JPPIPA 9(Special Issue) (2023)

Jurnal Penelitian Pendidikan IPA

Jurnal Penelitian Pendidikan Sains

http://jppipa.unram.ac.id/index.php/jppipa/index

Analysis of the Quality of Scientific Literacy Question Items on Biodiversity Material Using the ANATES Application

Fauziah Hervi[1], Syamsurizal[2*], Fitri Arsih[2], Linda Advinda[2], Mutia Sari[3]

[1]Graduate School, Padang State University, Indonesia.
[2]Department of Biology, Faculty of Mathematics and Natural Sciences, Padang State University, Indonesia.
[3]Biology Teacher, Adabiah 2 High School Padang, Indonesia.

Received: October 29, 2023
Revised: December 5, 2023
Accepted: December 25, 2023
Published: December 31, 2023

Corresponding Author:
Syamsurizal
syam_unp@fmipa.unp.ac.id

DOI: 10.29303/jppipa.v9iSpecialIssue.6304

Abstract: This research aims to analyze the quality of scientific literacy questions on biodiversity material using the ANATES application. ANATES (Test Analysis) is a computer application specifically developed to measure and evaluate the quality of exam questions. The research was conducted by collecting data from a collection of scientific literacy questions on biodiversity material and then analyzing it using ANATES. This type of research is descriptive quantitative. This research was carried out at Padang State University. The research subjects are Biology students in 2023 who are taking basic science courses. The population is all biology students in 2023 who are taking basic science courses. In this case, the sample used was only one class, namely biology education class C, totaling 35 people. The instrument used was scientific literacy questions on biodiversity material, totaling 25 questions. Next, the items are analyzed using ANATES to evaluate validity, reliability, level of difficulty, discriminating power, and distractor quality. The results of the analysis provide an in-depth picture of the quality of the scientific literacy questions on biodiversity material. It is hoped that the findings of this research can contribute to the development of better evaluation instruments for measuring students' understanding of biodiversity material. The implications of this research can be used by teachers, curriculum developers, and educational researchers to improve the quality of science learning at the secondary education level.

Keywords: ANATES; Biodiversity; Quality of Question Items; Scientific Literacy

Introduction

Advances in science and technology have positive and negative impacts on human life. Positive effects arise due to various conveniences that can improve the quality of human life. Issues regarding ethics, morals, and global affairs are negative impacts resulting from the development of science and technology (Sari et al., 2017). Students need to be equipped with the ability to care and be responsive to issues that are developing in society. The abilities in question are thinking critically and creatively to plan problem-solving, as well as having in-depth knowledge and understanding to apply in problem-solving (Eny & Wiyarsi, 2019). This can be achieved if students have scientific literacy.

Scientific literacy refers to the multiplicity of literacies related to the use of digital technology in the field of science that supports the learning process (Widiatmo et al., 2019). Scientific literacy skills play a role in preparing a generation that can solve challenges and problems in society scientifically and responsibly (Dirman & Mufit, 2022). Science education has a strategic role in developing students' understanding of the environment and making them citizens who are more aware of sustainability issues. One key aspect of the education system is assessment (Putri et al., 2022), which can provide an overview of students' understanding and skills in a subject.

The development of appropriate evaluation instruments such as scientific literacy questions is very important to measure students' understanding and skills

How to Cite:
Hervi, F., Syamsurizal, S., Arsih, F., Advinda, L., & Sari, M. (2023). Analysis of the Quality of Scientific Literacy Question Items on Biodiversity Material Using the ANATES Application. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 1062–1069. https://doi.org/10.29303/jppipa.v9iSpecialIssue.6304

in understanding biological concepts (Huryah et al., 2017). In the context of scientific literacy, especially in biodiversity material, it is important to ensure that the questions used in the assessment are of good quality. The quality of the question items greatly influences the validity and reliability of the evaluation instrument (Abdul, 2015).

Previous research has shown that the quality of scientific literacy questions has a direct impact on student learning outcomes (Adawiyah & Wisudawati, 2017). Therefore, there is a need to make efforts to identify, analyze, and improve the quality of scientific literacy questions, especially in the context of biodiversity material. Previous research (Hervi et al., 2023) also shows that the use of technology can make a significant contribution to the analysis process and improve the quality of question items.

