

## Empirical Analysis of the Validity and Reliability of an Energy Literacy Instrument for Indonesian Junior Secondary School Students

Musawwir Usman<sup>1</sup>, Yulianti Yusal<sup>2\*</sup>, Zulqifli Alqadri<sup>3</sup>  
Universitas Negeri Makassar

**Corresponding Author:** Yulianti Yusal [yulianti.yusal@unm.ac.id](mailto:yulianti.yusal@unm.ac.id)

---

### ARTICLE INFO

*Keywords:* Empirical Analysis, Validity, Reliability, Energy Literacy Instrument, Indonesian Junior Secondary School

*Received:* 19, February

*Revised:* 20, March

*Accepted:* 30, April

©2026 Usman, Yusal, Alqadri:  
This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).



### ABSTRACT

This study aimed to empirically evaluate the validity and reliability of an energy literacy assessment instrument for junior secondary school students. The instrument covers cognitive, affective, and behavioral domains of energy literacy. A quantitative approach was employed using classical test theory analysis. Data were collected from 134 Grade VIII students from four junior secondary schools. Item difficulty ( $p$ , proportion of correct responses) and item discrimination ( $r_{pbis}$ , point-biserial correlation) were analyzed for the cognitive domain, while item-total correlation ( $r$ , Pearson correlation) was used to examine the affective and behavioral domains. Internal consistency reliability was estimated using Cronbach's alpha ( $\alpha$ ) and KR-20. The results showed that most items demonstrated acceptable difficulty, discrimination, and item-total correlation values, and the overall instrument exhibited high reliability. It can be concluded that the refined energy literacy instrument is psychometrically sound and suitable for assessing students' energy literacy in junior secondary education

---

## INTRODUCTION

The success of energy education, as a part of sustainable education, is closely linked to how well students are able to understand, value, and apply energy-related concepts in their everyday lives (Hu & Yang, 2024). In this regard, energy literacy has become a central outcome of energy education, representing students' capacity to comprehend energy concepts, develop responsible attitudes, and engage in energy-related behaviors that support sustainability (DeWaters & Powers, 2011; U.S. Department of Energy, 2017; Usman, Syakur, et al., 2025). As energy literacy is inherently multidimensional, its assessment must address cognitive, affective, and behavioral aspects in an integrated manner.

To meaningfully evaluate students' energy literacy, instruments must be supported by empirical evidence demonstrating that they operate effectively in real classroom contexts. Energy literacy instruments are often developed based on well-established theoretical frameworks (Akitsu & Ishihara, 2018; DeWaters & Powers, 2013; Usman, Pramudawardani, et al., 2025). However, without empirical testing, it remains unclear whether individual items accurately represent the intended domains or whether the instrument can reliably distinguish between different levels of student energy literacy (Hartelt & Martens, 2025; Zhang et al., 2023). Empirical analysis provides essential information about item performance and internal consistency, ensuring that the instrument yields interpretable and trustworthy results when applied to students (Setiadi, 2021; Sudaryono et al., 2019; Sukmawati et al., 2023). Without this empirical analysis, conclusions drawn from assessment results may lack empirical robustness.

The need for empirical testing becomes more pronounced given the multidimensional structure of energy literacy. Each domain—cognitive, affective, and behavioral—possesses distinct characteristics that influence how it should be measured and evaluated (Chen et al., 2013; Hendinata et al., 2022). The cognitive domain requires objective items capable of distinguishing variations in students' conceptual understanding, while affective and behavioral domains are typically assessed through self-report measures that must demonstrate meaningful item-total relationships and internal consistency (Mukherjee & Kittur, 2025). These differences necessitate domain-specific analytical approaches to confirm that each component contributes appropriately to the overall measurement of energy literacy.

Existing research on energy literacy in Indonesia often emphasizes descriptive outcomes of students' literacy levels, with insufficient reporting of item difficulty, discrimination, or reliability indices (Putri et al., 2022; Usman et al., 2021). Within this body of research, empirical studies that rigorously evaluate the psychometric quality of energy literacy instruments have predominantly focused on university students, with limited evidence available at the junior secondary school level (Fiskawarni et al., 2024; Yusup, 2021). As a result, there are few well-validated energy literacy instruments available for junior secondary school students, which limits systematic evaluation and comparative research at this educational level.

