Jurnal Ilmiah Profesi Pendidikan
Volume 11, Nomor 1, Februari 2026
ISSN (Print): 2502-7069; ISSN (*Online*): 2620-8326

# Measuring Students' Critical Thinking Skills in Human Reproductive System Learning: Validity and Reliability Evidence

**Nurisma Farhani*, Dios Sarkity, Nevrita**
Biology Education Department, Universitas Maritim Raja Ali Haji, Tanjungpinang, Kepulauan Riau, Indonesia
*Corresponding Author: 2103030007@student.umrah.ac.id

**Article History**
Received : December 22th, 2025
Revised : January 20th, 2026
Accepted : February 06th, 2026

**Abstract:** Critical thinking has become a fundamental competency in twenty-first-century education and must be carefully cultivated and assessed within biology instruction, particularly in topics related to the human reproductive system. Accordingly, the availability of valid and reliable assessment instruments is essential to ensure accurate measurement of students' critical thinking skills. This study aimed to develop and evaluate the feasibility of an instrument designed to measure critical thinking in the context of human reproductive system content, with emphasis on empirical validity and reliability. A quantitative approach with a descriptive research design was employed in this study, involving 24 students from class XII.3 at SMA Negeri 6 Tanjungpinang. Data were collected using a test instrument consisting of 12 essay items constructed based on Ennis's critical thinking indicators. Empirical validity was examined using the product–moment correlation technique, while reliability was determined through Cronbach's Alpha coefficient with the assistance of SPSS software. The findings revealed that 10 of the 12 items met the criteria for empirical validity, whereas 2 items were deemed invalid. The reliability analysis produced a Cronbach's Alpha value of 0.799, indicating a high level of internal consistency. These results suggest that the developed instrument is appropriate and dependable for assessing senior high school students' critical thinking skills on the topic of the human reproductive system.

**Keywords:** Critical Thinking Skills, Human Reproductive System, Reliability, Validity

## INTRODUCTION

Advances in science and digital technology have significantly reshaped educational demands in the 21st century, requiring learners to develop higher-order thinking capacities. Among these capacities, critical thinking has emerged as a core competency that enables students to respond effectively to complex and dynamic challenges. In educational settings, the development of critical thinking skills should be accompanied by appropriate assessment mechanisms to ensure that these skills are measured accurately using valid and reliable instruments (Apriliani et al., 2021; Scott, 2019; Sudirman et al., 2020).

Critical thinking skills are essential competencies that every individual should possess. According to Ennis (2011), critical thinking refers to a rational and reflective process used to determine what actions to take or beliefs to hold. Along with technological advancements that enable rapid and widespread access to information, critical thinking skills are required to filter, evaluate, and process relevant information whose validity can be justified (Aulia et al., 2025). Critical thinking is also categorized as a higher-order thinking skill that serves as a foundation for effective problem-solving and decision-making processes (Ariadila et al., 2023). Through critical thinking skills, individuals are able to understand problems in depth, integrate various pieces of information, and construct logical assumptions or perspectives to identify appropriate solutions (Susianti et al., 2024). Therefore, critical thinking skills need to be developed through learning activities, but also systematically assessed using appropriate assessment instruments so that students' abilities can be accurately identified.

Considering the essential role of critical thinking, efforts to cultivate these skills are necessary to encourage students to actively use and refine their thinking abilities. Consequently, the enhancement of critical thinking skills has emerged as a key priority in learning, especially

in biology education (Santia & Hidayati, 2024). In biology learning, students are frequently confronted with complex problems that are not only conceptual in nature but also related to real-life phenomena and issues (Sahil et al., 2022). Biology instruction extends beyond the acquisition of concepts by engaging students in analyzing information, assessing arguments, and formulating logical conclusions (Mardliyah et al., 2023). Through critical thinking skills, students are able to distinguish between facts and opinions, relate concepts to real-world situations, and construct logical and well-structured arguments (Pranandsa et al., 2025). However, fostering critical thinking skills in biology education requires the support of assessment instruments that can measure these abilities in an objective and standardized way.