To improve the quality of the test items, the ANATES application emerged as a tool that could make a significant contribution. ANATES offers the ability to efficiently analyze the validity, reliability, level of difficulty, discriminating power, and distractor quality of questions (Alpusari, 2014). ANATES can provide in-depth information about the characteristics of test items, thereby enabling educators and researchers to make breakthrough improvements in the development of evaluation instruments (Fietri et al., 2021). Therefore, this article aims to look at the quality of scientific literacy questions using ANATES, especially on biodiversity material. Thus, this research is not only a step forward in updating evaluation methods but also opens up the potential for continued improvement in science education in the future.

Method

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

This research was carried out at Padang State University. The research subjects are Biology students in 2023 who are taking basic science courses. The population is all biology students in 2023 who are taking fundamentals of science courses. Sampling was carried out using a simple random sampling technique, that is, every individual in a population has the same rights and opportunities to be selected as a research sample. In this case, the sample used was only one class, namely biology education class C, totaling 35 people.

The instrument used was scientific literacy questions regarding biodiversity material, totaling 25 questions. The questions consist of objective questions, essay questions, and true-false questions. The parameters measured in this research consist of an analysis of the validity of the question items, an analysis of the reliability of the question items, an analysis of the level of difficulty of the question items, an analysis of the distinguishing power of each question item and analysis of the distracting power of each question item. This research analyzes scientific literacy questions regarding biodiversity material. The data analysis technique is descriptive statistics using the Anates program version 4.0 for Windows.

The procedure for using the Anates 4.0 for Windows program is to open the Anates program version 4.0 for Windows. Then click "Create New File" so that a dialog box appears. Next, fill in the dialog box according to the criteria for the question to be analyzed. Then, fill in the data to be analyzed such as the number of subjects, number of questions, and number of answer choices. After that, input the answers from each individual according to the question number. Click "save" so that the file is not lost and process the data by clicking "process all automatically". Several data analyses were carried out on scientific literacy questions with the help of the Anates version 4.0 for Windows program, that is:

Test Validity

Validation of the questions is carried out by empirical validation. The purpose of validation is to find out the validity value of the instrument in measuring learning outcomes (Ardi et al., 2023). Empirical validation analysis using Microsoft Excel software. A question item is said to be valid if the correlation between the question item score and the total score is significant or very significant in data processing (Hidayati et al., 2023).

Test Reliability

The reliability of the instrument can be determined from the analysis of the questions with Anates version 4.0. The interpretation of reliability values used by researchers is as follows (Sundayana, 2016).

Table 1. Reliability Value Criteria

Reliability Value	Criteria
$0.00 \leq r < 0.20$	Very Low
$0.20 \leq r < 0.40$	Low
$0.40 \leq r < 0.60$	Medium/Fair
$0.60 \leq r < 0.80$	High
$0.80 \leq r \leq 1.00$	Very High

Differentiating Power

Discriminative power analysis is used to examine test questions from the aspect of the test's ability to differentiate students who are in the low and high categories (Rokhim et al., 2023). The higher the value of the discriminating power of the questions, the better it will be in differentiating the abilities or achievements of students. A high discriminating power value indicates that the quality of the items is better in differentiating between students who achieve high achievement and students who achieve low achievement (Jusniar et al., 2020). The differentiating power of the questions was obtained from the results of Anates version 4.0. The classification of differentiating power interpretations used by researchers is as follows.

Table 2. Differentiating Power Criteria

Differential Power Index	Classification
0.00 - 0.20	Poor
0.21 – 0.40	Satistifactory
0.41 – 0.70	Good
0.71 – 1.00	Excellent

(Sudijono, 2018)

Difficulty Index

The difficulty level of the instrument was 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 Difficulty Levels

Difficulty Index	Classification
0.00 – 0.30	Hard
0.31 – 0.70	Currently
0.71 – 1.00	Easy

(Sudijono, 2018)

Distractor Qualities

Question distractors are said to be able to carry out their function well if they have such an allure that individuals who do not understand the concept feel confused and are ultimately tricked into choosing the distractor answer as the correct choice (Rahayu & Djazari, 2016). The following is a table of distractor index values and the distribution of the results of the analysis of the quality of distractors on the scientific literacy questions used.

