UNIVERSITAS SEBELAS MARET
BIOEDUKASI: JURNAL PENDIDIKAN BIOLOGI
https://jurnal.uns.ac.id/bioedukasi/
1693-265X (Print) | 2549-0605 (Online)

Analysis of the Questions Item for Biology Olympiad Participants

Ahmad Subari [a, 1, *], **Zulyusri** [a], **Violita** [a]

[a] Padang State University, Padang-Indonesia
[1] subary_ahmad@yahoo.com*
* Corresponding author

Submission : 29/12/2020
Revision : 28/01/2021
Accepted : 15/02/2021

ABSTRACT

Biology Olympiad is one of the national tests run annually. The quality of the item tests needs to be analyzed details. This study aims to determine the quality of the MAN (Public Islamic School) Model Biological Olympiad questions in Jambi City. The subjects in this study were 10 graders of MAN Model Jambi City who attended the Biology Olympiad training. The data analysis technique used in this study was a program called States 4.0.9. The results in the study showed that for the validity of the items at the coefficient level of 5%, there were 13 valid questions and 27 invalid questions. Meanwhile, for the reliability of the questions, the value was 0.46 which was categorized as sufficient. Furthermore, for the difficulty level of the questions, they were 3 questions in the easy category, 16 questions in the medium category, and 21 questions in the difficult category. Then for the results of the distinguishing power of the questions, 7 questions were categorized as very good, 12 questions were categorized as good, 7 questions were categorized as poor, and 14 questions were categorized as very poor. Meanwhile, for the results of the analysis of the quality of distracting items, there were 4 questions with quality was categorized as very good and good, while the rest were classified as poor, poor, and very poor criteria, or those criteria need to be revised.

Keywords: Question Item Analysis, Biology Olympiad, Islamic high school

INTRODUCTION

The High School Biology Olympiad is one of the government programs to develop student's talents and interests in the field of biological science which is held annually. Selection is carried out at the district/city, provincial and national levels. In the selection at the district/city level, the participants of the Olympics are high school/MA students who were selected from all schools in each district/city. There are a series of evaluation activities for students carried out by each school to represent each school in participating in the selection of biology Olympiad participants at the district/city level.

Learning evaluation is a process or activity that is systematic, continuous, and comprehensive in controlling, guaranteeing, and determining the quality of learning for various learning components based on certain considerations and criteria (Arifin, 2014). Teachers must evaluate learning outcomes and determine competency standards that must be achieved by students participating in the biology Olympiad. By evaluating learning outcomes, teachers can find out whether the instruments used are too easy or difficult, or whether the instruments are following the indicators of learning or not, and whether the learning (models, approaches, strategies, and methods) used by teachers in teaching the Olympics is appropriate. biology.

Item analysis is a process of examining the quality of the questions in each item. analyzing each item is very important, lest each item contains things that are not following what is the goal, when viewed from the level of difficulty, distracting patterns, discriminating power, and others. Item analysis can be done with the help of the animated program which is one of the software to analyze items (Haryanto, 2020).

Anates is an application developed to calculate item analysis quickly, easily, and accurately. This application can display features and calculations including weighted data scores, reliability, discriminatory power, level of difficulty, correlation of item scores with total, and quality of distractors.

Based on observations at the Man Model Jambi school, it was found that the teacher had not analyzed the items that would be tested on the school-level Olympiad selection participants. Because no research or trial has been conducted on these questions, whether the questions are of high quality and meet the standards or not. For this reason, it is important to analyze the selection of the Olympics. By using questions that have been analyzed and whose quality is known, makes the school more mature and optimal in preparation for the selection for the Biology Olympiad at the district/city level.

Based on the description of the background of the problem, then conducted a study that aims to determine the quality of the MAN Model Biology Olympiad questions in Jambi City.

METHODS

This research is quantitative descriptive. The subjects in this study were students of class X MAN Model Jambi City who took part in the biology Olympiad training. The object of this research was the Jambi Model MAN Biology Olympiad Participant Selection Exam. Data were obtained through test instruments in the form of multiple-choice questions, answer sheets, and answer keys. The data analysis technique used in this study used the program Anates, among others, to determine the validity, reliability, level of difficulty, discriminatory power, quality of distractors, and recap of item analysis.

Validity shows the extent to which the accuracy of measuring an item in a test of learning outcomes. Calculation of validity using software version 4.0.9 Anates that is, after the r-count is known then it is compared with the table value with a significant level (α) = 0.05 at the 95% confidence level with df = n -2. With the following criteria, if r-count > r-table then the item is

declared significant (valid), and if r-count < r-table then the item is declared insignificant (invalid).

