Comprehensive Analysis of Assessment Instruments: Validity, Reliability, Item Difficulty, and Discrimination Indices in Ecosystem Learning Materials

Dasari Waruwu[a*] | Hardikupatu Gulo[a] | Sehati Gea[a]

[a]Department of Biology Education, Faculty of Teacher and Training Education, Universitas Nias, Gunungsitoli, Indonesia

Received: 13 March 2026
Revised: 27 March 2026
Accepted: 28 April 2026
Published: 8 May 2026

Corresponding Author:
Author Name: Dasari Waruwu
Email: dasariwaruwu@gmail.com

DOI: 10.56207/genbionix.v4i1.813

Abstract: This study aimed to analyze the quality of assessment instruments on ecosystem material based on the aspects of validity, reliability, difficulty level, and discriminating power of test items. The study employed a quantitative descriptive research method. The research population consisted of all tenth-grade students at SMKN 1 Onohazumba, with the research sample comprising one agricultural major class. The instrument used was a multiple-choice test consisting of 30 items developed according to the learning indicators. Data collection techniques were carried out through learning achievement tests and documentation. Data analysis was conducted using validity testing with the Product Moment correlation formula, reliability testing using the KR-20 formula, difficulty level analysis, and discriminating power analysis of test items. The results showed that all test items were declared valid, with the highest r_{xy} value of 0.506 and the lowest value of 0.368, where all r_{xy} values were greater than the r-table value of 0.367. The reliability test produced a reliability coefficient of 0.844, which falls into the very high category. The difficulty level analysis indicated that all test items were categorized as moderate, with difficulty index values ranging from 0.32 to 0.68. Meanwhile, the discriminating power analysis showed that 20 test items were categorized as good, while 10 test items were categorized as excellent.

Keywords: assessment, validity, reliability, difficulty level, discriminating power

How to Cite:
Waruwu, D., Gulo, H., Gea, S. (2026). Comprehensive Analysis of Assessment Instruments: Validity, Reliability, Item Difficulty, and Discrimination Indices in Ecosystem Learning Materials. *Gen Bionix: Journal of Biology Education*, *4*(1), 8-15. https://doi.org/10.56207/genbionix.v4i1.813

Introduction

Learning evaluation is an essential component of the educational process, as it functions to determine the extent to which students achieve the intended learning objectives. Evaluation is conducted to identify the development of students' abilities after participating in the learning process (Sudijono, 2015). One form of learning evaluation that can be applied is the use of assessment instruments.

Assessment instruments are measurement tools used by teachers to obtain data regarding the outcomes of the learning process (Kushwah et al., 2024). These instruments aim to determine the achievement of learning objectives so that teachers can conduct accurate and appropriate assessments (Chen & Bonner, 2020). Learning evaluation may be carried out through both test and non-test assessment instruments to measure students' learning outcomes (Sudijono, 2015). In this context, the assessment instrument employed in the evaluation process is a test-based instrument.

A test instrument is an assessment tool used to measure students' learning outcomes. Test instruments are generally designed in the form of questions or test items that students must answer or complete according to predetermined instructions (Molnár, 2021a; Myoung & Liou, 2025). However, before a test instrument can be used as an effective measurement tool, the quality of the test items must first be analyzed.

Item analysis is a process used to determine the quality of test questions based on specific criteria, enabling educators to identify whether the items are appropriate for use (Molnár, 2021a, 2021b). Item analysis is conducted to identify good-quality items, items of insufficient quality, and items that need revision or elimination in order to ensure that the test instrument is truly effective and reliable (Arikunto, 2018). The purpose of item analysis is to evaluate the quality of each test item so that its appropriateness in measuring students' abilities can be ensured. Therefore, the quality of a test instrument can be examined through item analysis covering validity, reliability, difficulty level, and discriminating power. Item analysis is essential to ensure that every test item accurately measures students' abilities (Arikunto, 2018).

One of the primary aspects analyzed in item analysis is validity. Validity refers to the degree to which an instrument accurately measures what it is intended to measure (Aryadoust et al., 2020). A valid instrument is capable of accurately representing students' abilities, knowledge, or skills in accordance with the learning objectives. Validity demonstrates the precision of an instrument in assessing aspects that align with the intended measurement objectives, ensuring that

the test results genuinely reflect the abilities being measured (Arikunto, 2018). Thus, validity is a crucial aspect of item analysis because it determines the extent to which a test instrument effectively measures the intended competencies. Instruments with high validity produce accurate data that can serve as a reliable basis for assessing students' learning achievement.

