KPJ 9(2) (2025)

KAPPA JOURNAL
Physics & Physics Education

https://e-journal.hamzanwadi.ac.id/index.php/kpj/index

Validity and Reliability Analysis of Science Modules Using Problem-Based Learning Models Based on Ethno-STEM Assisted by Augmented Reality

Prelia Dwi Amanah[1*], Satutik Rahayu[2], Gunawan[3]

[2,3]Physics Education, Faculty of Teacher Training and Education, University of Mataram, Lombok, West Nusa Tenggara, Indonesia.
[1,2,3]Master of Science Education Program, University of Mataram, Lombok, West Nusa Tenggara, Indonesia.
[3]Doctor of Science Education Program, University of Mataram, Lombok, West Nusa Tenggara, Indonesia.

Received: 26 July 2025
Revised: 12 August 2025
Accepted: 15 August 2025

Corresponding Author:
Prelia Dwi Amanah
preliadwia@gmail.com

DOI:
https://doi.org/10.29408/kpj.v9i2.31825

Abstract: This study aims to analyze the validity and reliability of a Problem-Based Learning (PBL)-based science module integrated with the Ethno-STEM approach and assisted by Augmented Reality (AR) on the Motion and Force material for seventh-grade junior high school students. The module was developed as an innovative effort to improve the quality of science learning by linking local cultural contexts and the use of technology. The research method used includes validity and reliability testing involving seven expert validators, consisting of four lecturers and three science teachers, who assessed the feasibility of learning documents with five assessment scales, including the Learning Objective Flow (ATP), Learning Implementation Plan (RPP), Teaching Module, and Student Worksheet (LKPD). Validity was analyzed using Aiken's V formula, and the results of the analysis showed that all components obtained V values between 0.75 and 1.00, which was categorized as valid. Meanwhile, the inter-rater reliability test using the Intraclass Correlation Coefficient (ICC) showed an ICC value between 0.753 and 0.947. This indicates that the developed module has excellent inter-rater consistency and good to very good reliability. Therefore, this PBL-based science module, integrated with Ethno-STEM and AR, has been proven valid and reliable for use in the learning process at the junior high school level, making it a reliable learning resource.

Keywords: validity and reliability; science module; problem-based learning; ethno-STEM; augmented reality.

Introduction

The rapid development of information and communication technology has brought significant changes to various aspects of life, including education. In this increasingly dynamic era of education, a major challenge in science learning is the availability of modules that are not only informative but also engaging and capable of increasing student engagement. Many schools still struggle to access contextual and adaptive learning modules that meet student needs and current developments in educational technology. Science learning in junior high schools still faces numerous challenges related to the limited availability of contextual and innovative learning modules. Existing modules generally lack the ability to connect scientific concepts with local cultural contexts and utilize learning technology to optimally stimulate student interest in learning.

Successful learning is supported by teaching materials. Teaching materials also support the learning process (Setyaningsih et al. 2023). Modules are teaching materials that facilitate student interaction with the material presented and make them more active in the

How to Cite:
Amanah, P. D., Rahayu, S., & Gunawan, G. (2025). Validity and Reliability Analysis of Science Modules Using Problem-Based Learning Models Based on Ethno-STEM Assisted by Augmented Reality. *Kappa Journal, 9*(2), 213-222. https://doi.org/10.29408/kpj.v9i2.31825

learning process because they contain activities that engage students. A module is a type of learning media that is packaged comprehensively and systematically, containing a set of planned learning experiences designed to help students master specific learning objectives (Jusuf and Sobari 2021). Modules developed with the right approach can increase the efficiency and effectiveness of learning and facilitate students' deeper and more meaningful mastery of science concepts. A module is considered good and engaging if it exhibits self-instruction, self-directedness, self-containedness, adaptiveness, and user-friendly characteristics (Zendrato, Harefa, and Lase 2022).

Natural Science is a type of theoretical knowledge acquired through specific methods, namely through observation, experimentation, inference, theory formation, experimentation, observation, and so on (Suwartiningsih 2021). Natural Science is a branch of science that explores the regularities of nature and mastery of knowledge, including facts, concepts, principles, discovery processes, and scientific attitudes. Physics is one such branch (Gunawan 2015; Gunawan., Harjono, A. 2015; Amanah, Harjono, and Gunada 2017). In the context of science education, students are faced with complex challenges that require mastery of high-level thinking skills and the ability to integrate various representations to understand abstract material and solve scientific problems comprehensively (Amanah, Martin, and Rokhmat 2025).

The development of science learning modules based on Problem-Based Learning (PBL) is a crucial innovation in increasing student engagement and understanding of scientific concepts (Rahayu, Verawati, and Islamiah 2019). PBL encourages students to develop critical thinking skills, solve problems independently, and apply knowledge in real-world contexts (Hmelo-Silver 2004). This approach becomes even more relevant when combined with the integration of modern technologies such as Augmented Reality (AR), which can provide interactive and immersive learning experiences (Tüysüz, 2016).

Problem-Based Learning (PBL) was defined by Howard Barrows and his team, as an early PBL design at McMaster University Medical School in Canada in 1970. It is a student-centered teaching approach that uses real-world problems and their solutions as learning stimuli for students in small groups (Barrows 1996). PBL is suitable for helping students develop higher-order thinking skills (Rahayu and Al Hadi 2023), problem-solving skills, and attitudes needed in real life, such as being active, independent, and cooperative (Aryulina and Riyanto 2016).

