Analysis of Biology Summative Test Items at Odd Semester of Class 11 MIA in Methodist – 8 Medan Senior High School

Ivana Odelia Nainggolan, Ely Djulia and Golarden Purba

Analysis of Biology Summative Test Items at Odd Semester of Class 11 MIA in Methodist – 8 Medan Senior High School

Ivana Odelia*, Ely Djulia, Golarden Purba

Biology Education Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Willem Iskandar Psr. V Medan Estate, Medan, Indonesia, 20221.

*Author's Corespondence: ivanadel.11u@gmail.com

ABSTRACT - Analysis of the test items is an activity to determine the quality of the test items. The result can be used to improve the items that have been made. Some teachers rarely even never analyze the items that used for the test. The purpose of the research is to determine the quality of biology summative test items at odd semester of class 11 MIA in Methodist - 8 Senior High School Academic Year 2020/2021 and 2021/2022 by quantitative, qualitative and the level distribution of cognitive domain taxonomy bloom. This descriptive research including all the students of class 11 MIA academic year 2020/2021 and 2021/2022 as the subject and the biology summative test item as the object of the research. Based on the results of the analysis of test items 11 MIA Academic year 2020/2021, there are no test items of very good quality, there are 2 good quality test items. items, and there are 6 test items that have very bad quality. Then, based on the results of the analysis of the 11 MIA Academic year 2021/2022 test items, there were no test items that were of very good quality, a total of test items with good quality were included in the test items that had medium quality, and a total of bad quality test items. 13 test items, and there are 6 test items that have very bad quality. In qualitative analysis, all test items have entered the category of very good questions based on the content aspect and language with a percentage of results of 100%. But 2 test items (8%) fulfill all construction aspects and 23 test items (92%) not meet one construction aspect. In cognitive level of Taxonomy Bloom with academic year 2020/2021, C1 are 13 test items (52%), C2 are 7 test items (28%), C3 and C4 each with 2 test items (8%), C5 is 1 test item (4%). The test items of academic year 2021/2022, C1 are 10 test items (40%), C2 are 7 test items (28%), C3 are 2 test items (8%), C4 are 5 test items (20%), and C5 is 1 test item (4%).

I. Introduction

The quality of education is certainly influenced by the role of teachers in learning in schools. Teachers who direct the learning process follow the learning objectives that have been set. that teachers must also evaluate their students in addition to teaching and guiding students. It aims to determine the extent to which the learning process implemented has been successful. Evaluation is an activity that cannot be separated from a program, including learning programs in schools. Evaluation is a systematic process to determine the success and efficiency of the program concerned. Evaluation is usually using the test items. The test items that must give answers and responses to measure a person's level of ability. The test is a technique that teachers often use to determine the extent to which students are able to master the material. In addition to having the ability to teach or deliver material, teachers must also have the ability to develop test instruments so that the test includes good questions to measure the abilities of their students. A good test will provide an accurate picture of student learning outcomes. In other words, a good test must be a quality test. Learning outcomes test is said to be good if the test is valid, reliable, objective, and practical. A test items is said to be good if it is able to measure what is to be measured with an accurate value. If the test items are not analyzed and the question given is not yet known about the condition, it could be that the question is not suitable for use in the exam. Analyzing items is a process of collecting, summarizing, and using information from students' answers to make decisions about each assessment. Item analysis shows why an item does not work well and how much it works with analytical process quantitatively and qualitatively. This research has been done at the Methodist 8 Medan Senior High School. There are some problems in this research, first the biology summative test items have not been analyzed quantitatively including validity, reliability, discriminating index, difficulty index, and distractor function, have not been analyzed qualitative including subject aspects, construction aspects, and language aspects, and based on the level distribution of cognitive domain Taxonomy Bloom. It can be concluded that the summative test items that he has made cannot be said to be test items that have quality. The summative test consists of the items that are used as evaluation subjects. However, the items

often do not go through the quality analysis stage, so the quality of the questions prepared is not known. Questions that have qualified are the questions that can describe the actual abilities of students and provide precise information, and from the results of students, it can be seen that students who have and have not mastered the subject. Based on the description above, the researcher did the research with the title Analysis of Biology Summative Test Items at Odd Semester of Class 11 MIA in Methodist – 8 Medan Senior High School.

