Vol. X No. 2 December 2024
DOI: https://doi.org/10.19109/bioilmi.v10i2.25727
E-ISSN: 2527-3760
Page: 105 - 114

Student Cognitive Evaluation Instrument Protist Material: Anates Application Analysis Version 4.09

Istinganah Nur Rahayu*, Muhamad Jalil, Achmad Ali Fikri, Iseu Laelasari, Hashina Izzatin Nida

Biology Education Study Program, Tarbiyah Faculty, Kudus State Islamic Institute

**istinganahnurrahayu03@gmail.com* *(Corresponding Author)*

ARTICLE INFO

***Article history*:**
Received: 05/12/2024
Revised: 21/12/2024
Accepted: 26/12/2024

***Keywords*:**
Anates
Cognitive
Evaluation
Instruments
Protist

ABSTRACT

In the world of education, as an educator, of course you are expected to have instruments that meet standards to measure students abilities. However, in the field there are still many evaluation questions that do not meet these standards. The aim of this research is to analyze Biology evaluation questions for class X Protista material using the Anates program ver. 4.09, to ensure the questions meet the appropriate question quality standards.This research use a quantitative descriptive approach and use documentation techniques in data collection. The questions used were 10 questions in the form of multuple choices. The result of the analysis of items in the measurement of differentiating power contained 1 item in the very poor category, 5 items in the poor category, 2 items in the sufficient category and 1 item in the good category. Analysis of the difficulty level of questions in the difficult category contained 4 questions, 2 questions in the medium category and 4 questions in the easy category. Analysis of the validity of the questions found that 30% were declared valid and 70% were declared invalid. The reliability value of the questions is 0.02, in accordance with the criteria in the range 0.000-0.199. This means that the questions used have very low coefficients. The average question criteria are included in the poor criteria so changes and revisions need to be made for further use of the questions. It is hoped that this research can improve the quality of test questions on protist material.

INTRODUCTION

Learning activities are the main process in education, where students interact with material to develop knowledge, skills and attitudes. Learning development is the responsibility of an educator. Teachers not only deliver material, but ensure that the methods, media and learning models used are in accordance with students' needs, so that learning objectives are achieved as expected. In the learning process educators must carry out an activity called learning evaluation (Idrus 2019).

Evaluation is an important element in learning, because evaluation can provide a level of learning efficiency(Sodikin & Gumiandari, 2022). Every educator or teacher is required to have the ability to evaluate learning. The purpose of evaluation is to measure the extent to which the set learning objectives have been achieved and to identify areas that require improvement. With evaluation, educators are able to categorize students based on their level of ability, students who require special attention, and can consider graduation based on objective data (Bahtiar, 2020).

The cognitive aspect in learning evaluation has a role in developing students' intellectual abilities. This aspect is most often assessed by teachers in schools because it relates to students' abilities (Bella, 2018). The cognitive aspect includes students' ability to understand concepts and develop critical thinking skills. The purpose of learning evaluation in the cognitive aspect is to measure the level of achievement of learning related to knowledge and understanding (Phafiandita, Permadani, Pradani, & Wahyudi, 2022).

In Biology learning, especially protist material, cognitive evaluation is used to assess students' ability to recognize the characteristics of protists, types of protists and analyze their role in everyday life. Studying protist material often presents challenges related to understanding scientific terms and classification of complex organisms (Dwi Zunitasari, Sukarni Hidayati, 2016). Based on previous research conducted by (Melandi Wimudi et al. 2022) In Biology learning, protist material is considered difficult by students. Apart from that, studying Biology is considered a rote lesson, students tend to have difficulty building an understandi ng of the learning material, especially complex protist material. (Liza Yulia Sari 2013). To determine students' level of understanding, item analysis is a relevant tool to use (Wiwik Sulistyawati, Wahyudi, 2022).

The activy of analizing questions item is an important step for an educator. The purpose of item analysis is to obtain information about the characteristics of each item through examining the items (Elviana, 2020). Identification of each question item is carried out in the hope of finding information which is basically feedback to make improvements, improvements and re-fine the question items. So that the questions prepared can be used at a later time to measure what you want to measure (Fitrianawat, 2015).

