

DEVELOPMENT OF BIOLOGICAL ASSESSMENT BASED ON STUDENT CRITICAL THINKING SKILLS: a case study in eleventh grade at MAN 2 Palembang

Aseptianova¹, Sulton Nawawi², Resti Maya Astia³

**^{1,2,3} Program Studi Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Palembang
aseptianova@gmail.com¹, sulton.bio@gmail.com², restimayaa@gmail.com³**

Accepted: February, 25, 2020

Published: April, 30, 2020

DOI: 10.21107/jps.v7i1.5113

ABSTRACT

This study aims to determine the characteristics and feasibility of biology assessment based on critical thinking skills. This research is a research and development that adapts to the McIntire & Miller development model. The sample used in the first trial was 80 students, and in the second trial were 76 students. Data collection techniques using questionnaire sheets, questionnaire sheets, interview sheets, & test instruments in the form of essay questions of biology grade XI based on critical thinking from Facione with indicators: interpretation, analysis, evaluation, explanation, conclusions, and self-regulation. Techniques of data analysis using qualitative analysis and quantitative analysis. The results of this study were in the form of a biology assessment based on critical thinking skills in the form of essays in odd semester XI material and from the results of the 1st and 2nd trials totaling 24 items. The content validity of biology assessment based on critical thinking skills is qualified to very high. The construct validity the items are 28 questions valid (1st trial) with reliability value is 0.875. Furthermore, 24 questions are valid (2nd trial) with reliability value is 0.764. The results of the level of difficulty test are 0.332 (1st trial) and 0.537 (2nd Trial). The value of discriminating power is 1,541 (1st trial) and 2,021 (2nd trial). The data shows that the items can be used to measure students' critical thinking skills. We can involve the mental processes of pupils by the test based on critical thinking skills.

Keywords: *biological assessment, critical thinking skills.*

¹ Corresponding Author

Development of Biological Assessment

Introduction

Education is an attempt to improve the atmosphere of learning and the learning process so that students actively develop their potential to have the intelligence, noble character, and the skills needed by themselves. Quality education supports the creation of people who can think critically in the era of globalization. One of them is by developing students' skills, potential, creativity, and ability in education.

The survey results regarding student achievement in Indonesia are still low. Sulastri, Johar, & Munzir (2014) stated that the data from PISA (Program for International Student Assessment) ranked Indonesia was always in the top five in the lower group. That was caused by students not yet habitually solving high-level questions, especially critical thinking questions.

Learning currently has implemented the 2013 curriculum, which is part of 21st Century education. In the 21st Century competence, requested that the best quality of learning in critical thinking. Students are expected to think critically in identifying, understanding, solving problems, and applying learning material (Nawawi, 2017). Critical thinking is the ability to interpret and evaluate, skillfully, and actively through observation and communication, information, and argumentation (Fisher, 2008).

There are six indicators of critical thinking skills according to Facione (2013), namely 1) interpretation, namely understanding the meaning and significance of various situations, data or events; 2) analysis, i.e. identifying the true intentions and conclusions of the relationship between statements, concepts, descriptions, or forms of statements that are expected to express trust, despair, experience, reasons, information or opinions; 3) evaluation, which is the ability to assess the credibility of a statement or presentation by assessing or describing the perceptions of others, including: experiences, situations, decisions, beliefs, and assessing the logical strength of the expected inferential relationship or the actual inferential relationship between statements, descriptions, statements, or other forms of representation; 4) conclusions, namely the ability to identify and choose the elements needed in forming reasonable conclusions or to form hypotheses by paying attention to relevant information and reducing the consequences arising from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, and other forms of representation; 5)

explanation, namely stating one's position or justifying position based on evidence, criteria, or contextual aims to convince and use insight criteria that support the decision; 6) self-regulation, is the ability to state the results of a person's consideration process or the ability to justify that a reason is based on evidence, concepts, methodology, a certain criteria, and a reasonable balance; and is the ability to present one's reasons in the form of convincing arguments.

Critical thinking is important for students to have in the learning process because critical thinking can train students' ability to think more critically and more optimally. Students need to be encouraged to emerge critical thinking skills in themselves, and it is necessary to carry out measurements and assessments. Assessment is one of the main components to determine the potential of students from the learning process.