In addition to validity, reliability is another essential requirement in item analysis. Reliability refers to the consistency or stability of measurement results obtained from an instrument (Vlachopoulos, 2025). A reliable instrument can be trusted because it consistently produces stable measurement results, making it suitable as an accurate evaluation tool (Arikunto, 2018). Reliability analysis is conducted to determine the consistency of measurement outcomes when an instrument is used under relatively similar conditions. Therefore, reliability is an important aspect of item analysis because it reflects the stability of measurement results. Instruments with high reliability generate more dependable data, allowing evaluation results to be used accurately in assessing students' abilities.

Another important aspect that must be considered is the difficulty level of test items. Item difficulty level is one of the key analyses in test evaluation used to determine the quality of a test instrument (Sudijono, 2015). The difficulty level refers to the degree of ease or difficulty of a test item based on the proportion of students who answer the item correctly (Halawa & Gulo, 2023; Lahagu et al., 2023). Difficulty level analysis is conducted to classify items into easy, moderate, or difficult categories, which can then be used as a basis for evaluating the quality of test items in proportionally measuring students' abilities (Zebua, 2024). Therefore, the difficulty level of test items is a significant aspect in both the development and analysis of test instruments.

Besides difficulty level, discriminating power is also an important indicator in item analysis. Discriminating power refers to the ability of a test item to distinguish between high-achieving students and low-achieving students. It is used to determine how effectively a test item can identify differences in students' mastery of the material. A test item with good discriminating power can clearly differentiate students based on their levels of understanding, thereby enabling the quality of the test instrument to be evaluated more accurately and objectively (Arikunto, 2018). Consequently, discriminating power is an important aspect in determining item quality because it relates to the ability of test items to accurately distinguish students' varying abilities (Gulo, 2023; Harefa, 2023).

The analysis of test instruments is conducted in learning evaluation, particularly in ecosystem material, to ensure that the test items used possess good quality in terms of validity, reliability, difficulty level, and discriminating power. This ensures that evaluation results accurately and objectively reflect students' abilities. Test instrument analysis is necessary in ecosystem material because, in practice, various problems are still frequently found in the quality of test items and inaccuracy in measuring learning outcomes (Telaumbanua et al., 2023; Waruwu & Gulo, 2023; Zega & Gulo, 2023).

In ecosystem test instruments, one common issue is the mismatch between test items and learning indicators, causing the questions to fail in accurately measuring the expected competencies. This problem may arise from inadequate validity and inconsistent test results due to low reliability, making the assessment outcomes less accurate and less trustworthy (Arikunto, 2018). In addition, many test items are either too easy or too difficult, preventing them from providing optimal information about students' abilities (Zebua, 2026). This highlights the importance of difficulty level analysis to ensure that test items fall within proportional categories (Sudijono, 2015). Another issue is the low discriminating power of test items, namely the inability of items to differentiate between high-ability and low-ability students, resulting in evaluation outcomes that do not accurately reflect students' actual competencies. Therefore, test instrument analysis is essential to ensure that ecosystem-related test items possess high quality in terms of validity, reliability, difficulty level, and discriminating power, so that evaluation results truly represent students' abilities objectively and accurately.

Methods

This study employed a quantitative descriptive research design aimed at analyzing the quality of test items based on four major aspects: validity, reliability, difficulty level, and discriminating power. The research was conducted at SMKN 1 Onohazumba. The study utilized both primary and secondary data sources. Primary data consisted of students' learning achievement scores obtained directly from ecosystem learning tests administered to the participants. Meanwhile, secondary data were collected from relevant scientific articles, academic journals, and previous research findings related to educational evaluation and item analysis to support the study.

The population of this study included all tenth-grade students majoring in agriculture at SMKN 1 Onohazumba, while the research sample consisted of tenth-grade students selected from the same population. The instrument used in this study was a multiple-choice test on ecosystem material that had been developed according to predetermined learning indicators to ensure alignment with the instructional objectives. Data collection techniques involved administering

students' learning achievement tests on ecosystem material and conducting documentation activities, including collecting test questions, students' answer sheets, and other supporting documents relevant to the research process.

