

# Item Analysis of Teacher Made Test in Biology Subject

Devi Dwi Kurniawan<sup>1,\*</sup>, Abdullah Syifa<sup>2</sup>, Nurul Huda<sup>3</sup>, Mobinta Kusuma<sup>1</sup>

<sup>1</sup> Universitas Negeri Yogyakarta, Indonesia

<sup>2</sup> Institut Agama Islam Negeri Pontianak, Indonesia

<sup>3</sup> Universitas Islam Negeri Malang, Indonesia

\*Corresponding author. Email: [devidwikurniawan@gmail.com](mailto:devidwikurniawan@gmail.com)

## ABSTRACT

One of the teachers' competencies is the ability to compose tests that include difficulty, discriminating power, and the effectiveness of distractors. This research was conducted to evaluate the analysis of teacher-made tests, which can be seen from the three aspects of the test. The population of this study was 80 students of class X-MIPA MAN 1 Kubu Raya Regency, West Kalimantan Province, Indonesia, who had taken the final exam, using the total sampling technique. The research was conducted in December 2020. Data analysis using ITEMAN software to measure the level of difficulty, discriminating power, and effectiveness of distractors on 40 items. The results show that a). Difficulty level: 12,5% difficult, 45% moderate, 42,5% easy; b) discriminatory power: 30% poor, 15% unsatisfactory, 42,86% satisfactory, and 32,5% very satisfactory; c) distractor effectiveness: 22,5% without revision, 25% incorrect revision of 1 option, 22,5% incorrect revision of 2 options, 22,5% incorrect revision of 3 options, and 7,5% incorrect revision of 4 options. The evaluation results showed that only 37.5% of the items had good enough and good characteristics. There are 63.5% of items with poor and bad characteristics. Researchers recommend training organized by schools or education offices on the competence to make tests.

**Keywords:** *biology subject, item analysis, teacher, test,*

## 1. INTRODUCTION

The teacher is one of the determinants of students' success in learning. This is because the teacher is a facilitator, learning resource, scientific resource, and assessing student learning outcomes. We need teachers who are capable of learning in learning, especially in learning Biology subject. The teachers are expected to improve students' competence, including motivating students [1], diagnosing learning difficulties, collaboration skills, and guiding students to think critically [2].

The teacher's main task to students is not only how the competence of each student can be improved, but the teacher can also conduct appropriate class assessments on learning outcomes [3]. One of the schools' roles is to help students develop competencies in knowledge, skills, and understanding so that educators are expected to understand curriculum design, lesson planning design, and assessment design for

student success [4]. The classroom assessment practice carried out by the teacher goes through five stages: planning assessments, preparing assessment items, administering and assessing, reporting scores and assessments, and utilizing and evaluating assessment data [5]. The application of class assessment techniques carried out by Biology teachers must also be evaluated properly [6]. This is because the assessment results become the basis for the extent to which individual students' abilities are successful in classroom learning. To conduct an assessment of students, teachers must have the ability to compose a good test.

The test is a tool used to make a measurement, meaning that the test is structured to collect information on an object [7]. Measurement is the determination of numbers or quantification to describe the characteristics or conditions of individuals, both cognitive, affective, and psychomotoric, without violating the measurement methods [8]. The assessment interprets the data resulting from the measurement process based on the

rules and criteria. Two test theories have developed in education, namely classical test theory and item response theory. Classical test theory uses a simple mathematical model to show the relationship between observed scores, actual scores, and error scores. The formula is  $X = T + E$  [9]. The assumptions built in classical test theory can develop at least three analyzes, namely the level of difficulty, discriminating power, and distractor effectiveness [10]. Evaluation, assessment, and measurement have a strong relationship. Evaluation can be done if an assessment precedes it, and the assessment must be preceded by measurement. Evaluation is a process of determining the implications or values of behavior, so the three terms are hierarchical.

The function of the difficulty level analysis is to ensure that the questions that have been prepared are in accordance with the student's ability to take the exam [11]. The discriminatory power of the item is the ability of the item to distinguish between high-ability students and low-ability students [12]. Alternative answers that are wrong on multiple-choice questions can be used as a distractor. Distractors provide data to teachers about the ability of students who master the subject matter or vice versa [13].

