



## **VALIDITY AND RELIABILITY ANALYSIS OF BIOLOGY TEST AND NON-TEST INSTRUMENTS ON HUMAN RESPIRATORY SYSTEM LEARNING AT SMA NEGERI 3 PAREPARE**

Analisis Validitas dan Reliabilitas Instrumen Tes Dan Non-Tes Pada Materi Sistem Pernapasan Manusia di SMA Negeri 3 Parepare

**Joherlina<sup>1</sup>, Herlina<sup>2</sup>, Riang Sari<sup>3</sup>, Merkurinus S<sup>4</sup>, Hastuti<sup>5</sup>, Eka Apriyanti<sup>6\*</sup>**

<sup>1,2,3,4,5</sup> Program Pascasarjana, Universitas Patempo Makassar

<sup>6</sup>Program Pascasarjana, Universitas Negeri Makassar

Correspondence Author: Eka Apriyanti, [eka.apriyanti@unm.ac.id](mailto:eka.apriyanti@unm.ac.id), Departement of Pascasarjana, Universitas Negeri Makassar

### **ARTICLE INFO**

*Article History:* 12-12-2025

Received : 15-12-2025

Revised form : 16-12-2025

Accepted : 18-12-2025

Published online : 19-12-2025

#### **Keywords:**

Validity;

Reliability;

Test Instrument;

Non-test Instrument;

Human Respiratory System;

High School Biology

#### **Kata Kunci:**

Validitas;

Reliabilitas;

Instrumen Tes;

Instrumen Non-Tes;

Sistem Pernapasan Manusia;

Biologi SMA

### **ABSTRACT**

Learning evaluation plays a fundamental role in ensuring the quality of educational processes and outcomes. This study aimed to analyze the validity and reliability of test and non-test instruments used to evaluate Biology learning on the human respiratory system at SMA Negeri 3 Parepare. The instruments analyzed consisted of 20 multiple-choice test items and 20 student attitude questionnaire items. A descriptive quantitative research design was employed. Item validity was examined using the Pearson Product Moment correlation, while instrument reliability was analyzed using Cronbach's Alpha. The results showed that 15 out of 20 test items (75%) and 17 out of 20 questionnaire items (85%) met the validity criteria. The reliability coefficients were 0.86 for the test instrument and 0.88 for the non-test instrument, both classified as very high. These findings indicate that the instruments are suitable for evaluating Biology learning on the human respiratory system, although several items require revision. Continuous development and periodic testing of evaluation instruments are recommended to maintain the quality of learning assessment.

**How to Cite:** Joherlina, et.al. (2025). Validity And Reliability Analysis Of Biology Test And Non-Test Instruments On Human Respiratory System Learning At Sma Negeri 3 Parepare. *Journal Research Education Mamminasata*, Volume 1 (1), Page 115-121.

---

## ABSTRAK

Evaluasi pembelajaran merupakan komponen penting dalam menjamin mutu proses dan hasil pendidikan. Penelitian ini bertujuan untuk menganalisis validitas dan reliabilitas instrumen tes dan non-tes yang digunakan dalam evaluasi pembelajaran Biologi pada materi sistem pernapasan manusia di SMA Negeri 3 Parepare. Instrumen yang dianalisis terdiri atas 20 butir soal pilihan ganda dan 20 item angket sikap siswa. Penelitian ini menggunakan pendekatan deskriptif kuantitatif. Uji validitas dilakukan dengan korelasi Pearson Product Moment, sedangkan uji reliabilitas menggunakan Alpha Cronbach. Hasil penelitian menunjukkan bahwa 15 dari 20 butir soal (75%) dan 17 dari 20 item angket (85%) memenuhi kriteria validitas. Nilai reliabilitas instrumen tes sebesar 0,86 dan instrumen non-tes sebesar 0,88, keduanya tergolong sangat tinggi. Dengan demikian, instrumen tersebut layak digunakan untuk mengevaluasi pembelajaran Biologi pada materi sistem pernapasan manusia, meskipun beberapa butir masih memerlukan revisi. Pengembangan dan uji coba instrumen secara berkelanjutan direkomendasikan untuk menjaga kualitas evaluasi pembelajaran.

---

## INTRODUCTION

Learning evaluation is a crucial component of the educational system, functioning not only as a tool to measure learning outcomes but also as a mechanism to improve the quality of instruction [1]. Without valid and reliable instruments, evaluation results may be inaccurate and lead to inappropriate educational decisions [2].

In Biology learning, particularly at the senior high school level, the topic of the human respiratory system is fundamental for understanding physiological processes and maintaining body homeostasis. This topic is also highly relevant to contemporary health issues, making it essential for students to develop deep conceptual understanding rather than rote memorization [3]. Consequently, assessment of this topic requires well-designed instruments capable of measuring various dimensions of learning outcomes.

Current educational curricula emphasize holistic competency assessment, including higher-order thinking skills (HOTS), attitudes, and scientific dispositions [4]. However, several studies indicate that learning evaluation instruments used in schools often focus predominantly on cognitive outcomes and inadequately address affective aspects [5]. Moreover, assessment tools are frequently

adopted from existing sources without prior psychometric analysis, raising concerns about their validity and reliability [6].