## LITERATURE REVIEW

To address this gap, the present study conducts an empirical test of an energy literacy assessment instrument administered to Grade VIII students from several junior secondary schools. Using classical test theory analysis supported by Iteman software, this study examines item difficulty and discrimination for the cognitive domain, evaluates item validity for the affective and behavioral domains, and estimates the overall reliability of the instrument. The results are expected to provide empirical support for the use of the instrument in assessing students' energy literacy and to contribute to the improvement of energy education practices through more accurate and reliable measurement.

## METHODOLOGY

This study employed a quantitative approach to empirically examine the validity and reliability of an energy literacy assessment instrument developed for junior secondary school students. The focus of the study was on examining item performance and internal consistency using classical test theory analysis. The empirical evaluation was guided by the MEASURE Approach to instrument development (Kalkbrenner, 2021). The sequence of steps in the MEASURE approach is illustrated in Figure 1.

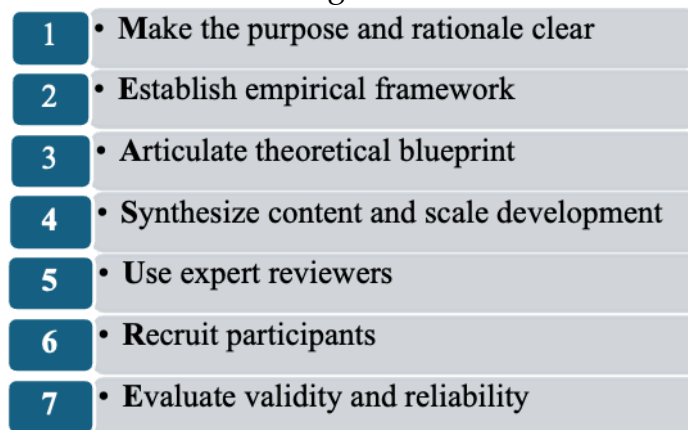


Figure 1. The Measure Approach to Instrument Development (Kalkbrenner, 2021)

In this study, attention was directed specifically to the final two stages of the MEASURE Approach, namely step 6: Recruiting participants (R) and step 7: Evaluating validity and reliability (E). The earlier stages – including framework synthesis, item construction, and expert-based content validation – were completed in a prior study that established the conceptual foundation and content validity of the energy literacy instrument (Usman, Pramudawardani, et al., 2025). The present study extends this work by providing empirical evidence of item quality and instrument reliability based on student response data.

The energy literacy instrument consisted of 59 items distributed across three domains: cognitive, affective, and behavioral. The cognitive domain comprised 30 multiple-choice items designed to assess students' conceptual understanding of energy-related topics. The affective and behavioral domains

were measured using 17 and 12 Likert-scale items, respectively, each employing a five-point response scale. The instrument was developed based on a synthesized theoretical framework of energy literacy and had undergone expert review to establish content validity before empirical administration (Usman, Pramudawardani, et al., 2025). The distribution of items across domains and aspects of energy literacy is presented in Table 1.

Table 1. Distribution of Energy Literacy Instrument Items by Domain

Domain	Aspect	Item No.
Cognitive	Basic energy knowledge	1 - 9
	Energy resource issue	10 - 16
	Common trend of energy in Indonesia	17 - 20
	Impact of individual and societal decisions related to energy	21 - 25
	Cognitive skills	26 - 30
Affective	Awareness/concern toward energy	1 - 5
	Positive attitudes, norms and values	6 - 9
	Strong efficacy beliefs	10 - 13
	Responsibility	14 - 17
Behavioral	Energy-saving behavior	1 - 8
	Thoughtful, effective decision making	9 - 12

Data were collected from 134 Grade VIII students drawn from four junior secondary schools in Indonesia, namely SMPN 1 Pakem, SMPN 3 Makassar, SMPN 1 Gempol, and MTs Sunan Kalijaga. The schools represented diverse educational contexts, and all participating students had previously received instruction related to energy concepts within the science curriculum. The sample size was considered adequate for classical test theory-based item analysis and internal consistency reliability estimation, which do not require the larger samples typically recommended for factor analytic procedures (Cappelleri et al., 2014; Kalkbrenner, 2021). Participation was voluntary, and responses were collected anonymously to ensure confidentiality.

