Gema WIRALODRA

Editor-in-Chief: Yudhi Mahmud

Publisher: Universitas Wiralodra

Analysis of science subject questions in class IX junior high school using Anates 4.0 computer program

Fauziah Hervi[a*], Oksa Afdarina[b], Syamsurizal[c], Lufri[d]
[a*]Universitas Padang, Indonesia,
fauziahhervi2@gmail.com
[b]Universitas Padang, Indonesia,
oksaafdarina25@gmail.com
[c]Universitas Padang, Indonesia,
syam_unp@fmipa.unp.ac.id
[d]Universitas Padang, Indonesia,
lufri_unp@yahoo.com

Orginal Article

Analysis of science subject questions in class IX junior high school using Anates 4.0 computer program

Fauziah Hervi[a*], Oksa Afdarina[b], Syamsurizal[c], Lufri[d]

[a*]Universitas Padang, Indonesia, fauziahhervi2@gmail.com
[b]Universitas Padang, Indonesia, oksaafdarina25@gmail.com
[c]Universitas Padang, Indonesia, syam_unp@fmipa.unp.ac.id
[d]Universitas Padang, Indonesia, lufri_unp@yahoo.com

[*]Corresponding Author: fauziahhervi2@gmail.com

Abstract
One of the activities to assess test quality is item analysis, which considers the general quality of the test and the quality of each item. The study was completed after tests were collected and run on all samples, and the results served as a guide for future quality improvement of the test. This study aims to analyze the validity, reliability, level of difficulty, discriminating power, and effectiveness of the distractors on Class IX Science questions at junior high school (SMP) Negeri 9 Batanghari. The data analysis technique used is a quantitative descriptive technique. This research was conducted at SMP Negeri 9 Batanghari. Class IX students of SMP Negeri 9 Batanghari were used as research subjects. The population is all class IX students at SMP Negeri 9 Batanghari, totaling 50 students. The results of this study indicate that, in terms of (1) validity, 25 items (83%) are valid, and five items (17%) are invalid. (2) Reliability: The questions have excellent reliability, with a reliability coefficient of 0.87. (3) Difficulty level: There are seven easy questions (23%), 18 moderate questions (60%), and five difficult questions. (4) The ability to discriminate the items is considered good (moderate, good, and very good) for up to 86% of the questions. (5) The quality of the distractor is considered very good for up to 100% of the questions. Therefore, this science question can be used to assess student learning outcomes at SMP Negeri 9 Batanghari because it fulfills the requirements of a learning evaluation instrument.
Keywords: Science Item Analysis, Anates Program, Validity, Reliability, Difficulty Level

1. Introduction

Education is characterized as a continuous process that involves many components. Understanding education correctly and the right introduction and appreciation greatly help prepare the learning design that will be implemented to improve the quality of education (Fuadiyah et al., 2022). The curriculum is a very important component of education, in which various systematic and complex learning plan designs are included. One of them is learning evaluation (Fadilla et al., 2019).

Evaluation is an assessment activity that goes through making a systematic plan and finding and analyzing important information needed to conclude. This important information must be appropriate and support the learning objectives contained in the lesson plan. Evaluation can provide information about student learning outcomes which can be seen from whether or not learning objectives are achieved in the classroom (Lestari et al., 2022). Learning evaluation is one of the components of the teaching system.

Evaluation serves as a controller for the achievement of student learning outcomes under the curriculum used. The assessment determines whether the previously set learning objectives have been achieved (Fietri et al., 2021). Assessment is generally categorized into three domains, namely the cognitive domain, the affective domain, and the psychomotor domain. Evaluation in learning can be done by planning assessments, compiling and making the required assessment instruments, implementing assessment activities, processing assessment results, and reporting assessment results.

Evaluation activities require much relevant information related to the object being evaluated. In learning activities, the data needed to evaluate students' behavior when

Orginal Article

participating in the learning process, daily test scores, skill scores, and final scores during semester exams. To see the achievement of student learning outcomes, it must follow the stages of the learning assessment (evaluation) process in the form of tests or other evaluation instruments that can measure the results of student learning outcomes about teaching materials and teaching methods practiced and learning objectives that have been set (Alpusari, 2014) and (Sudirman et al., 2014).