Based on previous research, it shows that item analysis has an important role in instrument development, so that many researchers have conducted research related to item evaluation, some of which is research conducted by (Irmalasari, Suratsih, & Wibowo, 2016) with the title "Even Semester End Exam Question Item Analysis Class. Another research by (Irmalasari et al., 2016) with the title "Analysis of Odd Semester Final Exam Question Items Using the Anbuso Program at SMA Negeri 1 Boyolali" in class about the bad category 34.28%. Another research was conducted by (Kusnani, Muldayanti, & Rahayu, 2016) with the title "Analysis of Odd Semester Final Test Question Items in Biology Learning for Class sufficient reliability, good level of difficulty and good discriminating power.

This research is different from previous research which focused more on analyzing questions for other subjects such as Physics or Biology in general. This research examines students' cognitive evaluation instruments, especially Protista material in class X, which has not been widely discussed in previous literature. Although previous research has used various analysis programs such as Quest (Irmalasari et al., 2016), Anbuso (Irawati, Ekawati, & Budiawanti, 2020)and Anates (Kusnani et al., 2016), this research uses the more recent Anates application version 4.09 to analyze the quality of the questions.

In the current digital era, utilizing technology in analyzing question items can be made easier, one of which is by using Anates Software Ver 4.09. Anates is software specifically used to analyze multiple choice and essay tests (Elviana, 2020). Anates can automatically check correct and incorrect answers quickly and practically. Apart from that, the use of Indonesian in this program makes it easier for users to run Anates (Wiguna, Syaukani, & Ananda, 2018). Anates' ability to analyze test items includes: validity, reliability, level of difficulty, distinguishing power and effectiveness of distracting questions(Bella, 2018).

Difficulty level analysis aims to determine the level of difficulty of the questions for students. The difficulty level index in obtained from the comparison between the number of students who answerd correctly and total number of students who answerd the questions. The more students who answerd correctly, the higher the difficulty level index. On the other hand, if the number of students who answer correctly is less, then the question is included in the difficult level of difficulty (Hartaty & Daulay, 2018).

Measuring the discriminating power of questions shows how effective the questions are in differentiating students who have a high level of ability from students who have low ability. The discriminating power index for questions is obtained from the difference between the proportion of answers from high ability students and the correct answers from low ability students(Khaerudin, 2015). The higher the discriminating power value of the item, the better the item is at differentiating the two groups. Measuring the discriminating power of questions can measure differences in student abilities, so that it can be used to make decisions regarding students' ability levels (Hanifah, 2014).

The activity of analyzing question items to measure the level of distracting questions is called a distractor. This measurement is to analyze answer patterns by counting the number of students who choose each alternative answer for each question item. An item is said to be good if the distractor is chosen equally by students who answer the question incorrectly. On the other hand, if the question items are not good then the distractors will be chosen unevenly. A distractor question is said to be good if the number students who choose the distrasctor is the same or close to the ideal number (Z. Arifin, 2012).

A question can be said to be of good quality if the assessment of the question items is in the valid and reliable category. The validity and reliability of each instrument item is very important. Validity testing is a step to measure the accuracy of the question items to be used. Reliability testing is to determine the level of consistency of an instrument used. An instrument is said to be reliable if it gives correct results if used repeatedly(Pahri, 2021).

Questions that are valid, reliable and meet standards are the key to measuring student abilities. However, in the field there are still many questions found that do not meet the validity and reliability criteria, so they are less effective in assessing students' abilities. This mismatch prevents educators from understanding students' abilities according to their respective levels. Therefore, this research aims to analyze the quality of Biology questions on Protista class Thus, it is hoped that the results of this analysis can ensure that the questions meet the required standards.

MATERIALS AND METHODS

1. Time and Place of Research

This research was conducted at MA NU Raudhotus Shibyan which is located in Peganjaran District, Kudus City. Research activities were carried out on December 2023.

2. Research Methods

The research used is a quantitative descriptive method. Quantitative descriptive research is research that uses quantitative methods and descriptive analysis techniques (Alfatih, 2021). is research that describes, examines and explains an event based on data without testing a particular hypothesis (Wiwik Sulistyawati & Wahyudi, 2022).