There are still many teachers who pay less attention to the preparation of the test so that the results obtained in the implementation of the test are not able to measure the actual abilities of students and are unable to distinguish between students who have mastered the subject matter and those who have not mastered the subject matter. The research results from Mulyani et al. [14] show that there are still many teacher-made tests that are not suitable for use because they do not meet the level of difficulty, discriminatory power, and are effective as a good distractor. Therefore, the focus of this research is to see the quality of teacher-made tests in schools used for students' final semester exams in biology subjects.

Some of the descriptions above confirm that the test has an important role in assessing teaching and learning activities in the classroom. This is because the question as a measuring tool has accurate information about the mastery of certain subject matter. Preparation of questions is one of the duties of a teacher, so the teacher must prepare quality exam questions. This article will reveal the extent of the teacher's ability to prepare the final exam questions for Biology subjects in the analysis of the level of difficulty, discriminating power, and the effectiveness of the distractor.

Many teachers have not carried out preparing good questions. Therefore, this research can provide something new in preparing good questions through appropriate item analysis. This research is expected to complement the results of previous studies that have not revealed students' overall ability. From the results of the analysis, it is hoped that it can provide constructive information to the school to pay attention to the competence of the teachers in preparing questions in the instruction.

## **2. MATERIAL & METHODOLOGY**

This research is a descriptive study using a quantitative approach. Descriptive approach was chosen because the researcher only wanted to explain the data on the existing facts without testing the hypothesis. The quantitative approach was chosen because the data collected in the form of numbers were then analyzed statistically with the help of a computer. The data was obtained from the answers to 40 multiple choice questions for the final semester of Biology material for the 2019/2020 school year. The researcher chose the research location at MAN 1 Kubu Raya, West Kalimantan. The research population was 80 students of class X-MIPA, with a total sample of 80 students. The sampling technique used is total sampling. This study was conducted in December 2019. Analysis of items in the analysis of the difficulty level, discriminating power, and effectiveness of distractors using ITEMAN 4.3 software.

The criteria for item difficulty level are  $0.3 > P$  is difficult,  $0.7 > P > 0.3$  is moderate, and  $P > 0.7$  is easy. The criteria for discriminating power are 0.4-1 is very satisfactory, 0.3-0.39 is satisfactory, 0.2-0.29 is unsatisfactory, and negative-0.19 is poor. The distractor is concluded to be effective if a minimum of 5% of all examinees are selected.

## **3. RESULT & DISCUSSION**

### **3.1. Result**

Analysis of the items characteristics uses the computer program ITEMAN version 4.3. The analysis produces information about items/sets of questions that are feasible and not suitable for later use based on the parameters of items/sets of questions, namely the level of difficulty, discriminating power, and the effectiveness of distractors. The analysis of the level of difficulty of the items using the ITEMAN computer program version 4.3 can be seen in column P, which is recapitulated in Table 1.

**Table 1.** Recap the level of difficulty of the biology items made by the teacher

Criteria	Total	Percentage
Difficult	5	12.5
Moderate	18	45.0
Easy	17	42.5
	40	100.00

Analysis of the level of difficulty of the items is done to see the degree of difficulty of the questions that will be tested on students. This was done to determine the proportion of the number of students who were able to answer incorrectly and answer correctly from the package or all the questions tested [15]. Based on Table 1, the distribution of the level of difficulty is evenly distributed because each criterion is represented. Questions with difficult criteria are number 13, 26, 28, 29, and 30. Questions with moderate criteria are numbers 02, 07, 09, 14, 15, 18, 21, 24, 25, 27, 32, 33, 35, 36, 37, 38, 39 and 40. Questions with easy criteria are numbers 01, 03, 04, 05, 06, 08, 10, 11, 12, 16, 17, 19, 20, 22, 23, 24 and 31.