II. Research Methods

The type of research used in this research is descriptive. The population was used is biology summative test items at odd semester of class 11 MIA class 11 MIA that was constructed by the biology teacher in Methodist – 8 Medan Senior High School. Sampling was done by total sampling. The subjects of this study were all students of class 11 MIA in Methodist – 8 Medan Senior High School for the Academic Year 2020/2021, totaling 79 students and 2021/2022 totaling 73 people and the objects of this research are the test items of summative tests at odd semester for the academic year 2020/2021 and 2021/2022. This research data obtained by analysis and validation. The data was analyzed based on quantitative aspects using *Microsoft excel*, cognitive domain of Taxonomy bloom and categories of the qualitative aspects by validator. The data analyzed based on :

A. Quantitative aspects :
a. Validity

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{P}{q}}$$

Information :
r_{pbi} = Biserial point correlation coefficient
M_p = Standard deviation of total score
M_t = The average score of the testee
SD_t = The average score of the total score
P   = The proportion of testees who answered correctly to the items being tested for the validity of the items
q   = The proportion of testees who answered incorrectly to the item being tested for the validity of the item

In the validity, the the value of the correlation coefficient/validity can be seen that the correlation coefficient is divided into five criteria :

Table 1. Correlation Coefficient/Validity

No	Value Range	Criteria
1	0,81 – 1,00	Very High
2	0,61 – 0,80	High
3	0,41 – 0,60	Medium
4	0,21 – 0,40	Low
5	0,00 – 0,20	Very Low

b. Reliability

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{S^2 - \sum pq}{S^2}\right)$$

Information :
r_{11} = Overall test reliability
p   = Proportion of subjects who answered the item correctly
q   = Proportion of subjects who answered the item incorrectly (q = 1 - p)
$\sum_{pq}$ = The sum of the products of p and q.
n   = Number of items
S^2 = The standard deviation of the test (the standard deviation is the root of the variance).

The reliability criteria are divided into five, namely :

Table 2. Reliability Criteria

No	Value Range	Criteria
1	0,90 - 1,00	Very High
2	0,70 - 0.89	High
3	0,40 - 0,69	Medium
4	0,20 - 0,39	Low
5	0,00 – 0,19	Very Low

c. Difficulty Index

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

Information :
P   : Difficulty index
B   : Number of participants who answered the test item correctly
JS   : Total number of students

The difficulty level category is divided into 3 categories as shown in the table above :

Table 3. Category of Difficulty index

No	Value of p	Category
1	$0,00 \leq 0,30$	Hard
2	$0,31 \leq p \leq 0,70$	Medium
3	$0,71 \leq 1,00$	Easy

d. Discriminating Index

$$D = \frac{n_{it}}{N_T} - \frac{n_{ir}}{N_R}$$

Information :

D : Discriminating index

n_{it} : the number who answered the test items correctly from the high group

N_T : the number who answered from the high group

n_{ir} : the number who answered the test items correctly from the low group

N_R : the number who answered from the low group

The discriminating index criteria are divided into four :

Table 4. Discriminating Index Criteria

No	Discriminating Index	Criteria
1	$0,00 – 0,20$	Poor
2	$0,21 – 0,40$	Fair
3	$0,41 – 0,70$	Good
4	$0,71 - 1,00$	Excellent

e. Distractor Function

$$IP = \frac{P}{\frac{N - B}{n - 1}} \times 100\%$$

Information :

IP = Distractor index

P = Number of students who have distractors

N = Number of students taking the test

B = The number of students who answered correctly on each question

n = Number of alternative answers (options)

l = Fixed number

The distractor criteria which are divided into 5 categories :

Table 5. Distractor function

No	IP Value	Criteria
1	76% - 125%	Very Good
2	51% - 75% or 126% - 150%	Good
3	26% - 50% or 151% - 175%	Less Good
4	0% - 25% or 176% - 200%	Bad
5	More than 200%	Very Bad

B. Qualitative aspects :

$$\frac{The\ number\ of\ suitable\ aspects}{The\ number\ of\ all\ aspects\ studied} \; x\ 100\%$$

Informations of the Category :

90% - 100%   : Exellent

80% - 89%   : Very Good

70% - 79%   : Good

60% - 69%   : Bad

$\leq$ 59%   : Very Bad

In qualitative research, there are 3 aspect, namely content aspect; the test items were in accordance with the indicators, the material asked was in accordance with competence, the answer choices were homogeneous and logical, and there was only one answer for each test item. Then language aspects; the subject matter is formulated briefly, concisely, and firmly, the formulation of the main question and the answer choices are statements that are multiple, the subject matter is free from statements that are double negative, pictures/graphics /tables/diagrams/etc. on the questions are clear and functional, the length of the answer choices is relatively the same, and the items do not depend on the answers to the previous questions. And language aspects; the test items have used a language that is in accordance with the Indonesian language rules, used communicative language and the answer choices did not repeat the same word or group of words, unless it was a unified understanding.