Empirical analysis was conducted using classical test theory supported by Itean version 4.3. Item difficulty and discrimination indices were calculated for the cognitive domain to evaluate the quality of multiple-choice items, with the corresponding interpretation criteria presented in Tables 2 and 3, respectively. For the affective and behavioral domains, item-total correlations based on Pearson coefficients were used to examine item validity, and the interpretation criteria are provided in Table 4.

Table 2. Criteria for Item Difficulty

Item Difficulty Value (p)	Category
$p > 0.70$	Easy
$0.30 \leq p \leq 0.70$	Moderate
$p < 0.30$	Difficult

Source: Guyer & Thompson (2013) and Huda & Wahyuni (2020)

Table 3. Criteria for Item Discrimination

<b>Item Discrimination Value (<math>r_{pbis}</math>)</b>	<b>Category</b>
0.40 – 1.00	Good
0.30 – 0.39	Moderate
0.20 – 0.29	Fair
< 0.20	Poor

Source: Guyer & Thompson (2013) and Huda & Wahyuni (2020)

Table 4. Criteria for Item-Total Correlation

<b>Item-Total Correlation Value (r)</b>	<b>Category</b>
0.50 – 1.00	Good
0.30 – 0.49	Moderate
0.20 – 0.29	Fair
< 0.20	Poor

Source: Guyer & Thompson (2013) and Huda & Wahyuni (2020)

Internal consistency reliability coefficients were computed using Cronbach’s alpha (KR-20) to assess the overall reliability of the instrument across all domains, with the reliability interpretation criteria summarized in Table 5. An overview of the validity and reliability evaluation procedures is illustrated in Figure 2. Items that did not meet the predetermined criteria were removed from the instrument.

Table 5. Criteria for Reliability Coefficients

<b>Reliability Coefficient Value (<math>\alpha</math>)</b>	<b>Category</b>
$\alpha \leq 0.20$	Very Low
$0.20 < \alpha \leq 0.40$	Low
$0.40 < \alpha \leq 0.60$	Moderate
$0.60 < \alpha \leq 0.80$	High
$0.80 < \alpha \leq 1.00$	Very High

Source: Guyer & Thompson (2013) and Huda & Wahyuni (2020)