The quality of the test instrument was analyzed through several procedures, including validity testing, reliability testing, difficulty level analysis, and discriminating power analysis. The first stage of instrument analysis was item validity testing. A test instrument is considered valid if it accurately measures the competencies or constructs intended to be assessed. The validity test was conducted using students' test results. According to Suharsimi Arikunto (2012), item validity can be measured using the *Product Moment Correlation* formula. Reliability analysis was then conducted to determine the consistency and stability of the instrument in measuring students' abilities. Furthermore, difficulty level analysis was carried out to classify test items into easy, moderate, or difficult categories based on the proportion of correct answers provided by students. Discriminating power analysis was also performed to determine the ability of each test item to distinguish between high-achieving and low-achieving students. Through these analyses, the study aimed to ensure that the test instrument used in ecosystem learning evaluation possessed good quality and was capable of measuring students' learning outcomes accurately and objectively.

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

Description:

r_{xy} = Correlation coefficient between variable X and variable Y, representing the two correlated variables
N = Number of subjects
$\sum X$ = Total score of each test item
$\sum Y$ = Total score obtained
$\sum XY$ = Total multiplication of item scores and total scores
$\sum X^2$ = Sum of squared item scores
$\sum Y^2$ = Sum of squared total scores

According to Arikunto (2012), the validity testing criteria are determined by comparing the calculated Product Moment correlation value with the value in the Product Moment table at a significance level of 5%. If $r_{xy} > r^{\text{table}}$, the test item is considered valid; however, if $r_{xy} < r^{\text{table}}$, the test item is considered invalid. After conducting the validity test, the next step is reliability testing. Reliability refers to the consistency or stability of a measuring instrument in assessing what it is intended to measure. Reliability analysis is conducted to determine the extent to which the measurement results can be trusted. According to Sukiman (2012), the formula used to calculate reliability is the KR-20 formula as follows.

$$\text{r11} = \left(\frac{n}{n-1}\right)\left(\frac{t \quad S^2 - \sum piqi}{S_t^2}\right)$$

Where:

r_{11} = Test reliability coefficient
n = Number of test items
1 = Constant value
S_t^2 = Total variance
p_i = Proportion of students who answered the corresponding test item correctly
q_i = Proportion of students who answered the corresponding test item incorrectly
$\sum p_i q_i$ = Total result of the multiplication between p_i and q_i

The calculated reliability coefficient is then compared with the r-table value using the following decision criteria: if $r_{11} > r$-table, the instrument is considered reliable; however, if $r_{11} < r$-table, the instrument is considered unreliable. The calculation results are subsequently interpreted based on the criteria presented in Table 1.

Table 1. Reliability Categories

Reliability Category	Interpretation
$0.81 < r \leq 1.00$	Very High
$0.61 < r \leq 0.80$	High
$0.41 < r \leq 0.60$	Moderate
$0.21 < r \leq 0.40$	Low
$0.00 < r \leq 0.20$	Very Low

(Source: Suharsimi Arikunto, 2013)

Analyzing the difficulty level of test items means examining which questions are categorized as easy, moderate, or difficult. According to Suharsimi Arikunto (2012), a good test item is neither too easy nor too difficult. The criteria for item difficulty level can be seen in Table 2. To calculate the difficulty index, the proportion correct (P)formula can be used as follows.

$$P = \frac{B}{JS}$$

Description:
(P) = Difficulty index
(B) = Number of students who answered correctly
(JS) = Total number of students taking the test

Table 2. Criteria for Item Difficulty Level

Difficulty Index	Criteria
(0.00 - 0.30)	Difficult
(0.31 - 0.70)	Moderate
(0.71 - 1.00)	Easy

(Source: Suharsimi Arikunto, 2013)

After determining the item difficulty level, the next step is to determine the discriminating power of the test items. The test items used in this study were in the form of multiple-choice questions. Following the calculation process, the discriminating power criteria of each item can be identified based on the established classification standards. The criteria for the discriminating power index of the developed test items are presented in Table 3. Furthermore, to determine the discriminating power of the multiple-choice items, the following formula can be used.