Based on preliminary data collection, it shows that teachers at the school still experience difficulties in conducting curriculum-based assessments in 2013. In addition, teachers, not many people, know how to make and use instruments that fit the critical thinking dimension. It is proven by the questions developed by the teacher in the school that still uses questions in the cognitive domain C1 – C3.

On the other hand, the mastery value of the material especially the first semester material at MAN 2 Palembang seen from the national exam data in the 2017/2018 academic year that the percentage of mastery of cell material obtained value of 65.04%, plant tissue system 57.72%, animal tissue system 42.28%, the motion system 47.97%, and the circulatory system with a value of 48.78%. Therefore, this research develops critical thinking evaluation in the material of the first semester. That is because the value is still low and has not yet reached the minimum completeness criteria value.

Ineffective assessment makes the lack of exploring the potential in measuring students' critical thinking skills, so that more effective assessment instruments are needed to explore the potential abilities of students' critical thinking skills. Therefore, it is necessary to develop essay questions that measure up to stage C4 – C5 so that students can express ideas from each given problem, and can develop a critical thinking ability possessed by each student.

Research Methods

The research used is Research and Development (R & D), while the one developed was in the form of a Biology assessment based on critical thinking skills. The procedure used in developing a Biology assessment that is following the development model of McIntire (Mulyatiningsih, 2011) consists of 10 stages, such as 1) explain the competencies and objectives of the test to be achieved by test takers, 2) develop a test design, 3) arranging question items, 4) write the test instructions, 5) perform trials on tests that have been prepared, 6) revise the test items, 7) conduct test item analysis, 8) validate test questions, 9) establish reference norms, and 10) complete the manual test.

The instruments used in this study can see in table 1.

Table 1. Research instrument

Research Instrument	Definition
Questionnaire sheet	The teacher questionnaire sheet is used for data retrieval regarding the information on how the level of questions the teacher uses in the learning process, and there are 15 question items.
Interview sheet	The interview sheet serves to obtain information about the way teachers in the assessment of critical thinking skills carried out during the learning process and find out the difficulties of teachers in measuring critical thinking skills.
Validation sheet	The validation sheet is shown to the material expert lecturer and the item development expert, which is useful to find out whether or not the researcher has made the questions. Validation is providing question scripts, assessment rubrics, and question lines.
Tests	Essay questions used to measure the critical thinking skills of students in class XII who have studied class XI material.

Data analysis techniques used in this study are qualitative and quantitative analysis.

Qualitative analysis

Based on data scores from expert judgment validation, an analysis was conducted using Aiken statistics (Azwar, 2017) (see formula 1).

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

Information:

s = r-lo

lo = the lowest rating number

c = highest rating score

r = rating number given by the expert

Table 2. Validity Criteria

No	Validity Result	Validity Criteria
1	$0,81 < V \leq 1,00$	Very High
2	$0,61 < V \leq 0,80$	High
3	$0,41 < V \leq 0,60$	Moderate
4	$0,21 < V \leq 0,40$	Low
5	$0,01 < V \leq 0,20$	Very Low

(Mardliya, Abdurachman, & Hartono, 2017)

Quantitative Analysis

Construct Validity

Analysis of the validity of the item items in this study using SPSS 16. The validity test assesses whether the item items are valid or not, through the significance test, that is the t-test using degrees of freedom 'n-2' where 'n' is the number of subjects. The use of a certain level of significance will be known as t-theoretical value through t tables, which are widely included in various statistical books and research methods (Aritonang, 2008).

Reliability

Reliability is the level or degree of consistency of an instrument. To calculate the reliability test in this study using Cronbach's Alpha statistics on SPSS 16. The reliability values were interpreted in Table 3.

Tabel 3. Alpha Cronbach's Value

Level of Reliability	Criteria
<0,5	Bad
0,5-0,6	Poor
0,6-0,7	Moderate
0,7-0,8	Good
>0,8	Very good

(Sumintono & Widhiarso, 2014)

Level of Difficulty

The level of difficulty is determined, according to Arifin (2016) (see formulas 2 and 3).

$$\text{Score Average} = \frac{\text{number of learners' scores per question}}{\text{number of learners}} \quad (2)$$

$$\text{Level of difficulty} = \frac{\text{score average}}{\text{maximum score per question}} \quad (3)$$

The determination of the difficulty index criteria in the items is determined by the values shown in Table 4.