In addition to the validity of the questions, a question must also have a measure that states the level of consistency of a test item (Reliability). The following are the criteria for interpreting the reliability of the questions in Table 1 (Putri & Ofianto, 2019).

Table 1. Question reliability criteria

Reliability Value (r)	Interpretation
0.800 - 1,000	Very high
0.600 - 0.799	High
0.400 - 0.599	Enough
0.200 - 0.399	Low
0.000 - 0.199	Very low

One of the requirements of a good test instrument is to have a level of difficulty that is not too difficult and not too easy. The criteria for interpreting the level of difficulty of the questions is presented in Table 2 (Arifin, 2012).

Table 2. Criteria for item difficulty level

Difficulty Level (%)	Interpretation
0 - 27	Difficult
28 -72	Medium
73 - 100	Easy

In addition to the level of difficulty, you must also pay attention to the distinguishing power of an item. To interpret the discriminating power coefficient, the criteria in Table 3 can be used (Elviana, 2020).

Table 3. Criteria for distinguishing power of items

Distinguishing Power Level (%)	Interpretation
Negative – 9	Very poor (should be discarded)
10 – 19	Poor (Better throw it away)
20 – 29	Fairly Good (Enough)
30 – 49	Good
50 and above	Very good

In multiple choice questions, there are several alternative answers (options) known as distractors. A good item is that the distractors are chosen evenly by students who answer incorrectly. The quality of the distractors based on the distractor index can be seen in Table 4 (Arifin, 2014).

Table 4. Criteria for the Detractor Index

Distraction Index (%)	Interpretation
76 – 125	Very good
51 – 75 or 126 – 150	Good
26 – 50 or 151 – 175	Not good
0 – 25 or 176 – 200	Poor
Over 200	Very poor

RESULTS AND DISCUSSION

Validity

Item validity is the accuracy of measuring a question to determine the support of an item on the total score. A measuring instrument can be said to be valid if it is really what is intended to be measured precisely (Arifin, 2012). The support for each item can be expressed in the form of a correlation so that to get the validity of each item the correlation formula is used. The results of the analysis of the validity of the selected items for the Biology Olympiad MAN Model Jambi City using the Anates computer program version 4.0.9 are displayed in Table 5.

Table 5. Results of Item Validity Analysis

No	Correlation	Significance	Category
1	0.422	Very Significant	Valid
2	0.653	Very Significant	Valid
3	0.484	Very Significant	Valid
4	0.289	-	Invalid
5	-0.192	-	Invalid
6	0.141	-	Invalid
7	0.094	-	Invalid
8	0.291	-	Invalid
9	-0.048	-	Invalid
10	-0.378	-	Invalid
11	0.498	Very Significant	Valid
12	NAN	NAN	Invalid
13	0.261	-	Invalid
14	-0.146	-	Invalid
15	0.244	-	Invalid
16	0.008	-	Invalid
17	0.410	Very Significant	Valid
18	0.439	Very Significant	Valid
19	0.291	-	Invalid
20	0.213	-	Invalid
21	0.350	Significant	Valid
22	0.475	Very Significant	Valid
23	0.233	-	Invalid
24	-0.086	-	Invalid
25	0.115	-	Invalid
26	0.336	Significant	Valid
27	0.116	-	Invalid
28	0.378	Significant	Valid
29	0.352	Significant	Valid
30	0.008	-	Invalid
31	-0.012	-	Invalid
32	0.005	-	Invalid
33	0.557	Very Significant	Valid
34	0.202	-	Invalid
35	0.315	Significant	Valid
36	NAN	NAN	Invalid
37	0.105	-	Invalid
38	0.261	-	Invalid
39	NAN	NAN	Invalid
40	0.143	-	Invalid

Based on the results of the analysis of the validity of the items, 13 questions are valid, the questions that have the highest validity are 8 questions, while the questions that have moderate validity are 5 questions (Table 5). This is indicated by a positive correlation value and the t-count is greater than the t-table at a significance level of 5% = 0.304. For items that are not valid, 27 questions are consisting of numbers 11, 36, and 39 whose analysis cannot be calculated because the correlation value is 0.000 so that it displays the word NAN, numbers 5, 9, 10, 14, 24, and 31 display the correct correlation value. negative and the remaining question numbers are questions that have a correlation value of less than 0.304. This is indicated by a negative correlation value and the t-count is less than the t-table (Asrul *et al*, 2014).

