



EVALUATION OF THE VALIDITY AND RELIABILITY OF BIOLOGY LEARNING INSTRUMENTS AT SMA NEGERI 6 BANTAENG

Evaluasi Validitas dan Reliabilitas Instrumen Pembelajaran Biologi di SMA Negeri 6 Bantaeng

Asdar Efendi¹, Ihsan², Try Suyanti³, Siska Bin Basso⁴, Hastuti⁵

^{1,2,3,4,5}Master's Program in Biology Education, Patompo University

Correspondence Author: Hastuti, hastuti19sumarno@unpatompo.ac.id, Patompo University

ARTICLE INFO

Article History: 26-11-2025

Received : 27-11-2025

Revised form : 27-11-2027

Accepted : 28-11-2025

Published online : 28-11-2025

Keywords:

Instrument Evaluation;
Validity, Reliability;
Biology Learning;
Multiple Choice Questions

Kata Kunci:

Evaluasi Instrumen;
Validitas, Reliabilitas;
Pembelajaran Biologi;
Soal Pilihan Ganda;

How To Cite: Eka Apriyanti. (2025). *Evaluation Of The Validity And Reliability Of Biology Learning Instruments At Sma Negeri 6 Bantaeng*. *Journal Research Education Mamminasata*, Volume 1 (1), Page. 69-74

ABSTRACT

This study aims to evaluate the quality of a multiple-choice Biology learning achievement test instrument through validity and reliability tests. The method used was quantitative descriptive, with 20 grade XII students of SMA Negeri 6 Bantaeng as research subjects. The research instrument consisted of 20 multiple-choice items. Item validity was tested using Pearson Product Moment correlation, while reliability was calculated using the KR-21 formula. The results showed that 11 items (55%) were declared valid with a correlation coefficient ≥ 0.602 , and 9 items (45%) were invalid. The instrument's reliability coefficient of 0.82 was classified as very high. The study concludes that the evaluated Biology learning assessment instrument has excellent reliability; however, revision of the invalid items is required to enhance the quality of measurement. The development of valid and reliable instruments is crucial for supporting learning effectiveness and the accuracy of learning outcome assessment.

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi kualitas instrumen tes hasil belajar Biologi berupa soal pilihan ganda melalui uji validitas dan reliabilitas. Metode yang digunakan adalah deskriptif kuantitatif dengan subjek penelitian 20 siswa kelas XII SMA Negeri 6 Bantaeng. Instrumen penelitian terdiri dari 20 butir soal pilihan ganda. Validitas butir diuji menggunakan korelasi Product Moment Pearson, sedangkan reliabilitas dihitung dengan rumus KR-21. Hasil penelitian menunjukkan bahwa 11 butir soal (55%) dinyatakan valid dengan koefisien korelasi $\geq 0,602$, dan 9 butir soal (45%) tidak valid. Koefisien reliabilitas instrumen sebesar 0,82 tergolong dalam kategori sangat tinggi. Simpulan penelitian ini adalah instrumen evaluasi pembelajaran Biologi yang diuji memiliki reliabilitas yang sangat baik, namun diperlukan revisi pada butir-butir yang tidak valid untuk meningkatkan kualitas pengukuran. Pengembangan instrumen yang valid dan reliabel sangat penting untuk mendukung efektivitas pembelajaran dan akurasi penilaian hasil belajar.

INTRODUCTION

The 21st century demands that young generations master critical, creative, communicative, and collaborative thinking skills, often known as the 4Cs (Partnership for 21st Century Skills, 2015). Science education, including Biology, plays a strategic role in preparing students to face these challenges, due to its characteristics which emphasize not only the mastery of concepts but also the process of discovery and problem-solving (OECD, 2018).

Biology, as a branch of Natural Sciences, studies all aspects of life and living organisms. A deep understanding of Biology provides a foundation for students to comprehend themselves, their environment, and global issues such as climate change, biotechnology, and biodiversity conservation (Campbell et al., 2018).

At the senior high school level, Biology learning aims not only to transfer knowledge but also to develop scientific literacy and shape students' scientific character (Zubaidah, 2016). This scientific literacy includes the ability to use scientific knowledge, identify questions, and draw evidence-based conclusions (PISA, 2018).

To determine the extent to which these Biology learning objectives are achieved, an appropriate and measurable evaluation mechanism is required. Learning evaluation is an integral component inseparable from the overall educational process (Gronlund & Waugh, 2009).

Evaluation instruments, in this case, learning achievement tests, function as tools to measure the attainment of student competencies in the cognitive, affective, and psychomotor domains (Anderson & Krathwohl, 2001). A good test can objectively map student abilities and provide valuable information for improving the learning process (Nitko & Brookhart, 2011).