A test is interpreted as a systematic activity or process that can measure students' abilities. In general, tests are interpreted as various questions that are answered correctly or incorrectly by students so that student's abilities can be measured, differentiated, and displayed in the form of scores or assessed based on a certain scale (Afrita & Darussyamsu, 2020). A learning outcome test is a measuring instrument (instrument) used to see the achievement of learning outcomes made and given by educators to their students within a certain time. It is better if the test questions used have gone through a process of standardization and analysis of each item in depth so that the questions can measure precisely and accurately the achievement of student learning outcomes.

Test analysis is a process of standardizing and analyzing each item to get a complete and detailed interpretation of the quality of the test questions; the analysis is carried out after the test questions are compiled and tested on the sample so that the results become material for improving the quality of further test questions (Akhir Gusti, 2021). The development of science and technology (IPTEK) in the current era of globalization can be used as a medium to support learning, such as using the Anates version 4.0 for Windows program to analyze test questions made by teachers at various levels of educational units.

Karno and Yudi Wibisono are the figures who designed the Anates Program version 4.0 for Windows. This program is categorized into software specifically developed to analyze test questions in the form of multiple-choice and description test questions. The Anates version 4 for Windows program can help and make it easier for teachers to calculate each item's reliability, variability, distinguishing power, difficulty level, and quality (Subari et al., 2021).

Analysis of question items is necessary because these questions will be used as instruments or tools to measure students' abilities. The instrument must be valid to get the actual measurement results. Currently, in general, schools need to carry out the process of analyzing the quality of questions, so teachers need help to know and differentiate the abilities of their students. According to (Elviana, 2020), analyzing the questions that will be given to students is very important, considering that giving tests in the form of questions is an evaluation process that a teacher must carry out on his students as an objective assessment.

Some research on item analysis has been carried out at several levels of education in several regions, such as the analysis of items on the Plumbungan elementary school level theme evaluation using the Anates version 4.0.9 for Windows program, which revealed that overall the questions analyzed still need improvement (Akhmadi, 2021). Test question quality analysis has also been conducted at the secondary school level, such as analyzing Class VIII Biology questions at Madrasah Tsanawiyah Negeri 6 Kerinci. The results of this study illustrate that the questions have met the criteria suitable for use as a learning evaluation tool (Fietri, Lufri, et al., 2021). Some junior high schools in the Batanghari district have never carried out the process of analyzing question items.

Based on observations made at SMP Negeri 9 Batanghari, Jambi Province, the learning evaluation process is usually carried out by teachers with descriptive or essay test questions and objective (multiple choice) tests. However, the quality analysis of the test questions needs to be carried out. This is reinforced by the results of interviews with science subject teachers at the school. He stated that analyzing test questions had never been carried out. Of course, this will affect and impact the quality of test questions used as an evaluation instrument for science learning because one of the characteristics of a good and standard test question is that it must

Orginal Article

meet at least two requirements, namely having validity and reliability values (Zulyusri et al., 2017). Therefore, it is necessary to conduct a study on the feasibility of test questions. Based on these problems, conducting research activities to analyze each item of science questions in class IX SMP Negeri 9 Batanghari using Anates 4.0 for Windows software is necessary.

2. Methods

This type of research is descriptive quantitative. Quantitative research is interpreted as research in which the data is presented as numbers. Then the analysis process is carried out with appropriate statistical methods and meets scientific principles such as objective, concrete, rational, measurable, and systematic. In descriptive research, the data is presented in sentences that describe the research object in full and in detail according to the facts in the field without any general or general conclusions (Zellatifanny & Mudjiyanto, 2018).

This research was conducted at SMP Negeri 9 Batanghari. The research subjects were ninth-grade students of SMP Negeri 9 Batanghari. The population is all class IX SMP Negeri 9 Batanghari students, which amounted to 50 students. The number of students in the population is as follows.

Table 1

Total Population

Class	Number of Students
IX A	25
IX B	25

The sample was taken using a simple random sampling technique; each individual in a population has the same rights and opportunities to be selected as a research sample. The instrument used was an objective or multiple choice test question for class IX IPA SMP Negeri 9 Batanghari, which amounted to 30 items. Objective tests are tests in the form of questions whose answer choices are already available. The parameters measured in this study consist of analyzing the validity of the items, analyzing the reliability of the items, analyzing the difficulty level of the items, analyzing the differentiating power of each item, and analyzing the power of each item. This study analyzed multiple-choice questions made by teachers in science subjects. The data analysis technique is descriptive statistics using the Anates version 4.0 for Windows program.