Reliability

Reliability is a measure that states the level of consistency of a question. Reliability relates to whether a test can be trusted according to predetermined criteria. Based on the results of the reliability analysis of the test items using the Anates 4.0.9 software, the reliability value of the test items was 0.46. This value, if interpreted with the reliability criteria, is included in the sufficient criteria. In this case, the test instrument used needs to be improved to produce high reliability. Added by Ratnawulan & Rusdiana (2014) one of the test requirements as an evaluation instrument is to have high reliability. Tests that have high reliability will give results that are fixed and do not change.

Difficulty Level

The level of difficulty is a measurement of how difficult a question is. The analysis of the level of difficulty of the items can be seen in Table 6. Based on the results of the analysis of the difficulty level of the questions, some questions have easy, medium, and very difficult levels of difficulty. Questions that have an easy level of difficulty consist of 3 questions, questions that have a moderate level of difficulty are 16 questions, and questions that have a difficulty level of difficulty are 21 questions, in which 9 questions are classified as very difficult categories.

Table 6. The Analysis of the difficulty of the item

No	Correct Amount	Difficulty Level (%)	Interpretation
1	15	65.22	Medium
2	9	39.13	Medium
3	6	26.09	Difficult
4	6	26.09	Difficult
5	3	13.04	Very Difficult
6	12	52.17	Medium
7	1	4.35	Very Difficult
8	5	21.74	Difficult
9	9	39.13	Medium
10	4	17.39	Difficult
11	14	60.87	Medium
12	0	0.00	Very Difficult
13	5	21.74	Difficult
14	1	4.35	Very Difficult
15	8	34.78	Medium
16	4	17.39	Difficult
17	10	43.48	Medium
18	2	8.70	Very Difficult
19	5	21.74	Difficult

20	10	43.48	Medium
21	5	21.74	Difficult
22	19	82.61	Easy
23	6	26.09	Difficult
24	1	4.35	Very Difficult
25	10	43.48	Medium
26	10	43.48	Medium
27	8	34.78	Medium
28	17	73.91	Easy
29	9	39.13	Medium
30	4	17.39	Difficult
31	8	34.78	Medium
32	18	78.26	Easy
33	12	52.17	Medium
34	9	39.13	Medium
35	3	13.04	Very Difficult
36	0	0.00	Very Difficult
37	7	30.43	Medium
38	6	26.09	Difficult
39	0	0.00	Very Difficult
40	5	21.74	Difficult

To describe the actual learning achievement, the teacher must pay attention to the level of difficulty of the questions tested to students. Problems with very easy and very difficult categories must be replaced and repaired. The items used are items that are not too difficult and not too easy. Added by Arikunto (2005) a good question is a question that is not too difficult and not too easy. Very easy questions cannot stimulate students to increase their efforts to solve problems on the items. While very difficult questions will make students not have the enthusiasm to try to solve problems on the items because they are out of reach

Distinguishing Power

Distinguishing power is the measurement of a question in distinguishing the ability of students between students who have high abilities and low abilities. The results of the analysis of the level of the discriminating power of the questions can be seen in Table 7.

Table 7. The Analysis of the level of distinguishing power of questions

No	Top Group	Bottom Group	Different	Distinguishing Power (%)
1	5	1	4	66.67
2	5	0	5	83.33
3	3	0	3	50.00
4	3	1	2	33.33
5	0	1	-1	-16.67
6	4	2	2	33.33
7	0	0	0	0.00
8	3	0	3	50.00
9	1	2	-1	-16.67
10	1	3	-2	-33.33
11	5	2	3	50.00
12	0	0	0	0.00
13	2	1	1	16.67
14	0	1	-1	-16.67
15	4	2	2	33.33

16	1	1	0	0.00
17	6	2	4	66.67
18	1	0	1	16.67
19	2	1	1	16.67
20	2	2	0	0.00
21	2	0	2	33.33
22	6	4	2	33.33
23	2	1	1	16.67
24	0	0	0	0.00
25	4	3	1	16.67
26	3	1	2	33.33
27	3	1	2	33.33
28	5	3	2	33.33
29	3	1	2	33.33
30	2	1	1	16.67
31	1	2	-1	-16.67
32	5	5	0	0.00
33	5	0	5	83.33
34	2	0	2	33.33
35	2	0	2	33.33
36	0	0	0	0.00
37	2	2	0	0.00
38	3	1	2	33.33
39	0	0	0	0.00
40	1	0	1	16.67

Based on the results of the analysis of the level of discriminating power of items, in Table 3, there are upper and lower groups on the discriminatory power, to distinguish smart students from less intelligent students. The discriminatory power of questions obtained from the test results of 40 items, namely, 7 items are belonging to the very good category, 12 items belonging to the good category, 7 items belonging to the poor category, and 14 questions belonging to the very poor category. The percentage of discriminatory power of very low to very low questions should be replaced and repaired. Added by Arifin (2014), the higher the coefficient of discriminating power of an item, the more capable the item is to distinguish between students who can master and those who are less able to master competence.