However, this strategic function of evaluation instruments can only be realized if the instrument meets certain quality standards. The two fundamental psychometric criteria that a measurement instrument must fulfill are validity and reliability (Arikunto, 2019; Sugiyono, 2017).

Validity refers to the extent to which an instrument accurately measures the construct or variable it is intended to measure (Creswell

& Creswell, 2018). A Biology test is considered valid if it can measure the understanding of Biological concepts, not merely reading comprehension or test-taking skills. Validity can be viewed from several aspects, including content, construct, and criterion validity (Azwar, 2015).

Meanwhile, reliability pertains to the consistency of the measurement results obtained from the instrument (Fraenkel et al., 2012). A reliable instrument will produce stable and trustworthy scores, which do not fluctuate significantly if administered at different times or under different conditions to the same subjects (Gay et al., 2012).

In the context of Biology learning evaluation, the assurance of validity and reliability is crucial to ensure that the assessment results truly reflect the students' actual abilities and are not due to instrument error (Mardapi, 2017).

Unfortunately, the reality in the field is often less than ideal. Many Biology teachers still use evaluation instruments, particularly test items, that have not undergone adequate validity and reliability testing (Puspa Chandra et al., 2021). This may be due to limitations in time, resources, or understanding of the importance of item quality analysis.

Consequently, the obtained learning outcome data is potentially biased and inaccurate. Invalid test items can be misleading and fail to distinguish between students who have mastered the material and those who have not (Sudjana, 2016). Furthermore, items that only measure rote memorization (C1) and rarely touch upon higher-order thinking skills (HOTS) are still commonly found, despite changing curriculum demands and era (Zubaidah, 2016; Brookhart, 2010).

Several previous studies have underscored the urgency of instrument testing. Research by Bari, Ibrahim, & Yuliani (2023) successfully developed a practical skills assessment instrument that was highly valid and reliable, confirming that systematic steps in instrument development are absolutely necessary.

Similarly, Habibah & Fuadiyah (2025) successfully produced a science process skills test instrument with very high content validity, demonstrating that involving experts in the validation process is key to success.

Studies by Marthiani (2024) and Handayani (2022) also provided methodological contributions by demonstrating the effectiveness of statistical analyses such as Pearson correlation and the Rasch model in revealing the quality of Biology test items.

The findings of Wilsa et al. (2023) further reinforce the importance of instruments that are not only valid and reliable but also possess good item characteristics for measuring scientific literacy in complex Biological material.

The currently implemented "Kurikulum Merdeka" emphasizes student-centered learning and authentic assessment (Kemendikbudristek, 2022). Within this framework, the role of quality evaluation instruments becomes even more central in providing meaningful feedback for both students and teachers (Andrade & Heritage, 2017).

A good instrument in the "Kurikulum Merdeka" must be able to measure learning outcomes comprehensively, encourage higher-order thinking skills, and be relevant to real-life contexts (Wiggins & McTighe, 2005).

Based on the above exposition, a gap is apparent between the importance of having quality evaluation instruments and actual practices in the field. Therefore, efforts to evaluate and improve the quality of Biology learning instruments through validity and reliability testing are imperative.

This research specifically aims to analyze the item validity and reliability of a multiple-choice test instrument used in Biology subjects at SMA Negeri 6 Bantaeng. The results of this study are expected not only to provide an objective overview of the existing instrument's quality but also to serve as a basis for the improvement and development of better instruments in the future, thereby ultimately contributing to the overall enhancement of Biology learning quality..

METHODS

The research approach used was quantitative descriptive. This study aimed to describe the characteristics of the test instrument based on numerical data from the trial results (Sugiyono, 2017).

The subjects of the research were 20 grade XII science students at SMA Negeri 6 Bantaeng, selected through purposive sampling. This selection was based on the consideration that grade XII students have studied all the tested material and possess an adequate level of ability heterogeneity for analysis (Creswell & Creswell, 2018).

The research instrument was a multiple-choice test consisting of 20 items, compiled based on basic competencies in the "Kurikulum Merdeka" (Kemendikbudristek, 2022) and covering cognitive levels C1 (remembering) to C4 (analyzing) according to the revised Bloom's taxonomy (Anderson & Krathwohl, 2001).

Data Collection Procedure: The instrument was trialed on the research subjects. Student answer sheets were then scored, and the score for each item and the total score were recorded in dichotomous form (1 for correct, 0 for incorrect) for analysis (Arikunto, 2019).