The procedure for using the Anates 4.0 for Windows program is as follows.

1. Open the Anates program version 4.0 for Windows.
2. Then click "Create New File" so that a dialog box appears.
3. Next, fill in the dialog box according to the criteria of the question to be analyzed.
4. Then, fill in the data to be analyzed, such as the number of subjects, questions, and answer choices.
5. After that, input each student's answers according to the question number.
6. Click "save" so that the file is not lost.
7. Process the data by clicking "process all automatically."

Some data analysis was carried out on science questions with the help of Microsoft Excel and the Anates program version 4.0 for Windows, namely:

1. Test Validity

Question validation is done by empirical validation. Validation aims to determine the instrument's validity in measuring student learning outcomes. Empirical validation analysis using Microsoft Excel software. The item is valid if the correlation of the item score with the total score is significant or very significant in the data processing.

2. Test Reliability

Instrument reliability can be known from item analysis with Anates version 4.0. The interpretation of the reliability value used by researchers is as follows (Sundayana, 2016).

Orginal Article

0,00 ≤ r < 0,20: Very low
0,20 ≤ r < 0,40: Low
0,40 ≤ r < 0,60: Medium
0,60 ≤ r < 0,80: High
0,80 ≤ r ≤ 1,00: Very High

3. Differentiating Power

The question's differentiating power is obtained from the results of Anates version 4.0. The classification of differentiating power interpretations used by researchers is as follows.

Table 2
A *Criterion of Differential Power*

Index of Differential Power	Classification
0,00 - 0,20	Poor
0,21 – 0,40	Satisfactory
0,41 – 0,70	Good
0,71 – 1.00	Excellent

(Sudijono, 2018)

4. Index of Difficulty

The instrument's difficulty level is obtained from item analysis using Anates version 4.0. The following are the criteria for interpreting the question difficulty index used.

Table 3
Criteria for Interpreting the Level of Difficulty

Index of Difficulty	Classification
0,00 – 0,30	Difficult
0,31 – 0,70	Medium
0,71 – 1,00	Easy

(Sudijono, 2018)

5. Quality of Outliers

A question examiner is said to be able to carry out its function well if it has a such allure that students who do not understand the concept feel confused and finally deceived into choosing the answer to the examiner as the right choice (Rahayu & Djazari, 2016). The following is a table of the examiner index value, and the distribution of the analysis results of the quality of the science question examiners used.

Table 4
Criteria for the Quality of an Outlier

Outlier Index	Quality
0	Very Good
1	Good
2	Less Good
3	Bad
4	Very Bad

(Arbiatin & Mulabbiyah, 2020)

3. Results And Discussion

Research Results

One of the activities to assess test quality is item analysis, which considers the general quality of the test and the quality of each item. This research is completed after the test is collected and administered to all samples, and the results serve as a guide for test quality improvement. This test analysis activity is necessary (Fietri, Zulyusri, et al., 2021). Item analysis is very important to improve the standard of answers given by students (Amelia et al., 2021). This study will

Orginal Article

scrutinize each item to ensure its high quality before use. The validity, reliability, differentiating ability, difficulty level, and quality of the item checkers are all examined in this study.

Validity Test

The results of the question validity analysis are presented in the Table below.

Table 5

Results of Validity Analysis of IPA Questions

No.	R-hitting	R-tabel	Criteria
1	0,717	0,2787	Valid
2	0,802	0,2787	Valid
3	0,575	0,2787	Valid
4	0,545	0,2787	Valid
5	0,333	0,2787	Valid
6	0,420	0,2787	Valid
7	0,598	0,2787	Valid
8	0,270	0,2787	Invalid
9	-0,297	0,2787	Invalid
10	-0,074	0,2787	Invalid
11	0,560	0,2787	Valid
12	0,662	0,2787	Valid
13	0,443	0,2787	Valid
14	0,543	0,2787	Valid
15	0,472	0,2787	Valid
16	0,570	0,2787	Valid
17	-0,274	0,2787	Invalid
18	0,689	0,2787	Valid
19	0,487	0,2787	Valid
20	0,649	0,2787	Valid
21	0,594	0,2787	Valid
22	0,587	0,2787	Valid
23	0,747	0,2787	Valid
24	0,816	0,2787	Valid
25	0,384	0,2787	Valid
26	0,691	0,2787	Valid
27	-0,310	0,2787	Invalid
28	0,494	0,2787	Valid
29	0,446	0,2787	Valid
30	0,614	0,2787	Valid

Based on the analysis results, 25 questions (83%) were classified as valid criteria, and five (17%) were classified as invalid.