3. Population ans Sample

The population of this study were students of class X MIPA MA NU Raudhotus Shibyan. The sample used was 50% of the number of students in class X MIPA. The total number of samples taken was 10 students. Sample selection was carried out by considering the proportion of the total population.

4. Research Procedure

The research procedures carried out were to measure the quality of the multiple choice questions on protist material, which included the stages of preparing the questions, trying out the questions with students, processing the data using the Anates application, and analyzing aspects of the level of difficulty of the questions, the distinguishing power of the questions, the validity and reliability of the questions. The steps taken in this research are as follows:

1) Preparation of questions

   Sole preparation is the initial stage in this research. The protist material consisted of 10 multiple choice questions. The questions are designed based on learning indicators that are relevant to the applicable curriculum.

2) Try out the questions

   The questions that had been prepared were then tested on 10 students of class X MIPA at MA NU Raudhotus Shibyan. This question trial was carried out to obtain data regarding the questions before analyzing the question items.

3) Processing question data

   The test results were then processed using the Anates application version 4.09. This application is used to process data obtained from testing questions and produce information about the quality of the questions.

4) Analyze the questions

Question analysis is carried out to analyze the level of difficulty of the questions, the distinguishing power of the questions, the validity of the questions and the reliability of the questions.

5) Interpretation of results

Based on the analysis result obtained from the anates application, interpretation is carried out to determine whether the questions are suitable for use or need improvement.

5. Data Collection

The data collection technique used is documentation tecnique. Documentation techniques are tracking and obtaining the required data through available data (Apriyanti, Lorita, & Yusuarsono, 2019)**.** In this research, documentation techniques were used to collect existing data, namely questions about protist material which were arranged in multiple choice form. Then the questions are distributed to students via Google Form to get student answers..

6. Data Analysis

Data analysis in this study was descriptive to evaluate protist material using the Anates application. Descriptive analysis aims to provide a quantitative description of the characteristics of each question item, including the difficulty level of the questions, the distinguishing power of the questions, the validity and reliability of the questions.

a. Discriminating power of questions

The discriminating power of a question is a measurement of how effective a question is in differentiating between students who have been able to master the competency and students who have not been able to master the competency. The number that shows the amount of discriminating power is called the test discrimination index or discriminating power (D). The formula used to determine differentiating power is as follows:

$$D = \frac{B_A - B_B \times 100\%}{N_A}$$

Information:

D = Differentiating Power
B_A = Number of test participants in the upper group who answerd correctly
B_B= Number of lower group test participants who answerd correctly
N_A= Number of students in either group A or B

The following is a table of criteria for determining the level of distinguishing power of questions:

Table 1. Criteria for level of differentiation of questions

Interval	Criteria
0% < DP ≤ 20%	Bad
21% < DP ≤ 40%	Enough
41% < DP ≤ 70%	Good
71 < DP ≤ 100%	Very Good

(Arikunto, 2005).

b. Question difficulty level

The difficulty level of the questions is calculated using the following equation:

$$TK = \frac{B}{N} \times 100\%$$

Information:

TK = Difficulty level
B = Number of students who answerd correctly
N = Total number of test participants

The following is a table of criteria for determining the level of difficulty of a question:

Tabel 2. Question difficulty level criteria

Intervals	Criteria
0% < TK ≤ 30%	Difficult
31% < TK ≤ 70%	Medium
71% < TK ≤ 100%	Easy

(Karnoto, 2003)

c. Test validity of the Question

The validity of the questions was analyzed using correlation analysis and valid and invalid results were obtained. Whether a question item is valid or not can be determined using correlation analysis techniques. Valid questions can be used for the next test, while invalid questions cannot be used. Calculations are carried out using the following correlation formula:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)\}\{N\sum Y^2 - (\sum Y)\}}}$$

Information:

R xy = Correlation coefficient between x and y
X = Score of the question item
Y = Total Score
N = Number of students

Interpretation of the magnitude of the correlation value can be seen in Table 3: (Arikunto, 2005)