Data Analysis:

1. **Item Validity Test:** Using Pearson Product Moment correlation with the assistance of Microsoft Excel software. An item was declared valid if the correlation coefficient (*rhitung*) ≥ 0.3 and significant ($p < 0.05$) (Fraenkel et al., 2012). The interpretation of the coefficient referred to Arikunto (2019): 0.80-1.00 (very high), 0.60-0.79 (high), 0.40-0.59 (moderate), 0.20-0.39 (low), 0.00-0.19 (very low).
2. **Reliability Test:** Using the *Kuder-Richardson 21 (KR-21)* formula due to the dichotomous multiple-choice question format (Gronlund & Waugh, 2009). The instrument was declared reliable if the coefficient ≥ 0.70 , with interpretation: 0.80-1.00 (very high), 0.60-0.79 (high), 0.40-0.59 (moderate), and < 0.40 (low) (Nitko & Brookhart, 2011).

RESULTS

1. Results of the Item Validity Test

Of the 20 items tested, the results were as follows:

- a. 11 items (55%) were declared valid with a correlation coefficient ≥ 0.602 .
- b. 9 items (45%) were declared invalid with a correlation coefficient < 0.602 . The distribution of the validity levels of the valid items can be seen in Table 1.

Table 1. Distribution of Validity Levels of Valid Items

Validity Level	Coefficient Range	Number of Items	Remarks
Very High	0.80 – 1.00	2	Very Good
High	0.60 – 0.79	9	Good
<i>Total Valid Items</i>		*11*	

2. Results of the Reliability Test

Based on calculations using the KR-21 formula, a reliability coefficient of 0.82 was obtained. This value falls into the very high category according to the interpretation by Nitko & Brookhart (2011).

DISCUSSION

Item Validity

The finding that 55% of the items (11 out of 20) were declared valid indicates that most of the instrument functioned well in measuring the intended Biology competencies. The nine valid items categorized as "High" to "Very High" suggest that these items have strong measuring power and consistently contribute to the total score (Azwar, 2015). This aligns with the research by Habibah & Fuadiyah (2025), which also reported a very high level of content validity (>90%) for a Biology instrument, affirming that careful planning and construction of items based on learning indicators are key to creating good validity.

However, the presence of 9 invalid items (45%) is a critical note in this study. This high proportion of invalid items confirms the findings of Puspa Chandra et al. (2021) that many instruments in the field do not meet psychometric quality standards. The invalidity of these items can be caused by several factors. First, the possibility of non-functioning distractors. Distractors that are too obviously wrong or conversely too close to being correct can make an item too easy or too difficult, thereby reducing the item-total correlation (Sudjana, 2016). Second, ambiguous or

confusing item phrasing can lead to varying interpretations among students, meaning the answers no longer purely reflect mastery of the material (Mardapi, 2017). Third, a misalignment between the item and the actual learning indicators and objectives. The item might measure aspects outside those planned, thus being invalid for measuring the intended construct (Anderson & Krathwohl, 2001). These items require thorough revision or even elimination to improve the overall validity of the instrument.

Instrument Reliability

The reliability coefficient of 0.82 is a very positive indicator. This value not only meets the reliability standard (≥ 0.70) but is also classified as "very high" (Nitko & Brookhart, 2011). This means that this test instrument has excellent internal consistency. The scores produced by this instrument are stable and trustworthy, so if the same test were administered again to a similar group of students at a different time, the results would be relatively consistent (Fraenkel et al., 2012). This finding is consistent with the research of Marthiani (2024), who also reported a reliability coefficient above 0.80 for a Biology concept understanding instrument, as well as the study by Bari et al. (2023), which found a reliability of 0.96 for a performance assessment instrument.

The high reliability in this study is supported by the number of items that have a high correlation with the total score (as seen in the 11 valid items). These well-functioning items "work together" in measuring the same construct, namely Biology learning achievement, thus creating measurement consistency (Gay et al., 2012). High reliability is an important prerequisite for validity. Although high reliability does not guarantee validity, an instrument cannot be valid unless it is reliable first (Creswell & Creswell, 2018). In other words, this instrument can be trusted to measure "something" consistently, and from the validity analysis, we know that this "something" is largely the intended Biology competency.

Implications for Learning and Assessment

The findings of this study have significant practical implications. First, teachers can have greater confidence in using

the 11 valid items as a tool to assess student understanding and evaluate the effectiveness of their teaching (Gronlund & Waugh, 2009). Data from these valid items can be used as a basis for providing targeted feedback to students (Andrade & Heritage, 2017).

Second, these findings highlight the importance of periodic item analysis as part of a teacher's professional reflective practice. The process of revising the 9 invalid items is a valuable professional development step. Teachers can involve colleagues or experts (*expert judgment*) to re-examine these items, improve distractors, clarify questions, and ensure their alignment with indicators (Azwar, 2015).