Table 4. Interpretation Criteria for Distractor Quality

Distraction Index	Quality
0	Excellent
1	Good
2	Not Good
3	Bad
4	So Bad

(Arbiatin & Mulabbiyah, 2020)

Result and Discussion

One of the activities to assess the quality of the test is question analysis which considers the general quality of the test and the quality of each question item. This study was completed after the tests were collected and carried out on all samples, and the results served to ensure the validity, reliability, discriminating power, difficulty index, and distractor quality of the instrument developed (Fietri, Zulyusri, dan Violita. 2021). Each question item will be examined in this research to ensure that the question items are of high quality before being used in research. Therefore, this test analysis activity is very necessary. Question analysis is very important to improve the standard of answers given (Amelia, Paridjo, dan Sina. 2021). Validity, reliability, discriminating ability, level of difficulty, and distractor quality of items were all examined in this study.

Validity Test

The validity of an instrument can be determined by looking at its validity level. A measurement or observation is included in the notion of validity, which also refers to the concept of reliability of the instrument in data collection (Rahmi et al., 2021). The tool must be able to measure what it is designed to detect (Alifah & Istiyono, 2023). Therefore, validity emphasizes measurement or observation methods. Validity is the level of conformity between information about the research object and the quality tested by the researcher. Results are said to be valid if the facts are comparable to the facts reported by the researcher and are facts about the research subject (Anazalia dkk. 2021).

Testing the validity of items on scientific literacy questions uses an r-calculation search using Microsoft Excel software. At a significance level of 5%, the results of this calculation are then compared with the r-table. With a sample size of 35 research subjects, at a significance level of 5% and n=35, an r-table value of 0.3338 was obtained. The question item is said to be valid if r-count > r-table, and conversely the question item is said to be invalid if r-count < r table. The results of the question validity analysis for objective questions and essay questions are presented in the table below.

Table 5. Results of Validity Analysis of Scientific Literacy Objective Questions

No. question	R-count	R-table	Conclusion
1	0.420	0.3338	Significant
2	0.348	0.3338	Significant
3	0.103	0.3338	-
4	0.639	0.3338	Significant
5	0.379	0.3338	Significant
6	0.595	0.3338	Significant
7	0.477	0.3338	Significant
8	0.319	0.3338	-
9	0.520	0.3338	Significant
10	0.603	0.3338	Significant
11	0.041	0.3338	-

Table 6. Results of Validity Analysis of Science Literacy Essay Questions

No. question	R-count	R-table	Conclusion
1	0.499	0.3338	Significant
2	0.512	0.3338	Significant
3	0.566	0.3338	Significant
4	0.626	0.3338	Significant
5	0.199	0.3338	-
6	0.336	0.3338	Significant
7	0.369	0.3338	Significant
8	0.437	0.3338	Significant
9	0.594	0.3338	Significant
10	0.367	0.3338	Significant
11	0.636	0.3338	Significant
12	0.607	0.3338	Significant
13	0.742	0.3338	Significant

Following the results of the analysis of scientific literacy questions, on objective questions, 8 questions, or 73% included valid criteria, and 3 questions or 27% included invalid criteria. Meanwhile, in the essay questions, 12 questions, or 92% included valid criteria and 1 item or 8% included invalid criteria. Invalid question items are discarded or revised so they can be used. According to (Kimberlin & Winterstein, 2008), validity refers to whether the information obtained from a test represents the actual understanding of the test taker. Based on the table above, most of the question items can be said to be valid. The validity of the test items is closely related to the discrimination index or differentiating power because if the test items can differentiate between high and low-achieving students, it means that the test items are trusted to measure conception (Jusniar et al., 2020). Based on the results of the analysis, it lead to the conclusion that the scientific literacy questions analyzed are valid and very good.

Reliability Test

Reliability testing aims to ensure that the instruments used are reliable and always consistent when repeated research is carried out (Hasan, 2014). The capacity of a measuring instrument to make precise measurements is related to reliability. Reliability refers to the accuracy and precision that an instrument produces when making measurements (Zein et al., 2013). The right size will be produced by an accurate measuring tool.

Based on calculations carried out with Anates version 4.0, the reliability test results for objective questions were 0.55 in the medium category and the reliability test results for essay questions were 0.72 in the high category. The results of these calculations show that the scientific literacy questions have a medium and high level of reliability with a range of $0.40 \leq r < 0.60$ (medium) and $0.60 \leq r < 0.80$ (high).