Reliability Test

Based on calculations carried out with Anates version 4.0, the reliability test results obtained were 0.87. The results of these calculations indicate that the ninth-grade science questions at SMP Negeri 9 Batanghari have a very high level of reliability, with a range of 0.80-1.00. The extent to which a test consistently measures something reliable is called test reliability (Loka Son, 2019). Reliability shows that the instrument is trustworthy enough to be used as a tool for collecting data (Syahril et al., 2019).

Differentiating Power

The calculation using Anates version 4.0 shows that thirteen questions (43%) have good differentiating power, ten questions (33%) have good differentiating power, four questions (13%) have poor differentiating power, and three questions (10%) have sufficient

Orginal Article

differentiating power. The distribution of findings from various power analyses of grade IX science questions at SMP Negeri 9 Batanghari is shown in the Table below.
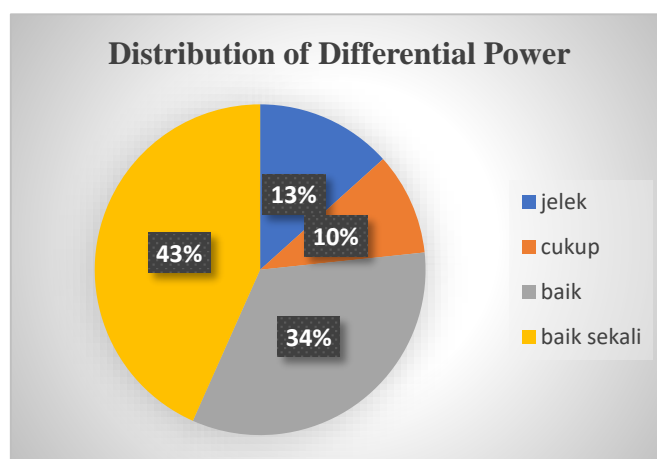
Table 6

Distribution of Differential Power

Differentiating Power	Question Number	Total	Percentage
0,00 – 0,20 (Bad)	9, 10, 17, 27	4	13%
0,21 – 0,40 (Satisfactory)	5, 8, 29	3	10%
0,41 – 0,71 (Good)	3, 4, 6, 7, 13, 14, 15, 19, 25, 28	10	34%
0,71 – 1,00 (Very Good)	1, 2, 11, 12, 16, 18, 20, 21, 22, 23, 24, 26, 30	13	43%

Figure 1

Distribution of Differential Power

Based on Table 4. and Figure 1. it can be explained that the science questions of class IX SMP Negeri 9 Batanghari have good differentiating power (sufficient, good, and very good), amounting to 86% of the total questions.
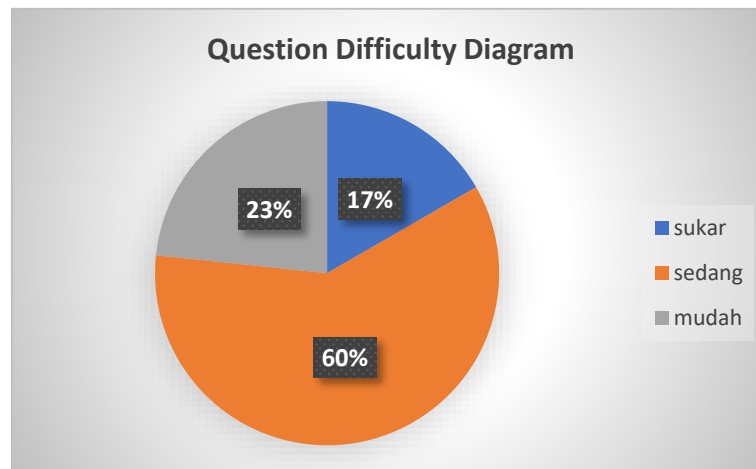
Difficulty Level

The results of the analysis of the difficulty level of class IX science questions at SMP Negeri 9 Batanghari are shown in the following Table and figure.