The use of AR technology in learning modules is expected to enhance students' learning experiences by providing more engaging and interactive visualizations. AR can help students understand abstract scientific concepts in a more concrete way, making learning more enjoyable and effective (Muktiani et al. 2022). Researchers hope that in this way, the developed science module will not only function as a learning aid, but also as a medium that can improve students' critical thinking skills and digital literacy.

The integration of local wisdom values through the Ethno-STEM approach also adds a new dimension to learning. Linking learning materials to local culture, such as Banyumulek pottery production, has the effect of enabling students to not only learn theory but also understand the relevance of science within their cultural context. This is expected to increase their motivation and engagement in learning, as well as strengthen their cultural identity amidst globalization (Idrus 2022). Furthermore, the application of the Ethno-STEM approach in science modules aims to link science, technology, engineering, and mathematics learning with local culture. This is believed to increase the relevance of learning, enabling students to connect scientific concepts with their own cultural values (Sleeter 2001). This approach also strengthens contextual learning, supporting the development of an appreciation for cultural heritage and scientific creativity (Yusuf, M., & Iskandar 2019).

Several previous studies have examined the effectiveness of PBL in science learning, demonstrating improvements in students' learning outcomes and critical thinking skills (Belland, B. R., Glazewski, K. D., & Richardson 2008). However, research integrating PBL with Ethno-STEM and AR technology is still relatively rare. Therefore, developing modules that combine these three aspects is crucial to provide new contributions to science learning practices (Bakar and Ismail 2020).

Validity and reliability are crucial aspects in developing research instruments, ensuring that the measuring instrument consistently and accurately measures what it is intended to measure. This ensures that the data obtained can be trusted and used for valid analysis (Amanah, P. D., Al Qayyim, T. M., Jauhari, M. T., Muliati, E., & Jufri 2025). The validity and reliability of modules as learning tools must be tested to ensure they meet quality standards and can be trusted as learning resources. Validity relates to the appropriateness of the content and presentation of the material in accordance with the learning objectives, while reliability measures the consistency of assessments between validators. The use of the Aiken's V method for validity and the Intraclass Correlation Coefficient (ICC) for inter-rater reliability have been widely applied in research on the development of learning instruments (Polit, D. F., & Beck 2006).

The main problem examined in this study is the validity and reliability of the PBL-based science module

integrated with Ethno-STEM and AR in supporting learning for seventh-grade junior high school students. This study focuses on the Motion and Force material, with assessments by seven expert validators consisting of lecturers and science teachers. This is crucial to ensure the module is not only theoretically effective but also practically applicable in the classroom (Creswell, J. W., & Creswell 2014). The primary objective of this study is to analyze the validity and reliability of the PBL-based Ethno-STEM-assisted science module with AR. With strong validity and reliability, this module is expected to be a credible and innovative learning resource to improve the quality of science learning in junior high schools (Gay, L. R., Mills, G. E., & Airasian 2012). The results of this study are expected to provide practical and academic contributions in the development of science learning media and modules that are adaptive to technological developments and local wisdom. In addition, this article also serves as a reference for researchers and education practitioners in developing contextual STEM-based modules that are based on the latest technology.

Method

This quantitative study aimed to test the validity and reliability of an AR-assisted Ethno-STEM-based PBL science learning module. A questionnaire was used as a measuring instrument, validating the Learning Objective Flow (ATP), Lesson Plan (RPP), Teaching Module, and Student Worksheet (LKPD). The designed questionnaire was then subjected to validity and reliability tests. This approach was chosen to ensure the developed module met academic and technical quality standards as an applicable learning medium.

The study subjects consisted of seven expert validators randomly selected based on different expertise criteria. The validators assessed the appropriateness of the content, presentation, and language. The instrument used was a validity questionnaire consisting of 31 statements (for each validation instrument) developed based on the guidelines of the National Education Standards Agency (BNSP). The instrument covered several assessment aspects, such as content suitability, material completeness, language clarity, and the technical quality of AR usage. All were measured on a 1-5 Likert scale to facilitate quantitative data analysis. The data collection procedure began with the creation of a module draft that integrated the topic of Motion and Force with content enrichment based on local cultural values through an Ethno-STEM approach and the inclusion of 3D visual materials using AR. The draft was then distributed to validators for assessment and feedback was collected via an online questionnaire.

Validity analysis was measured using Aiken's V validity coefficient. The Aiken's V value for 7 raters on a 5-point scale was 0.75. Therefore, it is considered valid and scientifically acceptable when the V value is greater than or equal to 0.75 (Aiken 1985). The formula used is according to Aiken (1985, as cited in Arafani, Mahrus, and Zulkifli 2025).

$$V = \frac{\sum s}{[\, n\,(c-1)]} \quad (1)$$

V = Expert agreement index
Σs = Total score
S = $r - L_0$
R = Value given by respondent experts
L_0 = Lowest validation assessment value
n = Number of experts/respondents
c = Highest validity assessment value

The obtained validity percentage data can then be matched against the validity criteria as stated in (Utaminingsih et al. 2024). The data is then interpreted based on the validity criteria (feasibility) using Aiken's V table. The validity level is determined based on Table 3.1 below.