The human reproductive system is one of the biology topics that demands the application of students' critical thinking skills. This topic is complex because it involves various interrelated biological concepts and The complexity of this topic often leads to misconceptions among students, which may stem from insufficient conceptual understanding as well as from inaccurate information circulating in society (Luzyawati & Hamidah, 2023; Tenzer et al., 2022). Topics such as reproductive hormone mechanisms, gametogenesis, fertilization, the menstrual cycle, and disorders of the reproductive system require students to critically analyze and evaluate information. Thus, the distinctive features of the human reproductive system provide an appropriate context for developing and assessing students' critical thinking skills

Various studies have emphasized the importance of critical thinking skills in biology learning; however, most of these studies still focus on the implementation of learning models or instructional strategies to enhance students' critical thinking skills (Sahil et al., 2022; Santia & Hidayati, 2024). Studies that specifically develop critical thinking assessment instruments and examine their empirical validity and reliability, particularly on the human reproductive system topic at the senior high school level, remain limited (Fakhri Ramadhan et al., 2024; Subhaktiyasa, 2024). Therefore, the novelty of this study lies in the development of an essay-based critical thinking skills test constructed based on Ennis' critical thinking indicators and examined through empirical validity and reliability testing.

The development of critical thinking skills requires accurate and appropriate measurement so that students' abilities can be objectively identified. One commonly used approach to assess critical thinking skills is through written tests, particularly essay-type questions that require students to analyze, evaluate, and construct logical arguments. Accurate and consistent measurement of critical thinking skills can only be achieved through the use of valid and reliable instruments. Instrument validity is necessary to ensure that the instrument accurately measures the intended construct (Kurniawan et al., 2022), whereas instrument reliability indicates the consistency of measurement results when the instrument is administered under relatively similar conditions (Akbar & Zahfa, 2025). Accordingly, this study seeks to develop an instrument with established validity and reliability for assessing students' critical thinking skills in the context of the human reproductive system.

## METHODE

The study adopted a quantitative approach with a descriptive design aimed at developing and evaluating the feasibility of an instrument for measuring critical thinking skills in the topic of the human reproductive system. Instrument feasibility was examined through empirical validity and reliability analyses (Sahir, 2021). This research was conducted in December 2025 at SMA Negeri 6 Tanjungpinang, located in Tanjungpinang City, Riau Islands Province. The participants consisted of 24 students from class XII.3, selected through purposive sampling based on the criterion that they had previously studied the human reproductive system.

The research instrument was a critical thinking skills test comprising 12 essay questions related to the human reproductive system. The procedure involved several stages of instrument development and testing: (1) conducting a literature review to identify critical thinking indicators, (2) establishing indicators based on Ennis's critical thinking framework (Ennis, 2011), (3) constructing essay items aligned with the specified indicators, (4) piloting the instrument with students to obtain research data, and (5) analyzing the trial results.

Empirical validity for each item was assessed using the product–moment correlation method, while instrument reliability was evaluated through Cronbach's Alpha coefficient (Sundayana, 2014). All statistical analyses were performed using SPSS software. The distribution of test items according to the primary indicators of critical thinking is presented in Table 1.

**Table 1.** Distribution of Items Based on Critical Thinking Skill Indicators

| No. | Indicators | Item Number |
|---|---|---|
| 1. | *Basic Clarification* | 1, 7 |
| 2. | *Bases for a Decision* | 2, 9 |
| 3. | *Inference* | 5, 8 |
| 4. | *Advanced Clarification* | 3, 10 |
| 5. | *Supposition and Integration* | 4, 11 |
| 6. | *Auxiliary Abilities* | 6, 12 |

Valid and reliable instruments are fundamental criteria that must be fulfilled in quantitative research (Sarkity et al., 2023; Sarkity et al., 2023; Soesana et al., 2023). Validity and reliability testing are essential to demonstrate the accuracy of research instruments in measuring specific variables as well as the consistency and stability of an instrument in producing relatively similar measurement results when administered at different times (Iba & Wardhana, 2023). In this study, empirical validity and reliability testing involved 24 eleventh-grade students of SMA Negeri 6 Tanjungpinang. Data analysis was conducted using SPSS software. This study was carried out through two stages, namely:

### 1. Empirical Validity Testing

Empirical validity testing was conducted by administering the critical thinking skills test instrument to 24 twelfth-grade students of SMA Negeri 6 Tanjungpinang. An instrument is considered valid when the score of each item shows a significant correlation with the total test score (Ambarwati & Ismiyati, 2021). Validity testing was conducted to evaluate the accuracy of each item in measuring students' critical thinking skills and to analyze the association between individual item scores and the overall test score. The empirical validity of the essay-based items was examined using product moment correlation analysis with the support of SPSS software. The resulting correlation coefficients were then compared with the critical r-value at a 5% significance level (0.05) to determine item validity (Hajaroh & Raehanah, 2021). With a total of 24 participants, the r-table value applied in this study was 0.404. An item was classified as valid when the calculated correlation coefficient exceeded the r-table value (Sugiyono, 2023).