Table 7

Question Difficulty Level

Indeks of Difficulty	Item Number	Total	Percentage
0,00 – 0,30 (Difficult)	2, 9, 10, 17, 27,	5	17%
0,31 – 0,70 (Medium)	1, 4, 6, 7, 8, 11, 12, 13, 14, 16, 18, 19, 20, 23, 24, 25, 26, 28,	18	60%
0,71 – 1,00 (Easy)	3, 5, 15, 21, 22, 29, 30	7	23%

Orginal Article

Figure 2
Distribution of Question Difficulty

Question Difficulty Diagram

They based on the difficulty level calculation analysis using Anates version 4 software in Table. 6, six conclusions were drawn, namely, five items (17%) had difficult criteria, 18 items (60%) had moderate criteria, and seven items (23%) had easy criteria. The items' difficulty levels must be analyzed as diagnostic tools for students with learning problems or to improve classroom-based evaluation. The difficulty level of the object, which can be determined by item analysis, also affects whether it is good or bad (Putri et al., 2022).

Quality of Outliers

The function of a test item's exception can be understood as a pattern that explains how learners choose their responses to the potential answers provided in each test item's answer choices. The more learners who choose the answer to an exception, the more effective it will be; conversely, if the exception answer is not chosen, then the exception does not function (Arbiatin & Mulabbiyah, 2020). The following are the results of analyzing the quality of science question exemptions.

Table 7
The result of the Quality of Outliers

Outlier Index	Question Number	Total	Percentage
0 (Very Good)	1-30	30	100%
1 (Good)	-	-	0%
2 (Less Good)	-	-	0%
3 (Bad)	-	-	0%
4 (Very Bad)	-	-	0%

Orginal Article

Gambar 3
Distribution of Quality Outliers
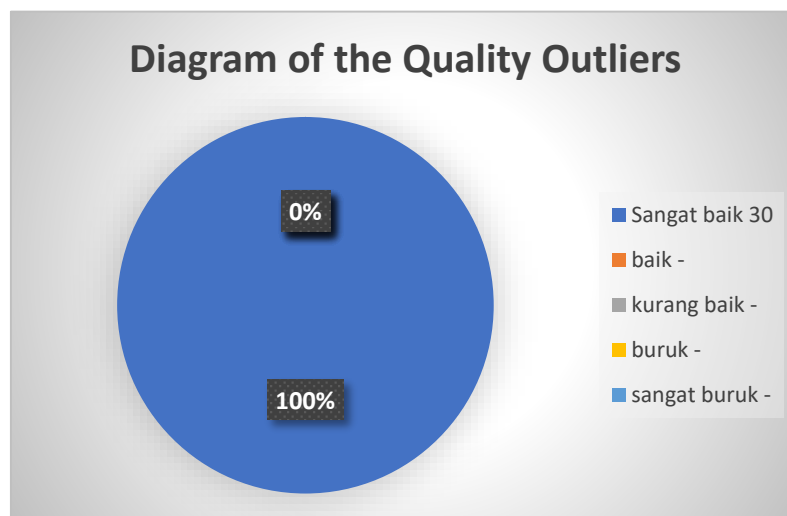
Based on calculations with Anates version 4.0, each science question for class IX at SMP Negeri 9 Batanghari has a high exception value of 100%.

Discussion

Based on the results of the analysis of science items in class IX SMP Negeri 9 Batanghari using the help of Microsoft Excel and Anates Software above, this discussion will discuss in detail the results of the analysis, including validity, reliability, difficulty level, distinguishing power, and the effectiveness of the examiner. The discussion of the results of the research that has been carried out is explained as follows.

Validity

Validity is the degree of conformity between information about the research object and the quality tested by the researcher. The results are valid if the facts are comparable to the facts reported by the researcher and the facts about the research subject (Anazalia et al., 2020).