There is a need for follow-up so that the questions that will be used as Biology semester exams are close to proportional. As for what can be done, namely, by: first, the items that fall into the difficult category with a percentage of 12.5% need to be used as material for evaluating the teacher's ability to prepare questions in the difficult category. This can be done by analyzing which questions are relevant to use on questions with easy criteria into questions with difficult criteria. Second, items that fall into the good category or show a moderate degree of difficulty with a percentage value of 45% can still be used and put into the question bank and later used as questions to be tested. Third, the analysis of items belonging to the easy category with a percentage of 42.5% needs to be followed up by compiling questions that have higher quality weights. The analysis of the discriminatory power of items using the ITEMAN computer program version 4.3 can be seen in the Rpbis column, shown in Table 2.

**Table 2.** Recap of the different power of Biology items made by the teacher

Criteria	Total	Percentage
Very Satisfactory	13	32.5
Satisfactory	9	22.5
Unsatisfactory	6	15
Poor	12	30
	40	100.00

ased on Table 2. that there are 22 (55%) Biology questions made by teachers who are able to distinguish students who have high and low abilities, while Biology questions made by teachers who are unable to distinguish students who have high and low abilities are 18 items (45%). Questions that have poor discriminating power are 12 questions, namely numbers 06, 10, 13, 14, 16, 19, 22, 26, 28, 29, 30, and 31, with a percentage value of 30%. Questions with unsatisfactory criteria with a percentage value of 15% are numbers 04, 15, 21, 25, 27, and 38. Questions with satisfactory criteria are numbers 01, 02, 05, 18, 20, 23, 35, 37 and 40 with scores percentage 22.5%. Questions with very satisfactory criteria with a percentage value of 32.5%, namely numbers 03, 07, 08, 09, 11, 12, 17, 24, 32, 33, 34, 36 and 39. The items' analysis is carried out to determine any students who lack or have not mastered the material and those who have mastered the material being tested. Questions with adequate discriminating power can be used to test students, while questions with inadequate power should not be used [16]. There are several reasons why these questions are inadequate, including incorrect answers, the material is tested being too difficult, the competence to be measured being unclear, or the distractors are not functioning.

The results of the analysis of the discriminatory power need to be followed up, among others: first, items that are proven to have adequate distinguishing power should be directly used as test questions because these items have feasibility in terms of distinguishing power. Second, for items that have distinguishing power with inadequate categories, if they can still be improved, the teacher should improve the quality of the questions. If the question cannot be repaired, it should be discarded or replaced with another item. Information on the effectiveness of the distractor can also be obtained from the computer program Iteman version 4.3 can be seen in the Prop column in Table 2.

**Table 3.** Recap the effectiveness of the teacher-made the question-answer options distractors

Decision	Total	Percentage
Without Revision	9	22.5
Revised 1 Answer	10	25.0
Revised 2 Answer	9	22.5
Revised 3 Answer	9	22.5
Revised 4 Answer	3	7.5
	40	100

Distractors are said to be effective and good if the alternative answers are chosen by students of at least 5% of all students taking the exam. The purpose of the distractor effectiveness analysis is to provide information on how well the wrong answer options are in deceiving students who do not know the answer key to the exam questions. The distractor effectiveness can be done by looking at the distribution pattern of the answers to the questions chosen by the students. Based on Table 3, the information obtained is questions that have answer options without revision are 9 (22.5%) questions with question numbers 15, 24, 26, 33, 35, 36, 37, 38, and 39. revised one answer option as many as 10 questions (25%) namely with question numbers 02, 07, 09, 17, 18, 27, 28, 32, 34 and 40. Questions that need to be revised are two answer options are nine questions (22, 5%), namely with the number of questions 03, 12, 16, 20, 21, 23, 25, 30, and 31. The questions that need to be revised are incorrect. Three answer options are nine questions (22.5%), namely with the numbers 05, 06, 08, 10, 11, 13, 14, 19, and 29. The questions that need to be revised are four answer options, three questions (7.5%) with question numbers 01, 04, and 22.

