Journal of Educational Research and Evaluation
Volume 8, Issue 4, 2024, pp. 675-681
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://doi.org/10.23887/jere.v8i4.83905

Quality of Argumentation Questions on Excretory System Materials Using the ANATES Application

Ruri Lorika[1*], Fitri Arsih[2], Heffi Alberida[3], Irdawati[4]

[1,2,3,4]Pendidikan Biologi, Universitas Negeri Padang, Padang, Indonesia

ARTICLE INFO

Article history:
Received August 06, 2024
Accepted November 10, 2024
Available online November 25, 2024

Kata Kunci :
Kualitas butir soal, soal argumentasi; evaluasi, sistem ekskresi, ANATES

Keywords:
Quality of question items, about argumentation, evaluation, excretory system, ANATES

ABSTRAK

Salah satu materi yang relevan dalam hal ini adalah sistem ekskresi dalam pembelajaran biologi, yang memerlukan pemahaman mendalam dan keterampilan pemecahan masalah dari siswa. Namun, pengembangan soal argumentasi yang valid, reliabel, dan berkualitas untuk materi ini masih menghadapi tantangan. Penelitian ini bertujuan untuk mengevaluasi kualitas soal-soal argumentasi pada materi sistem ekskresi dengan menggunakan aplikasi ANATES. Sebanyak 35 siswa kelas XI menjadi partisipan penelitian. Instrumen penelitian berupa 10 soal argumentasi yang dianalisis menggunakan ANATES untuk mengukur validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal. Hasil analisis memberikan penjelasan rinci tentang tingkat argumentasi soal-soal pada materi sistem ekskresi. Sebagai kesimpulan, penelitian ini menggarisbawahi peran penting kualitas soal dalam menilai pemahaman siswa secara akurat. Temuan ini menganjurkan penggunaan ANATES secara berkelanjutan untuk meningkatkan konstruksi butir soal. Implikasi dari penelitian ini meluas kepada para pendidik dan pengembang kurikulum dalam menyusun instrumen penilaian yang lebih efektif. Temuan penelitian ini diharapkan dapat membantu dalam mengembangkan instrumen evaluasi yang lebih tepat untuk mengukur pemahaman siswa terhadap materi sistem ekskresi. Selain itu, hasil analisis ini juga dapat digunakan untuk memperbaiki dan menyempurnakan jenis pertanyaan guna meningkatkan validitas dan reliabilitas alat evaluasi argumentasi. Untuk meningkatkan kualitas pengajaran sains di sekolah menengah, para pendidik, perancang kurikulum, dan peneliti pendidikan dapat memanfaatkan implikasi dari penelitian ini.

ABSTRACT

One of the relevant materials in this regard is the excretory system in biology learning, which requires students' deep understanding and problem-solving skills. However, developing valid, reliable, and quality argumentation questions for this material still faces challenges. This study evaluates the quality of argumentation questions on excretory system material using the ANATES application. A total of 35 grade XI students became research participants. The research instrument was in the form of 10 argumentation questions, which were analyzed using ANATES to measure the validity, reliability, difficulty level, and differentiating power of the questions. The analysis results provide a detailed explanation of the level of argumentation of questions on the material of the excretory system. In conclusion, this study underscores the critical role of item quality in accurately assessing student understanding. The findings advocate the continued use of ANATES to improve item construction. The implications of this study extend to educators and curriculum developers developing more effective assessment instruments. The findings of this study are expected to assist in creating more appropriate evaluation instruments to measure students' understanding of the excretory system material. In addition, the results of this analysis can also be used to improve and refine the types of questions to increase the validity and reliability of argumentation evaluation tools. To enhance the quality of science teaching in secondary schools, educators, curriculum designers, and educational researchers can utilize the implications of this study.