Table 1. Validity Level (Aiken 1985)

Value Range	Validation Level
$V \geq 0.75$	Valid
$V < 0{,}75$	Invalid

The results of this analysis are used to determine aspects of the module that need improvement. To test the reliability between validators, the Intraclass Correlation Coefficient (ICC) method is used with a two-way random effects model that measures the consistency of assessments between assessors. The two-way random ICC (2, k) model is used because the assessors have different backgrounds. The ICC (2, k) model provides an overview of reliability at the combined average level of several assessors so it is usually higher and more stable (Shrout and Fleiss 1979). When each sample of subjects is assessed on a quantitative scale by two or more observers, agreement is usually measured by the intraclass correlation coefficient (Fleiss, Joseph 1975). The intraclass correlation coefficient, defined as the ratio of variances, can be interpreted as a correlation coefficient (Bartko 1966). The formula used to calculate ICC (2, k) is as follows (Shrout and Fleiss 1979):

$$ICC = \frac{BMS - EMS}{BMS + \frac{(JMS - EMS)}{n}} \quad (2)$$

Description:
ICC = Interclass correlation coefficient/inter-rater reliability

BMS = Variance between assessed targets (between-targets mean square)
EMS = Residual variance of the measurement (error mean square)
JMS = Variance between raters in assessing objects (judge mean square)
k = Number of raters providing ratings
n = Number of objects or targets assessed

An ICC value above 0.75 is categorized as high reliability, indicating a consistent and trustworthy assessment (Polit, D. F., & Beck 2006). In this study, an ICC value less than 0.51 indicates poor reliability, a value between 0.51 and 0.75 indicates fair reliability, a value between 0.76 and 0.90 indicates good reliability, and a value greater than 0.91 indicates excellent reliability (Koo and Li 2016; Giuseppe 2018; Portney, L.G. and Watkins 2000). The ICC value criteria can be seen in Table 2 below (Portney, L.G. and Watkins 2000):

Table 2. Reliability Criteria Using ICC

ICC	Category
0,00 – 0,50	Poor Reliability
0,51 – 0,75	Fair (Moderate) Reliability
0,76 – 0,90	Good Reliability
0,90 – 1,00	Excellent Reliability

Data processing and analysis were carried out using Microsoft Excel and IBM SPSS Statistics 22.0 for Windows software, which facilitates the calculation of the Aiken's V index and ICC quickly and accurately. Microsoft Excel was used to find validity by calculating the Aiken's V index. IBM SPSS Statistics 22.0 for Windows was used to measure reliability by calculating the two-way random ICC. Evaluation of the analysis results then became the basis for the final revision of the module to achieve the quality standards of innovative and contextual science learning. With this method, it is expected to produce a valid and reliable PBL-based science module, as well as being applicable to improve the quality of science learning at the junior high school level by utilizing local cultural richness and current learning technology.

Result and Discussion

The results of this study comprehensively present data on the validity and reliability of various learning instruments developed within the context of a PBL-based science module with an Ethno-STEM approach modified with the aid of AR technology. The validated instruments include the Learning Objectives Flow (ATP), the Lesson Implementation Plan (RPP), the science module as the main teaching material, and the Student Activity Sheet (LKPD), all contained in a single teaching module. The module is intended for junior high school students and covers material on motion and force. The module uses PBL model syntax, is based on ethno-STEM (ethno-STEM related to Banyumulek pottery), and utilizes 3D AR technology (created using the *Assemblr Edu* application). The instruments were validated by seven experts: four lecturers and three junior high school science teachers.

Module Validity

Validity testing using the Aiken's V coefficient provides a strong indication of the appropriateness and adequacy of the content, presentation, and language used in each instrument. Reliability testing ensures the consistency of the instrument's measurements across a variety of conditions, increasing confidence in the instrument's overall reliability. The average Aiken's V value exceeded 0.80 for all instruments, thus concluding that all the learning products studied met excellent validity standards for use as learning media and resources. Validity is a measure of the accuracy or validity of a measuring instrument in measuring what it is intended to measure (Arikunto 2010). A high Aiken's V value indicates that the instrument has strong validity, accurately and relevantly reflecting learning objectives.

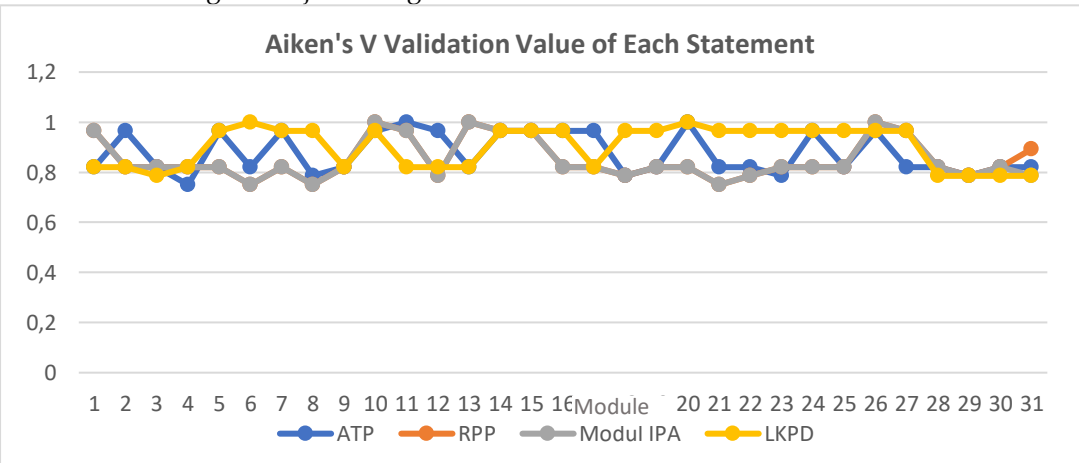

Figure 1. Validation Results of ATP, RPP, Science Module, and LKPD for Each Statement

