Vol10(2), 2021, 79-92 DOI: 10.23960/jppk.v10.i2.2021.06
Jurnal Pendidikan dan Pembelajaran Kimia
e-ISSN: 2714-9595| p-ISSN 2302-1772
http://jurnal.fkip.unila.ac.id/index.php/JPK/index

Development of an Assessment to Measure Critical Thinking Ability Materials for Interaction of Living Things with the Environment in Bengkulu City Junior High

Ahmad Walid[1] *Putri Marfhadella[2], Irwan Satria[3]

Program Studi Ilmu Pengetahuan Alam, Jurusan Sains dan Sosial, Fakultas Tarbiyah dan Tadris, Institut Agama Islam Negeri Bengkulu, Jl. Raden Fatah Street, Pagar Dewa, Bengkulu, Indonesia.
*e-mail:putrimarfhadella99@gmail.com

Received: Juni 5th, 2021 *Accepted: Juli 23rd, 2021* *Online Published: Agustus 23rd 2021*

Abstract: This study aims to develop an assessment to measure critical thinking skills in the interaction of living things with the environment in junior high schools in Bengkulu City. The research method uses the Borg & Gall development model which consists of 8 stages which include: potential and problems, product planning, initial product development, expert validation, expert revision, product trials, product revisions, and final products. Based on the results of the analysis of questions that empower students to think critically in SMP in Bengkulu City, it consists of SMPN A by 14%, SMPN B by 10%, and SMPN C by 6%. The data from the validation results of assessment experts, material experts, and linguists are 78.1%, 83.5%, and 88.5% are in the (adequate, very feasible, and very feasible categories). The level of readability of the assessment questions is 91.7% in the very good category. Analysis of critical thinking assessment data using the Rasch model with Ministep (Winsteps) software. The quality of the questions from the aspect of validity obtained 1 invalid question at SMPN A, 2 invalid questions at SMPN B and SMPN C. The item reliability values were 0.84, 0.87, and 0.82 were categorized as good. The average item difficulty index in the critical thinking assessment shows that there are four categories, namely 7 questions which are difficult, 2 are moderate, 3 are easy, and 1 is very easy. The discriminatory power of items was obtained by 4 groups of items based on the separation item index, namely 2.33, 2.58, and 2.48. The conclusion of the research shows that based on expert validation, analysis of validity, reliability, level of difficulty, and distinguishing power, essay questions are feasible to use.

Keywords: Assessment, Critical Thinking, Characteristics, Feasibility

Abstrak: *Penelitian ini bertujuan untuk mengembangkan penilaian untuk mengukur kemampuan berpikir kritis dalam interaksi makhluk hidup dengan lingkungan di SMP Negeri di Kota Bengkulu. Metode penelitian menggunakan model pengembangan Borg & Gall yang terdiri dari 8 tahapan yang meliputi: potensi dan masalah, perencanaan produk, pengembangan produk awal, validasi ahli, revisi ahli, uji coba produk, revisi produk, dan produk akhir. Berdasarkan hasil analisis soal yang memberdayakan siswa untuk berpikir kritis di SMP Negeri di Kota Bengkulu terdiri dari SMPN A sebesar 14%, SMPN B sebesar 10%, dan SMPN C sebesar 6%. Data hasil validasi ahli penilaian, ahli materi, dan ahli bahasa adalah 78,1%, 83,5%, dan 88,5% berada pada (kategori cukup, sangat layak, dan sangat layak). Tingkat keterbacaan soal penilaian sebesar 91,7% dengan kategori sangat baik. Analisis data penilaian berpikir kritis menggunakan model Rasch dengan software Ministep (Winsteps). Kualitas soal dari aspek validitas diperoleh 1 soal tidak valid di SMPN A, 2 soal tidak valid di SMPN B dan SMPN C. Nilai reliabilitas butir soal adalah 0,84, 0,87, dan 0,82 dikategorikan*

baik. Rata-rata indeks kesukaran butir soal dalam penilaian berpikir kritis menunjukkan bahwa terdapat empat kategori yaitu 7 soal yang sulit, 2 soal sedang, 3 soal mudah, dan 1 soal sangat mudah. Daya pembeda item diperoleh 4 kelompok item berdasarkan indeks keterpisahan item, yaitu 2,33, 2,58, dan 2,48. Kesimpulan penelitian menunjukkan bahwa berdasarkan validasi ahli, analisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, soal esai layak digunakan.

Kata kunci: *Assessment, Berpikir Kritis, Karakteristik, Kelayakan*

▪ INTRODUCTION

In accordance with the Regulation of the Minister of Education and Culture No. 64 of 2013 concerning Standards of Content for Primary and Secondary Education to meet future needs and to meet the Indonesian Golden Generation in 2045, Graduate Competency Standards have been set based on XXI Century Competencies, that learning in the 21st century must be able to develop competitive skills that focus on development of thinking skills tingkattinggi*(HigherOrder Thingking*skills)which one of them is critical thinking*(criticalthinking)*(Dharmawati et al, 2016: 266).

Critical thinking skills is the ability to think students need to be developed in educational institutions so that Indonesia can compete globally in the 21st century There are four competencies required of students to compete in the 21st century, the *critical thinking and problem solving* (critical thinking and solve problems), *creativity* (creativity), *communication skills* (the ability to communicate), and the *ability to work collaboratively* (the ability to work together). Every student has the potential to think critically but the problem is how a teacher develops these abilities in the learning process. The ability of students to think critically is quite low, which is inseparable from the teacher's habit of making assessments that only measure the level of basic abilities. Students will not have higher-order thinking skills if they are not given the opportunity to develop them and are not directed (Arifin, 2012: 6).

Critical thinking is part of *cognitive skills* where there are six indicators of critical thinking skills involved in the critical thinking process according to Facione (2015) the dimensions of critical thinking skills which include *interpretation* (interpretation), *analysis* (analysis), *evaluation* (evaluation), *inference* (inference). , *explanation* (explanation), and *self regulationself-regulation* () (Facione, 2015: 4-7).*Interpretation* is the ability to understand, explain and give meaning to data and information. *Analysis* (analysis) is the ability to identify the relationship of an information that is used to express thoughts or opinions. *Evaluation* is the ability to be able to access the credibility of a statement or representation and to be able to logically access the relationship between statements, descriptions, questions, and concepts. *Inference* is the ability to identify and obtain the elements needed to draw a conclusion. *Explanation* (explanation) is the ability to explain or state the results of thought based on evidence, methodology and context. *Self regulation* is the ability to regulate one's way of thinking, especially in the application of the ability to analyze and evaluate (Tanti Anggia Sari.dkk, 2018: 183-195).

Assessment learningis one of the elements used to determine the extent to which student competence is achieved and the effectiveness of the learning process carried out to achieve learning objectives. *assessment* Learning needs to be done in order to find out what they have learned or show what they have not learned. *Assessment is* important to 1) clarify, share, and understand learning objectives, 2) conduct discussions, effective learning and produce evidence of learning, 3) obtain feedback for better learning, 4) activate students as sources of learning from each other, and 5) activate students as owners of their own learning (Ahmad Kamal s. et al, 2018: 244).

Several studies that are relevant to this research are: 1) Wafa Amalina's

research, in 2020 with a research entitled "Development of Middle School Students' Self-Efficacy Assessment Instruments in the Form of Test Items in Science Learning". The difference between the research that will be conducted and previous research lies in the self-efficacy assessment instrument for junior high school students, while the research that will be carried out is the development of an assessment to measure the critical thinking of junior high school students. 2) Research by Tanti Anggiansari, et al, in 2018 with a study entitled "Analysis of Critical Thinking Skills of High School Students in Kalidoni and Ilir Timur II Districts". The difference in this study is only to analyze the critical thinking skills of high school students while the research that will be carried out is the development of an *assessment* to measure critical thinking of junior high school students, and this type of research uses quantitative descriptive while the type of research that will be carried out is Borg & Gall research. 3) Titi Meilasari, in 2018 in her thesis entitled "Development of Biological Assessment Based on Creative Thinking Skills in Class X Animalia at SMA Negeri Pangkalan Lampam Oki". which will be carried out based on critical thinking, then the differences in the material to be studied are the interaction of living things with the environment, while this research is animalia, and the different types of tests carried out in this study use multiple choice questions while the research to be conducted is in the form of essays.

Based on the results of interviews and observations that have been carried out with several teachers in junior high schools throughout the city of Bengkulu, it shows that the schools where they teach use the 2013 Curriculum system, where the curriculum seeks to improve students' mindsets to become more critical. In addition, other research shows also that the assessment*(assessment)*is used for this is still not measure students' critical thinking skills which can be seen from the questions that are used in evaluating student teachers only measure on aspects of knowledge alone is not at the level of critical thinking skills.

The results of observations that have been carried out in science learning material on the interaction of living things with their environment are various kinds of exam questions consisting of Daily Test (UH) questions, Mid-Semester Exams (UTS), Final Semester Exams (UAS), National Exams (UN) , as well as Package Books. Shows that the questions that empower students to think critically in junior high schools in Bengkulu City consist of 14% SMPN A, 10% SMPN B, and 6% SMPN C consisting of 100 questions in junior high schools throughout Bengkulu City. This is because very few indicators of critical thinking skills are found in the exam questions, most of them in the assessment turn out to be many questions that only measure basic level skills, such as the types of questions that measure aspects of student knowledge in the form of questions C1, C2 and C3 only. So this proves that the critical thinking ability of students in junior high schools in Bengkulu City is still categorized as low. One of the factors is the teacher's difficulty in making the questions for the critical thinking category, for this reason it is necessary to develop an *assessment* to measure critical thinking skills in the interaction of living things with the environment in junior high schools in Bengkulu City.

METHOD

Jenis this study is a research and development *(Research and* Development) by Borg and Gall were used in this study refers to Sugiyono (2015) which is adapted to the needs penelitianmelalui8 stages, ie potential problems, information collection, product development, validation expert, expert revision, product trial, product revision, and final product.

Development research *assessment* Thiswas conducted at 3 (three) Junior High Schools (SMP) in Bengkulu City, namely SMPN 1 Bengkulu City, SMPN 8 Bengkulu City, and SMPN 19 Bengkulu City. The time of this research was carried out from the preparation stage to the implementation stage, starting in early October 2020 until March 2021. The subjects of this research were class VII students of SMPN 1 Bengkulu City, SMPN 8 Bengkulu City, and SMPN 19 Bengkulu City for the 2021/202 academic year. totaling 75 students.

This research procedure uses the Borg & Gall development model developed by Sugiyono. Research and development according to Borg & Gall is a process used to develop and validate educational products. According to Borg and Gall which states that theapproach *Research and Development* (R&D) in education includes ten steps, namely: 1) potential and problems, at this stage the emergence of problems faced, namely in the learning assessment process, educators do not carry out*assessments* that measure critical thinking skills. students. 2) product planning, product planning in the form of an *assessment* to measure critical thinking skills in the interaction of living things with the environment in junior high schools in Bengkulu City. At this stage the researcher designs a plan, namely making the initial product format of the question. 3) initial product development, product design is a description of the product that will be developed by the researcher so that it can be used as a guide to assess and make the product. 4) expert validation, this stage is an activity process to assess the product used will be more effective than the previous one. 5) expert revision, in this process improvements are made to the weaknesses contained in the product. 6) product trials, conducted to determine the assessment of students and educators. 7) product revision, based on the results of product trials, if there are errors and weaknesses, the product needs to be revised again and if the product is correct and feasible, then the next stage will be carried out. 8) the final product, after all the steps have been taken and there are no further revisions, the final product produced is an *assessment* to measure critical thinking skills in the interaction of living things with the environment in junior high schools in Bengkulu City (Sugiyono, 2015: 30). The stages of the research and development process usually form a consistent cycle to produce a certain product according to needs, through the initial product design step, initial product testing to find various weaknesses, retested, improved until finally found the product that is considered ideal (Wina Sanjaya, 2007). 2013: 129-130).

The data collection instrument used in the development research *assessment* used the needs analysis phase, expert validation, initial trial, and small group tryout. The data collection instrument was carried out by analyzing questions consisting of UH, MID, UAS, US and UN questions based on critical thinking indicators according to Facione.

Data collection techniques are carried out by distributing validation questionnaires which are arranged in making an *assessment* based on critical thinking skills to get an assessment from the validator, which aims to test the feasibility by assessment experts, material experts, linguists, and user practitioners. The data obtained were analyzed and used to revise *assessment* the developed. In this technique, the questionnaire is processed in a percentage presentation using a Likert scale as a measurement scale. The Likert scale is a method of scaling attitude questions that uses the distribution of responses as the basis for determining the value of the scale. The Likert scale is used to measure attitudes, opinions, and perceptions of a person or group of people about social phenomena that have gradations from very positive to very negative (Sugiyono, 2015: 165).

Data analysis techniques were used to calculate data fromexperts *assessment,* material experts, linguists, and practitioners of the use of science questions on the interaction of living things with the environment with the following formula:

$$P = \frac{f}{N} \times 100\%$$

Information:

P : percentage of data questionnaire

f : total score obtained

N : maximum score

The data instrument in this study tested the validity, reliability, level of difficulty, and distinguishing power which aims to determine the feasibility of the items in theproduct *assessment* developed by using the Ministep (Winsteps) Rasch program. The validity test is carried out to determine the validity of a question item, an item is said to be valid if the outfit mean square value received is 0.5 < MNSQ < 1.5, the outfit Z-standard value is -0.2 < ZSTD < +0.2 , and the point measure correlation value is 0.4 < Pt Measure Corr < 0.85. Reliability is a measuring tool to determine the constancy of an item, reliability testing is carried out using the Rasch model with reliability criteria as follows:

Table 1. Reliability Criteria

Value Reliability	Category
>0,94	Special
0,91-0,94	Very Good
0,81-0,90	Good
0,67-0,80	Quite Good
<0,67	Weak

The level of difficulty of a test subject or question is the proportion of all students who answered correctly on the test subject or question. The difficulty level of the test results is seen from the combination of the standard deviation value and the logit mean value, so that the difficulty level of the items can be grouped using Rasch modeling as follows:

Table 2. Category Level of difficulty

Value *Measure* (Logit)	category Item Difficulty Problem
***Measure* logit < -SD logit**	Very Easy Items
-SD logit ≤*Measure* logit ≤ 0	Easy Items
0 ≤*Measure* logit ≤ SD logit	Difficult Items
***Measure* logit > SD logit**	Very Difficult Items

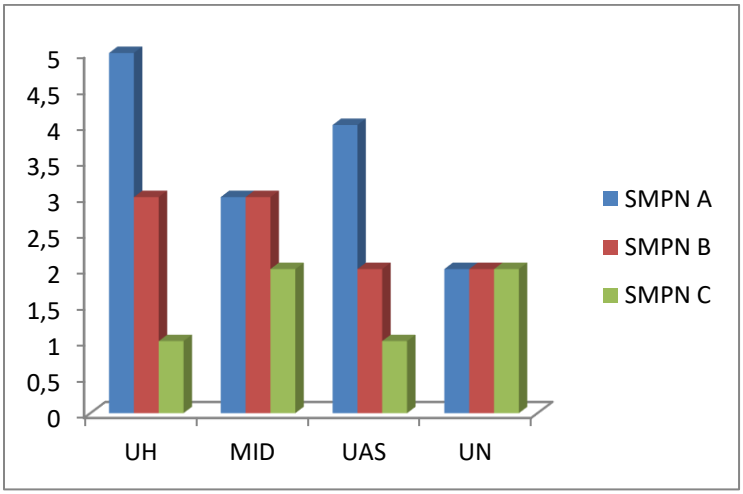
Then for the discriminatory power of an item, it is seen from *PTMEASURE-AL COOR* in Rasch modeling with the interpretation of the discriminatory index as follows:

Table 3. Distinguishing Power Index

Power Index Distinctions	Category
DP ≥ 0, 70	Very Well
0, 40 ≤ DP < 0, 70	Well
0, 20 ≤ DP < 0, 40	Pretty Well
DP < 0, 20	Bad

▪ RESULTS AND DISCUSSION

Research and development, namely *Assessment* to measure critical thinking skills matter interaction of living beings with the environment. Based on the results of interviews found in this research and development problem, there is no *assessment* to measure critical thinking skills in the interaction of living things with the environment that can train students' critical thinking patterns. In addition, as for the results of observations that have been made, various types of exam questions are obtained which consist of UH, MID, UAS, and UN questions. After collecting questions, the next step is to analyze questions based on critical thinking indicators.



. **Figure 1**. Critical Thinking Problem Analysis Diagram

Tabel 4. Hasil Analisis Soal pada Tiga SMPN di Kota Bengkulu

No	Type of Question	Non-Critical ThinkingCritical Thinking			IndicatorIndicator		
		SMPN A	SMPN B	SMPN C	SMPN A	SMPN B	SMPN C
1	UH	20	7	9	5	3	1
2	MID	3	11	7	3	3	2
3	UAS	4	7	8	4	2	1
4	UN	-	-	-	2	2	2
Total		27	25	24	14	10	6
Percentage		-	-	-	14%	10%	6%

Based on Figure 4.1 and Table 4.1 shows that the questions categorized as critical thinking in the three schools in Bengkulu City Junior High School The results are different, with the results of SMP A getting a percentage result of 14%, SMP B getting a percentage result of 10%, and SMP C getting a percentage result of 6%. From the results of the analysis of the questions that have been carried out, it can be concluded that the questions in SMP in Bengkulu City are still categorized as low.

Before conducting a small group trial, the researcher first conducted a limited trial which aimed to determine the students' readability of the questions developed in accordance with critical thinking skills. Good results are obtained from the quality of the questions and question naires given. In this study, the SPSS 26 program was used to determine the validity of a statement or question. In the results of the questionnaire, the students' responses were declared valid with the value of r table > 0.55. It can be seen that the highest total score is 0.791 with item number 4. The field trial was conducted on 25 students in three schools. Students in field trials are asked to work on questions and the results will be analyzed to determine the validity, reliability, level of difficulty, and differentiating power of the questions developed. the test results of validity, reliability, level of difficulty, and differentiating power in the *assessment* analysis using *software* Rasch's Ministep (Winsteps)which will automatically provide values of validity, reliability, level of

difficulty, and distinguishing power. The number of question items is 10 essays.

The results of the feasibility test for the validity of SMPN A showed that from the 10 question items there were 9 items that were declared valid, namely numbers 1, 3, 4, 5, 6, 7, 8, 9, 10 and 1 item that was not valid, namely number 2. SMPN B shows that from the 10 question items there are 8 items that are declared valid, namely numbers 1, 4, 5, 6, 7, 8, 9, 10 and 2 items that are not valid, namely numbers 2 and 3. that of the 10 question items there are 8 items that are declared valid, namely numbers 1,3,4,5,7,8,9,10 and 2 items that are not valid, namely numbers 2 and 6. An evaluation technique is said to have high validity (valid) if the evaluation technique can fully measure a certain expected ability (Arikunto, 2006).

In addition to testing the validity of theresearch *assessment*, a reliability test was also carried out to determine whether the items were reliable or not. At SMPN A, it shows that the results of the calculation of the reliability of the questions using the Ministep application (*Winsteps)* Rasch obtained a *person reliability* of 0.85. This shows that the results of students' answers to questions in thein *assessment are* the good category. *Item Reliability* got a value of 0.86 which indicates that the quality of the statements on thequestions is in *assessment* the good category. *Croncach Alpha* obtained a value of 0.82, which indicates that the interaction between respondents and items is in the good category. SMPN B shows that the results of the calculation of the reliability of the questions using the Ministep application (*Winsteps)* Rasch obtained a *person reliability* of 0.83. This shows that the results of students' answers to questions in thein *assessment are* the good category. *Item Reliability* obtained a value of 0.87 which indicates that the quality of the statements in thein *assessment is* the good category. *Croncach Alpha* obtained a value of 0.81, which indicates that the interaction between respondents and items is in the good category. And SMPN C shows that the results of the calculation of the reliability of the questions using the Ministep application (*Winsteps)* Rasch obtained a *person reliability* of 0.85. This shows that the results of students' answers to questions in thein *assessment are* the good category. *Item Reliability* got a value of 0.86 which indicates that the quality of the statements on thequestions is in *assessment* the good category. *Croncach Alpha* obtained a value of 0.82, which indicates that the interaction between respondents and items is in the good category. The reliability of a test essentially tests the constancy of test questions in which it is in the form of a set of questions when given repeatedly to the same object (M. Erfan, 2020).

Furthermore, to test the level of difficulty by analyzing the *Person Item Map is* carried out to determine the level of thinking of class VII students. The high and low level of difficulty of an item can be caused by the complexity of the subject matter and the condition of the students' answers (Bistok Sirait in Nani Hanifah, 2014).

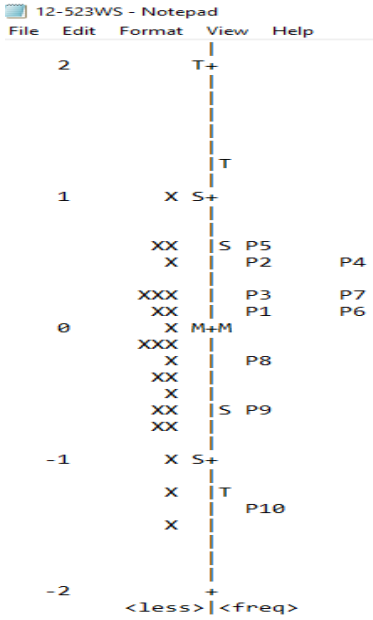


Figure 2. The results of the analysis of the *Person Item Map of* SMPN A

The level of students' critical thinking skills has several criteria. The grouping of students' critical thinking ability criteria can be seen from thevalue *Separation* seen from the *Summary Statistics*. The grouping of criteria can be seen from the results of the discriminating power of questions. SMPN A categorizes questions that are classified as difficult in questions number 1, 2, 3, 4, 5, 6 and 7, questions that are classified as easy are found in questions number 8 and 9, while questions that are classified as It is very easy to find in question number 10.

The results of the difficulty level at SMPN B show that students at SMPN B categorize questions that are classified as difficult in questions number 1, 2, 3, 4, 5, and 7, questions that are classified as moderate are found in questions number 8, relatively easy questions are found in questions number 6 and 9, while questions that are classified as very easy are found in question number 10 as shown in the image below.