*Corresponding author.
E-mail addresses: rurilorika5@gmail.com (*Ruri Lorika*)

1. INTRODUCTION

Education is the main pillar in the development of a nation. The quality of education can be measured through various aspects, one of which is the quality of assessment (Gladushyna & Strietholt, 2023; Ibarra-Sáiz et al., 2021). A good assessment is able to provide accurate information about student learning achievement. One important component of assessment is test questions, especially argumentation questions that require higher-order thinking skills. The effectiveness of evaluation is paramount in the educational system, particularly in ensuring that students achieve their learning objectives (Elfira et al., 2023; Magdalena et al., 2021). A comprehensive evaluation not only measures the extent to which students have met their goals but also provides immediate feedback on their progress and adaptation to the learning process (Putri et al., 2022; Virginia et al., 2021). Despite its importance, the creation of high-quality assessment tools, specifically argumentation questions, remains a challenge in evaluating students' understanding of complex subjects like the excretory system (Huryah et al., 2017) (Mumpuni et al., 2022). Excretory system is one of the important materials in learning biology. A good understanding of the excretory system is necessary to maintain a healthy body. This can support the importance of developing quality argumentation questions related to the excretory system to measure students' critical thinking and problem solving skills (Omolafe Babalola et al., 2023; Rungkat et al., 2023). In addition, existing research underlines that in assessing a problem, a reliable and valid assessment tool is needed to accurately measure students' cognitive abilities. The quality of test questions is very important, because it directly affects the validity and reliability of the assessment. Therefore, conducting a thorough analysis of test questions is essential to improve their quality and ensure that they meet educational standards (Fietri, Lufri, et al., 2021; Ida & Musyarofah, 2021).

In the context of developing quality questions, the use of information technology is increasingly crucial. The use of question analysis software such as ANATES (National Test Analysis) offers an effective and efficient solution (Al Kamil et al., 2023; Fitriyah & Heri, 2023; Hanifah et al., 2020). ANATES is specifically designed to assist educators in evaluating the quality of questions objectively, beyond manual assessments that tend to be subjective. With its various advanced features, ANATES is able to provide in-depth analysis of the characteristics of each item, such as difficulty level, differentiating power, and reliability. This allows educators to identify poor items, so that improvements or replacements can be made. In addition, ANATES can also provide recommendations regarding the development of better questions in the future, so that the overall quality of assessment can be improved. Previous research has demonstrated a significant correlation between the quality of argumentation questions and student learning outcomes. The introduction of technology in the analysis process, such as the use of the ANATES application, has proven to enhance the quality of these assessment tools, thereby improving their effectiveness (Adawiyah & Wisudawati, 2017; Hervi, Afdarina, et al., 2023). This study aims to address the gap in the existing evaluation methods by assessing the quality of argumentation questions related to the excretory system using the ANATES application. The focus is on ensuring that these questions are of the highest standard in terms of validity, reliability, difficulty level, and discriminative power. By doing so, this research contributes to the ongoing efforts to improve science education, providing educators with tools that more accurately measure students' understanding and fostering better learning outcomes.

2. METHODS

This study employs a descriptive quantitative research design, which is well-suited for analyzing data in the form of numbers while adhering to the principles of objectivity, concreteness, rationality, measurability, and systematic analysis. The approach involves the collection and statistical analysis of numerical data to describe the characteristics of the research subject accurately. The research was conducted at SMAN 2 Payakumbuh, with the research subjects being students of class XI F12. The sampling technique used was simple random sampling, ensuring that every member of the population had an equal chance of being selected. A total of 35 students were chosen as the research sample, representing the entire class XI F12. The primary data collection tool consisted of ten argumentation questions related to the excretory system. These questions were designed in essay format to evaluate students' reasoning abilities. The questions were carefully constructed to assess various cognitive levels associated with the excretory system's content. The quality of these questions specifically their validity, reliability, difficulty level, and discriminative power was of primary interest in this research. The research procedure began with the administration of the argumentation questions to the selected sample of students. The students' responses were then collected and analyzed using the ANATES software version 4.0 for Windows. This software was chosen due to its effectiveness in providing detailed statistical analyses of test items, including validation,

reliability testing, difficulty level assessment, and differentiation (Putri et al., 2022; Yusanto, 2020; Zellatifanny & Mudjiyanto, 2018).

The process of using ANATES 4.0 software involves several key steps. First, the ANATES version 4.0 application is launched on Windows to begin the analysis. After that, the user creates a new file and enters the analysis criteria in the dialog box, including the number of subjects, questions, and answer choices. Next, student responses were entered individually according to the question number to ensure data entry accuracy. Once all the data were entered, the "process everything automatically" option was selected to begin data analysis. ANATES 4.0 software performs various analyses, including empirical validation, to assess the instrument's validity in evaluating learning outcomes. In addition, the reliability of the questions was tested, and specific reliability levels grouped results. ANATES also calculates the differentiating power of each question to determine the extent to which each item can differentiate between students with different understandings. Finally, the software evaluates the difficulty index to categorize questions as high, medium, or low.