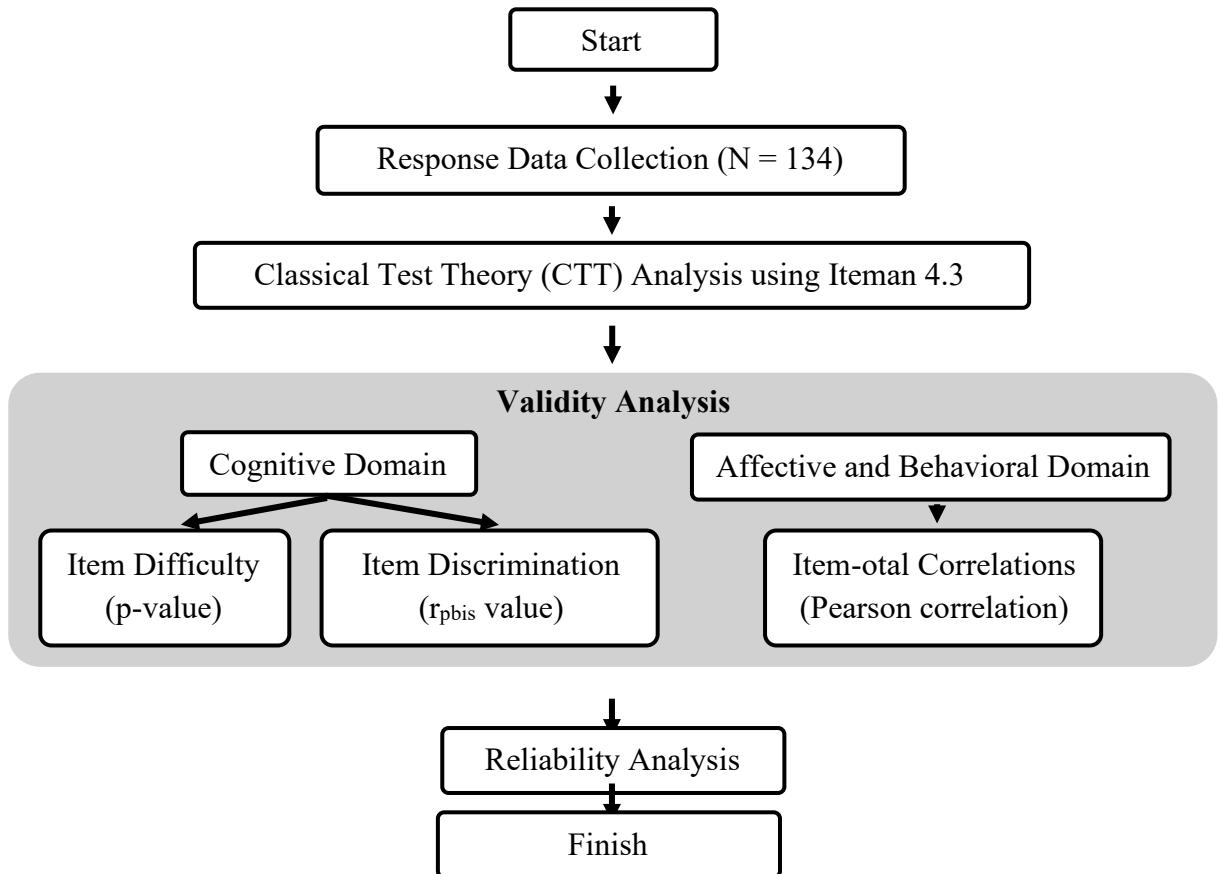


Figure 2. Flow of Validity and Reliability Evaluation Procedures

## RESULT AND DISCUSSION

### Item Analysis of the Cognitive Domain

The cognitive domain of the energy literacy instrument was evaluated to examine the quality of multiple-choice items in terms of item difficulty (p-value) and item discrimination indices (rpbis) using classical test theory. This analysis aimed to determine whether the cognitive items were appropriate for assessing Grade VIII students' understanding of energy concepts and whether the items were capable of distinguishing between students with different levels of energy literacy. The results of the item difficulty analysis are presented in Table 6.

Table 6. Item Difficulty Analysis of Cognitive Domain

Difficulty Index (p) Range	Number of Items	Item Numbers	Category
0.00 - 0.20	2	18, 21	Very Difficult
0.21 - 0.40	7	13, 14, 15, 16, 24, 29, 30	Difficult
0.41 - 0.60	13	1, 2, 5, 6, 8, 10, 12, 19, 22, 25, 26, 27, 28	Moderate
0.61 - 0.80	6	3, 9, 11, 17, 20, 23	Easy
0.81 - 1.00	2	4, 7	Very Easy

As shown in Table 6, the cognitive items were distributed across difficulty categories, with 2 items classified as very difficult, 7 as difficult, 13 as moderate, 6 as easy, and 2 as very easy. This distribution indicates that most cognitive items were concentrated in the moderate difficulty range, suggesting that the instrument was generally well matched to the ability level of junior secondary school students. This finding is consistent with (Kurniawan et al., 2024), who emphasize that well-designed assessment instruments should exhibit appropriate levels of item difficulty to support accurate measurement. When this condition is met, assessment results are more likely to reflect students' actual performance (Ahmed & Ishtiaq, 2021). Items with moderate difficulty are particularly valuable in educational assessment because they provide meaningful information about students' conceptual understanding and enhance score differentiation (Mura et al., 2025; Rabiatal Adawiah, 2020).