Development of Biological Assessment

Table 4. Level of Difficulty Criteria

Score	Description
0,00–0,30	Difficult
0,31–0,70	Moderate
0,71–1,00	Easy

(Arifin, 2016)

Test of Discriminating Power

According to Arifin (2017), for the form of description of the techniques used to calculate the discriminating power with formula 4.

$$DP = \frac{\bar{X}KA - \bar{X}KB}{SM} \cdot 4$$

Information:

DP = Discriminating power

$\bar{X}KA$ = the average of the upper group

$\bar{X}KB$ = the average of the lower group

SM = Maximum Score

Table 5. The Criteria of Discriminating Power

Score	Description
0,00–0,19	Poor
0,20–0,39	Satisfactory
0,40–0,69	Good
0,70–1,00	Excellent
Tanda Negatif (-)	Not Good

(Arikunto, 2018)

Result and Discussion

The results obtained from this study include the validity test of the lecturers of material experts as well as expert lecturers on critical thinking skills. The results of the validation from the lecturer are presented in Table 6. The trial is conducted two times with students in MAN 2 Palembang. At the trial stage, the results obtained from the students' answers are analyzed to get the results of the validity test, reliability test, the level of difficulty, and discriminating power.

Table 6. The Result of Expert Judgment Validation

No	Description	No. Item	Total
1	Very High	2,3,7,8,11,12, 13,14,15,16,17,19,21, 22,23,24,25,26,28,29	20
2	High	1,4,6,9,27,30	6
3	Moderate	5,10,18,20	4

The data in table 6, it can be concluded that of the 30 questions that have been validated by the expert. It shows that the questions that have very high information are 20 items with a statistical value of 1.00, there are six items with high criteria with a value of 0.75. There are 4 Item items with moderate criteria that are 0.50. It can be said that the overall questions in the criteria are very high, which means the questions are valid according to what they want to be measured and deserve to be tested on students. However, there are some

suggestions and input from experts that need to be corrected, as in Table 7.

Table 7. Suggestions from expert judgments

No.	Suggestion from Expert Judgments	
	Expert in Test Development	Expert in Subject Matter
1	Pay attention in making sentences	The image improvement must be clear and understandable
2	Note that punctuation must know when to use	Improvements in language because they do not use communicative language
3	The picture in the question gives its source	The assessment instrument matrix has fulfilled the valid rating scale without revision and hopefully useful
4	Do not use attractive stimuli.	
5	The answer is not implicit in the stimulus.	
6	There are still questions that are made that do not yet fit the material used.	

Suggestions from experts are used to improve items that have been developed before being trialed to students. Trials carried out twice, trials 1 and 2.

Table 8. Construct Validity - 1st Trial

No. Valid Item	Number of item valid	No. Item Invalid	Number of Item Invalid
1,2,3,4,5,6, 9,10,11,12, 13,14,15,16, 17,18,19,20,2 1,22,23,24,25 ,26,27,28,29 & 30	28	7 & 8	2

Based on Table 8, the results of the validity test items of critical thinking skills tested in MAN 2 Palembang that the validity of items from 30 items categorized as valid amounted to 28 questions and questions categorized as invalid amounted to 2 questions. Invalid question items are caused because of the difficulty level of the questions that are classified as difficult to make students difficult to answer these questions, and the results are not optimal. According to Bagiyono (2017), outcomes are stated as good if the item is neither too difficult nor too easy. Therefore, items that cannot be answered correctly by all trainees because they are too difficult can be declared as bad and invalid items.

Furthermore, the data are analyzed to determine the value of reliability. The reliability

Aseptianova, Nawawi, & Astia

test is used to determine the extent to which the results of a measurement process can be consistent or can be trusted from this study. The reliability test will be presented in table 9.

Table 9. Reliability – 1st Trial

Alpha Cronbach's Value	Number of Items	Criteria
0,879	30	Very Good

Based on Table 9, it can be seen that overall of the item items developed, these questions are reliable with an alpha Cronbach's value of 0.879 with very good criteria, which means that the test instruments developed are authentic.