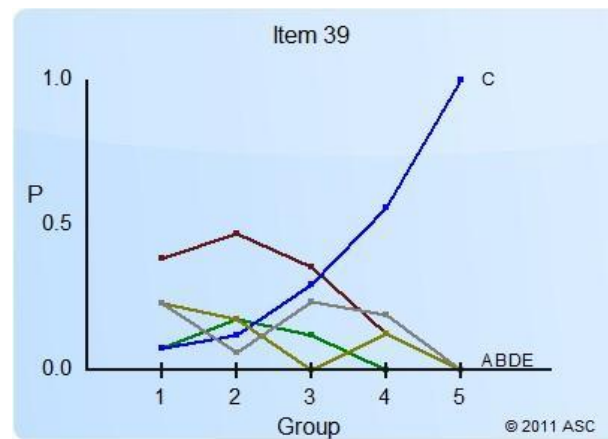
The results of the analysis of the effectiveness of the distractor options for the answers to Biology questions made by the teacher require follow-up, including: first, the items that are proven to have distracting effectiveness with categories without revision or revision of one of the options should be used directly as exam questions, because these items have feasibility in aspects distractor effectiveness. Second, for items that have distracting effectiveness with the wrong revision category of 2, 3, and even four options, the teacher should improve the quality of the answer options. If the answer option cannot be repaired, it should be discarded or replaced with another answer option. Based on the items characteristics (difficulty level, discriminating power, and effectiveness of distractors), the teacher-made Biology questions can be categorized in Table 4.

**Table 4.** Recap of the characteristics of the biology items made by the teacher

Conclusion	Total	Percentage
Good	6	15.0
Fairly Good	9	22.5
Poor	14	35.0
Bad	11	27.5
	40	100

Based on table 4, only 15 items (37.5%) of Biology questions made by the teacher have quite good and good characteristics, so they can be included in the question bank. The total number of questions that have poor and bad categories is 25 questions (63.5%). The questions have good characteristics, namely no. 24, 33, 35, 36, 37, and 39, with a percentage of 15%. The questions have fairly good characteristics with a percentage of 22%, namely numbers 02, 07, 09, 15, 18, 32, 34, 38, and 40. The questions have poor characteristics, namely numbers 01, 03, 05, 08, 11, 12, 14, 17, 20, 21, 23, 25, 26 and 27, with a percentage of 35%. The questions have bad characteristics with a percentage of 27.5%, namely numbers 04, 06, 10, 13, 16, 19, 22, 28, 29, 30 and 31. These results provide information that there is still a high percentage of questions that are not suitable for use. This is, of course, an evaluation material and even a concern for schools and the education office in an effort to improve the quality of teacher competence in preparing exam questions both from the aspect of difficulty level, discriminating power, and the effectiveness of distractors.

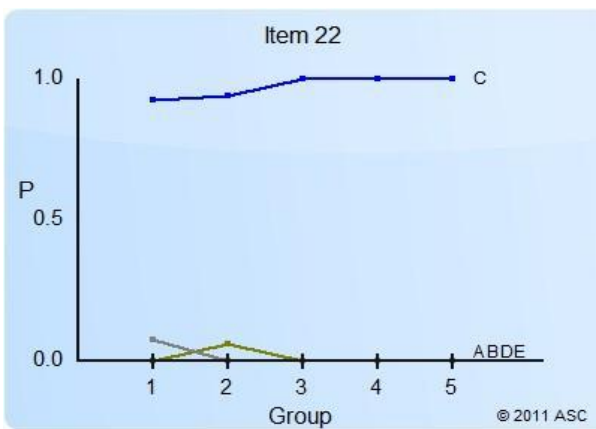
Figure 1 explains the state of the test participants in choosing the answer options. There are two question items that will be displayed as representative items that have good and good enough categories, namely questions number 24 and 15.