According to (Creswell, 2012), test instrument results have high internal consistency or regularity. This is supported by research from (Utari et al., 2018) regarding the development of a four-level misconception diagnostic test instrument with a reliability test result of 1.067 which is included in the very reliable category. The extent to which a test consistently measures something reliably 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

Calculations on objective questions using Anates version 4.0 show that five questions (45%) have excellent discriminating power, two questions (19%) have good discriminating power, one question (9%) has sufficient discriminating power, and three questions (27%) had poor discriminating power. The distribution of findings from various power analyses of scientific literacy objective questions is shown in Table 7 and Figure 1.

Table 7. Distribution of Different Powers on Objective Questions

Differentiating Power	Question Item Numbers	Total	Percentage (%)
0.00 – 0.20 (Bad)	2, 3, 11	3	27
0.21 – 0.40 (Enough)	7	1	9
0.41 – 0.71 (Good)	1, 8	2	19
0.71 – 1.00 (Exellent)	4, 5, 6, 9, 10	5	45

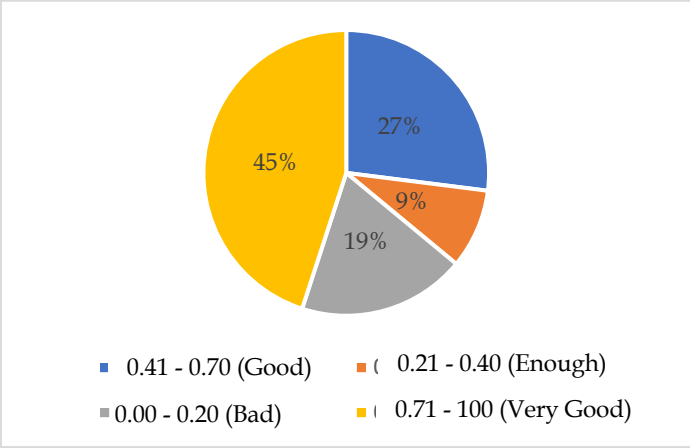

Figure 1. Percentage of Distinguishing Power on Science Literacy Objective Questions

Calculations on essay questions using Anates version 4.0 shows that one question (10%) has very good differentiating power, two questions (14%) have good differentiating power, nine questions (66%) have sufficient differentiating power, and one question (10%) had poor discriminating power. The distribution of findings from various power analyses of scientific literacy questions is shown in Table 8 and Figure 2.

Table 8. Differentiating Power in Essay Questions

Differentiating Power	Question Item Numbers	Total	Percentage (%)
0.00 – 0.20 (Bad)	5	1	10
0.21 – 0.40 (Enough)	1, 2, 3, 4, 6, 7, 8, 10, 11	9	66
0.41 – 0.71 (Good)	9, 12	2	14
0.71 – 1.00 (Excellent)	13	1	10

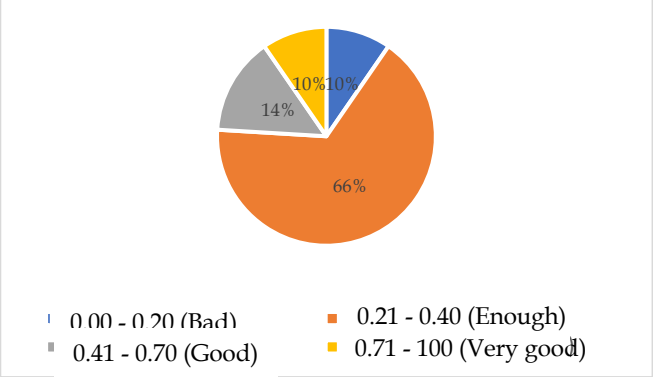

Figure 2. Percentage of Distinguishing Power on Science Literacy Essay Questions

The results of the analysis of the discriminating power of scientific literacy questions show that the questions have good discriminating power with criteria (fair, good, and very good) for a total of 73% of the objective questions and 90% of the essay questions. Based on research conducted by (Ayubi et al., 2023), if a question has low discrimination, it means that the question is not effective in measuring students' abilities. On the other hand, if a question has a high differential power, then the question is more efficient in measuring students' abilities. This scientific literacy question is worthy of being used as an instrument that can be used to assess scientific literacy abilities in learning.