### 2. Reliability Testing

After obtaining the results of the empirical validity testing, the items that met the validity criteria were subsequently subjected to reliability testing. Reliability testing of the critical thinking skills test instrument, which consisted of essay-type items, was conducted using Cronbach's Alpha coefficient with the assistance of SPSS software. *An instrument was regarded as reliable when its Cronbach's Alpha coefficient exceeded the threshold value of 0.60* (Darma, 2021).

## FINDINGS AND DISCUSSION

### 1. Empirical Validity Testing

After the critical thinking skills test instrument was administered to 24 twelfth-grade students at SMA Negeri 6 Tanjungpinang, the resulting data were subsequently analyzed through empirical validity testing to identify valid and invalid items. Table 2 presents the results of the empirical validity analysis of the critical thinking skills test instrument developed for the human reproductive system topic. An item was regarded as valid when the calculated r-value exceeded the r-table value or when the significance level was below 0.05 (Sugiyono, 2019). Based on the results presented in Table 2, 10 items were found to be valid, while the remaining two items were classified as invalid.

**Table 2.** Results of Empirical Validity Testing of the Critical Thinking Skills Test Instrument

| **No.** | **r-calculated** | **r-table** | **significance** | **category** |
|---|---|---|---|---|
| 1. | 0.530 | 0.404 | 0.008 | **V** |
| 2. | 0.363 | | 0.081 | **NV** |
| 3. | 0.605 | | 0.002 | **V** |
| 4. | 0.587 | | 0.003 | **V** |
| 5. | 0.491 | | 0.015 | **V** |
| 6. | 0.308 | | 0.143 | **NV** |
| 7. | 0.512 | | 0.010 | **V** |

| No. | r-calculated | r-table | significance | category |
|---|---|---|---|---|
| 8. | 0.769 | | 0.000 | **V** |
| 9. | 0.643 | | 0.001 | **V** |
| 10. | 0.763 | | 0.000 | **V** |
| 11. | 0.583 | | 0.003 | **V** |
| 12. | 0.679 | | 0.000 | **V** |

Note: **V = Valid, NV = Not Valid**

Among the total items developed, 10 items demonstrated r-calculated values that exceeded the r-table threshold, indicating their validity and suitability for use. These valid items were then included in the reliability testing process. The empirical validity analysis showed that items 1, 3, 4, 5, 7, 8, 9, 10, 11, and 12 satisfied the validity requirements. Conversely, items 2 and 6 failed to meet the validity criteria due to r-calculated values below the r-table value. Item 2 recorded an r-calculated value of 0.363, whereas item 6 obtained an r-calculated value of 0.308. Overall, the results of the empirical validity testing indicate that 80% of the 12 items developed for the critical thinking skills test instrument on the human reproductive system topic were valid and appropriate for use.

The empirical validity analysis revealed that the majority of the developed items satisfied the validity criteria, indicating that the items were capable of assessing students' critical thinking skills based on the specified indicators. Good item validity reflects the alignment between the content of the instrument and the critical thinking construct being measured (Kurniawan et al., 2022). However, two items were found to be invalid, which may be attributed to suboptimal alignment between the cognitive demands of the items and the critical thinking indicators or to unclear item stimuli for students. This finding is consistent with the study by Ambarwati & Ismiyati (2021), which reported that unclear wording and inappropriate item difficulty levels can affect item validity.

## 2. Reliability Testing

After the ten items were confirmed to meet the validity criteria, a reliability test was conducted using Cronbach's Alpha coefficient. The results of this analysis are shown in Figure 1.

**Reliability Statistics**

| Cronbach's Alpha | N of Items |
|---|---|
| .799 | 10 |

**Figure 1.** Screenshot of the Reliability Test Results of the Critical Thinking Skills Test Instrument

An instrument is regarded as reliable when its Cronbach's Alpha coefficient exceeds the value of 0.60 (Darma, 2021). Based on the reliability analysis presented in Figure 1, the critical thinking skills test instrument obtained a reliability coefficient of 0.799. Since this value is higher than the established minimum threshold, the instrument developed for assessing critical thinking skills on the human reproductive system topic can be considered reliable.