Testing the validity of items on science questions at SMP Negeri 9 Batanghari uses the r-count search using Microsoft Excel software. At a significance level of 5%, the results of this calculation are then compared with the r-table. With the number of students as research subjects, 50 people, at a significance level of 5% and n = 50, the r-table value is 0.2787. The item is said to be valid if r-count> r-table, and vice versa. The item is said to be invalid if r-count < r table. A measurement or observation falls under the notion of validity, which also refers to the concept of reliability of the instrument in data collection. The tool must be able to measure what it is designed to detect. Therefore, validity emphasizes the measurement or observation method. Following the findings of the analysis of scientific questions at SMP Negeri 9 Batanghari, 25 questions, or 83%, included valid criteria, and five items, or 17%, included invalid criteria. The description concludes that the science exam questions at SMP Negeri 9 Batanghari are valid and very good questions.

Reliability

The reliability test aims to ascertain whether the instrument is reliable and always consistent during repeated research (Hasan, 2014). The capacity of a measuring instrument to make precise measurements is related to reliability. Reliability refers to the accuracy and precision that the instrument produces when taking measurements (Zein et al., 2013). An accurate measuring instrument will produce a precise measurement. Produces consistent findings when tested on the same group multiple times. With a reliability value of 0.87, the instrument used for the 9th-grade science exam of SMP Negeri 9 Batanghari is very trustworthy. If r is 0.70 or

	Orginal Article

higher, the test is considered reliable (highly reliable) (Dewi et al., 2020). With a rating of 0.87, the ninth-grade science test items of SMP Negeri 9 Batanghari are very reliable.

Differentiating Power

The analysis of the differentiating power of science questions shows that the questions have good differentiating power with criteria (sufficient, good, and very good), a total of 86% of all questions. This science question is suitable as an instrument that can be used for learning evaluation. Questions with insufficient discriminating power should be improved thoroughly by identifying the root cause of the problem's failure. Questions with good, excellent, and sufficient discriminating power should be saved by adding them to the question bank. One way to prevent learners who can answer questions from becoming confused is to revise questions with unclear wording. Questions should be able to distinguish between learners who understand the subject and those who do not (Muluki et al., 2020). Based on the analysis results, the questions meet the requirements as an instrument that can be used to assess learning.

Difficulty Level

Science questions in class IX SMP Negeri 9 Batanghari have a moderate level of difficulty; according to the analysis of calculations using Anatès version 4 for Windows, several items, or 60% of them, have a moderate level of difficulty. The question's difficulty level is a number that describes how challenging the question item is. It is considered good if the test question needs to be more complex and more manageable (Arbiatin & Mulabbiyah, 2020). Learners need to be motivated to increase their effort in answering questions that are too easy and simple (Iskandar & Rizal, 2017). On the other hand, questions that are too challenging will demoralize learners and make them less motivated to try again (Solichin, 2017). Analysis of the difficulty level of the questions is done so that it is easy to identify the standard of superior questions (Jumrodah et al., 2023). According to (Abdul, 2015), it is emphasizing that test instruments need to classify questions into easy, medium, and difficult levels.

Research (Pratiwi et al., 2013) stated that question adjustments must be made to increase the difficulty level of the questions. In the next exam, questions with a high difficulty level (moderate or moderate) can be added to the question bank. For questions with a high level of difficulty (difficult), one of three possible follow-up procedures will be used, namely discarding the question, re-examining it, or revising it again. Questions with an easy difficulty level also have three potential follow-up actions that can be taken. Namely, the question can be discarded, re-examined, or revised again. These items can be used for very lenient tests because most participants pass the test, so the test is considered successful.

The Power of an Outlier

Science questions in class IX SMP Negeri 9 Batanghari have a very high quality of checking. Therefore, this science question instrument meets the requirements of questions that can be used to evaluate learning. Questions that are highly effective as distractors can be stored in the question bank or reused in future exams. Solutions for questions with poor tracer scores can be revised for reuse in future tests; There is no need to reuse questions of substandard or very poor quality. Exception items whose efficacy is below average or very low indicate that the tool needs to be able to work properly and have an attraction for students who need help understanding the concept or subject matter (Rahayu & Djazari, 2016).

4. Conclusion

Based on the research findings, it can be concluded that the ninth-grade science questions of SMP Negeri 9 Batanghari have a very high item reliability of 0.87, a good level of difficulty or 60% have a medium level of difficulty, the quality of the question exemption is 100% of all questions. In addition, the distinguishing power of the items is considered good (sufficient, good, and very good) in as much as 86% of the total questions. This science question can be used to assess student learning outcomes at SMP Negeri 9 Batanghari because it meets the requirements as a learning evaluation instrument.

