this assessment instrument does not require a lot of equipment to work on and gives students the freedom to answer questions that they feel can be answered first.

Judging from the aspect of examining the questions, the instrument for assessing high-order thinking skills that has been made is categorized as practical with a value of 100%. Questions

are equipped with answers and answer sheets. This will make it easier for teachers to check students' answers.

Judging from the aspect of question instructions, the instrument for assessing high-order thinking skills that has been made is categorized as practical with a value of 100%. This is obtained because the test is equipped with clear instructions to guide students to be able to work on the questions.

a. Level of difficulty and Difference

The results of the analysis of the difficulty level of the questions in the research carried out had a difficulty level of questions between 0.31 to 0.70 with moderate criteria. Arikunto (2012) states that a good question is a question that is neither easy nor too difficult. Furthermore, according to Tika et al, (2014) also states that the level of difficulty of an item is marked by the percentage of students who answered correctly on the item in question. In line with this, according to Rofiah (2013) states that the easier an item is, the more students will answer the question correctly, whereas if the question is too difficult, only a few students are able to answer the item correctly. From some of these opinions, it can be said that a good question is a question that has a medium difficulty level criterion.

Julistiawati (2013) states that the distinguishing power of a question is the ability of a question to differentiate between high-ability students and low-ability students. Based on the results of the analysis of the difference in power obtained at the time of the study, it was found that the different power of the questions owned was in the sufficient criteria According to Nofiana (2016) states that the distinguishing power of good questions is the distinguishing power of questions that are categorized as sufficient, good and excellent. In line with that, according to Ita (2018) states that an assessment instrument that has sufficient, good and excellent distinction power can differentiate between low-ability students and high-ability students.

E. Conclusion

The development of instruments for assessing high-level thinking skills on environmental pollution material in the 2013 curriculum for grade VII students of MTsN 3 Kota Pariaman which resulted in this study was stated to be logically valid (3.33) and empirical validity (30 valid questions out of 40 questions), reliable (0.77 with high criteria), practical (94.79% assessment from the teacher and 80.80% from students), moderate difficulty level and sufficient differentiation power. This assessment instrument can be used as an assessment instrument that can measure students' higher order thinking skills. Higher order thinking skills can be developed if educators use the right assessment. Therefore, higher order thinking assessment instruments are needed by educators, in this case the teacher, as an evaluation that can stimulate and measure students' higher order thinking skills. Another part of being able to stimulate and train students' higher order thinking skills is to carry out activity-based learning activities, so as to encourage students to build creativity and critical thinking.

The importance of 21st century skills that emphasize HOTS implementation efforts among students, where these skills are very important to produce human resources who are able to apply knowledge to face various challenges, have creative and innovative thinking styles and high competitiveness. Researchers encountered several obstacles during the process of making this higher order thinking ability assessment instrument. The first difficulty faced is in making a question stimulus. Sometimes the question stimulus that is made cannot present the questions and does not function to be able to answer the questions, so that without any stimulus the questions can be answered by students.

Acknowledgement

The authors would like to thank the students and teachers of class VII MTsN 3 Pariaman City and those who took part in contributing to this research.

F. References

Anderson, L. W., & Krathwohl, D.R. (2001). *A taxonomy of learning, teaching, and assessing. A revision of bloom'staxonomy of educational objectives, New York Longman. 41*(4), 212-218.

Arifin, Z. (2012). *Penelitian pendidikan metode dan paradigma baru*. Bandung: Remaja Rosda Karya.

Arikunto, S. (2012). *Evaluasi pembelajaran (prinsip, teknik, prosedur).* Bandung: Remaja Rosdakarya Offset.

Barnett, J. E & Francis, A.L. (2012).Using higher order thinking question to foster critical thinking: a classroom study. Educational Psychology: *An International journal of Experimental Educational Psychology, 3*(2), 209-216.

Budiani, S., Sudarmin., & Syamwil, R. (2017). Evaluasi implementasi kurikulum 2013 di Sekolah Pelaksana Mandiri. *Innovative Journal of Curriculum and Educational Technology (IJCET)*. *6*(1), 45-57.

Hamzah, A. (2013). *Evaluasi pembelajaran matematika.* Jakarta: PT Raja GrafindoPersada.

Heong, M. Y. (2011). The level of marzano higher thingking skills among technical education students. *International Jurnal of Social Science and Humanity, 1*(2), 121-125.

Julistiawati., Rini., & Bertha Y. (2013). Keterampilan level C4,C5 dan C6 revisi taksonomi bloom siswa kelas X.3 SMAN 1 semenep pada penerapan model pembelajaran inkuiri pokok bahasan larutan elektrolit dan non elektrolit. *UNESA Journal of Chemical Education*, *2*(2), 37-62.

Kunandar. (2015). *Penilaian autentik (Penilaian hasil belajar peserta didik berdasarkan kurikulum 2013).* Jakarta : PT Raja GrafindoPersada.

Kurniati, D. (2016). Kemampuan berpikir tingkat tinggi siswa smp di kabupaten jember dalam menyelesaikan soal berstandar PISA. *Penelitian dan Evaluasi Pendidikan, 20*(2): 142-155

Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.

Muliyardi. (2006). *Pengembangan model pembelajaran matematika menggunakan komik di kelas 1 Sekolah Dasar.* Surabaya: UniversitasNegeri Surabaya (UNNESA).

Nofiani, M., Sajidan., and Puguh. (2016). Pengembangan Instrumen Evaluasi Higher Order Thingking Skills Pada Materi Kingdom Plantae. *Jurnal Pendagogi Hayati, 1*(1), 67-80.

Nurwanah. (2019). Pengembangan Butir Soal Kemampuan Berpikir Tingkat Tinggi Pada Mata Pelajaran Biologi Kelas XI SMA Negeri 3 Pangkep. *Skripsi.* Makasar. Universitas Islam Negeri Alauddin Makasar.

Putri, S. R. (2018). Pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi pada materi keanekaragaman hayati untuk peserta didik SMA/MA kelas X. *Skipsi.* Padang. Universitas Negeri Padang (UNP).

Permendikbud No 53 Tahun 2015. Penilaian Hasil Belajar Pendidik pada Pendidikan Dasar dan Pendidikan Menengah.

Purwanto. (2009). *Evaluasi hasil pembelajaran*. Yogyakarta: Pustaka Pelajar

Riduwan.2012. *Pengantar statistika sosial*. Bandung: Alfabeta

Rahayuni, G. (2016). Hubungan Keterampilan Berpikir Tingkat Kritis dan Literasi Sains pada Pembelajaran IPA Terpadu dengan Model PBM dan STM. *Jurnal penelitian dan pembelajaran IPA, 2*(2), 131-146.

Rofiah, Emi, Nonoh, S. A., & Elvin, Y. (2013). Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Peserta Didik SMP. *Jurnal Pendidikan Fisika*, *1*(2), 17-22.

Rosidah, N. A., Ramalis, T. R., & Suyana, I. (2018). Karakteristik tes kemampuan berpikir kritis (KBK) berdasarkan pendekatan teori respon butir. *Jurnal Inovasi dan Pembelajaran Fisika, 1*(3),54-63.

Safitri, W. R. (2017). pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi pada materi virus untuk peserta didik kelas X. *Skripsi*. Padang. Universitas Negeri Padang (UNP).

Siskandar & Basrowi. (2012). *Evaluasi belajar berbasis Kinerja*. Bandung: Karya Putra Darwati

Sudarmin. (2012). Meningkatkan kemampuan berpikir tingkat tinggi mahasiswa melalui pembelajaran kimia terintegrasi kemampuan generik sains. *Varia pendidikan, 42*(1), 97-103.

Supardi. (2015). *Penilaian Autentik Pembelajaran Afektif, Kognitif, Psikomotor*. Jakarta: PT Raja Grafindo Persada.

Supranata. (2005). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Test Implementasi Kurikulum 2004*. Bandung: Remaja Rosda Karya.

Thiangarajan, S., et al. (1974). *Instructional Development for Training Teachers of Exceptional Children*. Indiana: Indiana University

Thomas, L. (2012). *Educating for character: Mendidik untuk membentuk karakter*. Jakarta: Bumi Aksara

Tika, D. R., Bambang, H. R., & Sukidin. (2014). The Analyis of Difficulaties and Distinguishing Power on The Middle Test With From of Multiple Choice on Odd Semester at Economic Subjects on The Tenth Grade of SMA Negeri 5 Jember in 2012/2013 Academic Year. *Jurnal Edukasi UNEJ*, *1*(1), 39-43.

Wati, E.R.. (2016). *Evaluasi pembelajaran*. Yogyakarta: Kata Pena.

Widodo, T., & Kadarwati, S. (2010). Higher order thingking berbasis pemecahan masalah untuk meningkatkan hasil belajar berorientasi pembentukan karakter siswa. *Cakrawala Pendidikan*, *32*(1), 161-171.