Empirical validity and reliability testing are essential stages in evaluating the feasibility of an instrument and ensuring that the developed instrument is capable of accurately measuring the intended research variables. Research instruments play a crucial role as tools for measuring and collecting data related to the variables under investigation (Ramadhan et al., 2024). In research, the accuracy of variable measurement is determined by the use of valid and reliable instruments (Zayrin et al., 2025). Therefore, research instruments must be rigorously tested to ensure their accuracy and consistency in meeting validity and reliability criteria (Subhaktiyasa, 2024). A valid instrument is able to accurately measure and represent the investigated variables, while a reliable instrument demonstrates consistency and stability of measurement results when administered repeatedly over time (Fitriyani et al., 2025).

The findings of this study demonstrate that the developed instrument fulfills the requirements of empirical validity and reliability, indicating that it is suitable for measuring students' critical thinking skills within the context of the human reproductive system.

## CONCLUSION

Based on the conducted study, it can be concluded that the results of the empirical validity and reliability testing indicate that the developed instrument is valid and reliable for measuring students' critical thinking skills on the human reproductive system topic. The empirical validity results show that out of the 12 developed items, 10 items were found to be valid and

therefore eligible for reliability testing. The reliability testing yielded a coefficient of 0.799, indicating that the critical thinking skills test instrument on the human reproductive system topic is reliable.

## ACKNOWLEDGMENT

The authors gratefully acknowledge all parties who supported the conduct of this study, especially SMA Negeri 6 Tanjungpinang for providing permission and assistance during the research process. Appreciation is also extended to the twelfth-grade students of SMA Negeri 6 Tanjungpinang whose participation as respondents contributed to the successful completion of this study.

## REFERENCES

Akbar, A., & Zahfa, F. (2025). Validitas and Reliabilitas. *Jurnal Intelek Insan Cendikia*, *2*(5), 8781–8787.

Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. *Measurement in Educational Research*, *1*(2), 64–75.

Apriliani, E., Afandi, A., & Marlina, R. (2021). Memberdayakan Keterampilan Berpikir Kritis di Era Abad 21. *Prosiding Seminar Nasional Pendidikan 2020*, 1045–1052.

Ariadila, S. N., Silalahi, Y. F. N., Fadiyah, F. H., Jamaluddin, U., & Setiawan, S. (2023). Analisis Pentingnya Keterampilan Berpikir Kritis Terhadap Pembelajaran Bagi Siswa. *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 664–669.

Aulia, Q. N., Al Ayubi, S., & Rosyadi, S. (2025). Critical Thinking Dalam Al-Qur'an: Studi Tafsir Tematik dan Implementasinya di Era Digital. *Al-Fahmu: Jurnal Ilmu Al-Qur'an Dan Tafsir*, *4*(1), 131–149.

Darma, B. (2021). *Statistika Penelitian Menggunakan SPSS*. Guepedia.

Ennis, R. (2011). Critical thinking: Reflection and perspective Part II. *Inquiry: Critical Thinking across the Disciplines*, *26*(2), 5–19.

Fakhri Ramadhan, M., Siroj, R. A., Win Afgani, M., Raden Fatah Palembang, U., H Zainal Abidin Fikri, J. K., Kemuning, K., Palembang, K., & Selatan, S. (2024). Validitas and Reliabilitas. *Journal on Education*, *06*(02), 10967–10975.

Fitriyani, Sinaga, M. H., Hulu, R., & Wardani, W. (2025). Uji Validitas dan Reliabilitas Instrumen Persepsi Mahasiswa terhadap Penugasan Artikel dalam Pembelajaran di Jurusan Matematika. *Jurnal Penelitian Nusantara*, *1*(12), 714–719.

Hajaroh, S., & Raehanah. (2021). *Statistik Pendidikan Teori dan Praktik*. Sanabil Publisher.

Iba, Z., & Wardhana, A. (2023). *Metode Penelitian*. Eureka Media Aksara.