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Description:
(D) = Discriminating power of the test item
(B_A) = Number of students in the upper group who answered correctly
(B_B) = Number of students in the lower group who answered correctly
(J_A) = Total number of students in the upper group
(J_B) = Total number of students in the lower group

Table 3. Discriminating Power Criteria

Category	Criteria
(0.00 - 0.20)	Poor
(0.21 - 0.40)	Fair
(0.41 - 0.70)	Good
(0.71 - 1.00)	Excellent

Result and Discussion

Test Instrument Validity Results

The validity test was conducted on 30 test items. The validity of the test instrument was tested on 30 tenth-grade agricultural students at SMKN 1 Onohazumba. Based on the data analysis obtained from the instrument validity test, the results are summarized in Table 4.

Table 4. Validity Test Results

Component	Results
Number of trial participants	30 students
Number of test items	30 items
(r)-table value	0.367
Highest (r_{xy}) value	0.506
Lowest (r_{xy}) value	0.368

Number of valid items	30 items
Number of invalid items	0 items

Based on the validity analysis results, all test items were declared valid. The validity results were obtained by comparing the (r_{xy}) values with the (r)-table value at a 5% significance level, which was 0.367. All test items had (r_{xy}) values greater than the (r)-table value, indicating that they met the validity criteria and were appropriate for use as evaluation instruments in the learning process. The analysis showed that the highest validity value was 0.506, while the lowest value was 0.368. Nevertheless, all validity values remained above the (r)-table value, meaning that no invalid items were found. This indicates that each test item had a good ability to measure the predetermined learning indicators. Overall, the test instrument demonstrated a high level of accuracy in measuring students' abilities. A valid instrument is one that accurately measures what it is intended to measure (Suharsimi Arikunto, 2018). Therefore, all test items used in this study were aligned with the learning indicators and were suitable for use in learning evaluation.

Reliability Test

After the validity test was completed, the next stage involved reliability testing to determine the consistency of the instrument. Based on the reliability analysis, the results are summarized in Table 5.

Table 5. Summary of Reliability Test Results

Component	Results
Testing method	KR-20
Reliability value	0.844
Reliability category	Very High

Based on calculations using the KR-20 formula, the reliability coefficient obtained was 0.844. This value falls within the "very high" category because it is within the range of 0.81–1.00. A very high reliability value indicates that the instrument has a strong level of consistency in measuring students' abilities. In other words, if the instrument is administered again under similar conditions, the measurement results are likely to remain relatively consistent. This finding suggests that the test instrument is reliable and appropriate for use as a learning evaluation tool. According to Suharsimi Arikunto (2018), reliability reflects the consistency of an instrument in producing stable measurement results. Instruments with high reliability indicate that the test results are dependable because they are stable and not influenced by random factors (Sudijono, 2015).

Item Difficulty Level Test

The difficulty level test was conducted to determine the level of difficulty of each test item. The results of the item difficulty analysis are presented in Table 6.

Table 6. Item Difficulty Level Results

Category	Number of Items
Difficult	0
Moderate	30
Easy	0

Based on the analysis of the 30 test items, all items were categorized as moderate in difficulty. The difficulty index values ranged from 0.32 to 0.68. The results indicate that the test items were neither too easy nor too difficult for the students. Items classified as moderate are considered more effective in measuring students' abilities because they provide a better distribution of learning outcomes. A proportional level of difficulty also indicates that the instrument was developed in accordance with the students' level of ability.

Discriminating Power Test

The discriminating power test was conducted to determine the ability of each test item to distinguish between high-achieving students and low-achieving students. Based on the discriminating power analysis, the results are summarized in Table 7.