3. RESULT AND DISCUSSION

Results

To ensure the quality of a test, an in-depth analysis of each item is required, as well as an evaluation of the overall test quality. Before use, each item in this study will be carefully examined to ensure optimal quality. Item analysis plays an important role in improving the standard of answers given by students. After administering the test to each sample, research is conducted, and the findings are used as a roadmap to improve the quality of future tests. Therefore, test analysis is a very important step (Fietri, Zulyusri, dan Violita. 2021). This research explores the level of difficulty, validity, reliability, and differentiating power of each question. The validity of an instrument can be determined by looking at its level of validity. Validity refers to the extent to which an instrument is able to provide an accurate description of the characteristics to be studied. An instrument is said to be valid if the resulting data truly represents the concept or variable that is the focus of the research. A measurement or observation is included in the notion of validity, which also refers to the concept of instrument reliability in data collection. The tool must be able to measure what it is designed to detect. Therefore, validity emphasizes more on the method of measurement or observation. Using ANATES, r-count search is used to evaluate the validity of each item of science literacy questions. Then, the calculation results were compared with the r-table value at a significance level of 5%. The study involved 35 students as research subjects, with a significance level of 5% and n = 35. In the event that the calculated r value is less than the table r value, then the question item is considered invalid; On the other hand, it is considered valid if the calculated r value is greater than the table r value. The results of the validity analysis for the essay questions are shown in Table 1 below.

Table 1. Analysis of the Validity of Argumentation Questions

Question number	Significance Value	Conclusion
1	0.095	-
2	0.793	Very significant
3	0.652	Significant
4	0.800	Very significant
5	0.775	Very significant
6	0.832	Very significant
7	0.825	Very significant
8	0.871	Very significant
9	0.700	Significant
10	0.853	Significant

Only 10% of the argumentative essay questions contain erroneous criteria, according to the study's findings, while up to 90% of the questions or nine questions had valid criteria. Question items deemed invalid are rectified and restored to functionality. The analysis's conclusions support the assertion that the argumentation questions under consideration are sound and legitimate. The reliability test aims to assess whether an instrument is reliable and consistent in its use for further research. The ability of a measuring instrument to make accurate measurements is closely related to reliability. Reliability refers to the accuracy and precision of the instrument in taking measurements. An accurate measuring instrument will produce the right size. Based on calculations carried out with ANATES version 4.0, the reliability test of the questions showed a very high category, which was 0.89. The results of this calculation indicate a very high level of

reliability for argumentation problems, with a reliability range of 0.80 ≤ r < 1.00 (very high). The reliability test resulted in a very high category of 0.89 based on calculations performed using ANATES version 4.0. This computation's results show that argumentation issues have an extremely high degree of reliability, with a reliability range of 0.80 ≤ r < 1.00 (very high). Test reliability, according to (Loka Son, 2019), quantifies how well the test is able to consistently assess the intended construction. These findings show that up to 60% of all questions have strong discriminating power, and 20% of all questions have excellent discriminating power. But 10% of the questions had adequate discriminating power, and 10% of the questions had insufficient discriminating power. As a result, there are differences in the distribution of argumentation questions' discriminating power; however, the majority of the questions have a good or very good level of discriminating power. The results of the Power of differentiation in argumentation questions are presented in Table 2 below.

Table 2. The Power of Differentiation in Argumentation Questions

Differentiation	Question Number	Total	Persentase
0.00 – 0.20 (Ugly)	1	1	10%
0.21 – 0.40 (Enough)	9	1	10%
0.41 – 0.71 (Good)	2, 3, 4, 5, 7, 10	6	60%
0.71 – 1.00 (Very good)	6, 8	2	20%