Table 3. Validity test interval

Coefficient	Category
$0.80 < r_{xy} \leq 1.00$	Very high
$0.60 < r_{xy} \leq 0.80$	Tall
$0.40 < r_{xy} \leq 0.60$	Enough
$0.20 < r_{xy} \leq 0.40$	Low
$0.00 < r_{xy} \leq 0.20$	Very Low

(Arikunto, 2005)

d. Reliability test

Reliability is a measurement that states the level of consistency of a question, which means the extent to which a question item can produce the same results in repeated trials (Warju, Sudirman Rizki Ariyanto, Soeryanto, 2020). To find reliability, the following formula is used: (Arikunto, 2005)

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

Information:

r11 = Overall test reliability
p = Proportion of subjects who answered correctly
q = Proportion of subjects who answerd incorrectly
n = Number of items
S = Standard deviation of the test

The following is a table of reliability criteria:

Tabel 4. Reliability test interval

Intervals	Criteria
0.800-1.00	Very high
0,600- 0,799	High
0,400- 0,599	Enough
0,200-0,399	Low
0,000-0,199	Very low

RESULTS AND DISCUSSION

The question item analysis activity uses the Anates ver.4.09 application in the initial stage, namely entering student answer sheet data into the Anates program. This answer sheet data includes the answer chosen by the student for each question item and the correct answer key for that question. After the data is entered, a menu will appear to calculate the analysis of different power levels, difficulty level of questions, validity and reliability of questions. The following are the measurement results using Anates ver 4.09. Based on the data obtained from student answers, there are several questions that need to be revised, to see detailed data from the analysis results as follows.

1. Discriminating power of questions

Based on table 1, the results of the differentiating power analyzed using the Anates ver.4.09 application obtained data for the upper group and lower group. The differentiating power obtained from the test results can be seen in the following table:

Tabel 1. The results of the differentiating power of the questions

No.	Upper Group	Lower Group	Different	Differentiating Power(%)
1	3	3	0	0.00
2	3	0	3	100.000
3	2	1	1	33.33
4	3	3	0	0.00
5	1	1	0	0.00
6	0	0	0	0.00
7	0	0	0	0.00
8	2	3	1	-33.33
9	2	1	1	33.33
10	3	1	2	66.67

The differentiating power of questions was obtained from class X MA NU Raudhotus Shibyan students, totaling 10 questions. The discriminatory power item with a value of -33.33 in item number 8 is classified as the worst category item because it is a negative number. There are 5 questions that are categorized as poor, there are 5 questions in questions number 1, 4, 5, 6 and 7 so that these questions are considered not to have good differentiating power and a revision of the questions is needed. There are two questions that are in the fair category with a distinguishing power value of 33.33%, then there is one question item that is in the good category with a differentiating power value of 66.67% and one question item that is in the very good category with a differentiating power value of 100%. There are five questions that have a differentiating power of 00.00, this shows that these questions have no differentiating power at all.

The results of the analysis of 10 multiple choice questions included 50% of the questions including poor criteria, adequate categories with a proportion of 20%, good criteria with a proportion

of 10% and very good criteria with a proportion of 10%. Based on this, the questions require revision because their differentiating power is low or have no differentiating power at all, only a small number of questions meet the good and very good criteria. Items with a low category index indicate that the questions are not effective in distinguishing between students who have mastered the material and those who have not. This means that the question fails to identify differences in levels of understanding between students (Rahmasari & Ismiyati, 2016).

2. Question difficulty level

The difficulty level a questions is a measure of the proportion of students who answerd the questions correctly. The more students who answer questions correctly, the higher the difficylty level index. On the other hand, if the number of students who answer correctly is less, then the question is included in the difficult level of difficulty (Hartaty and Daulay 2018). The following are the results of calculating the difficulty level of the questions:

Table 3. Results of the difficulty level of the questions

No	correct amount	level of difficulty	Interpretation
1	10	100.000	Easy
2	6	60.000	Currently
3	7	70.000	Currently
4	10	100.000	Easy
5	3	30.000	Difficult
6	0	0.00	Difficult
7	0	0.00	Difficult
8	9	90.000	Easy
9	3	30.000	difficult
10	8	80.000	Easy

Based on the calculation results in Table 3, there are four question items number 5, 6, 7 and 9 which are included in the difficult category, there are two question items number 3 and 4 with percentages of 60% and 70%, and four question items number 1,4,8 .10 is included in the easy category, with percentages of 80%, 90% and 100%. The calculation of the difficulty level of a question is a measure of the level of difficulty of a question. The level of difficulty is one of the question parameters, namely the ratio between correct answers and the number of students who answer the question. The closer to zero the difficulty level of the question, the more difficult it is. On the other hand, the closer you get to one question, the easier it is (Restiyawati, Wayan Suama, 2020).