However, the presence of very easy and very difficult items indicates that some items may have been too simple or too challenging for the target population, thereby offering limited diagnostic value (Gul et al., 2022; Ibrahim & Ismail, 2025). These findings highlight the necessity of empirical item analysis to identify items that may not function optimally in practice. The results of the item discrimination analysis are summarized in Table 7.

Table 7. Item Discrimination Analysis of Cognitive Domain

Discrimination Index ( <i>rpbis</i> ) Range	Number of Items	Item Numbers	Category
0.40 - 1.00	8	4, 8, 9, 10, 11, 23, 25, 27	Good
0.30 - 0.39	2	1, 13	Moderate
0.20 - 0.29	15	2, 3, 6, 7, 14, 15, 16, 17, 18, 19, 21, 24, 26, 29, 30	Fair
< 0.20	5	5, 12, 20, 22, 28	Poor

The analysis revealed that 8 items demonstrated good discrimination, 2 items showed moderate discrimination, 15 items had fair discrimination, and 5 items exhibited poor discrimination. These findings indicate variability in item quality across the cognitive domain. In this study, 25 (83%) items had fair to good discriminating power ( $DI \geq 0.20$ ), which were effective in differentiating between students with higher and lower levels of energy literacy (Kumar et al., 2021; Ridwan et al., 2021). Additionally, five items (17%) demonstrated poor discrimination indices ( $DI < 0.20$ ), indicating that they failed to differentiate between students with different ability levels and were consequently excluded from the instrument (Karim et al., 2021; Quaigrain & Arhin, 2017). In accordance with the predetermined criteria, cognitive items with poor discrimination indices and extreme difficulty levels were identified for revision and subsequently removed to enhance overall test quality.

From a classical test theory perspective, high-quality cognitive items are expected to exhibit moderate difficulty levels and sufficient discrimination power to effectively differentiate learners and enhance measurement precision

(Rohmatdi et al., 2024; Vincent & Shanmugam, 2020). The overall pattern observed in this study is consistent with these theoretical expectations, providing empirical support for the psychometric soundness of the refined cognitive domain of the energy literacy instrument.

**Item Validity of the Affective and Behavioral Domain**

The affective and behavioral domains of the energy literacy instrument were analyzed to examine the validity of Likert-scale items using item-total correlation coefficients (Pearson correlation). This analysis aimed to evaluate the extent to which each item was consistent with its respective construct within the multidimensional framework of energy literacy. The results of the item-total correlation analysis for the affective and behavioral domains are presented in Table 8.

Table 8. Item-Total Correlation Analysis of the Affective and Behavioral Domains

Item-Total Correlation Range (r)	Affective Domain		Behavioral Domain		Category
	Number of Items	Item Numbers	Number of Items	Item Numbers	
0.50 – 1.00	13	1, 2, 3, 5, 6, 9, 11, 12, 15, 16, 17, 18, 20	3	5, 6, 7	Good
0.30 – 0.49	2	8, 14	7	1, 8, 10, 11, 12, 13, 14	Moderate
0.20 – 0.29	0	-	0	-	Fair
< 0.20	2	4, 19	2	4, 15	Poor

For the affective domain, most items (88%) showed good to moderate item-total correlations, with only a small number exhibiting poor correlations. Similarly, in the behavioral domain, 10 (83%) items demonstrated acceptable correlations, while two items showed weak alignment with the total score. Overall, most affective and behavioral items exhibited positive and acceptable item-total correlation values, indicating consistency with their intended constructs (Cai et al., 2024). In accordance with the predetermined criteria, items with item-total correlation values below 0.30 ( $r < 0.30$ ) were removed to enhance construct clarity and measurement accuracy.

Item-total correlation is widely used to evaluate the consistency of items in polytomous instruments that assess attitudes and behaviors (Izah et al., 2023). The majority of items in this study demonstrated acceptable correlations, supporting the conceptualization of energy literacy as a multidimensional construct that includes affective and behavioral components alongside cognitive understanding.

**Reliability of the Energy Literacy Instrument**

Internal consistency reliability analysis was conducted to evaluate the stability and consistency of the energy literacy instrument across domains after the removal of poorly performing items. Reliability analysis aimed to determine whether the retained items functioned coherently in measuring energy literacy

among junior secondary school students. The results of the internal consistency reliability analysis are presented in Table 9.