Kurniawan, A., Febrianti, A. N., Hardianti, T., Ichsan, Desy, Risan, R., Sari, D. M. M., Sitopu, J. W., Dewi, R. S., Sianipar, D., Fitriyah, L. A., Zulkarnaini, Jalal, N. M., Hasriani, & Hasyim, F. (2022). *EVALUASI PEMBELAJARAN* (A. Yanto & T. P. Wahyuni, Eds.; 1st ed.). PT. GLOBAL EKSEKUTIF TEKNOLOGI. www.globaleksekutifteknologi.co.id

Luzyawati, L., & Hamidah, I. (2023). Miskonsepsi peserta didik pada pembelajaran online sistem reproduksi manusia. *Venn: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences*, *2*(2), 37–51.

Mardliyah, S. Z., Salsabilla, A. P., & Herianingtyas, N. L. R. (2023). Strategi Peningkatan Kemampuan Berfikir Kritis Siswa Melalui Learning Community. *Awwaliyah: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *6*(2), 102–109.

Pranandsa, H., Istiqomah, C., Sarastri, F. A., & Utami, S. (2025). Peningkatan Kemampuan Membedakan Fakta, Opini, dan Asumsi dalam Teks Rekon melalui Model Role Playing. *Jurnal Pendidikan Tambusai*, *9*(2), 17055–17060.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975. https://doi.org/10.31004/joe.v6i2.4885

Sahil, J., Hasan, S., Haerullah, A., & Saibi, N. (2022). Penerapan pembelajaran abad 21 pada mata pelajaran biologi di SMA Negeri Kota Ternate. *BIOSFER: Jurnal Biologi Dan Pendidikan Biologi*, *7*(1), 13–19.

Santia, R., & Hidayati, N. (2024). Profil keterampilan berpikir kritis dalam pembelajaran biologi siswa kelas XI. *Bio-Pedagogi: Jurnal Pembelajaran Biologi*, *13*(2), 78–84.

Sarkity, D., Fernando, A., & Hindrasti, N. E. K. (2023). Designing and Content Validity of Instrument for Measuring Marine Environmental Care Attitude Through Integrated Science Learning. *BIO Web of Conferences*, *70*. https://doi.org/10.1051/bioconf/20237002011

Sarkity, D., Fernando, A., Pratama, Y., & Intasir, M. P. (2023). Expert Judgement on Instrument for Measuring Healthy Lifestyle Knowledges and Behaviors of Senior High School Students in Coastal Area of Tanjungpinang. *BIO Web of Conferences*, *79*, 5006. https://doi.org/10.1051/bioconf/20237905006

Scott, L. A. (2019). *21st CENTURY LEARNING FOR EARLY CHILDHOOD GUIDE*.

Soesana, A., Subakti, H., Karwanto, Fitri, A., Kuswandi, S., Sastri, L., Falani, I., Aswan, N., Hasibuan, F. A., & Lestari, H. (2023). *Metodologi Penelitian Kuantitatif*. Yayasan Kita Menulis.

Subhaktiyasa, P. G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka. *Journal of Education Research*, *5*(4), 5599–5609.

Sudirman, Anggereni, S., Marlinda, N. L. P. M., Silalahi, E. K., Fitrian, A., Siregar, H. T., Pa, R. H. B., Azizah, N. N., Hidayat, Saputri, M., Wirda, Nasrianty, & Karim, S. (2020). *IMPLEMENTASI PEMBELAJARAN ABAD 21 PADA BERBAGAI BIDANG ILMU PENGETAHUAN* (S. Haryanti, Ed.). CV. Media Sains Indonesia.

Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan RND* (Ke 2). Alfa Beta.

Sugiyono. (2023). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Sundayana, R. (2014). *Statistika Penelitian Pendidikan*. Alfabeta. Jakarta.

Susianti, L., Kania, N., Marliana, N. M., Listiani, H., Inayah, S., Rahmawati, F., Yulianto, E., & Rusli, T. S. (2024). PENDIDIKAN ABAD 21: Sebuah Tinjauan Kritis. *EDUPEDIA Publisher*, 1–116.

Tenzer, A., Handayani, N., & Daniarsih, A. (2022). Identifikasi miskonsepsi materi sistem reproduksi pada buku teks SMA kelas XI di Kota Malang. *Jurnal Pendidikan Biologi*, *13*(1), 11–23.

Zayrin, A. A., Nupus, H., Maizia, K. K., Marsela, S., Hidayatullah, R., & Harmonedi, H. (2025). Analisis Instrumen Penelitian Pendidikan (Uji Validitas Dan Relibilitas Instrumen Penelitian). *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, *3*(2), 780–789. https://doi.org/10.61104/jq.v3i2.1070