Distracting Quality

A good question distractor is a distractor that is chosen evenly by students, on the contrary, if it is chosen unevenly, it is considered less good. The results of the analysis of the level of distractors are shown in Table 8.

Table 8. Results of the distraction level of questions

No	**Distractor Rate (%)**					**Interpretation**
	a	**b**	**c**	**d**	**e**	
1	100	-	100	150	50	Revised
2	57	-	200	114	29	Revised
3	-	188	94	94	24	Revised
4	0	-	306	94	0	Revised
5	160	60	40	-	140	Revised
6	109	-	145	73	73	Used
7	109	127	127	36	-	Revised
8	44	-	133	178	44	Revised
9	171	-	86	114	29	Revised

10	232	126	42	0	-	Revised
11	89	133	89	-	89	Used
12	243	-	87	70	0	Revised
13	89	111	-	200	0	Revised
14	55	-	164	127	55	Revised
15	107	-	133	80	80	Used
16	-	0	189	42	168	Revised
17	31	31	-	308	31	Revised
18	114	0	152	-	133	Revised
19	22	89	-	111	178	Revised
20	338	0	-	62	0	Revised
21	200	111	-	44	44	Revised
22	0	300	100	-	0	Revised
23	118	165	118	0	-	Revised
24	18	200	73	109	-	Revised
25	62	-	0	92	246	Revised
26	31	277	-	62	31	Revised
27	160	-	80	80	80	Revised
28	67	67	67	200	-	Revised
29	114	-	143	29	114	Revised
30	232	63	-	21	84	Revised
31	53	293	-	27	27	Revised
32	80	-	240	80	0	Revised
33	-	218	145	36	0	Revised
34	143	57	114	86	-	Used
35	60	-	100	20	220	Revised
36	-	52	243	52	52	Revised
37	50	100	-	150	100	Revised
38	71	165	118	47	-	Revised
39	-	157	243	0	0	Revised
40	0	111	222	67	-	Revised

Based on the results of the quality analysis of distractors, there are 4 questions whose distractors are categorized as very good and good, namely numbers 6, 11, 15, and 34. These questions can be used because the distractors are chosen evenly by students. In question number 6, the percentage of distractors choices a, c, d, and e, are 109%, 145%, 73% and 73%. Choice a belongs to very good criteria, while c, d, and e, belong to good criteria so that the questions can be used. Meanwhile, questions that have distracting qualities that are classified as poor, poor, and very poor criteria, need to be improved. Added by Arifin (2014) distractors are considered good if the number of students who choose the distractor is the same or close to the ideal number.

Item Analysis Recap

The item analysis recap is an overall analysis of whether the item is suitable for use or not. There are several criteria for the recap of item analysis as follows, the questions are used if they are valid, the discriminatory power is good, and the level of difficulty is moderate, the questions are not used if they are not valid, the discriminatory power is not good and the difficulty level is too difficult and too easy and the questions used are corrected if the questions are categorized as valid, but one of the distinguishing features or the level of difficulty is not good, so improvements need to be made so that they can be used (Table 9).