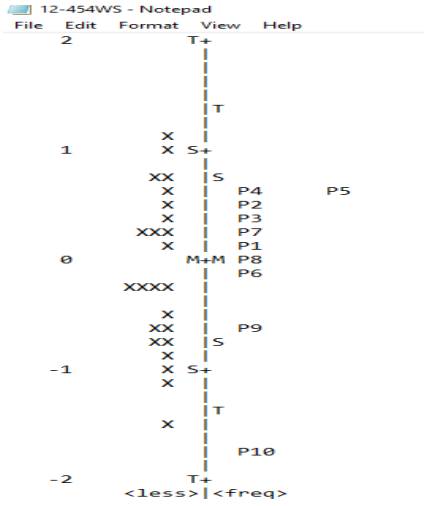


Figure 3. The results of theanalysis at *Person Item Map* SMPN B

Meanwhile for SMPN C, it shows that students at SMPN C categorize questions that are classified as difficult in questions number 1, 2, 3, 4, 5, and 7, questions that are classified as moderate are found in questions number 6 and 8, questions that are classified as easy are found in question number 10 as shown in the following picture.

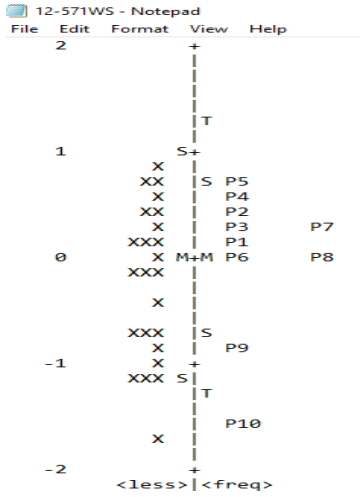


Figure 4. The results of the analysis of the *Person Item Map of* SMPN C

The discriminatory power test was carried out to determine the difficulty criteria for the items. The discriminating power of items is used to distinguish the ability of students who are able to answer high questions with students who have low ability to answer questions. In Junior A, Junior B and Junior C shows that distinguishing items that are in good category with an index of 0.40 distinguishing ≤ DP <0.70. Distinguishing power is used to identify groups of respondents based on the respondent separation index. The greater the item separation value, the better the quality of the instrument in terms of overall respondents and items (Muhammad Erfan, 2020).

▪ CONCLUSIONS

Assessment tests critical thinking are used for teaching science in the material living creatures and environment interaction calculation of validity, reliability, level of difficulty, and different power using Rasch models. This is shown in the results of the feasibility *assessment* which consists of 20 questions, there are 3 questions that are not valid. The results of the reliability analysis for the three schools in Bengkulu City obtained results of 0.82, 0.81, and 0.82 with good categories, it states that the product *assessment* developed is in good category. Thedifficulty level *item* for the three schools in Bengkulu City which is considered the most difficult is found in *items* P1, P2, P3, P4, P5, and P7, while the questions that are considered very easy are *item* P10. Results distinguishing the three secondary schools in the city of Bengkulu showed that all items categorized items either with an index of 0.40 distinguishing ≤ DP <0.70, so that the product *assessment* in the form of items can be tested.

▪ REFERENCES

Akbar, S. (2013). "*Instrumen perangkat pembelajaran*". Bandung: PT Remaja Rosdakarya.

Amalina, W. (2020). *"Pengembangan Instrumen Penilaian Efikasi Diri Siswa SMP dalam Bentuk Tes Butir Soal pada Pembelajaran IPA*". (Doctoral dissertation, IAIN Ponorogo).

Arikunto Suharsimi. (2013). "*Dasar-Dasar Evaluasi Pendidikan Edisi II*". Jakarta: Bumi Aksara.

Asmawati Sari Yuli Eka. (2018). *"Pengembangan Instrumen Asesmen Kemampuan Berpikir Kritis Siswa Pada Pembelajaran Fisika SMA Dengan Model Creative Problem Solving"*. (Doctoral dissertation, Universitas Lampung Bandar Lampung.

Astuti, W. (2019). *"Pengembangan Instrumen Tes Keterampilan Berpikir Kritis pada Konsep Fungi*. (Bachelor's thesis, Jakarta: FITK UIN Syarif Hidayatullah Jakarta).

Budiman, A & Jailani. (2014). *"Pengembangan Instrument Asesmen Higher Order Thinking Skill (HOTS) Pada Mata Pelajaran Matematika SMP kelas VIII Semester 1".* Jurnal riset pendidikan matematika volume 1 nomor 2, doi.http://journal.uny.ac.id/index.php/jrpm/article/viewFile/2671/2224).

Burlian. E. Z. (2020). *"Ilmu Alamiah Dasar, Ilmu Budaya Dasar, dan Ilmu Sosial Dasar"*. Malang: Inteligensia Media.

Departemen Agama RI. (2005). *"Al-Qur'an dan Terjemahannya".* Bandung: Syaamil Al-Qur'an.

Dharmawati, D., Rahayu, S., & Mahanal, S. (2016). "*Pengembangan instrumen asesmen berpikir kritis untuk siswa SMP kelas VII pada materi interaksi makhluk hidup dengan lingkungan*". Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan, *1*(8), 1598-1606.

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). "*Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Klasik dan Model Rasch*". Indonesian Journal Of Education Research and Review, 1(3), 11-19.

Facione, P. A. (2015). "*Critical Thinking: What It Is and Why It Counts*". Insight Assessment.

Fisher, A. (2009). "*Berpikir Kritis Sebuah Pengantar".* Jakarta: Penerbit Erlangga.

Fitriani, E. (2019). *Pengembangan Instrument Assessment Hots (High Order Thinking Skill) pada Mata PelajaraN IPS Terintegrasi Nilai-Nilai Pembangunan Karakter Kelas V SD/MI dI Bandar Lampung*. (Doctoral dissertation, UIN Raden Intan Lampung).

Fitriyani Eka. (2017). *"New Edition Big Book IPA SMP Kelas VII, VIII IX*". Jakarta: Cmedia.

Hanifah, N. (2017). *"Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi"*.Sosio e-Kons, 6(1).http://dx.doi.org/10.30998/sosioekons.v6i1.1715.

Hanifah, Nurdinah. (2019). "*Pengembangan Instrumen Penilaian Higher Order Thinking Skill (HOTS) di Sekolah Dasar*". Conference Series. 1(1), p. 005.https://ejournal.upi.edu/index.php/crecs/issue/archive.

Hernawan, Anggi. (2018). "*Pengembangan Instrumen Asesmen Berpikir Kritis pada Pembelajaran Tematik KelAS IV Sekolah Dasar di Kecamatan Tumijajar*".Journal of Chemical Information and Modeling, 7(1). http://dx.doi.org/10.23960/jpp

Ipin, A. (2018). "*Pengembangan Soal-Soal Pilihan Ganda untuk Mengukur Kemampuan Berpikir Kritis Siswa pada Konsep Sistem Regulasi Manusia untuk Jenjng SMA*". Jurnal Mangifera Edu, 3(1), 26-39.https://doi.org/10.31943/mangiferaedu.v3i1.10.

Jamaluddin, Jamaluddin, dkk. (2020) ."*Pengembangan Instrumen Keterampilan Berpikir Kritis Pada Pembelajaran IPA di SMP*." Jurnal Pijar Mipa, 15(1), 13-19. https://doi.org/10.29303/jpm.v15i1.1296.

Jasin, M. (2009). "*Ilmu Alamiah Dasar*". Jakarta: Rajawali Press.

Kartika, A. T., Eftiwin, L., Lubis, M. F., & Walid, A. (2020). "*Profil Kemampuan Berpikir Kritis Siswa Kelas VIII SMP Pada Mata Pelajaran IPA*". Jurnal Riset Teknologi dan Inovasi Pendidikan(JARTIKA), 3(1), 1-10.

Khasanah, N., Sajidan, S., Sutarno, S., Prayitno, B. A., & Walid, A. (2019). "*Critical Thinking Ability and Student's Personal Religious Beliefs: An Analysis of DBUS Model Implementation*". Tadris: Jurnal Keguruan dan Ilmu Tarbiyah, 4(1), 41-49. doi.10.24042/tadris.v4i1.4101.

Kurniawan Dian. (2020). "*Assessement Learning (AFL) dalam Pendidikan Matematika".* Yogyakarta: CV Budi Utama.

Kusaeri & Suprananto. (2012). "*Pengukuran dan Penilaian Pendidikan*". Yogyakarta: Graha Ilmu.

Kusairi, S. (2012). "*Analisis asesmen formatif fisika sma berbantuan komputer*". Jurnal Penelitian dan Evaluasi Pendidikan, 16, 68-87.

Mayasari, E. (2019). *"Pengembangan Instrumen Assesmen Berfikir Kreatif Pada Mata Pelajaran Pkn Terintegrasi Pada Nilai-Nilai Pendidikan Karakter Kelas IV SD/MI*". (Doctoral dissertation, UIN Raden Intan Lampung).

Meilasari, T. (2018). *"Pengembangan Asesmen Biologi Berbasis Keterampilan Berpikir Kreatif Pada Materi Animalia Kelas X DI SMA Negeri I Pangkalan Lampam Oki*". (Doctoral dissertation, UIN RADEN FATAH PALEMBANG).

Muhammad Erfan, dkk. (2020). "*Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Tes Klasik dan Model Rasch*". Indonesian Journal Of Education Research and Review, 3(1), 2621-8984. http://dx.doi.org/10.23887/ijerr.v3i1.24080.

Nuswowati Murbangun, dkk. (2010). "*Pengaruh Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Kimia Terhadap Pencapaian Kompetensi*". Jurnal Inovasi Pendidikan Kimia, 4(1), 566-573.

Parmin, (2012). *"Pengembangan Modul Pembelajaran IPA Terpadu Berwawasan Sains, Lingkungan, Teknologi dan Masyarakat"*. Jurnal Penelitian Pendidikan, 29(2), 132.https://doi.org/10.15294/jpp.v29i2.5654.

Priyadi, R., Mustajab, A., Tatsar, M. Z., & Kusairi, S. (2018). "*Analisis Kemampuan Berpikir Kritis Siswa SMA Kelas X MIPA Dalam Pembelajaran Fisika*". JPFT (Jurnal Pendidikan Fisika Tadulako Online), 6(1), 53-55.

Ryda. (2018). *"Silabus Materi Interaksi Makhluk Hidup Dengan Lingkungan".* https://id.scribd.com/document/379122440/Silabus-Materi-Interaksi-Makhluk-Hidup-Dengan-Lingkungan. Pada tanggal 20 Februari 2020 pukul 21.48.

Sanjaya Wina. (2013). "*Penelitian Pendidikan*". Jakarta: Kencana.

Saparudin Yudhi. (2018). *"Kemampuan Guru SMA Dalam Membuat Silabus dan Rencana Pelaksanaan Pembelajaran Berbasis Problem Based Learning*". Jurnal Penelitian Kebijakan Pendidikan, 11(3), 131-145.

Sari, N. P., Suhirman, S., & Walid, A. (2020). "*Pengembangan Modul Pembelajaran IPA Berbasis Etnosains Materi Interaksi Makhluk Hidup dengan Lingkungannya untuk Menanamkan Jiwa Konservasi Siswa Kelas VII SMP*". BIO-EDU: Jurnal Pendidikan Biologi, *5*(2), 62-73.https://doi.org/10.32938/jbe.v5i2.554.

Sari, T. A., Hidayat, S., & Harfian, B. A. A. (2018). "*Analisis Keterampilan Berpikir Kritis Siswa SMA Di Kecamatan Kalidoni Dan Ilir Timur II*". Bioma: Jurnal Ilmiah Biologi, 7(2), 183-195.https://doi.org/10.26877/bioma.v7i2.2859.

SISWANDANI, S. (2015). *"Penerapan Penilaian Berdasarkan Standar Penilaian Pendidikan (Permendikbud Nomor 66 Tahun 2013) Di Kelas Iv Sdn Junrejo 02 Kota Batu*". (Doctoral dissertation, University of Muhammadiyah Malang).

Stobaugh, R. (2013). *"Assessing Critical Thinking in Middle and High Schools".* New York London: Routledge Taylor & Francis Group.

Sugiyono. (2015). *"Metode Penelitian & Pengembangan Research And Development"*. Bandung: Alfabeta, cv. h. 30-166.

Sumintono Bambang, dkk (2015). "*Aplikasi Pemodelan Rasch pada Assesment Pendidikan*". Bandung: Trim Komunikata.

Uno, Hamzah B, & Satria Koni. (2014). "*Assessment Pembelaj1aran*". Jakarta: Bumi Aksara.

Walid, A., Sajidan, S., Ramli, M., & Kusumah, R. G. T. (2019). "*Construction of the assessment concept to measure students' high order thinking skills*". Journal for the Education of Gifted Young Scientists, 7(2), 237-251.https://doi.org/10.17478/jegys.528180.

Wijayanti, A. (2014). "*Pengembangan Autentic Assesment Berbasis Proyek Dengan Pendekatan Saintifik Untuk Meningkatkan Keterampilan Berpikir Ilmiah Mahasiswa*." Jurnal Pendidikan IPA Indonesia, Vol. 3, No. 2, h. 102–8. https://doi.org/10.15294/jpii.v3i2.3107.

Wiyoko, T. (2019). "*Analisis Profil Kemampuan Berpikir Kritis Mahasiswa PGSD Dengan Graded Response Models Pada Pembelajaran IPA*". IJIS Edu: Indonesian Journal of Integrated Science Education, 1(1), 25-32.http://dx.doi.org/10.29300/ijisedu.v1i1.1402.