Orginal Article

5. References

Abdul, K. (2015). Menyusun dan menganalisis tes hasil belajar. *Al-Ta'dib*, *8*(2), 70–81.

Afrita, M & Darussyamsu, R. (2020). Validitas Instrumen Tes Berpikir Tingkat Tinggi (HOTS) pada Materi Sistem Respirasi di Kelas XI SMA. *Mangifera Edu*, *4*(2), 129–142. https://doi.org/10.31943/mangiferaedu.v4i2.83

Akhir Gusti, U. (2021). *Quality analysis of natural science subject questions at SMP Islam Khaira Ummah*, *3*(2), 81–87.

Akhmadi, M. N. (2021). Analisis butir soal evaluasi tema 1 kelas 4 sdn plumbungan menggunakan Program Anates. *Ed-Humanistics*, *6*(1), 799–806. https://doi.org/10.33752/ed-humanistics.v6i1.1464

Alpusari, M. (2014). Analisis butir soal, program komputer anates versi 4.0 for windows mahmud alpusari. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Universitas Riau*, *3*(2), 106–115.

Amelia, N., Paridjo, & Sina, I. (2021). Analisis Butir soal penilaian akhir tahun mata pelajaran matematika menggunakan Software Anates. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika*, *12*(3), 447–456. https://doi.org/10.26877/aks.v12i3.8356

Anazalia, M., Darussyamsu, R., Lufri, L., Syamsurizal, S., & Atifah, Y. (2020). Validitas instrumen tes berpikir tingkat tinggi (Hots) tentang materi sistem pencernaan untuk peserta didik kelas XI SMA/MA. *BIOTIK: Jurnal Ilmiah Biologi Teknologi Dan Kependidikan*, *8*(2), 232–244. https://doi.org/10.22373/biotik.v8i2.7090

Arbiatin, E., & Mulabbiyah, M. (2020). Analisis kelayakan butir soal tes penilaian akhir semester mata pelajaran matematika kelas VI di SDN 19 Ampenan Tahun Pelajaran 2019/2020. *El Midad: Jurnal PGMI*, *12*(2), 146–171. https://doi.org/10.20414/elmidad.v12i2.2627

Dewi, N. P., Rahmi, Y. L., Alberida, H., & Darussyamsu, R. (2020). Validitas dan reliabilitas instrumen penilaian kemampuan berpikir tingkat tinggi tentang materi hereditas untuk peserta didik SMA/MA. *Jurnal Eksakta Pendidikan (Jep)*, *4*(2), 138. https://doi.org/10.24036/jep/vol4-iss2/512

Elviana. (2020). Analisis Butir soal evaluasi pembelajaran pendidikan agama islam menggunakan Program Anates. *Jurnal Mudarrisuna*, *10*(2), 58–74. http://dx.doi.org/10.22373/jm.v10i2.7839

Fadilla, I., Maharani, B., Usmeldi, & Festiyed. (2019). Meta-analisis pengembangan bahan ajar berbasis riset pada pembelajaran IPA dan Fisika. *Jurnal Penelitian Pembelajaran Fisika*, *5*(2), 188–195. https://doi.org/10.24036/jppf.v5i2.107489

Fietri, W. A., Lufri, L., Syamsurizal, S., & Zulyusri, Z. (2021). Analisis butir soal biologi kelas VIII Madrasah Tsanawiyah Negeri 6 Kerinci. *Jurnal Pendidikan Biologi Undiksha*, *8*(2), 50–60. https://ejournal.undiksha.ac.id/index.php/JJPB/article/view/35064

Fietri, W. A., Zulyusri, & Violita. (2021). Analisis Butir Soal Biologi Kelas XI Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, *7*(1), 28–35.

Fuadiyah, S., Widora, S., Alberida, H., & Fadilah, M. (2022). Analisis kebutuhan pengembangan penilaian bilogi berbasis kerangka quellmalz pada materi jaringan tumbuhan untuk peserta didik SMA/MA Kelas XI. *Biodidaktika: Jurnal Biologi dan Pembelajarannya*, *17*(2), 95–107. http://dx.doi.org/10.30870/biodidaktika.v17i2.16502

Hasan, I. (2014). *Analisis Data Penelitian dengan Statistik*. Jakarta: Bumi Aksara.