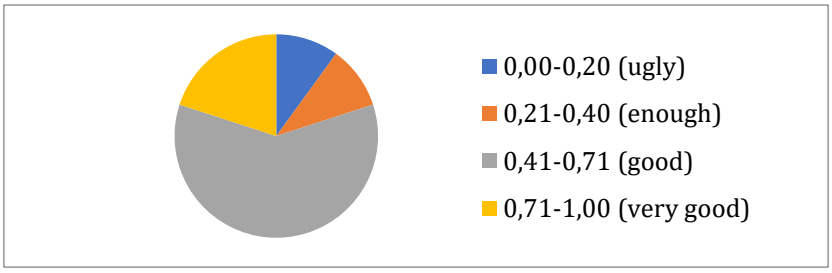

Figure 1. Percentage of Difference in Argumentation Questions

Based on the evaluation of the argumentation questions, it is known that 100% of the essay questions have good discriminating power based on the assessment categories of "quite good", "good", and "very good". This argumentation question meets the required standards as a tool to develop students' reasoning skills during the learning process. Analyzing the reasons behind the questions' initial failure to achieve the required level is crucial to improving the quality of questions that lack discriminating power. In order to serve as future reference material, questions that have demonstrated a sufficient, good, or great degree of discriminating power should be saved and added to the question bank. Questions that use ambiguous or unclear language can be revised to make it easier for students who have high ability to answer them. The difficulty level of the question aims to help in evaluating the complexity of a question. Questions that are considered good are those that have a moderate level of difficulty, not too easy but also not too difficult. The assessment of the difficulty level of the question should be based on the student's ability to answer the question, not from the perspective of the teacher as the compiler of the question. Exam questions are considered good if the number of demands is right (not too easy or too difficult). When presented with too easy of a problem, students often become demotivated since they don't feel challenged to advance their skills. Conversely, too-hard questions might dampen pupils' enthusiasm and discourage them from attempting to answer them. The analysis of the difficulty level of the questions is carried out to facilitate the identification of questions that have good standards. The results of the level of difficulty in argumentation questions are presented in Table 3.

Table 3. Difficulty Level in Argumentation Questions

Difficulty Index	Question Number	Total	Persentase
0.00 – 0.30 (high)	9	1	10%
0.31 – 0.70 (medium)	1, 2, 3, 4, 5, 6, 7, 8, 10	9	90%
0.71 – 1.00 (low)	-	-	-

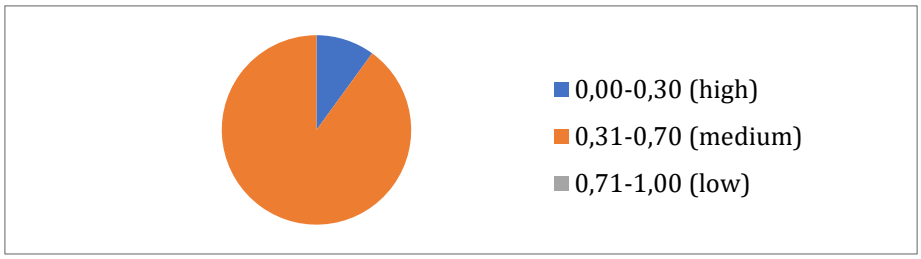

Figure 2. Percentage of Difficulty of Argumentation Questions

About 90% of the questions have a moderate difficulty, according to estimates conducted with Anates version 4 for Windows on the results. Therefore, argumentation questions can be said to have a moderate level of difficulty.

Discussion

Based on the results of the study, the quality of argumentation questions about the excretory system generally meets the criteria of good validity and reliability. The percentage of valid questions reached 90%, while the Cronbach's alpha reliability coefficient of 0.89 showed a high level of internal consistency. This indicates that the items are able to measure the same construct consistently and reliably to assess students' understanding of the concept of the excretory system (Jowinis & Siew, 2024; Yahfizham et al., 2021). These findings have positive implications for the development of better assessment instruments in the future, especially in the context of learning complex biological concepts. The results of the analysis show that the level of difficulty of the questions is generally in the medium category, with a percentage of 90%. The results of this study are in line with several previous studies which show that a moderate level of question difficulty tends to produce more reliable and valid data (Hervi, Syamsurizal, et al., 2023; Krieglstein et al., 2022; Riyanti et al., 2023). This level of difficulty is considered optimal in the context of educational assessment, as it is able to balance between high cognitive demands and student engagement. In addition, the differentiating power of the questions showed a fairly good variation (Kamalov et al., 2023; Rasul et al., 2023). The majority of questions have good to very good differentiating power, indicating the effectiveness of questions in distinguishing students with different abilities. However, there were some questions that had low differentiating power, especially those that contained ambiguity or lacked clarity in their formulation. This suggests the need to revise these questions to improve their validity.