In terms of ATP validity, the evaluation results showed an average Aiken's V score of 0.88, clearly confirming the valid category. A detailed analysis of each validation statement showed that the majority of statements received a perfect score of 1.00, indicating strong agreement among validators that the ATP was designed systematically, clearly, and effectively directed the learning stages in accordance with the desired objectives. Although there were several indicators with a lower validity score of 0.75, this was still included in the valid category, indicating that the ATP was generally good and only needed minor adjustments or refinements for optimal implementation. The ATP received the lowest validation score (0.75) for statement number 4, related to the accuracy of learning outcomes and objectives. This is in line with validator input that some learning objectives should be separated into two parts to be more specific and have a clearer focus. Meanwhile, the highest validity score (1.00) was for statements number 10 and 20, which discuss contextual learning objectives and student engagement. This indicates that the ATP created has contextual learning objectives and will create a high level of student engagement.

Table 3. Validation of Learning Objective Flow (ATP)

Aspect	V	Category
Content Validity	0.89	Valid
Validity of Presentation	0.89	Valid
Language Validity	0.85	Valid
Average	**0.88**	**Valid**

The measured content validity aspect obtained a value of 0.89, the presentation aspect obtained a value of 0.89 and language validity of 0.85 confirming that the RPP is not only complete in terms of content, but also compiled and communicated in clear, easy-to-understand language and in accordance with the modern learning context. The clarity and systematicity of this ATP are very important because they are the initial foundation for developing a directed and measurable learning process, so that the expected learning outcomes can be achieved effectively. This is in accordance with the principles of constructivist learning which encourage active participation of students (Piaget 1972).

Table 4. Results of Validation of the Learning Implementation Plan (RPP)

Aspect	V	Category
Content Validity	0.85	Valid
Validity of Presentation	0.85	Valid
Language Validity	0.86	Valid
Average	**0.85**	**Valid**

The validity calculation results for the Lesson Implementation Plan (RPP) indicate that this document meets standards, with an average Aiken's V score of 0.85. A breakdown of the scores shows that several aspects of the RPP received a perfect score of 1.00, while several others remained at the minimum score of 0.75, indicating room for improvement but not doubting the overall quality of the RPP. Statements 6, 8, and 21 received the lowest score of 0.75, addressing time allocation, terminology, and the interrelationships between activities. This aligns with the validator's input, which stated that detailed time allocation per meeting should be explicitly stated in the Identity (General Information) section for easy reference. The highest scores (1.00) were found for statements 13 and 20, which discuss the sequence of learning activities and student engagement.

The measured content validity aspect received a score of 0.85, implying that the material contained in the RPP contains important elements and is theoretically comprehensive. The presentation of the material, which achieved a score of 0.85 and a language validity of 0.86, confirms that the lesson plan is not only comprehensive in terms of content, but also structured and communicated in clear, easy-to-understand language, and appropriate for modern learning contexts. This demonstrates that the lesson plan is ready for educators to effectively and efficiently guide the teaching and learning process, meeting the needs of the 21st century.

Table 5. Science Module Validation Results

Aspect	V	Category
Content Validity	0.85	Valid
Validity of Presentation	0.85	Valid
Language Validity	0.85	Valid
Average	**0.85**	**Valid**

The evaluation of the science module also demonstrated very satisfactory validity results, with an average Aiken's V score of 0.85, indicating that the module is considered valid for use as the primary learning medium in science instruction. The science module in question includes a module description, a user guide, learning activities, and a summary and glossary. The module's content, presentation, and language aspects received a relatively consistent score of 0.85, indicating that each aspect has been carefully considered and ensures that the teaching content is presented in an engaging, logical, and easily understood manner by students.

Several indicators, including statements 6, 8, and 21, still meet the minimum validity threshold of 0.75. The statements with the lowest scores address the accuracy of examples and cases, terminology, and the interconnectedness of learning activities. This score

provides researchers with an opportunity to revise the module to enhance the interactivity and relevance of the material, further supporting innovative and contextual learning approaches, particularly in the use of AR technology and the application of Ethno-STEM. This aligns with input from the validator, such as the need for improvements in the consistency of terminology. The highest validity score, 1.00, was found in items 10, 13, and 26, which discuss examples and cases, conceptual coherence, and understanding of messages and information. This aligns with comments from validators, who stated that the module is generally good and can be used in research. Other validators stated that the module is ready for use in the next research phase.