Questions with poor discriminating power scores must be corrected completely by identifying the root cause of question failure. Meanwhile, questions with good, very good, and sufficient discriminatory power scores must be saved by adding them to the question bank. One way to prevent students who are very capable of answering questions from becoming confused is to revise questions whose wording is unclear. Questions must be able to differentiate between students who understand the subject and those who do not (Muluki et al., 2020).

Difficulty Level

The difficulty level of a question is a number that describes how challenging the question item is. If the test questions are quite difficult and not too simple or too difficult, then it is considered good (Arbiatin & Mulabbiyah, 2020). Students are not motivated to increase their efforts in answering questions that are too easy and simple (Iskandar & Rizal, 2017). On the other hand, questions that are too challenging will make students lose enthusiasm and make them less motivated to try again (Solichin, 2017). Analysis of the level of difficulty of the questions is carried out to make it easy to identify superior question standards (Jumrodah et al., 2023). According to (Abdul, 2015), he emphasized that exam instruments need to classify questions into easy, medium, and difficult levels.

The results of the analysis of the level of difficulty of scientific literacy questions are shown in the table and figure below.

Table 9. Distribution of Difficulty Levels of Objective Questions

Difficulty Index	Question Item Numbers	Total	Percentage (%)
0.00 – 0.30 (Hard)	1, 3, 8	3	27
0.31 – 0.70 (Medium)	2, 4, 5, 6, 7, 10, 11	7	64
0.71 – 1.00 (Easy)	9	1	9

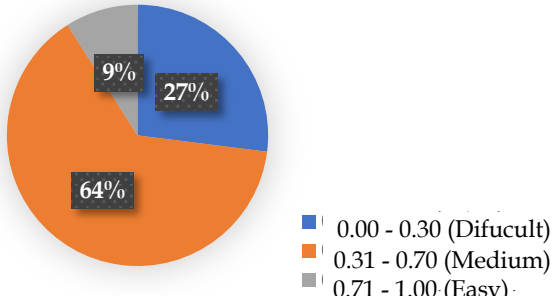

Figure 3. Percentage of Objective Question Difficulty Level

Table 10. Distribution of Difficulty Levels of Essay Questions

Difficulty Index	Question Item Numbers	Total	Percentage (%)
0.00 – 0.30 (Hard)	-	-	-
0.31 – 0.70 (Medium)	1, 2, 3, 4, 5, 6, 8. 9, 10, 11, 12, 13	12	92
0.71 – 1.00 (Easy)	7	1	8

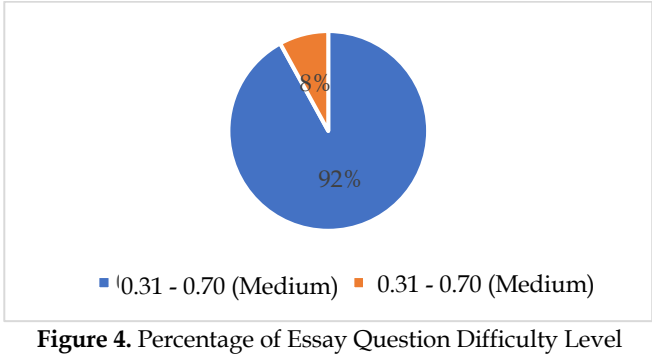

Figure 4. Percentage of Essay Question Difficulty Level

Based on the results of the analysis, it can be concluded that the scientific literacy questions have a moderate level of difficulty. According to calculation analysis using Anatès version 4 for Windows, several items, or 64% were obtained in the objective questions and 92% in the essay questions, some of which had a medium difficulty level.

Distractor Qualities

The distractor function of questions can be understood as a pattern that explains how the sample chooses its response to the potential answers that have been provided in the answer choices for each test item. parameters for. The quality of distractors also functions to determine whether distractors or incorrect answers function effectively or not (Rokhim et al., 2023). The more individuals choose a distractor answer, the more effective it becomes; conversely, if the distractor answer is not selected, then the distractor does not work (Arbiatin dan Mulabbiyah. 2020). The following are the results of the analysis of the quality of distractors on science questions.