Table 7. Discriminating Power Analysis Results

No	BA	BB	D	Category
1	6	3	0.75	Excellent

2	5	1	0.50	Good
3	6	2	0.50	Good
4	5	1	0.75	Excellent
5	6	2	0.50	Good
6	7	1	0.75	Excellent
7	5	2	0.50	Good
8	7	2	0.63	Good
9	6	2	0.50	Good
10	5	1	0.50	Good
11	6	1	0.75	Excellent
12	4	1	0.50	Good
13	4	0	0.50	Good
14	6	1	0.75	Excellent
15	7	3	0.50	Good
16	5	3	0.50	Good
17	5	0	0.75	Excellent
18	5	2	0.50	Good
19	6	1	0.75	Excellent
20	5	1	0.75	Excellent
21	7	3	0.50	Good
22	7	3	0.50	Good
23	6	3	0.50	Good
24	5	1	0.75	Excellent
25	8	5	0.50	Good
26	6	3	0.50	Good
27	7	3	0.50	Good
28	6	2	0.50	Good
29	5	2	0.50	Good
30	7	3	0.75	Excellent

Based on the discriminating power analysis of the 30 test items, most items were categorized as "good." The analysis showed that 20 items fell into the "good" category, while 10 items were classified as "excellent." These findings indicate that the majority of the test items were capable of accurately distinguishing between students with high and low levels of ability. Test items with good discriminating power are able to clearly demonstrate differences in students' mastery of the subject matter (Suharsimi Arikunto, 2013).

The study conducted by Nugraha and Sari (2024), entitled *Analysis of Validity, Reliability, and Item Characteristics of Senior High School Biology Examination Questions*, found that the test instrument demonstrated high reliability (>0.80), and most of the test items were categorized as valid. However, several items required revision because they were unable to effectively distinguish between high-achieving and low-achieving students (Zebua, 2026). The study emphasized that the quality of the instrument greatly influences the accuracy of learning evaluation results.

Another study conducted by Hutabarat (2025), entitled *Evaluation of the Quality of Science Test Instruments on Ecosystem Material Based on Higher-Order Thinking Skills (HOTS)*, showed that the development of HOTS-based questions tended to improve the discriminating power of test items. The findings revealed that test items with moderate to high difficulty levels were more effective in measuring students' higher-order thinking skills (Zebua & Ziraluo, 2026; Ziraluo, 2026). In addition, the analyzed instrument demonstrated a very high level of reliability, indicating that it was appropriate for use in learning evaluation.

Conclusion

Based on the results of the study, the assessment instrument demonstrated good quality in terms of validity, reliability, difficulty level, and discriminating power. The validity analysis showed that all 30 test items were declared valid, with (r_{xy}) values ranging from 0.368 to 0.506, exceeding the (r)-table value of 0.367. The reliability analysis using the KR-20 formula produced a coefficient of 0.844, which falls into the very high category, indicating that the instrument was consistent and reliable. Furthermore, the difficulty level analysis revealed that all 30 test items were categorized as moderate, with difficulty indices ranging from 0.32 to 0.68, demonstrating that the questions were neither too easy nor too difficult for students. In addition, the discriminating power analysis indicated that 20 items were categorized as good and 10 items as excellent, with discriminating power values ranging from 0.50 to 0.75. Overall, these findings indicate that the assessment instrument is appropriate and effective for measuring students' learning outcomes objectively and accurately.

Funding

This research received no external funding and was conducted independently by the authors without financial support from any institution or organization.

References

Arikunto, S. (2013). Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara.

Arikunto, S. (2018). Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi). Jakarta: Bumi Aksara.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2020). A Comprehensive Review of Rasch Measurement in Language Assessment: Recommendations and Guidelines for Research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Chen, P. P., & Bonner, S. M. (2020). A framework for classroom assessment, learning, and self-regulation. *Assessment in Education: Principles, Policy & Practice*, *27*(4), 373–393. https://doi.org/10.1080/0969594X.2019.1619515

Gulo, H. (2023). Penerapan Model Pembelajaran Probing Prompting Berbasis Active Learning Untuk Meningkatkan Hasil Belajar IPA SMP Negeri 4 Hiliserangkai. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, *1*(1), 55–63. https://doi.org/10.56207/genbionix.v1i1.217

Halawa, K. K., & Gulo, H. (2023). Development of Flashcard Learning Materials for the Excretory System Curriculum for 8th Grade Students at SMP Negeri 3 Lolofitu Moi Academic Year 2020/2021. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, *1*(2), 71–78. https://doi.org/10.56207/gb.v1i2.770

Harefa, A. R. (2023). Pengembangan Lembar Peserta Didik IPA Berbasis Inkuiri Pada Materi Pencemaran Lingkungan Kelas VII SMP. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, *1*(1), 32–40. https://ojs.unias.ac.id/index.php/gb/article/view/226

Hutabarat, J. (2025). Evaluasi kualitas instrumen tes IPA pada materi ekosistem berbasis HOTS. Jurnal Evaluasi Pendidikan.