Table 6. Validation Results of Student Worksheets (LKPD)

Aspect	V	Category
Content Validity	0.88	Valid
Validity of Presentation	0.94	Valid
Language Validity	0.88	Valid
Average	**0.90**	**Valid**

The Student Activity Sheet (LKPD) has the highest validity advantage of all the instruments tested, namely with an average Aiken's V value of 0.90. This value reflects that the LKPD is not only appropriate in terms of content and presentation, but also very effective in facilitating active and independent student learning activities. With a content validity score of 0.88, presentation of 0.94, and language validity of 0.88, the LKPD was successfully designed holistically to support the problem-based learning process and the integration of the STEM learning process contextually using AR assistance. This indicates that the LKPD is very suitable for stimulating student engagement in the learning process, honing critical thinking skills, and strengthening the understanding of science concepts in an applicable manner. The quality of this LKPD is important in meeting the needs of adaptive and innovative learning today. The LKPD validity value is quite high compared to other instruments because the lowest value is 0.78 while the highest is 1.00. The highest values are found in statements number 6 and 10, which discuss the accuracy of the case and using the case in everyday life. This is because the student worksheet (LKPD) created incorporates Banyumulek ethno-pottery content and is contextualized for students. Nearly all activities in the LKPD include local-themed cases and are contextualized to students' daily lives.

The successful validation opens up opportunities for further development of similar measuring tools and teaching materials in other subjects, using platforms that adapt to modern technology and a student-centered educational approach. This module allows for more standardized treatment in pedagogical development that can be widely adopted. The instrument is able to consistently measure the variables studied without being affected by assessment variations, ensuring scientifically credible results (Koo and Li 2016).

Module Reliability

Reliability is a crucial aspect in the development of research and learning instruments because it ensures the consistency and stability of the measurement results provided by the instrument (Creswell, J. W., & Creswell 2014). The reliability of all tested instruments confirms that the measurement tool is highly consistent and can be relied upon to measure the intended aspects continuously and consistently. This reliability testing strengthens the validity of the content and presentation, ensuring that the measurement results are not due to chance or variation, but rather reflect the instrument's stable quality across various testing conditions and across different groups of evaluators. Learning instruments that have been tested for reliability provide a strong foundation for repeated use and broader adaptation in science education contexts using PBL models, Ethno-STEM-based learning, and AR technology. Reliability is a coefficient that reflects the consistency of a measuring instrument's measurement results, so a reliable instrument will produce stable and consistent scores when retested (Anastasi, A., & Urbina, 2017). In the context of developing learning instruments, reliability is a crucial aspect that ensures that the evaluation results of instruments such as the Learning Objectives Flow (ATP), Lesson Plans (RPP), Science Modules, and Student Worksheets (LKPD) are not merely coincidental but reflect reliable quality.

In this study, reliability was measured using the Intraclass Correlation Coefficient (ICC) using IBM SPSS Statistics 22.0 for Windows, a popular statistical method for assessing consistency between assessors or continuous measurements. The ICC provides an indication of the level of agreement or correlation between repeated measurements, which ideally should be in the "good" or "excellent" category (Portney, L.G. and Watkins 2000; Koo and Li 2016). Based on data processing, the five instruments tested—ATP, Lesson Plans (RPP), Learning Modules, Student Worksheets, Critical Thinking Skills, and Digital Literacy—showed overall ICC values ranging from 0.75 to >0.90, categorizing them as good to excellent. This indicates that the instrument is capable of producing consistent data when assessed by different validators or in repeated measurement situations.

Table 7. Results of Reliability Calculations Using SPSS

Instrumen	ICC	Kategori	ICC (Tiap Aspek)		Kategori (Tiap Aspek)
ATP	0.847	Reliabilitas Baik (Good)	Isi	0.847	Reliabilitas Baik (Good)
			Penyajian	0.845	Reliabilitas Baik (Good)
			Bahasa	0.836	Reliabilitas Baik (Good)
RPP	0.788	Reliabilitas Baik (Good)	Isi	0.753	Reliabilitas Baik (Good)
			Penyajian	0.812	Reliabilitas Baik (Good)
			Bahasa	0.924	Reliabilitas Sangat Baik (Excellent)
Modul	0.79	Reliabilitas Baik (Good)	Isi	0.753	Reliabilitas Baik (Good)
			Penyajian	0.812	Reliabilitas Baik (Good)
			Bahasa	0.947	Reliabilitas Sangat Baik (Excellent)
LKPD	0.851	Reliabilitas Baik (Good)	Isi	0.81	Reliabilitas Baik (Good)
			Penyajian	0.777	Reliabilitas Baik (Good)
			Bahasa	1	Reliabilitas Sangat Baik (Excellent)

The reliability of each aspect—content, presentation, and language—was also analyzed to provide a more detailed picture of the quality of each section of the instrument. In the ATP section, the ICC value for the content aspect was 0.847, the presentation aspect 0.845, and the language aspect 0.836—all of which fall into the "good reliability" category according to ICC interpretation standards. In the Lesson Plan (RPP), the language aspect even demonstrated very good reliability with a value of 0.924, indicating that the language aspect of the instrument was highly consistent across validators. The Module and Student Worksheet (LKPD) also showed a similar pattern, with the language aspect receiving the highest ICC value, reaching 0.947 (module) and a perfect 1.00 (LKPD), indicating that the uniformity of language assessment was well maintained.

Reliability can be maintained or improved by ensuring that instrument items are interconnected and accurately represent the constructs intended to be measured. The high language reliability value in this instrument indicates that the language used in the measurement tool is easy to understand, consistent, and standardized, thus reducing ambiguity in assessment. Meanwhile, the content and presentation aspects have been systematically designed to maintain consistency across measurements. This is crucial in learning instruments that must accommodate tolerance between assessors or an unbiased evaluation approach. Furthermore, high internal consistency in these instruments provides confidence that the achieved validity also has a strong basis in measurement. Reliability and validity together build the credibility of a measurement instrument; if an instrument is valid but unreliable, the data obtained cannot be used as a strong reference. Conversely, good reliability supports validity because it indicates that the instrument can produce stable and reliable measurements across time and conditions.