The categories of questions that are included in the very easy and very difficult categories need to be revised if the questions are to be used, both revising the questions and the grammar of the questions. The standardization of the questions used is questions that are not too difficult and questions that are not too easy. Very easy questions do not encourage students to solve the questions. Questions that are very difficult will cause students to not be enthusiastic about trying to do the questions because they are beyond their reach(Mustaqim, 2024).

There are several factors that cause these questions to be very difficult for students to answer, including checking the answer key again. If the answer key is correct, the question is difficult because the material has not been taught or the learning has not been completed, so students have not mastered it well. Other factors that influence the discrepancy between the material measured and the form of questions given are questions that are too long (Rahmasari & Ismiyati, 2016).

3. Test the validity of the questions

The validity test of the questions analyzed using the Anates program obtained the results in the following table:

Table 5. Validity test results

Question	Correlation	Significance	Remarks
1	NAN	NAN	Invalid
2	0.681	Significant	Valid
3	0.599	Significant	Valid
4	NAN	NAN	Invalid
5	0.257	-	Invalid
6	NAN	NAN	Invalid
7	NAN	NAN	Invalid
8	-0.131	-	Invalid
9	0.043	-	Invalid
10	0.784	very significant	Valid

Based on result of the validity test, it shows that for questions number 1, 6 and 7, the correlation from the question analysis cannot be calculated because the correlation is 00.00 so the word NAN appears, while question number 8 has a negative correlation. That way, the question automatically becomes a valid question and cannot be used. As said (Ulfah Zahiroh, 2021) That the validity of a test is influenced by the validity of the items that support it. If the validity of the question items is high, then the validity of the test as a whole is also high. Conversely, if a test produces irrelevant data then it is said that the test has low validity.

Valid questions have a strong relationship with the indicators you want to measure, so they can be used in tests. Conversely, if the question is invalid it does not provide relevant information and should be discarded (Mahmud Alpusari, 2014). Based on the results of the validity calculations shown in Table 5, it was found that of the total questions tested, 3 questions were considered valid and could be used for the test, while the other 7 questions were invalid. Thus, only questions that are proven to be valid can be maintained, while invalid questions need to be discarded or corrected.

4. Test the reliability of the questions

The result data obtained through reliability testing, the reliability value of the questions is 0.02, in accordance with the criteria in the range 0.000-0.199. This means that the questions used have very low coefficients and are less reliable in measuring student abilities. Reliability measurements must be reliable, meaning they have a level of capability. A test is said to be reliable if it is trustworthy, consistent, or constant. Measurements can be trusted if several measurements on the same group produce relatively similar results, as long as the aspect being measured does not change (Bhakti, 2015).

The level of reliability of a question is influenced by: 1) The length of the test, the longer the test, the higher the reliability. However, adding questions to the test does not always improve the quality of the test and in some cases is actually detrimental or insignificant; 2) Testee, the number of students tested will reflect variations in results which indicate the level of test reliability, whether high or low; 3) Implementation of the test which includes work instructions, supervision during the test, as well as environmental conditions and the place where the test is carried out. (Rahmasari & Ismiyati, 2016)

In relation to the discussion above, it is known that the unreliability of the questions is probably due to students' lack of preparation in working on questions on protist material because there is a lot of material that students must master, besides that the teacher does not provide a grid of questions on protist material first. This is the same as research by (Rizka Intansari, 2007), that other factors that cause unreliable question reliability scores are students' lack of experience, students' abilities and lack of preparing the material.