**Figure 1** The state of the participants in choosing the answer option on question no. 39 (good question category)

Based on Figure 1, 39 has a moderate level of difficulty (0.425), the discriminatory power is very satisfactory (0.589), and all answer options are effective enough to deceive the participants. There are 80 respondents divided into five groups with a minimum correct score of 9 and a maximum of 34. Group 1 has 13 participants with a score of 9-16, group 2 has 17

participants with a score of 17-22, group 3 has 17 participants with a score of 23-25, group 4 has 16 participants with a score of 26-28, and group 5 has 17 participants with a score of 29-34. In group 1, at most 5 participants chose the answer option A, meaning that group 1 participants cannot answer the questions. In group 2, at most 8 participants chose the answer option A, meaning that the group 2 participants cannot answer the questions. In group 3, at most choose the answer option A as many as 6 participants, meaning that group 3 participants cannot answer the questions. In groups 4 and 5, at most, 9 and 17 participants chose the answer option C, respectively, meaning that the participants in these groups are able to answer the questions.



**Figure 2** The state of the participants in choosing the answer option on question no. 22 (the category of questions is not good)

Based on Figure 2, No. 22 has an easy level (0.975), poor discriminating power (0.143), and four answer options A, B, D, and E are ineffective in deceiving participants because the answer key is C. In groups 1 to 5, the most many chose the answer option C in a row as many as 12, 16, 17, 16, and 17 participants, meaning that participants could answer the questions.

### 3.2. Discussion

The results of the item analysis of the level of difficulty, discriminatory power, and effectiveness of the distractors show that many teacher-made tests are not yet suitable for use, so they need to be revised. In addition, the test has not been able to measure the condition of students in school. The results of this study are in line with the results of research from Indriyani et al. [17], showing that 79 items (99%) are very good and 1 item (1%) is good. From the results of this study, the test instruments that are feasible to use meet suitable parameters. A good instrument can consistently provide data according to reality [18]. The results of the study show that teachers do not all understand and master

thpreparation of items well, so that they have not been able to obtain a good level of student development because the mastery of the actual student material cannot be seen, meaning that at the Bloom's Taxonomy level, it is still at the primary level, namely knowledge. In addition, the appropriate learning objectives contained in the curriculum have not been realized. The researcher recommends the need for training organized by schools or the education office on the competence of preparing exam questions so that in the future, the tests made by teachers have good quality and are suitable for measuring student learning outcomes appropriately.

### 4. CONCLUSION

Based on the results of research and discussion on the evaluation and quality of 40 items of teacher-made Biology items, it can be concluded as follows: Characteristics of teacher-made Biology items in aspects: a) difficulty level: 5 items (12.5%) difficult, 18 items (45.0% ) moderate, 17 items (42.5%) easy; b) discriminatory power: 12 items (30.0%) poor, six items (15.0%) unsatisfactory, nine items (42.86%) satisfactory, and 13 items (32.5%) very satisfactory; c) effectiveness of distractors: 9 items (22.5%) without revision, ten items (25.0%) incorrect revision of 1 option, nine items (22.5%) incorrect revision of 2 options, nine items (22.5%) incorrect revision of 3 options, and three items (7.5%) incorrect revision of 4 options; d) 15 items (37.5%) of Biology questions made by the teacher have quite good and good characteristics, so they can be included in the question bank. This conclusion has implications for the teacher's inaccurate conclusions in measuring the quality of students in Biology material. This is because the questions quality is still low, so there is a need for training organized by schools or the education office on the competence of preparing exam questions. The results of this evaluation are expected to contribute empirically information in improving the quality of questions and the quality of education, especially in the subject of Biology. It is necessary to carry out the same evaluation and analysis on all subject exam questions to obtain information both from the aspect of difficulty level, discriminating power, and the effectiveness of distractors.

### REFERENCES

[1] E. N. Nacionales, M. Pelagio, J. Muyong, & J. C. Gavanoan, The Impact of Motivation and Learning Strategies as Predictors of Biology Performance among Non- Science Majors. *Asia Pacific Higher Education Research Journal (APHERJ)*, Vol. 3(2). 2016.