Variations in reliability values between aspects indicate that although the instrument is generally of good quality, special attention needs to be paid to aspects with relatively lower reliability values for continued improvement. For example, the specific content aspect of the lesson plan and module has an ICC value of around 0.75, although considered good, indicating room for improvement to increase measurement consistency. Reliability above 0.90 is highly desirable for instrumentation used for important decisions or widespread dissemination (often used in medical research). Therefore, the findings of good to very good reliability prove that the instrument developed has met the criteria suitable for widespread use in the world of education.

Validity and Reliability Analysis

The next discussion focuses on how the validity and reliability of the instruments compare. Based on the validity and reliability data obtained from the ATP, RPP, Module, and LKPD instruments, all instruments demonstrated high validity values, ranging from 0.85 to 0.9, indicating that each instrument was able to measure what it was supposed to measure very well. Furthermore, the reliability values for each instrument also demonstrated good reliability, with scores ranging from 0.788 to 0.851, indicating consistency and stability of measurement results across time and conditions. Specifically, the LKPD had the highest validity value of 0.9 and the best reliability of 0.851, indicating that this instrument was the strongest in terms of accuracy and reliability compared to the others. Therefore, it can be concluded that all instruments are valid and reliable, indicating that the data collected using these instruments is suitable for further analysis in research.

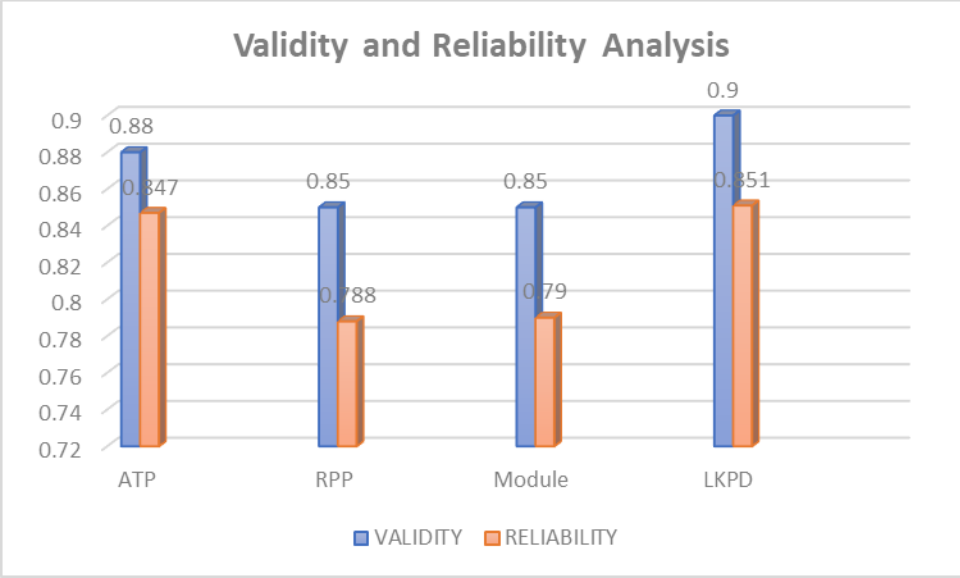

Figure 2. Comparison of Instrument Validity and Reliability Values

Based on validity and reliability data, the Student Worksheet (LKPD) instrument demonstrated the highest performance in both validity and reliability. The overall validity score for the LKPD reached 0.90, with the presentation aspect (0.94) the highest among the other instruments, and perfect language reliability (ICC = 1.00) categorized as Excellent. This indicates that the LKPD has optimal consistency and accuracy in measuring learning aspects and delivering material clearly and accurately. In contrast, the Lesson Plan (RPP) instrument had a slightly lower validity score, at 0.85, and the lowest reliability score for the content aspect (ICC = 0.753), although still in the "Good" category. This indicates that while the RPP is suitable for use, there is room for improvement in the content aspect to increase the instrument's accuracy and reliability.

Meanwhile, the ATP and Module instruments had relatively balanced quality, with validity scores ranging from 0.85 to 0.88 and reliability also in the "Good" category. The module excelled in its language aspect, with a reliability score of 0.947, which falls into the excellent category. However, the content score for this instrument was the lowest compared to the other instruments (0.753). The ATP had consistent scores across all aspects, both validity and reliability, but did not reach the peak of the LKPD. In conclusion, the LKPD excelled as the instrument with the best validity and reliability performance, while the RPP showed the lowest score and warrants further improvement to support more valid and reliable learning evaluations.

Reliability testing using the ICC in this study convincingly demonstrated that the various learning instruments developed had a high level of measurement stability. The combination of instrument development theory (validity and reliability) and pedagogical approaches (PBL, Ethno-STEM, AR technology) in this study not only empirically demonstrated the product's robustness but also paved the way for interdisciplinary learning innovation in the 21st century. This approach aligns with 21st-century learning theory that supports the development of critical thinking, collaboration, and problem-solving skills in science education, as well as the use of technology as a learning medium (Rahayu and Artayasa 2025). Practically, high validity and reliability enable educators to adopt this product as an effective learning medium and resource that is adaptive to technological developments and student needs. This reinforces the concept of learning that facilitates intrinsic motivation and active student engagement, as expressed in Self-Determination Theory (Deci, E. L., & Ryan 1985). Based on the analysis results from 7 diverse validators, the study successfully produced a science learning instrument that meets validity and reliability standards, making it ready to be used as a solution in learning.