Evaluation instruments can be classified into test instruments, which measure cognitive aspects, and non-test instruments, which assess affective and psychomotor domains. Both types are complementary and necessary to obtain a comprehensive picture of students' learning development [7]. The quality of these instruments is primarily determined by two psychometric characteristics: validity, which refers to the accuracy of measurement, and reliability, which indicates the consistency of measurement results [2,8].

Validity encompasses several dimensions, including content validity and construct validity. Content validity reflects the alignment between test items and curriculum objectives, while construct validity indicates whether an instrument truly measures the intended theoretical construct [1,2]. Reliability, on the other hand, ensures that an instrument yields stable and consistent results when administered under similar conditions [9].

Despite the importance of valid and reliable instruments, teachers often face practical constraints in developing and testing evaluation tools. As a result, many instruments

used in classroom assessment have uncertain psychometric quality [6]. Therefore, analyzing the validity and reliability of instruments already in use is essential as a formative effort to improve assessment quality and support continuous improvement in learning evaluation.

Based on these considerations, this study focuses on analyzing the validity and reliability of test and non-test instruments used to assess Biology learning on the human respiratory system at SMA Negeri 3 Parepare. The findings are expected to provide empirical evidence regarding the feasibility of the instruments and serve as a reference for improving evaluation practices in Biology education.

## METHODS

This study employed a descriptive quantitative approach to analyze the validity and reliability of test and non-test instruments on the human respiratory system material using a deep learning approach. The research was conducted during the second semester of the 2024/2025 academic year at SMA Negeri 3 Parepare.

The research subjects were eleventh-grade science students who had received Biology instruction on the human respiratory system using a deep learning approach. A total of 30 students were selected as samples through purposive sampling, based on their participation in formative assessments and completion of attitude questionnaires.

The instruments analyzed consisted of:

1. Test Instrument:

Twenty multiple-choice items designed to measure students' conceptual mastery of the human respiratory system, developed based on deep learning taxonomy.

2. Non-Test Instrument:

A Biology learning attitude questionnaire consisting of 20 statements using a four-point Likert scale to measure affective aspects within a deep learning framework.

Data analysis techniques included:

*Validity Testing:*

Pearson Product Moment correlation was used to assess the correlation between item scores and total scores. Items were considered valid if  $r_{count} > r_{table}$  at a 5% significance level.

*Reliability Testing:*

The KR-21 formula was applied for the test instrument, while Cronbach's Alpha was used for the non-test instrument. Reliability criteria followed Arikunto (2010):

0.80–1.00 : Very high reliability  
 0.60–0.79 : Adequate reliability  
 0.40–0.59 : Moderate reliability  
 < 0.40 : Low reliability

Data processing was conducted using Microsoft Excel and SPSS version 25 to facilitate validity and reliability calculations.

## RESULTS

**Table 1.** Validity Test Results of the Test Instrument (Multiple-Choice Items)

No.	Description	Number of Items	Percentage	Remarks
1	Valid items	15	75%	Valid
2	Invalid items (Items 3, 7, 12, 15, 18)	5	25%	Require revision or replacement
Total		20	100%	—

**Table 2.** Validity Test Results of the Non-Test Instrument (Attitude Questionnaire)

No.	Description	Number of Items	Percentage	Remarks
1	Valid items	17	85%	Valid
2	Invalid items (Items 5, 11, 16)	3	15%	Require revision
Total		20	100%	—

**Table 3.** Reliability Test Results of Research Instruments

No.	Instrument Type	Cronbach's Alpha	Reliability Category
1	Test Instrument	0.86	Very High
2	Non-Test Instrument	0.88	Very High

*a. Validity Test of the Test Instrument (Multiple-Choice Items)*

Of the 20 test items analyzed, 15 items (75%) were declared valid, having  $r_{count} \geq 0.361$ . The remaining five items (items 3, 7, 12, 15, and 18) were invalid due to correlation coefficients below the critical value and thus require revision or replacement. This level of validity indicates that most items adequately measure the intended construct of respiratory system understanding.

*b. Validity Test of the Non-Test Instrument (Attitude Questionnaire)*

Out of 20 questionnaire items, 17 items (85%) were declared valid. Three items (items 5, 11, and 16) showed low correlations with total scores and were therefore invalid. These items may contain ambiguous wording or may not accurately represent the intended attitude dimensions.

**c. Reliability Test**

The reliability analysis showed Cronbach's Alpha coefficients of 0.86 for the test instrument and 0.88 for the non-test instrument. Both values fall within the very high reliability category, indicating strong internal consistency and stable measurement properties [9].

**DISCUSSION**

Beyond classical item analysis, the validity and reliability findings in this study can also be interpreted within the broader framework of contemporary educational measurement. Recent literature emphasizes that instrument quality is not merely a technical requirement but a critical determinant of meaningful learning evaluation and instructional decision-making [12]. Valid assessment instruments enable teachers to accurately diagnose students' misconceptions and learning progress, particularly in conceptually demanding topics such as human physiology.