Table 9. Reliability Analysis of the Energy Literacy Instrument

Domain	Reliability Coefficient ( $\alpha$ / KR-20)	Category
Cognitive	0.637	High
Affective	0.857	Very High
Behavioral	0.695	High
Overall Instrument	0.855	Very High

The reliability analysis results presented in Table 9 indicate that the energy literacy instrument demonstrated high to very high internal consistency across all domains. The cognitive domain showed a high reliability coefficient ( $\alpha = 0.637$ ), while the affective and behavioral domains exhibited high to very high reliability coefficients ( $\alpha = 0.857$  and  $\alpha = 0.695$ , respectively). The overall instrument reliability coefficient ( $\alpha = 0.855$ ) indicates that the instrument produces consistent result when carried out by different individuals, across different testing times, using alternative test forms, and across variations in item construction (Ningsih et al., 2024). This consistency suggests that the instrument is stable and dependable in measuring students' energy literacy, as variations in administration conditions do not substantially affect the measurement outcomes.

In line with classical test theory, the removal of poorly performing items contributed to improved internal consistency reliability of the instrument (Chyung et al., 2018; Haroz et al., 2020). These finding highlights that reinforcing the role of reliability analysis as a critical step in the empirical evaluation of educational assessment instruments. Collectively, these results provide empirical support for the use of the refined energy literacy instrument for educational assessment and research purposes at the junior secondary school level.

Overall, this study provides an overview of the empirical validation procedures used to evaluate the quality of an energy literacy instrument for junior secondary school students. The results offer empirical support for the use of the instrument as a measurement tool in classroom-based energy education and educational research contexts. Educators may utilize the instrument to assess students' energy literacy across cognitive, affective, and behavioral domains, while researchers may apply it in studies related to energy education and sustainability. Nevertheless, this study has limitations, particularly with respect to the sample size and school representation. Further empirical testing involving larger samples and more diverse educational contexts is recommended to strengthen the generalizability of the findings and to support broader application of the instrument.

## CONCLUSIONS AND RECOMMENDATIONS

This study presents the results of the empirical validation testing of an energy literacy instrument designed for junior secondary school students. The findings indicate that the retained items across the cognitive, affective, and behavioral domains met the established psychometric criteria. The instrument demonstrated acceptable validity and reliability, with appropriate levels of item difficulty, discrimination, and internal consistency. These results suggest that the refined instrument functions effectively in measuring students' energy literacy. Therefore, this energy literacy instrument can be used by educators and researchers as a representative and high-quality tool for assessing students' understanding, attitudes, and behaviors related to energy use.

As a recommendation, future studies are encouraged to implement this instrument with larger and more diverse student populations to further strengthen its generalizability. Additional empirical testing using advanced psychometric approaches, such as factor analysis or item response theory, may also be conducted to examine the internal structure of the instrument in greater depth. Furthermore, the instrument may be adapted and applied in different educational contexts or integrated into instructional evaluation to support the development of effective energy education programs

## REFERENCES

- Ahmed, I., & Ishtiaq, S. (2021). Reliability and Validity: Importance in Medical Research. *Journal of the Pakistan Medical Association*, 71(10), 2401–2406.
- Akitsu, Y., & Ishihara, K. N. (2018). An Integrated Model Approach: Exploring the Energy Literacy and Values of Lower Secondary Students in Japan. *International Journal of Educational Methodology*, 4(3), 161–186.
- Cai, Y., Li, Y., Zou, J., Zhang, Jie, Luo, W., Zhang, Jingping, & Qu, C. (2024). Cross-cultural adaptation and reliability of the inventory of vicarious posttraumatic growth and research of its influencing factors: a cross-sectional study. *BMC Nursing*, 23(1), 763.
- Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clinical Therapeutics*, 36(5), 648–662.
- Chen, K. L., Huang, S. H., & Liu, S. Y. (2013). Devising a framework for energy education in Taiwan using the analytic hierarchy process. *Energy Policy*, 55, 396–403. <https://doi.org/10.1016/j.enpol.2012.12.025>
- Chyung, S. Y. Y., Barkin, J. R., & Shamsy, J. A. (2018). Evidence-Based Survey Design: The Use of Negatively Worded Items in Surveys. *Performance Improvement*, 57(3), 16–25. <https://doi.org/10.1002/pfi.21749>