Conclusion

This study successfully demonstrated that the Problem-Based Learning (PBL) science module, combined with the Ethno-STEM approach and supported by Augmented Reality (AR) technology, has high validity and reliability. Validity testing using the Aiken's V coefficient indicates that all aspects of the module, from the Learning Objectives Flow (ATP), Lesson Plan (RPP), teaching materials module, to Student Activity Sheets (LKPD), meet the valid category with an average value above 0.80 for content, presentation, and language. This indicates that the module has been systematically and comprehensively designed to meet academic standards

and science learning needs. In addition to validity, reliability testing using the Intraclass Correlation Coefficient (ICC) confirmed the high consistency and stability of the measurement instruments, with categories ranging from good to excellent. Consistent ICC values across all instruments strengthen confidence in the module's accuracy in measuring the achievement of learning objectives and support its practical use in the classroom. The implications of these findings are significant: this validated and reliable module can be used as an effective alternative in science learning, particularly one that integrates problem-based learning, local cultural context, STEM, and the use of AR technology. This module not only improves the quality of learning but also encourages active student engagement and the development of 21st-century skills. Further development is highly recommended, extending the module's application to other subjects and educational levels and adapting evolving technologies to enhance learning outcomes.

Acknowledgments

The authors would like to express their sincere gratitude to the validators, including lecturers and teachers, who have generously contributed their time and expertise to assist in this research. Their valuable feedback and support were instrumental in enhancing the quality and accuracy of the instruments used. This research could not have been completed without their dedicated involvement and collaboration.

References

Aiken, L. R. 1985. "Three Coefficients for Analyzing the Reliability and Validity of Ratings." *Educational and Psychological Measurement* 45(1):131–42. doi: https://doi.org/10.1177/0013164485451012.

Amanah, P. D., Al Qayyim, T. M., Jauhari, M. T., Muliati, E., & Jufri, A. W. 2025. "Validity and Reliability of the INSPIRE: Instrument Perceptions Science Inquiry and Nature of Science." *International Journal of Contextual Science* 2(1):9–16. doi: 10.29303/ijcse.v2i1.936.

Amanah, Prelia Dwi, Ahmad Harjono, and I. Wayan Gunada. 2017. "Generatif Berbantuan Scaffolding Dan Advance Organizer." *Jurnal Pendidikan Fisika Dan Teknologi* III(1):84–91.

Amanah, Prelia Dwi, Nurhidayat Martin, and Joni Rokhmat. 2025. "Research Trends on the Multiple Representation Learning Model , A Systematic Review (2016-2025)." 03(1):1–14.

Arafani, Wahyuni, Mahrus, and Lalu Zulkifli. 2025. "Validitas E-Modul IPA Berbasis PBL Berbantuan Heyzine Flipbooks Untuk Meningkatkan Kemampuan Penalaran Ilmiah Dan Literasi Sains Peserta Didik." *Journal of Classroom Action Research* 7(1):149–55. doi: 10.29303/jppipa.v6i1.264.

Arikunto, Suharsimi. 2010. *Dasar-Dasar Evaluasi Pendidikan*. edited by Jakarta. Jakarta: Bumi Aksara.

Aryulina, Diah, and Riyanto Riyanto. 2016. "A Problem-Based Learning Model in Biology Education Courses To Develop Inquiry Teaching Competency of Preservice Teachers." *Jurnal Cakrawala Pendidikan* 1(1):47–57. doi: 10.21831/cp.v1i1.8364.

Bakar, N. A., and H. N. Ismail. 2020. "Integrating Local Wisdom into STEM Education: A Review." *Journal of Technical Education and Training* 12(2):42–51. doi: https://doi.org/10.30880/jtet.2020.12.02.006.

Barrows, Howard S. 1996. "Problem-Based Learning in Medicine and beyond: A Brief Overview." *New Directions for Teaching and Learning* 1996(68):3–12. doi: 10.1002/tl.37219966804.

Bartko, J. J. 1966. "The Intraclass Correlation Coefficient as a Measure of Reliability." *Psychological Reports* 19(1):3–11. doi: 10.2466/pr0.1966.19.1.3.

Belland, B. R., Glazewski, K. D., & Richardson, J. C. 2008. "Problem-Based Learning and Argumentation: Testing a Scaffolded Collaboration Model." *Educational Technology Research and Development* 56(4):401–22. doi: https://doi.org/10.1007/s11423-007-9070-6.

Creswell, J. W., & Creswell, J. D. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (5th Ed.)*. SAGE Publications.

Deci, E. L., & Ryan, R. M. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Fleiss, Joseph, L. 1975. "Measuring Agreement between Two Judges on the Presence or Absence of a Trait." *Biometrics* 31(3):651–59.

Gay, L. R., Mills, G. E., & Airasian, P. 2012. *Educational Research: Competencies for Analysis and Applications (10th Ed.)*. Pearson.

Giuseppe, P. 2018. "Tatips Part Iv: Selection, Interpretation And Reporting Of The Intraclass Correlation Coefficient." *South European Journal of Orthodontics and Dentofacial Research* 1(5):3–5.