Table 11. Analysis of the Quality of Distracters on Science Literacy Questions

Distraction Index	Question Item Numbers	Total	Percentage
0 (Excellent)	1-11	11	100%
1 (Good)	-	-	0%
2 (Not Good)	-	-	0
3 (Bad)	-	-	0
4 (So Bad)	-	-	0

Based on calculations carried out with Anates version 4.0, the scientific literacy objective questions had a very high distractor value of 100%. A good distractor is chosen by at least 5% of students; otherwise, the distractor is considered a bad distractor (Dibattista & Kurzawa, 2011). This parameter also determines students' low conceptual understanding due to choosing the wrong answer or reason (Habiddin & Page, 2019).

Based on the results of Anates' analysis, it can be concluded that this scientific literacy question instrument meets the requirements for questions that can be used to measure the level of scientific literacy in learning. Research (Ningsih & Kamaludin, 2023) states that valid question instruments can be used to measure students' abilities in learning. Apart from that, according to (Mulyanti et al., 2022) question instruments that have a medium level of question difficulty, high validity, and high reliability can be used in further research.

Conclusion

Based on the research results, it can be concluded that scientific literacy questions have a very good level of validity, medium and high item reliability, namely 0.55 on objective questions and 0.72 on essay questions, a good level of difficulty with a percentage of 64% on objective questions and 92% on the essays have a medium level of difficulty, the quality of the questions is 100% of all questions. In addition, the discriminating power of the items was rated as good (fair, good, and very good) in a total of 73% of the objective questions and 90% of the essay questions. This scientific literacy question can be used to see the level of scientific literacy ability because it fulfills the requirements as a learning evaluation instrument.

Acknowledgments

The author would like to thank all parties who have supported and helped carry out this research.

Author Contributions

The authors of this article are five people, namely F.H., S.S., F.A., L.A., and M.S. All members of the research group were involved in each stage.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abdul, K. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Al-Ta'dib*, *8*(2), 70–81. http://dx.doi.org/10.31332/atdb.v8i2.411

Adawiyah, R., & Wisudawati, A. W. (2017). Pengembangan Instrumen Tes Berbasis Literasi Sains : Menilai Pemahaman Fenomena Ilmiah Mengenai Energi. *Indonesian Journal of Curriculum and Technology Studies*, *5*(2), 112–121. http://dx.doi.org/10.15294/ijcets.v3i1.8675

Alifah, F., & Istiyono, E. (2023). Development Of Test Instruments to Improve Students' Multirepresentational Ability. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 10212–10216. https://doi.org/10.29303/jppipa.v9i11.5286

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. http://dx.doi.org/10.33578/jpfkip.v3i2.2501

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

Ardi, A., Hervi, F., & Mudjiran, M. (2023). Validity and Reliability Questionnaire of Students' Critical Thinking Skills in General Biology Course. *Jurnal Penelitian Pendidikan IPA*, *9*(3), 1436–1444. https://doi.org/10.29303/jppipa.v9i3.2761

Ayubi, M., Arthamena, V. D., & Aznam, N. (2023). Development of HOTS (High Order Thinking Skill) Question Instrument on Thermochemistry Material. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 9945–9954. https://doi.org/10.29303/jppipa.v9i11.5162

Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.

Dibattista, D., & Kurzawa, L. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests Examination of the Quality of Multiple-choice Items on Classroom Tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, *2*(2), 1–23. http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4

Dirman, H. M., & Mufit, F. (2022). Disain dan Validitas E-Instrument Five Tier-Multiple Choice Test Menggunakan iSpring Quiz Maker Sebagai Asesmen Literasi Sains Siswa Pada Pembelajaran Abad-21. *Jurnal Penelitian Pendidikan IPA*, *8*(6), 3170–7179. https://doi.org/10.29303/jppipa.v8i6.2300

Eny, H. A., & Wiyarsi, A. (2019). Students' Chemical Literacy on Context-Based Learning: A Case of Equilibrium Topic. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1397/1/012035

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. Retrieved from https://ejournal.uinib.ac.id/jurnal/index.php/naturalscience/article/view/2329

Habiddin, & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

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

Hervi, F., Afdarina, O., Syamsurizal, & Lufri. (2023). Analysis of Science Subject Questions in Class IX Junior High School using Anates 4.0 Computer Program. *Gema Wiralodra*, *14*(2), 638–648. Retrieved from https://gemawiralodra.unwir.ac.id/index.php/gemawiralodra/issue/view/22