- DeWaters, J. E., & Powers, S. E. (2011). Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior. *Energy Policy*, 39(3), 1699–1710.
- DeWaters, J., & Powers, S. (2013). Establishing Measurement Criteria for an Energy Literacy Questionnaire. *Journal of Environmental Education*, 44(1), 38–55. <https://doi.org/10.1080/00958964.2012.711378>
- Fiskawarni, T. H., Rahmawati, R., Widiasih, W., & Saad, R. (2024). The Design and Validation of the Four-Tier Test Instrument for Energy Literacy Using the Rasch Model Analysis. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(1), 74.
- Gul, N., Shagufta, S., & Parveen, S. (2022). Item Difficulty in Item Analysis of Intelligence Test Items. *Pakistan Social Sciences Review*, 6(II).
- Guyer, R., & Thompson, N. A. (2013). *User's Manual for Iteman 4.3*. Assessment Systems Corporation.
- Haroz, E. E., Kane, J. C., Nguyen, A. J., Bass, J. K., Murray, L. K., & Bolton, P. (2020). When less is more: reducing redundancy in mental health and psychosocial instruments using Item Response Theory. *Global Mental Health*, 7, e3. <https://doi.org/10.1017/gmh.2019.30>
- Hartelt, T., & Martens, H. (2025). Development and validation of an instrument measuring students' self-efficacy regarding explaining evolutionary changes. *Journal of Biological Education*, 1–35.
- Hendinata, L. K., Ardiwinata, T., & Pratama, F. K. T. (2022). The Role of Energy Literacy in Supporting Energy Conservation: Perspective from Indonesian Citizens. *Indonesian Journal of Energy*, 5(2).
- Hu, J.-L., & Yang, P.-S. (2024). Interactive Cycles between Energy Education and Energy Preferences: A Literature Review on Empirical Evidence. *Energies*, 17(20), 5092. <https://doi.org/10.3390/en17205092>
- Huda, N., & Wahyuni, T. S. (2020). PENGGUNAAN APLIKASI ITEM AND TEST ANALYSIS (ITEMAN) PADA SOAL TRY OUT UN IPA SMP TAHUN 2019. *Jurnal Pembelajaran Sains*, 4(1), 1–8.
- Ibrahim, S. E., & Ismail, M. R. (2025). Item Analysis of Information Literacy Test for Open, Distance and Digital Education New Learners. *E-Learning Innovations Journal*, 3 (2), 31–54.
- Izah, S. C., Sylva, L., & Hait, M. (2023). Cronbach's Alpha: A Cornerstone in Ensuring Reliability and Validity in Environmental Health Assessment. *ES Energy & Environment*, 23. <https://doi.org/10.30919/esee1057>