Table 9. Recap of the analysis of the questions

No	Distinguishing Power	Difficulty Level	Distracting Quality	Validity	Action
1	66.67	Medium	Revised	Valid	Need revision
2	83.33	Medium	Revised	Valid	Need revision
3	50.00	Difficult	Revised	Valid	Need revision
4	33.33	Difficult	Revised	Invalid	Not Used
5	-16.67	Very Difficult	Revised	Invalid	Not Used
6	33.33	Medium	Used	Invalid	Not Used
7	0.00	Very Difficult	Revised	Invalid	Not Used
8	50.00	Difficult	Revised	Invalid	Not Used
9	-16.67	Medium	Revised	Invalid	Not Used
10	-33.33	Difficult	Revised	Invalid	Not Used
11	50.00	Medium	Used	Valid	Used
12	0.00	Very Difficult	Fixed	Invalid	Not Used
13	16.67	Difficult	Fixed	Invalid	Not Used
14	-16.67	Very Difficult	Fixed	Invalid	Not Used
15	33.33	Medium	Worn	Invalid	Not Used
16	0.00	Difficult	Fixed	Invalid	Not Used
17	66.67	Medium	Fixed	Valid	Need revision
18	16.67	Very Difficult	Fixed	Valid	Need revision
19	16.67	Difficult	Fixed	Invalid	Not Used
20	0.00	Medium	Fixed	Invalid	Not Used
21	33.33	Difficult	Fixed	Valid	Need revision
22	33.33	Easy	Fixed	Valid	Need revision
23	16.67	Difficult	Fixed	Invalid	Not Used
24	0.00	Very Difficult	Fixed	Invalid	Not Used
25	16.67	Medium	Fixed	Invalid	Not Used
26	33.33	Medium	Fixed	Valid	Need revision
27	33.33	Medium	Fixed	Invalid	Not Used
28	33.33	Easy	Fixed	Valid	Need revision
29	33.33	Medium	Fixed	Valid	Need revision
30	16.67	Difficult	Fixed	Invalid	Not Used
31	-16.67	Medium	Fixed	Invalid	Not Used
32	0.00	Easy	Fixed	Invalid	Not Used
33	83.33	Medium	Fixed	Valid	Need revision
34	33.33	Medium	Worn	Invalid	Not Used
35	33.33	Very Difficult	Fixed	Valid	Need revision
36	0.00	Very Difficult	Fixed	Invalid	Not Used
37	0.00	Very easy	Fixed	Invalid	Not Used
38	33.33	Difficult	Fixed	Invalid	Not Used
39	0.00	Very Difficult	Fixed	Invalid	Not Used
40	16.67	Difficult	Fixed	Invalid	Not Used

Based on the results of the recap of the item analysis in Table 9, overall, the questions that can be used are 1 question that can be used immediately, 12 questions that can be used but need to be improved because they have a level of difficulty, discriminatory power or quality of distractors that are not good, while 27 The remaining questions are classified as questions that are not used.

Questions that are in the question category are used, then the items are directly used and entered in the question bank that is in the teacher's hands, and these items can be used again in future learning outcomes tests Items included in the category are used to be corrected, the questions are researched and their grammar is corrected, whether the sentence questions are unclear or difficult for students to understand. While the items that are not used can be directly discarded and not used for the next test of learning outcomes.

CONCLUSION

Based on the results of the item analysis of the selection of Biology Olympiad participants in the MAN Model Jambi using Anates 4.0.9 software, the quality of the questions for the MAN Model Biology Olympiad in Jambi City is not good enough. This is indicated by the results of the overall item analysis recap of the 40 questions, there is 1 question that can be used immediately, 12 questions that can be used but need to be improved because they have a level of difficulty, discriminatory power, or quality of distractors that are not good, while the remaining 27 questions is a question that cannot be used.

Based on the results of the study, it is necessary to conduct training on making good and correct questions to teachers at the MAN Model school in Jambi City, then it is important to do further research on the Analysis of the Biology Olympiad Selection Questions in Various SMA/MA Schools in Jambi Province to see the quality of the questions comprehensively thorough.

REFERENCES

Arifin, Z. (2012). *Writer's Learning Evaluation*. In the Directorate General of Islamic Education, Ministry of Religion of the Republic of Indonesia.

Arifin, Z. (2014). *Learning evaluation*. In Pipih Latifah (Ed.), Bandung: Rosdakarya Youth. PT Youth Rosdakarya.

Arikunto, S. (2005). *Basics of Educational Evaluation (Revised Edition)*. Earth Literature.

Asrul, Ananda, R., & Rosinta. (2014). *Learning Evaluation*. In Cita Librarian Media. Media Citations.

Elviana. (2020). Analysis of Items for evaluation of islamic religious education learning using the Anates Program. *Mudarrisuna Journal*, 10(2), 58–74.

Haryanto. (2020). *Learning evaluation; Concepts and Management*. In UNY Press.

Putri, RH, & Ofianto. (2019). The Effectiveness of item analysis using anajohn, anates, and iteman study problems USBN Padang City History Lessons. *Student Journal of History and Education*, 1(2), 1–11.

Ratnawulan, E., & Rusdiana, A. (2014). *Learning Evaluation*. Bandung Faithful Library. http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf