JPPIPA 11(11) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Psychometric Properties of a Local Wisdom-Based Science Literacy Instrument for Primary School Students: An Item Response Theory Approach

Nurhasanah[1*], Tri Astari[1], Elvin Cahyanita[1], Nuriman[1]

[1] Department of Primary Teacher Education, Universitas Jember, Jember, Indonesia.

Received: September 19, 2025
Revised: October 20, 2025
Accepted: November 25, 2025
Published: November 30, 2025

Corresponding Author:
Nurhasanah
nurhasanah.fkip@unej.ac.id

DOI: 10.29303/jppipa.v11i11.12973

© 2025 The Authors. This open access article is distributed under a (CC-BY License)

Abstract: Science literacy is an essential competency for primary school students to face 21st-century challenges. However, few assessment tools combine strong psychometric quality with cultural relevance. This study aimed to examine the psychometric properties of a local wisdom–based science literacy instrument for primary school students within the Item Response Theory (IRT) framework. The instrument comprised 34 multiple-choice items aligned with science literacy indicators and was administered to 300 fourth-grade students from several schools in Jember, Indonesia. Expert judgment ensured content validity, while empirical analyses using QUEST and R Studio evaluated item performance. The results showed good model fit, excellent reliability (0.97), and well-balanced item difficulty levels, confirming the instrument's ability to measure a wide range of student abilities. Person–item maps and item characteristic curves indicated high discriminative power and measurement precision. Embedding local wisdom elements enhanced cultural authenticity and student engagement in science learning. Overall, the instrument demonstrated strong psychometric and contextual validity, offering a reliable tool for assessing science literacy in primary education. Its application can help teachers identify learning needs and design culturally responsive instruction. Future studies should extend validation to broader cultural contexts and refine items for students with lower ability levels.

Keywords: Assessment instrument; Item response theory; Primary education, Psychometric; Science literacy

Introduction

In the 21st century, scientific literacy has become an essential skill for tackling intricate global issues, such as climate change, environmental deterioration, and swift technological progress, and has increasingly attracted attention within the academic community. It empowers individuals to understand scientific concepts, utilize scientific methods, and make informed decisions in daily situations, thereby facilitating active engagement in enhancing quality of life in a dynamic environment (Ayu et al., 2025; Nurhasanah et al., 2020; Suryanti et al., 2021). In primary education, science instruction is essential for developing competencies by enabling children to interact meaningfully with scientific concepts and real-world events from an early age. The immediate environment is essential for children's learning, enabling them to gain knowledge through direct and authentic experiences. The facilities and resources present in this area affect the learning experiences children acquire, rendering local wisdom a crucial aspect that enhances scientific comprehension and contextual significance (Astari et al., 2024). Moreover, scientific literacy not only improves

How to Cite:
Nurhasanah, Astari, T., Cahyanita, E., & Nuriman. (2025). Psychometric Properties of a Local Wisdom-Based Science Literacy Instrument for Primary School Students: An Item Response Theory Approach. *Jurnal Penelitian Pendidikan IPA*, *11*(11), 965–975. https://doi.org/10.29303/jppipa.v11i11.12973

conceptual understanding but also fosters critical thinking and informed decision-making, skills that are increasingly essential in an age characterized by the swift proliferation of misinformation and disinformation. As modern technologies expedite the dissemination of misinformation, it is essential to cultivate students' capacity to critically assess and address science-related concerns from the early stages of schooling (Faizin et al., 2024; Kelp et al., 2023). Enhancing scientific literacy equips students to examine evidence, assess claims, and engage in scientific discourse, all of which are vital for fostering personal development and safeguarding societal welfare.

The notion of scientific literacy has evolved, building upon prior PISA science evaluations. The PISA 2025 framework, once seen solely as a learning outcome, now underscores the practical application of scientific knowledge in real-world scenarios, taking into account personal, local/national, and global settings (OECD, 2023). Improving scientific literacy allows pupils to address everyday issues through scientific reasoning (Shofiyah et al., 2020). Initiatives to improve scientific literacy must also take into account the wider educational framework. Frameworks like PISA emphasize that scientific literacy includes both cognitive skills and the ability to connect science with daily life and socio-cultural contexts. This comprehensive viewpoint emphasizes the necessity of integrating science education with students' real-life experiences to guarantee its significance and longevity (Roy et al., 2025).