The moderate proportion of invalid cognitive test items identified in this study is consistent with recent findings showing that science assessments often struggle to balance conceptual depth with item clarity [13]. Biology concepts related to physiological systems frequently involve abstract processes and interrelated mechanisms, which may reduce item discriminability if not supported by precise

wording and well-functioning distractors [14]. This reinforces the need for iterative item refinement based on empirical evidence rather than reliance on face validity alone.

From an affective assessment perspective, recent studies highlight that student attitude instruments demonstrate higher psychometric quality when grounded in clear theoretical models of motivation and engagement [15]. The relatively high validity and reliability of the non-test instrument in this study suggest that attitude measurement can achieve robust psychometric standards when items are explicitly aligned with observable learning behaviors and learning contexts. This finding supports the growing consensus that affective outcomes are measurable constructs rather than purely subjective attributes [16].

Furthermore, the high reliability coefficients obtained for both instruments align with recent meta-analytic evidence indicating that classroom-based assessments can reach strong internal consistency when item homogeneity and construct focus are maintained [17]. However, reliability alone should not be interpreted as evidence of validity. Contemporary measurement theory emphasizes the importance of integrating multiple sources of validity evidence, including response processes and internal structure, to strengthen instrument interpretation [18].

In the context of competency-based curricula, such as those emphasizing deeper learning and higher-order thinking, valid and reliable instruments are essential to ensure alignment between learning objectives and assessment practices [19]. Instruments that lack psychometric rigor risk underestimating students' higher-level reasoning abilities or misrepresenting affective engagement, which may ultimately affect instructional planning and curriculum evaluation.

Finally, recent advances in educational assessment recommend combining classical test theory with modern measurement approaches to enhance instrument quality [20]. While this study relied on classical validity and reliability analysis, future research may benefit from incorporating Item Response Theory or computerized assessment systems to obtain more detailed information about item difficulty, discrimination, and student response patterns. Such integration would contribute to more

adaptive and evidence-based assessment practices in Biology education.

## CONCLUSION

Based on the research findings, both test and non-test instruments used to assess the human respiratory system through a deep learning approach at SMA Negeri 3 Parepare demonstrate very good validity and reliability and are therefore suitable for Biology learning evaluation. Teachers are encouraged to revise invalid items and continuously develop evaluation instruments in accordance with deep learning competency demands. Future research may integrate Item Response Theory and technology-based assessment approaches for more comprehensive instrument analysis.

## REFERENCES

- Arikunto S. *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara; 2012.
- Azwar S. *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar; 2023.
- Khairunnisa A, Aeni AN, Ismail A. The role of interactive media in learning the human respiratory system. *Pendas J Ilm Pendidik Dasar*. 2025;10(1).
- Magdalena I, Nurchayati A, Suhirman DP, Fathya NN. Implementation of cognitive development theory in science learning. *Anwarul*. 2023;3(5):960–969.
- Arfan Delar D, Reinita R, Arwin A, Mansurdin M. Cognitive, affective, and psychomotor assessment in learning. *J Pendidik Tambusai*. 2022;6:8390–8400.
- Azmi N, Prastowo P, Maslena M. Analysis of Biology student worksheets. *J Pelita Pendidik*. 2018;6(2):65–70.
- DeVellis RF. *Scale development: Theory and applications*. 4th ed. Thousand Oaks: Sage; 2017.
- Taherdoost H. Validity and reliability of research instruments. *Int J Acad Res Manag*. 2016;5(3):28–36.
- Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53–55.
- Janna NM, Herianto H. Validity and reliability testing using SPSS. *J Darul Dakwah Wal-Irsyad*. 2021;1:1–12.
- Khasanah M, Supriatno B, Anggraeni S. Analysis of worksheets in developing HOTS. *Bioscientist*. 2021;9(2):362–377.
- Brookhart SM. *How to create and use rubrics for formative assessment and grading*. Alexandria: ASCD; 2013.
- Krathwohl DR. A revision of Bloom's taxonomy: An overview. *Theory Pract*. 2002;41(4):212–218. doi:10.1207/s15430421tip4104\_2
- Nehm RH, Ha M, Mayfield E. Transforming biology assessment with machine learning. *CBE Life Sci Educ*. 2012;11(2):103–118. doi:10.1187/cbe.11-08-0060
- Schunk DH, Meece JL, Pintrich PR. *Motivation in education: Theory, research, and applications*. 4th ed. Boston: Pearson; 2014.
- Duckworth AL, Yeager DS. Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educ Res*. 2015;44(4):237–251. doi:10.3102/0013189X15584327
- Bonner SM. Validity in educational assessment. *Rev Educ Res*. 2013;83(4):469–503. doi:10.3102/0034654313497616
- Kane MT. Validation. In: Brennan RL, editor. *Educational measurement*. 4th ed. Westport: Praeger; 2006. p. 17–64.
- Pellegrino JW, Hilton ML. *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington DC: National Academies Press; 2012.
- Embretson SE, Reise SP. *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum Associates; 2000.