C. Taxonomy Bloom

$$SKD = \frac{\sum s}{\sum Js} \; x\ 100\ \%$$

Information :

SKD : Cognitive domain category scores

$\sum s$: The number of items that give rise to the types of cognitive domain categories

$\sum Js$: The total number of Biology summative test items

There are 6 (six) levels of thinking processes in the cognitive aspects of Bloom's taxonomy, starting from the lowest level to the highest level, namely Remembering (C1), Understanding (C2), Applying (C3), Analyzing (C4), Evaluating (C5), and Creating (C6). The data research will inventory into a checklist table for cognitive analysis of Taxonomy bloom.

III. Result and Discussion

A. Quantitative Research

Based on the results of the analysis of test items 11 MIA Academic year 2020/2021 on quantitative research, there are no test items of very good quality, there are 2 good quality test items. items, and there are 6 test items that have very bad quality. Then, based on the results of the analysis of the 11 MIA Academic year 2021/2022 test items, there were no test items that were of very good quality, a total of 3. 3 test items with good quality were included in the test items that had medium quality, and a total of bad quality test items. 13 test items, and there are 6 test items that have very bad quality.

Table 5. Quality Criteria of the Test Items

No	11 MIA Academic Year	Quality Criteria of the Test Items	Number of Test Items	Total	Percentage
1	2020/2021	Very Good	0	0	0%
		Good	5,17	2	8%
		Medium	18	1	4%
		Bad	1,2,3,4,6,8,9,11, 12,14,15,16,20, 22,24,25	16	64%
		Very Bad	7,10,13,19,21,23	6	24%
2	2021/2022	Very Good	0	0	0%
		Good	4,11,15	3	12%
		Medium	1,20,24	3	12%
		Bad	3,5,6,7,9,12,13, 14,16,19,21,22, 23	13	52%
		Very Bad	2,8,10,17,18,25	6	24%

Based on the table, the summative test items of 11 MIA with academic year 2020/2021 and 2021/2022 have not been able to carry out their functions properly. The failure of test items in carrying out their functions is caused by not fulfilling one of the parameters of the question. The cause of failure can be identified through aspects of validity, difficulty index, discriminating index, and effectiveness of distractors. The following is a description of the reasons for the failure of these items.

Table 6. Reason of Rejected of The Test Items

No	11 MIA Academic Year	Reason of Rejected	Total	Percentage
1	2020/2021	Validity	21	84%
		Difficulty Index	23	92%
		Discriminating Index	23	92%

		Distractor Function	17	68%
2	2021/2022	Validity	13	52%
		Difficulty Index	24	96%
		Discriminating Index	22	88%
		Distractor Function	12	48%

Based on the table above, it can be seen that the biggest causes of failure are the difficulty index, discriminating index. This shows that the question has not been able to distinguish students who have high abilities and low abilities. Furthermore, the items used are still relatively easy. Based on the discussion above, it can be concluded that the summative test items of 11 MIA with academic year 2020/2021 and 2021/2022 do not have good quality. This is because 9 test items need to be revised, 41 test items cannot be used or discarded, and there are no test items of good quality.

B. Qualitative Research

Based on the analysis of the validator, that all questions meet all aspects of the contain. This shows that all test items have entered the category of very good questions based on the content aspect and language aspect with a percentage of results of 100%. But, that not all questions meet all aspects of construction. There are only 2 test items that meet the 4th point aspect, namely numbers 3 and 17. In addition, there are no other test items that meet this aspect. This shows that only 2 test items (8%) fulfill all construction aspects and 23 tests items (92%) that do not fulfill one aspect of construction, namely pictures / graphs / tables / diagrams / etc. on the questions are clear and functional, the length of the answer choices is relatively the same. This is because only test items with numbers 3 and 17 show the images.

Table 7. Result of Researched Aspect of Contain

No.	Researched Aspect	Total of The Test Items
1	Question according to the indicator	25
2	The subject asked is in accordance with competence	25
3	Answer choices are homogeneous and logical	25
4	There is only one correct answer key	25

Table 8. Result of Researched Aspect of Construction

No.	Researched Aspect	Total of The Test Items
1	Question points are formulated briefly, clearly, and concisely	25
2	The formulation of the main questions and the answer choices are the only	25

	questions that are needed	
3	Question points are free from double negative questions	25
4	The pictures / graphs / tables / diagrams / discourses / etc in the questions are clear and functional	2
5	The length of the answer choices is relatively the same	25
6	Test items do not depend on previous test answers	25