Hidayati, H., Ningsi, A. W., & Iskandar, A. M. (2023). Design and Validity of Student Worksheet Integrated Scientific Literacy for The Use of Physics Practicum KIT. *Jurnal Penelitian Pendidikan IPA*, *9*(1), 384–389. https://doi.org/10.29303/jppipa.v9i1.2887

Huryah, F., Sumarmin, R., & Effendi, J. (2017). Analisis Capaian Literasi Sains Biologi Siswa SMA Kelas X di Kota Padang. *Jurnal Eksakta Pendidikan (Jep)*, *1*(2), 72–79. https://doi.org/10.24036/jep.v1i2.70

Iskandar, A., & Rizal, M. (2017). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Pplikasi 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), 172–177. https://doi.org/10.30598/biopendixvol9issue2page172-177

Jusniar, J., Effendy, E., Budiasih, E., & Sutrisno, S. (2020). Developing a three-tier diagnostic instrument on chemical equilibrium (TT-DICE). *Educacion Quimica*, *31*(3), 84–102. https://doi.org/10.22201/fq.18708404e.2020.3.72133

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*(23), 2276–2284. https://doi.org/10.2146/ajhp070364

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

Mulyanti, S., Suwahono, S., Setiowati, H., & Ningrum, L. S. (2022). Validity Analysis Using the Rasch Model in the Development of Alkane Concept Test Instruments. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1142–1147. https://doi.org/10.29303/jppipa.v8i3.1383

Ningsih, N. R., & Kamaludin, A. (2023). Development of Higher Order Thinking Skills-Based Assessment

Instrument on Acid-Base Materials in High School. *Jurnal Penelitian Pendidikan IPA*, *9*(1), 13–19. https://doi.org/10.29303/jppipa.v9i1.1457

Putri, M. E., Syamsurizal, S., & Lufri, L. (2022). Analyzed Question Quality of School Examinations Class IX Science Subjects. *Bioeducation Journal*, *6*(1), 40–47. Retrieved from http://bioeducation.ppj.unp.ac.id/index.php/bioedu/article/view/318?articlesBySameAuthorPage=2#articlesBySameAuthor

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

Rahmi, Y. L., Miatidini, N. A., Alberida, H., Darusyamsyu, R., Ichsan, I. Z., Sigit, D. V., Titin, Koc, I., & Sison, M. H. (2021). HOTS Assessment of Biology Cell: Validity, Practicality and Reliability. *Jurnal Penelitian Pendidikan IPA*, *7*(3), 481–487. https://doi.org/10.29303/jppipa.v7i3.742

Rokhim, D. A., Widarti, H. R., & Sutrisno, S. (2023). Five-Tier Diagnostic Test Instrument Validation on Reaction Rate Materials: To Identify the Causes of Misconception and Student Representation. *Jurnal Penelitian Pendidikan IPA*, *9*(3), 1380–1385. https://doi.org/10.29303/jppipa.v9i3.2952

Sari, D. N. A., Rusilowati, A., & Nuswowati, M. (2017). Pengaruh Pembelajaran Berbasis Proyek terhadap Kemampuan Literasi Sains Siswa. *PSEJ (Pancasakti Science Education Journal)*, *2*(2), 114. https://doi.org/10.24905/psej.v2i2.741

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. https://doi.org/10.26594/dirasat.v2i2.879

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

Sundayana, R. (2016). *Statistika Penelitian Pendidikan*. 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

Utari, J. I., U, F., & Ermawati, U. (2018). Pengembangan Instrumen Tes Diagnostik Miskonsepsi Berformat Forur-Tier untuk Materi Suhu, Kalor dan Perpindahannya. *Inovasi Pendidikan Fisika*, *07*(03), 434–439. https://doi.org/10.26740/ipf.v7n3.p%25p

Widiatmo, T., Jufri, A. W., & Jamaluddin, J. (2019). Analysis of Validation of Instruments to Measure Student's Critical Thinking Ability and Science Literation. *Jurnal Penelitian Pendidikan IPA*, *5*(2), 212–218. https://doi.org/10.29303/jppipa.v5i2.272

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. Retrieved from 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