Third, in the context of the "Kurikulum Merdeka," which emphasizes authentic and meaningful assessment (Kemendikbudristek, 2022), sorting valid and invalid items helps teachers compile a quality item bank. Validated items can be reused or combined with other forms of assessment, such as performance assessments or portfolios, to create a more holistic picture of student learning progress (Wiggins & McTighe, 2005).

Limitations and Suggestions for Further Research

This study has limitations, particularly the relatively small sample size (20 students). Although validity and reliability analysis with KR-21 can still be performed, the findings should be generalized cautiously. For future research, it is recommended to use a larger sample (a minimum of 30 as recommended by Gay et al., 2012) to make the analysis results more stable and representative. Furthermore, subsequent research could conduct a more in-depth analysis using Item Response Theory (IRT) such as the Rasch model, which considers not only correlation but also item difficulty and student ability simultaneously (Handayani, 2022). Qualitative analysis of invalid items, for instance through *think-aloud* interviews with students, could also provide richer insights into the reasons for an item's invalidity.

CONCLUSION AND SUGGESTIONS

Conclusion

Based on the research results and discussion, it can be concluded that:

1. The evaluated Biology learning assessment instrument has 11 valid items (55%) and 9 invalid items (45%), indicating the need for significant improvement in nearly half of the total items.
2. The instrument has very high reliability with a coefficient of 0.82, which guarantees measurement consistency and its suitability as a tool for measuring learning outcomes.

Suggestions

Based on the conclusions, the following suggestions are proposed:

1. For Teachers: Conduct a thorough revision of the 9 invalid items by checking distractor functionality, clarity of wording, and alignment with learning indicators. Build an item bank consisting of items that have been tested for validity and reliability.
2. For Future Researchers: Replicate the study with a larger sample size and use more comprehensive analysis methods such as the Rasch model to obtain a deeper understanding of the instrument's quality.

REFERENCES

- Andrade, H. L., & Heritage, M. (2017). *Using Formative Assessment to Enhance Learning, Achievement, and Academic Self-Regulation*. Routledge.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Arikunto, S. (2019). *Dasar-dasar Evaluasi Pendidikan* (Edisi Revisi). Bumi Aksara.
- Azwar, S. (2015). *Penyusunan Skala Psikologi* (Edisi 2). Pustaka Pelajar.
- Bari, A. J. I., Ibrahim, M., & Yuliani, Y. (2023). Pengembangan Instrumen Penilaian Kinerja Keterampilan Membuat Alat

- Laboratorium Respirometer Sederhana. *JPBIO (Jurnal Pendidikan Biologi)*.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD.
- Campbell, N. A., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Reece, J. B. (2018). *Biology: A Global Approach* (11th ed.). Pearson.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to Design and Evaluate Research in Education* (8th ed.). McGraw-Hill.
- Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational Research: Competencies for Analysis and Applications* (10th ed.). Pearson.
- Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of Student Achievement* (9th ed.). Pearson.
- Habibah, S. N., & Fuadiyah, S. (2025). Validitas Instrumen Tes Keterampilan Proses Sains pada Pembelajaran Biologi Materi Sistem Pencernaan Kelas XI SMA/MA. *Jurnal Jendela Pendidikan*, 5(02), 353–360.
- Handayani, S. (2022). Analisis Kualitas Soal Penilaian Tengah Semester Biologi menggunakan Teori Tes Klasik dan Model Rasch. *Bio-Pedagogi: Jurnal Pembelajaran Biologi*, 11(2), 76–90.
- Kemendikbudristek. (2022). *Panduan Pembelajaran dan Asesmen di Sekolah Menengah Atas (SMA)*. Direktorat Jenderal Pendidikan Anak Usia Dini, Pendidikan Dasar dan Menengah.
- Mardapi, D. (2017). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Parama Publishing.
- Marthiani, I. (2024). Uji Validitas dan Reliabilitas Instrumen Penelitian Pemahaman Konsep Biologi. *Jurnal Yudistira: Publikasi Riset Ilmu Pendidikan dan Bahasa*, 2(2), 351–356.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students* (6th ed.). Pearson.
- OECD. (2018). *The Future of Education and Skills: Education 2030*. OECD Publishing.
- PISA. (2018). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing.
- Puspa Chandra, D., et al. (2021). Analisis Kualitas Butir Soal Ujian Akhir Semester Genap Mata Pelajaran Biologi Kelas X di SMA Negeri 1 ... [Jurnal Ilmiah Pendidikan Biologi], 10(1), 45-56.
- Sudjana, N. (2016). *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design* (2nd ed.). ASCD.
- Wilsa, A. W., et al. (2023). Pengembangan Instrumen Literasi Pengetahuan tentang Materi Sel dan Jaringan. [Jurnal Pendidikan Sains], 15(1), 112-125.