Iskandar, A., & Rizal, M. (2017). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 12–23. https://doi.org/10.21831/pep.v22i1.15609

Jumrodah, J., Ningrum, S. W., & Lestari, S. (2023). Analisis butir soal keterampilan berpikir tingkat tinggi siswa XI pada materi sistem gerak manusia. *Jurnal Biopendix*, *9*(2), 172–

Orginal Article

177.

Lestari, N. I., Razak, A., Lufri, L., & Zulyusri, Z. (2022). Analisis pelaksanaan evaluasi untuk meningkatkan mutu pembelajaran biologi di sekolah. *Bioilmi: Jurnal Pendidikan*, *8*(2), 114–127. https://doi.org/10.19109/bioilmi.v8i2.11629

Loka Son, A. (2019). Instrumentasi kemampuan pemecahan masalah matematis: analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema Wiralodra*, *10*(1), 41–52. https://doi.org/10.31943/gemawiralodra.v10i1.8

Muluki, A., Bundu, P., & Sukmawati. (2020). Analisis kualitas butir tes semester ganjil mata pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, *4*(1), 86–96. https://doi.org/10.23887/jisd.v4i1.23335

Pratiwi, D., Zakso, A., & Ibrahim, M. Y. (2013). Analisis kualitas soal semester ganjil pelajaran sosiologi kelas XI SMA. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, *Vol 2*(1).

Putri, M. E., Syamsurizal, S., & Lufri, L. (2022). Analyzed question quality of school examinations class IX Science Subjects. *Bioeducation Journal,* *6*(1), 40-47. http://bioeducation.ppj.unp.ac.id/index.php/bioedu/article/view/318

Rahayu, R., & Djazari, M. (2016). Analisis kualitas soal pra ujian nasional mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, *14*(1), 85–94. https://doi.org/10.21831/jpai.v14i1.11370

Solichin, M. (2017). Analisis daya beda soal taraf kesukaran,butir tes, validitas butir tes, interpretasi hasil tes valliditas ramalan dalam evaluasi pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*, *2*(2), 192–213.

Subari, A., Lufri, L., & Syamsurizal, S. (2021). Analisis butir soal ujian akhir semester ganjil biologi kelas XI MAN 2 Kota Jambi. *Bioscientist: Jurnal Ilmiah Biologi*, *9*(1), 45–53. https://doi.org/10.33394/bjib.v9i1.3612

Sudijono. (2018). *Pengantar Evaluasi Pendidikan*. Rajawali Pers.

Sudirman, S., Kusumah, Y. S., & Martadiputra, B. A. P. (2021). Augmented reality blended learning instruction: the impact on growing motivation, attitudes, and knowledge in 3D geometry. *Turkish Journal of Computer and Mathematics Education*, *12*(4), 674-683.

Sundayana, R. (2016). *Statistika Penelitian Pendidikan*. Bandung: Alfabeta.

Syahril, Hadiyanto, Adi, N., & Ningrum, T. A. (2019). Pemberdayaan aset sekolah dan instansi pemerintah melalui workshop sistem manajemen aset. *Jurnal Halaqah*, *1*(4), 433–438. https://doi.org/10.5281/zenodo.3522456

Zein, A., Fadillah, M., & Novianti, R. (2013). Hubungan Antara validitas butir, reliabilitas, tingkat kesukaran dan daya pembeda soal ujian semester genap bidang studi biologi kelas XI SMA/MA Negeri di Kota Padang Tahun Pelajaran 2010/2011. *Prosiding Semirata FMIPA Universitas Lampung*, 39–47. https://jurnal.fmipa.unila.ac.id/semirata/article/download/572/392

Zellatifanny, C. M., & Mudjiyanto, B. (2018). Tipe Penelitian deskripsi dalam ilmu komunikasi. *Diakom : Jurnal Media dan Komunikasi*, *1*(2), 83–90. https://doi.org/10.17933/diakom.v1i2.20

Zulyusri, Z., Sumarmin, R., & Miswati, M. (2017). Development of biology item test based on science literacy for senior high school students class X Semester 1. *Bioeducation Journal*, *1*(1), 88–94. https://doi.org/10.24036/bioedu.v1i1.34