The implementation of ANATES software in this study has demonstrated its effectiveness as a tool for analyzing the quality of exam questions in depth. Through comprehensive statistical analysis, ANATES enables precise identification of aspects of questions that need improvement. By using ANATES to calculate the difficulty index of each item, researchers can find out how difficult or easy a question is (Hervi, Afdarina, et al., 2023; Ma'rifah et al., 2021; Manalu et al., 2019). In addition, ANATES can also be used to determine the ability of questions to distinguish between students who have high and low abilities. With its complete features, ANATES can be used to test the validity and reliability of questions, and provide valuable information to improve the quality of questions in the future (Dayanto et al., 2023; Kim How et al., 2022; Marsevani, 2022). The use of this technology is in line with recent research trends that emphasize the importance of utilizing computer-based software in improving the quality of educational assessment instruments. In conclusion, this study underscores the critical role of question quality in accurately assessing student comprehension. The findings advocate for the sustained utilization of ANATES to enhance test item construction. The implications of this research extend to educators and curriculum developers in crafting more effective assessment instruments. Future research should investigate the generalizability of these findings across diverse subject areas and educational contexts.

4. CONCLUSION

The findings of this study underscore the importance of using well-validated and reliable assessment tools in educational settings. The high validity and reliability of the argumentation questions related to the excretory system reflect the robustness of the ANATES application in improving the quality of educational assessment. These tools not only ensure that student understanding is accurately measured but also contribute to the broader goal of improving educational outcomes by providing educators with accurate data to inform instruction. The moderate difficulty level of the questions, coupled with their strong discriminatory power, highlights their effectiveness in differentiating between students with varying levels of understanding. This is critical in ensuring that assessments are fair and challenging, which promotes deeper understanding of the subject matter among students. The results of this study underscore the need

for ongoing refinement of assessment tools, ensuring that they remain relevant and effective in measuring students' cognitive abilities.

5. REFERENCES

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. https://www.learntechlib.org/p/209353/.

Al Kamil, M. A. Z., Wafa, I. A., & Basori, M. A. (2023). *Indonesian Cognitive Test for Educational Psychology*. Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-032-9_9.

Dayanto, R. A. N., Sodiq, S., & Amri, M. (2023). *Analysis of Multiple Choice Questions for Japanese Language Mid-Term Exam at Public Senior High School*. Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-008-4_194.

Elfira, I., Sari, P. N., Syamsurizal, S., & Lufri, L. (2023). Analisis Butir Soal Uts Projek Ipas Kelas X Smk N1 Banda Aceh Memanfaatkan Anates 4.0 for Windows Pc Program. *Jurnal IPA Terpadu*, *7*(2), 156. https://doi.org/10.35580/ipaterpadu.v7i2.48346.

Fietri, W. A., Lufri, Syamzurizal, & Zulyusri. (2021). Analisis Butir Soal Biologi Kelas VIII Madrasah Tsanawiyah Negeri 6 Kerinci. *Jurnal Pendidikan Biologi Undiksha*, *8*(2), 50–60. https://ejournal.uinib.ac.id/jurnal/index.php/naturalscience/article/view/2329.

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

Fitriyah, I. M., & Heri, R. (2023). *Analysis of the distractor of the multiple-choice test using classical test theory (CTT) and item response theory (IRT)*. *Vi*, 196–203. https://doi.org/10.31643/2023.23.

Gladushyna, O., & Strietholt, R. (2023). Measuring education: Do we need a plethora of assessment studies or just a single score? *International Journal of Educational Research Open*, *5*(September), 100281. https://doi.org/10.1016/j.ijedro.2023.100281.

Hanifah, W., Rustaman, N., & Rusyati, L. (2020). The Validity and Reliability of Science Virtual Test on Levels of Organization and Cell Transport Topic (SVT-LOCT) to Assess Students' Science Process Skills. *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019*. https://doi.org/10.4108/eai.12-10-2019.2296435.

Hervi, F., Afdarina, O., Syamsurizal, S., & Lufri, L. (2023). Analysis of science subject questions in class IX junior high school using Anates 4.0 computer program. *Gema Wiralodra*, *14*(2), 638–648. https://doi.org/10.31943/gw.v14i2.415.

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.

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.

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2021). The quality of assessment tasks as a determinant of learning. *Assessment and Evaluation in Higher Education*, *46*(6), 943–955. https://doi.org/10.1080/02602938.2020.1828268.

Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'Arrib: Journal of Arabic Education*, *1*(1), 34–44. https://doi.org/10.32923/al-muarrib.v1i1.2100.

Jowinis, A., & Siew, N. M. (2024). Validity and Reliability of Multiple-Choice Biology Test: a Rasch Analysis Perspective. *International Journal of Modern Education*, *6*(21), 131–139. https://doi.org/10.35631/ijmoe.621010.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability (Switzerland)*, *15*(16), 1–27. https://doi.org/10.3390/su151612451.

Kim How, R. P. T., Zulnaidi, H., & Rahim, S. S. B. A. (2022). Development of Higher-Order Thinking Skills test instrument on Quadratic Equations (HOTS-QE) for Secondary School Students. *Pegem Egitim ve Ogretim Dergisi*, *13*(1), 379–394. https://doi.org/10.47750/pegegog.13.01.41.

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. In *Educational Psychology Review* (Vol. 34, Issue 4). Springer US. https://doi.org/10.1007/s10648-022-09683-4.

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.

Ma'rifah, U., Algiovan, N., & Sutarsyah, C. (2021). An Item Analysis of English Test During Online Learning. *International Journal of Multicultural and Multireligious Understanding*, *8*(12), 647–654. https://ijmmu.com/index.php/ijmmu/article/view/3396.

Magdalena, I., Ningsih, Y. A., Berliana, N., & Gunawan, A. N. (2021). Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran IPS Kelas IV SDS Sari Putra. *Jurnal Pendidikan Dan Sains*, *3*(1), 15–22.

Manalu, D., Sipayung, K. T., & Lestari, F. D. (2019). an Analysis of Students Reading Final Examination By Using Item Analysis Program on Eleventh Grade of Sma Negeri 8 Medan. *JETAL: Journal of English Teaching & Applied Linguistic*, *1*(1), 13–19. https://doi.org/10.36655/jetal.v1i1.98.

Marsevani, M. (2022). Item Analysis of Multiple-Choice Questions: an Assessment of Young Learners. *English Review: Journal of English Education*, *10*(2), 401–408. https://doi.org/10.25134/erjee.v10i2.6241.

Mumpuni, K. E., Priyayi, D. F., & Widoretno, S. (2022). How do Students Perform a Peer Assessment? *International Journal of Instruction*, *15*(3), 751–766. https://doi.org/10.29333/iji.2022.15341a.

Omolafe Babalola, E., Ayoola, D. A., & Omolafe, E. V. (2023). Analysis of Experts' Opinion on the Human Excretory System Model for Teaching Biology in Nigeria. *ASEAN Journal of Science and Engineering Education*, *3*(1), 19–26. https://doi.org/10.17509/ajsee.v3i1.45457.

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.

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. de O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56. https://doi.org/10.37074/jalt.2023.6.1.29.

Riyanti, V., Hamdiyati, Y., & Purwianingsih, W. (2023). The Effect of Argument-Driven Inquiry (ADI) on Argumentation Skills and Students' Concept Mastering of Human Excretory System Materials. *Journal of Science Learning*, *6*(2), 125–135. https://doi.org/10.17509/jsl.v6i2.54634.

Rungkat, J. A., Jeujanan, A., Wola, B. R., & Warouw, Z. W. M. (2023). Development of STEM-based Science E-Module on the Human Excretory System Topic. *Jurnal Penelitian Pendidikan IPA*, *9*(8), 6548–6556. https://doi.org/10.29303/jppipa.v9i8.4437.

Virginia, S., Angraini, W., Pratesya, W., & Walid, A. (2021). Analisis Butir Soal Ulangan Harian IPA Terpadu Kelas VII SMP 05 Kota Bengkulu Tahun Pelajaran 2020/2021. *Jurnal PTK Dan Pendidikan*, *6*(2), 61–65. https://doi.org/10.18592/ptk.v6i2.4173.

Yahfizham, Yusti, I., & Hamzah, M. L. (2021). The Test of Construct Validity for One-Factor Model. *Journal of Educational Research and Evaluation*, *5*(3), 429–435.

Yusanto, Y. (2020). Ragam Pendekatan Penelitian Kualitatif. *Journal of Scientific Communication (Jsc)*, *1*(1), 1–13. https://doi.org/10.31506/jsc.v1i1.7764.

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.