CONCLUSION

The result data obtained through reliability testing, the reliability value of the questions is 0.02, in accordance with the criteria in the range 0.000-0.199. This means that the questions used have very low coefficients and are less reliable in measuring student abilities. Reliability measurements must be reliable, meaning they have a level of capability. A test is said to be reliable if it is trustworthy, consistent, or constant. Measurements can be trusted if several measurements on the same group produce relatively similar results, as long as the aspect being measured does not change.

ACKNOWLEDGEMENTS

The author would like to thank the MA Raudhotus Shibyan madrasah for their cooperation during the implementation of the research. The author also expresses his gratitude to the supervisor who has provided direction, guidance and input. Thanks to the support of all parties, this research was completed successfully.

REFERENCES

Alfatih, A. (2021). *Panduan Praktis Penelitian Deskriptif Kuantitatif*. *Universitas Sriwijaya*.

Apriyanti, Y., Lorita, E., & Yusuarsono, Y. (2019). Kualitas Pelayanan Kesehatan Di Pusat Kesehatan Masyarakat Kembang Seri Kecamatan Talang Empat Kabupaten Bengkulu Tengah. *Profesional: Jurnal Komunikasi Dan Administrasi Publik*, *6*(1).72-80. https://doi.org/10.37676/professional.v6i1.839

Arifin, Z. (2012). *Evaluasi Pembelajaran*.

Arikunto, S. (2005). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Penerbit Bumi Aksara.

Bahtiar. (2020). *Evaluasi Pembelajaran Sains*. (S. Arifin, Ed.) (1st ed.). Sanabil.

Bella, O. N. (2018). Analisis Butir Soal Ujian Akhir Semester Genap Matapelajaran Danakas Bank Kelas Xi Item. *Kajian Pendidikan Akuntansi Indonesia*, *Volume 7(1)*, 1–18.

Bhakti, Y. B. (2015). Pengaruh Jumlah Alternatif Jawaban Dan Teknik Penskoran Terhadap Reliabilitas Tes. *Research and Development Journal Of Education 5*(1). 1-13.

Dwi Zunitasari, Sukarni Hidayati, T. (2016). Identifikasi Kesulitan Belajar Protista Pada Siswa Kelas X Semester 1 Sma Negeri 1 Muntilan Tahun Ajaran 2015 / 2016. *Jurnal Pendidikan Biologi*, *5(6)*, 17–27.

Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal MUDARRISUNA*, *10*(2), 58–74.

Fitrianawat, M. (2015). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, *5*(3), 282–295.

Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS*, *6*(1), 41–55.

Hartaty, L., & Daulay, S. D. (2018). Analisis Tingkat Kesukaran Dan Daya Pembeda Pada Butir Tes Soal Ujian Tengah Semester Bahasa Indonesia Kelas Xii Sma Negeri 7 Medan Tahun Pembelajaran 2016/2017. *Jurnal Unimed*, *6*(1), 15–24. https://doi.org/10.24114/kjb.v6i1.10814

Intansari, R. (2007). Analisis Perangkat Tes Ulangan Akhir Semester Gasal Mata Pelajaran Produktif Administrasi Perkantoran Kelas XII Kompetensi Keahlian Administrasi Perkantoran SMK Negeri 9 Semarang Tahun Pelajaran 2013/2014". *Universitas Negeri Semarang*.

Irawati, R., Ekawati, E. Y., & Budiawanti, S. (2020). Analisis Butir Soal Ujian Akhir Semester Gasal Menggunakan Program Anbuso di SMA Negeri 1 Boyolali Tahun Ajaran 2019/2020. *Jurnal Materi Dan Pembelajaran Fisika*, *10*(1) 11-19. https://doi.org/10.20961/jmpf.v10i1.42084

Irmalasari, K., Suratsih, & Wibowo, Y. (2016). Analisis butir soal ulangan akhir semester genap kelas X Biologi tahun ajaran 2015/2016. *Jurnal Pendidikan Biologi*, *5*(8), 10–18.

Karnoto. (2003). *Mengenal Analisis Tes. Bandung*. UPI.

Khaerudin. (2015). Analisis kualitas instrumen tes hasil belajar. *Jurnal Madaniyah*, *2*(1), 125–140.