This was also expressed by Nurjanah & Marlianingsih (2015), that measurements that have high reliability or are very good are called reliable or good measurements. Next, the level of difficulty to find out whether the items are categorized as difficult, moderate, or easy. The Level of Difficulty will be presented in table 10.

Table 10. Level of Difficulty – 1st Trial

Level of Difficulty	No. Items	Number of Items
Difficult	2,3,4,8,9,10,14,15,16,19,20,21,24,25,26,29,30	17
Moderate	1,5,6,7,11,12,13,17,18,22,23,27,28	13
Easy	-	-

Based on Table 10, there are several categories of levels of difficulty at the trial stage, which are divided into three categories, namely questions that are classified as easy, medium, and difficult. The level of difficulty of the critical thinking skills test instrument is known that of the 30 questions tested, and there are 13 questions in the medium category and 17 questions in the difficult category. Good questions must have a balance between difficult, moderate, and easy difficulty categories.

Furthermore, the discriminating power test is performed to see whether the questionability can distinguish between high-ability students and low-ability students. The discriminating power will be presented in table 11.

Table 11. Discriminating Power – 1st Trial

Level of Discriminating power	No. Items	Number of Items
Poor	-	-
Satisfactory	8 & 16	2
Good	18	1
Excellent	1,2,3,4,5,6,7,9,10,11,12,13,14,15,17,19,20,21,22,23,24,25,26,27,28,29,30	27

Based on Table 11, there are several discriminating power, such as bad, satisfactory, good, and excellent. The results of the analysis of distinguishing questions on the first test were 30 items, and it was found that there were no bad items, there were two items with enough categories, 1 item with good category and 27 questions with a very good category.

Developing a test product that will be trained on students must have good items, good items if they have excellent differentiator value categories, to determine whether or not an item can distinguish between high-ability and low-ability test takers. After the 1st trial, a second trial is conducted, which aims to determine the level of consistency in the item. The results of the validation of the second trial construct can be shown in Table 12.

Table 12. Construct Validity – 2nd trial

No. Valid Item	Number of items valid	No. Item Invalid	Number of Item Invalid
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25	24	18, 26, 27, 28, 29, 30	6

Based on the results of the validity test questions about critical thinking skills in MAN 2 Palembang in the second trial, it can be seen that there are 24 valid item items while six questions are declared invalid. When a test does not have good validity or the question is invalid, many things can affect the problem to be invalid, namely items that cannot test the ability of students. Then from the difficulty level of questions that are too easy, and some questions are considered severe, so students are challenging to answer that question.

The invalid items are eliminated or discarded the items, while valid items can be used for further trials. Furthermore, a reliability test will be conducted on the second trial using class XII samples, namely XII IPA 1 and XII IPA 2. The following results of the second trial will be presented in Table 13.

Table 13. Reliability - 2nd trial

Alpha Cronbach's Value	Number of Items	Criteria
0,694	24	Cukup

Based on Table 13, the results of the analysis using the Alpha Cronbach's formula obtained a score that is categorized sufficiently with a value of 0.694, meaning that the items developed are quite real and feasible to use. It is in line with the opinion of Purwanti (2014), which states that a good question as an evaluation tool is that it has high reliability or with sufficiently reliable information. The difficulty level test is performed on the second trial, which can be presented in Table 14.

Table 14. Level of Difficulty – 2nd Trial

Level of Difficulty	No. Items	Number of Items
Difficult	25	1
Moderate	2,3,4,6,7,8,9,10,12,13, 14,15,16, 17,19,20,21,22, 23	19
Easy	1,5,11,24	4

Based on the results of the second trial in Table 14 shows that there are several categories of difficulty levels, namely, the questions are classified as easy, moderate, and difficult. About four items are categorized as easy categories. About 19 items are categorized as moderate. The difficult questions category amounts to 1 question. A good question if it's not too easy and too difficult. According to Bagiyono (2017), in general, an item about evaluating learning outcomes is stated to be useful if the item is neither too difficult nor too easy. Therefore, questions that cannot be answered correctly by all participants because it is too complicated can be stated as items that are not good.

Conversely, items that all participants can answer correctly because too easily can also be stated as items that are not good. The preparation of test instruments requires a balance of the level of difficulty of the test. The balance between easy, medium, and difficult categorized questions must be proportional. Then a re-analysis of the test of discriminating power was performed in the second trial in table 15.

Table 15. Discriminating Power – 2nd Trial

Level of Discriminating power	No. Items	Number of Items
Poor	-	-
Satisfactory	-	-
Good	-	-
Excellent	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25	24

Based on the results of Table 15, there are discriminating categories, namely the poor, satisfactory, good, and excellent. In the second trial, items that were in the bad, enough, and good categories were absent. The results of the second test of discriminating power found that the item has very good. It shows that the questions developed can distinguish students' abilities.

After obtaining the results of the first and second trials, the final results of the questions will be made into the final product, amounting to 24 questions out of 30 items in class XI odd semester material. The results of this final product are in the form of a question cover, preface, table of contents, history of the development of the assessment, background, and instructions for using the test, questions, and bibliography. An overview of the products that have been designed can be seen in Figure 1.

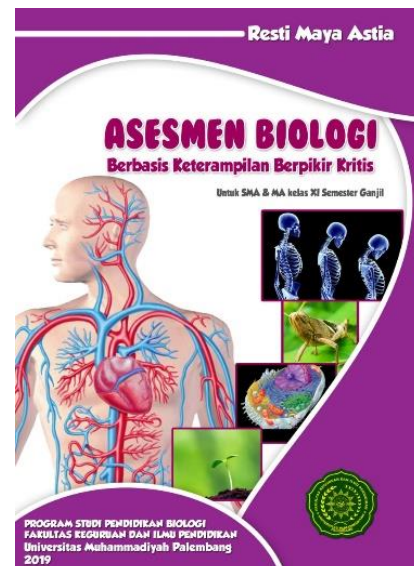


Figure 1. Cover of Biological Assessment

Conclusion

The development of Biology assessment based on critical thinking skills is obtained the final results; namely, 24 valid essay questions can

Aseptianova, Nawawi, & Astia

use as final products in research. That has fulfilled the requirements of validity, reliability, level of difficulty, and different matter power. The suggestions for further researchers are expected to pay more attention to the questions to be made and the time to be used later.

References

- Arifin, Z. (2016). *Evaluasi Pembelajaran*. Bandung.
- Arifin, Z. (2017). Kriteria Instrumen dalam Suatu Penelitian. *Jurnal Theorems (The Original Research of Mathematics)*, 2 (1): 28
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- Aritonang, R. (2008). Akademika. *Jurnal Pendidikan Universitas Tarumanagara*, 10 (2): 165-166.
- Azwar, S. (2017). *Metode Penelitian Psikologi*. Yogyakarta: Pustaka Pelajar.
- Bagiyono. (2017). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat 1. *Jurnal Pelatihan* , 16 (1): 3-4.
- Facione, P. A. (2013). Critical Thinking: What It Is and Why It Counts. *Insight Assessment*.
- Fisher, A. (2009). *Berpikir Kritis Sebuah Pengantar*. Jakarta: Erlangga
- Mardliya, Abdurachman, & Hartono. (2017). Pengembangan Instrumen Penilaian Keterampilan Proses Sains Dasar Mata Pelajaran Kimia pada Kompetisi Dasar Kelarutan dan Hasil Kali Kelarutan di SMA, *Prosiding Seminar Nasional Pendidikan IPA*.
- Mulyatiningsih, E. (2011). *Riset Terapan Bidang Pendidikan dan Teknik*. Yogyakarta: UNY Press.
- Nawawi, S. (2017). Developing of module challenge based learning in environmental material to empower the critical thinking ability. *Jurnal Inovasi Pendidikan IPA*, 3(2), 212-223.
- doi:<https://doi.org/10.21831/jipi.v3i2.15988>
- Nurjanah & Marlianingsih. (2015). Analisis Butir Soal Pilihan Ganda dari Aspek Kebahasaan. *Jurnal Ilmu Pendidikan*, 2 (1): 74.
- Purwanti, M. (2014). Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010. *Jurnal Pendidikan Akuntansi Indonesia*, 12 (1): 88.
- Sulastri, R., Johar, R., & Munzir, S. (2014). Kemampuan Mahasiswa Program Studi Pendidikan Matematika FKIP Unsyiah Menyelesaikan Soal PISA Most Difficult Level. *Jurnal Didaktik Matematika*. 1 (2): 14.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model RASCH untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: Trim Komunikata Publishing House.