Kushwah, L., Navrouzoglou, P., & Cheng, W. (2024). *Embedding Sustainability in Existing Curriculum by Using Authentic Assessments, Rubrics, and Formative Assessments: A Case Study* (pp. 241–254). https://doi.org/10.1007/978-3-031-55996-9_16

Lahagu, N., Harefa, A. O., & Telaumbanua, D. (2023). Pengaruh Model Pembelajaran Contextual Teaching and Learning (CTL) terhadap Motivasi dan Hasil Belajar IPA Siswa SMP. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, *1*(2), 84–89. https://doi.org/10.56207/genbionix.v1i2.799

Molnár, G. (2021a). Challenges and developments in technology-based assessment: possibilities in science education. *Europhysics News*, *52*(2), 16–19.

Molnár, G. (2021b). Challenges and developments in technology-based assessment: possibilities in science education. *Europhysics News*, *52*(2), 16–19.

Myoung, E., & Liou, P.-Y. (2025). Open Classroom Climate in International Large-Scale Assessments: Individual Construct or Climate Construct? *SAGE Open*, *15*(1). https://doi.org/10.1177/21582440251323560

Nugraha, A., & Sari, M. (2024). Analisis validitas, reliabilitas, dan karakteristik butir soal ujian biologi SMA. Jurnal Penelitian Pendidikan IPA.

Sari, D. P., & Wahyuni, R. (2023). Analisis kualitas butir soal biologi materi ekosistem menggunakan aplikasi ANATES. Jurnal Pendidikan Biologi Indonesia.

Sudijono, A. (2015). Pengantar Evaluasi Pendidikan. Jakarta: Raja Grafindo Persada.

Sukiman. (2012). Pengembangan Sistem Evaluasi. Yogyakarta: Insan Madani.

Telaumbanua, J. M., Gulo, H., & Zebua, N. (2023). Validity Test of the Growth and Development Module Based on Local Potential for Students at SMA Negeri 1 Afulu. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, 55–61. https://doi.org/10.56207/genbionix.v1i2.768

Vlachopoulos, D. (2025). *Authentic Assessment: Bridging Higher Education And Real-World Skills Through Diverse Methodologies*. 2225–2234. https://doi.org/10.21125/inted.2025.0628

Waruwu, T., & Gulo, H. (2023). Pengembangan Buku Saku Pada Materi Pemanasan Global Kelas VII SMP. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, 8–17. https://doi.org/10.56207/genbionix.v1i1.222

Zebua, N. (2024). Evaluating the Impact of GenAI in High School Education: A Critical Review. *Polygon: Jurnal Ilmu Komputer Dan Ilmu Pengetahuan Alam*, *2*(6), 79–85. https://doi.org/10.62383/polygon.v2i6.296

Zebua, N. (2026). Strengthening Partnerships Between Universities, Schools, and Government Agencies. In *Teacher Education for a Changing World* (pp. 289–312). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-4536-9.ch011

Zebua, N., & Ziraluo, Y. P. B. (2026). Bridging the Gap: Integrating GenAI into Secondary Science Curricula. In *Harnessing Generative Artificial Intelligence for Science Education* (pp. 137–162). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-2913-0.ch005

Zega, N. A., & Gulo, H. (2023). Pengembangan Lembar Kerja Peserta Didik Model Contextual Teaching and Learning Pada Materi Sistem Ekskresi Manusia Kelas VIII SMP Negeri 3 Lolofitu Moi. *GEN BIONIX: Jurnal Ilmiah Pendidikan Biologi*, *1*(1), 41–54. https://doi.org/10.56207/genbionix.v1i1.221

Ziraluo, Y. P. B. (2026). Enhancing Biology Students' Learning Outcomes through Microlearning at Universitas Nias Raya. *Jurnal Pendidikan Biologi Undiksha*, *12*(3), 223–237. https://doi.org/10.23887/jjpb.v12i3.110347