Kusnani, Muldayanti, N. D., & Rahayu, H. M. (2016). Analisis Butir Soal Ulangan Akhir Semester Ganjil Pada Mata Pelajaran Biologi Kelas X Mia Sma Negeri 1 Sungai Raya Tahun Ajaran 2014/2015. *Jurnal Bioeducation*, *3*(2), 42–52. https://doi.org/10.29406/185

Idrus. (2019). Evaluasi Dalam Proses Pembelajaran 1. *Evaluasi Dalam Proses Pembelajaran*, *9*(2), 920-935.

Mahmud Alpusari. (2014). Analisis Butir Soal Konsep Dasar Ipa 1 Melalui Penggunaan Program Komputer Anates Versi 4.0 for Windows. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan Dan Ilmu Pendidikan Universitas Riau*, *3*(2), 106-115. https://doi.org/10.33578/jpfkip.v3i2.2501

Melandi Wimudi, Sa'diatul Fuadiyah, Zulyusri, Helsa Rahmatika, A. (2022). Pengembangan E-Lkpd Bernuansa Esq (Emotional Spiritual Quotient) Pada Materi Protista Kelas X Sma. *Bioilmi: Jurnal Pendidikan*, *8*(2), 78–90. https://doi.org/10.19109/bioilmi.v8i2.13938

Mustaqim, H. S. (2024). Analisis Butir Soal Pas Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, Dan Kualitas Pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, *3*(1), 44–56.

Pahri, E. D. (2021). Validitas Dan Tingkat Reliabilitas Soal Buatan Pilihan Ganda Tes Sumatif Siswa Kelas Xi Akl 3 Smk Negeri 4 Pekanbaru. *Pendidikan Bahasa Dan Sastra Indonesia, Universitas Riau Pendahuluan*, 1–9.

Phafiandita, A. N., Permadani, A., Pradani, A. S., & Wahyudi, M. I. (2022). Urgensi Evaluasi Pembelajaran di Kelas. *JIRA: Jurnal Inovasi Dan Riset Akademik*, *3*(2), 111–121. https://doi.org/10.47387/jira.v3i2.262

Rahmasari, D., & Ismiyati. (2016). Analisis Butir Soal Mata Pelajaran Pengantar Administrasi Perkantoran. *Economic Education Analysis Journal*, *5*(1), 317–330. http://journal.unnes.ac.id/sju/index.php/eeaj

Restiyawati, Wayan Suama, P. S. (2020). Analisis tingkat kesukaran dan daya pembeda soal ulangan biologi jurusan MIA MAN 1 Buton Selatan. *Suparyanto Dan Rosad (2015*, *5*(3), 248–253.

Sari, L. Y. (2013). Analisis Proses Pembelajaran Biologi Pada Materi Protista Di Kelas X SMA Negeri 1 Batang Anai Kabupaten Padang Pariaman. *Prosiding Semirata FMIPA Universitas Lampung*, 53–58.

Sodikin, & Gumiandari, S. (2022). Analisis SWOT Mutu Evaluasi Pembelajaran. *JDMP (Jurnal Dinamika Manajemen Pendidikan)*, *6*(1) 59-69. https://doi.org/10.26740/jdmp.v6n1.p59-69

Ulfah Zahiroh, P. S. R. (2021). Analisis Kualitas Butir Soal Pilihan Ganda Mata Pelajaran Kimia Pada Ujian Akhir Semester (Uas) Kelas Xi Man 2 Kepulauan Meranti. *Jedchem (Journal Education and Chemistry)*, *3*(1), 11–20. https://doi.org/10.36378/jedchem.v3i1.780

Warju, Sudirman Rizki Ariyanto, Soeryanto, R. A. T. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *17*(1), 95-104. https://doi.org/10.23887/jptk-undiksha.v17i1.22914

Wiguna, S., Syaukani, & Ananda, R. (2018). Kemampuan Guru PAI Merancang Tes (Analisis Aplikasi Anates Ganda Di Sekolah SMA Negeri 1 Hinai). *Edu Riligia*, *2*(1), 1–10.

Wiwik Sulistyawati, Wahyudi, S. T. (2022). KADIKMA : Jurnal Matematika dan Pendidikan Matematika. *KADIKMA: Jurnal Matematika Dan Pendidikan Matematika*, *13*(1), 67–72.