- Kalkbrenner, M. T. (2021). A Practical Guide to Instrument Development and Score Validation in the Social Sciences: The MEASURE Approach. *Practical Assessment, Research, and Evaluation*, 26(1).
- Karim, S. A., Sudiro, S., & Sakinah, S. (2021). Utilizing test items analysis to examine the level of difficulty and discriminating power in a teacher-made test. *EduLite: Journal of English Education, Literature and Culture*, 6(2), 256. <https://doi.org/10.30659/e.6.2.256-269>
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85-S89.
- Kurniawan, A., Istiyono, E., & Naba, S. D. (2024). Item Quality Analysis of Physics Concept Understanding Test with Rasch Model. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(3), 474. <https://doi.org/10.26737/jipf.v9i3.5692>
- Mukherjee, S., & Kittur, J. (2025). Learning through cognitive, affective, and psychomotor domains: Understanding undergraduate engineering students' perspectives in the United States. *Engineering Education Review*, 3. <https://doi.org/10.54844/eer.2025.0916>
- Mura, K. B., Sujito, Taufiq, A., Irawan, I. D. A., & Basri, N. A. (2025). Empirical Validation of a Conceptual Understanding Test on Dynamic Electricity for High School Students. *Jurnal Penelitian Pendidikan IPA*, 11(11), 229-235.
- Ningsih, S. N., Istiyono, E., & Widiastuti. (2024). An Application of Classical Test Theory for Item Characteristic Analysis of Chemical Literacy Instruments. *Jurnal Pendidikan Kimia Indonesia*, 7(2), 58-68.
- Putri, I. R., Setiawan, A., & Nasrudin, D. (2022). Energy Literacy Profile of Vocational High School Teacher Candidates for Renewable Energy Engineering Expertise Program. *Jurnal Ilmiah Pendidikan Teknik Dan Kejuruan*, 15(2), 99. <https://doi.org/10.20961/jiptek.v15i2.67623>
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1301013.
- Rabiatul Adawiah. (2020). THE APPLICATION-BASED ANALYSIS OF QUESTIONS ITEM QUALITY IN JUNIOR HIGH SCHOOL. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 6(2), 140-148.
- Ridwan, M. R., Istiyono, E., & Widiastuti, W. (2021). Test Items Analysis of Mathematical Problem-Solving Ability using a Classical Test Theory Approach. *Jurnal Pendidikan MIPA*, 22(1), 98-111.

- Rohmatdi, A., Martono, M., Zafrullah, Z., & Safitri, R. (2024). Optimizing Question Quality in Junior High School Exams: Classical Test Theory Evaluation with ITEMAN 4.3. *AL-ISHLAH: Jurnal Pendidikan*, 16(4), 4363–4375. <https://doi.org/10.35445/alishlah.v16i4.5671>
- Setiadi, I. (2021). Empirical Test Of Content Validity And Construct Validity of the Mathematical Self-Efficacy Scale. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 10(2), 64–72.
- Sudaryono, Rahardja, U., Aini, Q., Isma Graha, Y., & Lutfiani, N. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1364(1), 012050. <https://doi.org/10.1088/1742-6596/1364/1/012050>
- Sukmawati, Sudarmin, & Salmia. (2023). DEVELOPMENT OF QUALITY INSTRUMENTS AND DATA COLLECTION TECHNIQUES. *Jurnal Pendidikan dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, 6(1), 119–124. <https://doi.org/10.55215/jppguseda.v6i1.7527>
- U.S. Department of Energy. (2017). *Energy Literacy: Essential Principles and Fundamental Concepts for Energy Education*.
- Usman, M., Pramudawardani, H., Umam, R., & Takahashi, H. (2025). Synthesizing A Framework and Establishing Content Validity of An Energy Literacy Instrument for Indonesian Students. *Islamic Journal of Integrated Science Education (IJISE)*, 4(3), 175–185.
- Usman, M., Suyanta, Pujianto, & Huda, K. (2021). Energy Literacy of Junior High School Students in Indonesia: A Preliminary Study. *Proceedings of the 6th International Seminar on Science Education (ISSE 2020)*, 609–614.
- Usman, M., Syakur, A., & Sangkala, N. R. (2025). IDENTIFIKASI PENCAPAIAN LITERASI ENERGI PESERTA DIDIK PADA PEMBELAJARAN BERBASIS STEM DENGAN BANTUAN SIMULASI PhET. *Jurnal Biogenerasi*, 10(4), 2354–2362.
- Vincent, W., & Shanmugam, S. K. S. (2020). The Role of Classical Test Theory to Determine the Quality of Classroom Teaching Test Items. *Pedagogia: Jurnal Pendidikan*, 9(1), 5–34.
- Yusup, M. (2021). Using the Rasch model for the development and validation of an energy literacy assessment instrument for prospective physics teachers. *Journal of Physics: Conference Series*, 1876(1), 012056.
- Zhang, L., Liu, X., & Feng, H. (2023). Development and validation of an instrument for assessing scientific literacy from junior to senior high

school. *Disciplinary and Interdisciplinary Science Education Research*,  
5(1), 21. <https://doi.org/10.1186/s43031-023-00093-2>