Gunawan., Harjono, A., dan Sutrio. 2015. "Multimedia Interaktif Dalam Pembelajaran Konsep Listrik Bagi Calon Guru." *Jurnal Pendidikan Fisika Dan Teknologi* 1(1):9.

Gunawan. 2015. *Model Pembelajaran Sains Berbasis ICT*. Mataram: FKIP UNRAM.

Hmelo-Silver, C. E. 2004. "Problem-Based Learning: What and How Do Students Learn?" *Educational Psychology Review* 16(3):235–66. doi: https://doi.org/10.1023/B:EDPR.0000034022.16470.f3.

Idrus, Syarifa Wahidah Al. 2022. "Implementasi STEM

Terintegrasi Etnosains (Etno-STEM) Di Indonesia: Tinjauan Meta Analisis." *Jurnal Ilmiah Profesi Pendidikan* 7(4):2370–76. doi: 10.29303/jipp.v7i4.879.

Jusuf, Heni, and Ahmad Sobari. 2021. "Pelatihan Pembuatan Modul Pembelajaran Untuk Mendukung Pembelajaran Online." *Jurnal Pengabdian Masyarakat* 2(1):33–37.

Koo, Terry K., and Mae Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012.

Muktiani, Nur Rohmah, Soegiyanto, Siswantoyo, Setya Rahayu, and Hedi Ardiyanto Hermawan. 2022. "Augmented Reality Mobile App-Based Multimedia Learning of Pencak Silat to Enhance the Junior High School Students' Learning Outcomes." *Cakrawala Pendidikan* 41(2):553–68. doi: 10.21831/cp.v41i2.49217.

Piaget, J. 1972. *The Psychology of the Child*. New York: Basic Books.

Polit, D. F., & Beck, C. T. 2006. "The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations." *Research in Nursing & Health* 29(5):489–97. doi: https://doi.org/10.1002/nur.20147.

Portney, L.G. and Watkins, M. P. 2000. *Oundations of Clinical Research: Applications to Practice*. 2nd ed. edited by P. H. Health. Upper Saddle River.

Rahayu, Satutik, and I. Putu Artayasa. 2025. "Trends in Project-Based Learning for Developing Critical Thinking Skills in Science Education : A Bibliometric Review." 2(1):26–34. doi: 10.56566/ijses.v2i1.258.

Rahayu, Satutik, and Kasnawi Al Hadi. 2023. "Analisis Kemampuan Berpikir Kritis Siswa: Studi Pembelajaran Menggunakan Media E-Book IPA Berbasis PBL." *Jurnal Ilmiah Profesi Pendidikan* 8(4):2795–99. doi: 10.29303/jipp.v8i4.1118.

Rahayu, Satutik, Ni Nyoman Sri Putu Verawati, and Andi Fatimatul Islamiah. 2019. "Effectiveness of Problem Based Learning Model with Worksheet Assisted on Students' Critical Thinking Ability." *Lensa: Jurnal Kependidikan Fisika* 7(2):51. doi: 10.33394/j-lkf.v7i2.2686.

Setyaningsih, Ani, Maryati Maryati, Insih Wilujeng, and Arina Zaida Ilma. 2023. "Electronic Science Module in the Merdeka Curriculum: Teacher and Student Perspectives." *Jurnal Penelitian Pendidikan IPA* 9(SpecialIssue):1172–78. doi: 10.29303/jppipa.v9ispecialissue.6591.

Shrout, Patrick E., and Joseph L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86(2):420–28. doi: 10.1037/0033-2909.86.2.420.

Sleeter, C. E. 2001. *Cultural Diversity and Education: Foundations, Curriculum, and Teaching*. Allyn & Bacon.

Suwartiningsih, Suwartiningsih. 2021. "Penerapan Pembelajaran Berdiferensiasi Untuk Meningkatkan Hasil Belajar Siswa Pada Mata Pelajaran IPA Pokok Bahasan Tanah Dan Keberlangsungan Kehidupan Di Kelas IXb Semester Genap SMPN 4 Monta Tahun Pelajaran 2020/2021." *Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI)* 1(2):80–94. doi: 10.53299/jppi.v1i2.39.

Tüysüz, C. 2016. "Augmented Reality in Science Education: Effects on Academic Achievement and Motivation." *Journal of Baltic Science Education* 15(5):681–94. doi: https://doi.org/10.33225/jbse/16.15.681.

Utaminingsih, Esty Setyo, Belita Yoan Intania, Hanik Aida, and Izza Melati Sukma. 2024. "Designing an Innovative Educational Framework for 'How We Live and Grow' Using the 4D Model." *103.23.102.168* 12(1):42–52.

Yusuf, M., & Iskandar, I. 2019. "Implementasi Pendekatan Etno-STEM Berbasis Kearifan Lokal Dalam Pembelajaran IPA." *Jurnal Pendidikan IPA Indonesia* 8(3):320–28. doi: https://doi.org/10.15294/jpii.v8i3.21239.

Zendrato, Ekhalis Dedi Kasih, Agnes Renostini Harefa, and Natalia Kristiani Lase. 2022. "Pengembangan Modul IPA Berbasis Contextual Teaching and Learning Pada Materi Sistem Pernapasan Manusia." *Educativo: Jurnal Pendidikan* 1(2):446–55. doi: 10.56248/educativo.v1i2.61.