[2] D. T. Boleng, S.V.T. Lumowa, E. Palenewen, & A.D. Corebima, The Effect Of Learning Models

- On Biology Critical Thinking Skills Of Multiethnic Students At Senior High Schools In Indonesia. *Problems of Education in the 21st Century*, Vol. 75(2), 2017, pp. 136–143.
- [3] E.T. Amua-Sekyi, Assessment, Student Learning and Classroom Practice: A Review. *Journal of Education and Practice*, Vol. 7(21), 2016, pp. 1–6.
- [4] C.A. Tomlinson, T. Moon, & B. Imbeau, Assessment and student success in a differentiated classroom. *ASCD Professional Learning Services*, Vo. 17. 2015.
- [5] Hasnida Che Md Ghazali, N., Mat Rabi, Md Hassan, & N. Abdul Wahab, A Confirmatory Factor Analysis of Classroom Assessment Practises scale in a Malaysian context. *International Journal of Academic Research in Progressive Education and Development*, Vol. 7(3), 2018, pp. 516–529. DOI: <https://doi.org/10.6007/IJARPED/v7-i3/4570>
- [6] K. Fomichova, & T. Misonou, Teachers Assessment of Japanese Lower Secondary School Biology Education Teachers' Assessment of Biology Education in Comparison to Other Areas of Science in Lower Secondary Schools of Japan- A Case Study. *Asian Journal of Biology Education*, Vol. 8, 2015.
- [7] D. Lebagi, S. Sumardi, & S. Sudjoko, The Quality of Teacher-Made Test in Efl Classroom At the Elementary School and Its Washback in the Learning. *Journal of English Education*, Vol. 2(2), 2017, pp. 97–104. DOI: <https://doi.org/10.31327/jee.v2i2.289>
- [8] A. Muhson, B. Lestari, Supriyanto, & K. Baroroh, The Development of Practical Item Analysis Program for Indonesian Teachers. *International Journal of Instruction*, Vol. 10(2), 2017, pp. 199–210, DOI: <https://doi.org/https://doi.org/10.12973/iji.2017.10213a>
- [9] M.J. Allen, & W.M. Yen, Introduction to measurement theory. California: Brooks. 1979.
- [10] R. L. Ebel, & D.A. Frisbie, Essentials of Educational Measurement. Bergen: Prentice Hall, Inc. 1986.
- [11] O. Sunbul, & S. Yormaz, Effects of Test Level Discrimination and Difficulty on Answer-Copying Indices. *International Journal of Evaluation and Research in Education (IJERE)*, Vol. 7(1), 2018, pp. 32. DOI: <https://doi.org/10.11591/ijere.v7i1.11488>
- [12] J.A. Odukoya, O. Adekeye, A.O. Igbino, & A. Afolabi, Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Quality and Quantity*, 52(3), 2018, pp. 983–997. <https://doi.org/10.1007/s11135-017-0499-2>
- [13] S.J. Haberman, Y. Liu, & Y.H. Lee, Distractor Analysis for Multiple-Choice Tests: An Empirical Study with International Language Assessment Data. *ETS Research Report Series*, 2019. DOI: <https://doi.org/10.1002/ets2.12275>
- [14] Mulyani, Heni., Tanuatmodjo, Heraeni., Iskandar, Rangga., Mulyadi, Ajang., Purnamasari, Imas., & Triantoro, Arvian. Quality Analysis Of Teacher Made Tests In Financial Accounting Subjects. *International Journal of Advanced Science and Technology*, Vol. 29(02), 2020, pp. 530-537.
- [15] S. Deshpande, & R.K. Prajapati, *Item Analysis of Mid-Trimester Test Paper and Its*. Vol. (2), 2018, pp. 27–31.
- [16] A. R. Menon, & P. N. Kannambra, Item Analysis to Identify Quality Multiple Choice Questions. *National Journal of Laboratory Medicine*, 2017, pp. 7–10. <https://doi.org/10.7860/njlm/2017/25690:2215>
- [17] M.S.D. Indriyani, A.A.I.N. Marhaeni, A.A.G.Y. Paramartha, & L.G.E. Wahyuni, The Analysis of the Teacher-Made Multiple-Choice Tests Quality for English Subject. *Journal of Educational Research and Evaluation*, Vol. 4(3), 2020, pp. 274–280.
- [18] Kusaeri dan Suprananto Educational Measurement and Assessment. Yogyakarta: Graha Ilmu, 2012.