Table 9. Result of Researched Aspect of Language

No.	Researched Aspect	Total of The Test Items
1	Using language that is in accordance with the rules of the Indonesian language	25
2	Use communicative language	25

| 3 | Answer choices do not repeat the same word/word group | 25 |

C. Taxonomy Bloom

The results of the analysis of the distribution of cognitive aspects of Bloom's taxonomy that have been carried out by validator. Regarding the summative test items with the academic year 2020/2021, it is known that the test items that have been used in the summative test are not good. This is because the cognitive level of the dominant test items is C1 (Remembering) with 13 test items (52%) and C2 (Understanding) with 7 test items (28%) and this level is included in the easy category. Cognitive level in the medium category, namely C3 (Applying) C4 (Analyzing) each with 2 test items (8%). And the cognitive level with the hard category, namely C5 (Evaluating) as much as 1 test item (4%), C6 (Creating) as much as 0 (0%). Then, regarding the summative test items with the academic year 2021/2022, it is known that the test items that have been used in the summative test are also not good. This is because the cognitive level of the dominant test items is C1 (Remembering) with 10 test items (40%) and C2 (Understanding) with 7 test items (28%) and this level is included in the easy category. Cognitive level in the medium category, namely C3 (Applying) with 2 test items (8%) and C4 (Analyzing) with 5 test items (20%). And the cognitive level with the hard category, namely C5 (Evaluating) as much as 1 test item (4%), C6 (Creating) as much as 0 (0%).

Table 10 . Result of the Analysis Biology Summative Test Items of 11 MIA with Academic Year 2020/2021 based on Taxonomy Bloom

Cognitive level	Category	Number of Item Test	Total	Percentage
C1	Easy	3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 16, 22, 25	13	52%
C2		1, 8, 14, 17, 19, 23, 24	7	28%
C3	Medium	18, 21	2	8%
C4		2, 20	2	8%
C5	Hard	15	1	4%
C6		-	0	0%
Total Item Test			25	

Table 11. Result of the Analysis Biology Summative Test Items of 11 MIA with Academic Year 2021/2022 based on Taxonomy Bloom

Cognitive level	Category	Number of Item Test	Total	Percentage
C1	Easy	5, 6, 7, 11, 12, 15, 16, 19, 21, 23	10	40%
C2		1, 13, 14, 17, 22, 24, 25	7	28%
C3	Medium	3, 18	2	8%
C4		2, 4, 8, 10, 20	5	20%
C5	Hard	9	1	4%
C6		-	0	0%
Total Item Test			25	

IV. Conclusions

Based on the results of the analysis of test items 11 MIA Academic year 2020/2021, there are no test items of very good quality, there are 2 good quality test items. items, and there are 6 test items that have very bad quality. Then, based on the results of the analysis of the 11 MIA Academic year 2021/2022 test items, there were no test items that were of very good quality, a total of test items with good quality were included in the test items that had medium quality, and a total of bad quality test items. 13 test items, and there are 6 test items that have very bad quality. In qualitative analysis, all test items have entered the category of very good questions based on the content aspect and language with a percentage of results of 100%. But 2 test items (8%) fulfill all construction aspects and 23 test items (92%) not meet one construction aspect. In cognitive level of Taxonomy Bloom with academic year 2020/2021, C1 are 13 test items (52%), C2 are 7 test items (28%), C3 and C4 each with 2 test items (8%), C5 is 1 test item (4%). The test items of academic year 2021/2022, C1 are 10 test items (40%), C2 are 7 test items (28%), C3 are 2 test items (8%), C4 are 5 test items (20%), and C5 is 1 test item (4%).

References

Fauzie, M., Pada, A. U., & Supriatno, S. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the COVID-19 pandemic. *Jurnal Penelitian dan Evaluasi Pendidikan*, *25*(2).

Lubis, N. S., & Prastowo, P. (2016). Analisis Kualitas Soal Ujian Mata Pelajaran Biologi Di Sma Negeri 3 Medan Tahun Pembelajaran 2016/2017. *Jurnal Pelita Pendidikan*, *5*(1), 149-157.

Manfaat, B., Nurazizah, A., & Misri, M. A. (2021). Analysis of mathematics test items quality for high school. *Jurnal Penelitian dan Evaluasi Pendidikan*, *25*(1), 108-117.

Salamah, U. (2018). Penjaminan Mutu Penilaian Pendidikan. *Jurnal Evaluasi Pendidikan*, *2*(1), 274-293.

Utami, I. P., & Aryeni, A. (2018). Analisis soal ujian akhir semester mata pelajaran biologi berdasarkan dimensi proses kognitif taksonomi Anderson. *Jurnal Pelita Pendidikan*, *6*(3).