Consistent with this viewpoint, there is an increasing acknowledgment of the importance of incorporating local knowledge into scientific education. The integration of local knowledge into education has emerged as a pivotal topic in modern discussions, especially with the advancement of learning that fosters sustainable development (Arjaya et al., 2024). Local wisdom comprises the information, behaviors, and values transmitted through generations within local communities (Lestari et al., 2024). It functions as a cultural compass that directs human conduct according to collective norms, ethical standards, and a profound dedication to environmental sustainability (Anau et al., 2019). In education, local knowledge offers valuable, contextually pertinent resources that can enhance the significance of learning and strengthen its connection to students' real-life experiences.

Culture-based learning has demonstrated a beneficial impact on student engagement and achievement by connecting educational content with students' cultural backgrounds. This alignment enhances enculturation and increases the incorporation and adaptation of new knowledge (Suprapto et al., 2021). Moreover, local wisdom, as a manifestation of experiential cultural knowledge, plays a crucial role in sustainable development by promoting ecologically aware and community-focused values (Wahyuni & Tandon, 2024). The Local Wisdom Approach serves as a conduit between traditional values and modern concerns, reconciling historical context with contemporary issues and linking local identities to global viewpoints. This approach in education not only safeguards cultural heritage but also enriches students' identity and facilitates intergenerational knowledge transfer. This technique promotes inclusive, contextual, and culturally attuned learning settings (Asmayawati et al., 2024).

When skillfully integrated, local wisdom can enhance science education by contextualizing abstract scientific concepts, reinforcing cultural identity, and promoting deeper student participation. In multicultural contexts like Indonesia, science education grounded in local wisdom facilitates the integration of contemporary scientific knowledge with traditional systems, enhancing accessibility and relevance, particularly at the primary school level. However, evaluation tools have not adequately reflected the potential of incorporating local knowledge into science education. Many cultural contexts utilize current methods, which frequently rely on global or standardized frameworks, without sufficient modification or validation (Coppi et al., 2023). Assessment is a crucial element in the evaluation and monitoring of students' learning development. Accurate and reliable assessments necessitate the use of suitable and effective assessment instruments (Nurhasanah et al., 2024; Setiawati et al., 2024). Among the diverse assessment types, written examinations are the most prevalent in educational contexts, especially for evaluating students' academic accomplishments.

The creation of objective and high-quality evaluation tools entails numerous essential phases, one of which is item analysis. This process is essential for assessing the instrument's quality and the functionality of each test item, thus acting as a crucial element in ensuring the test's validity, reliability, suitable difficulty levels, and discriminative ability (Jafar & Ridwan, 2024; Nisfatulsanah & Sugiharto, 2024). Initial testing is crucial to verify if the instrument satisfies these psychometric standards, thus endorsing the precision and efficacy of the assessment outcomes.

Modern psychometric methodologies, such as Item Response Theory (IRT), have been progressively utilized in educational research to improve the quality of assessment instruments. Item Response Theory (IRT) is extensively employed in the human sciences to model individual reactions to a collection of items that assess one or more latent dimensions (Bürkner, 2021). Within the IRT framework, the Rasch model facilitates a

comprehensive study by allowing for the identification of students' skill levels, the discovery of anomalous response patterns, and the assessment of potential bias in test items (Nisfatulsanah & Sugiharto, 2024). The Rasch model is a contemporary theory of item assessment developed to address many shortcomings of Classical Test Theory (CTT). The fundamental difference between the two lies in the data analysis process. In CTT, raw scores are analyzed directly as integer data, whereas in the Rasch model, raw scores must be converted to odds ratios and transformed into logit units to determine the probability of respondents answering an item. Therefore, the Rasch model can restore data according to its continuous nature and provide more accurate measurements based on probability principles (Darman et al., 2024).

The Rasch model, as a component of Item Response Theory (IRT), is used to generate interval scales and ensure measurement invariance. This model assesses items based on difficulty level and respondent ability while also modeling individual response probabilities. Its advantages lie in its ability to identify problematic items (misfits), estimate item difficulty, and measure respondents' abilities more accurately compared to Classical Test Theory (CTT) (Bimastari et al., 2025). Rasch analysis facilitates the evaluation of students' competencies via item analysis, yielding a more thorough assessment of each test item's quality (Nisfatulsanah & Sugiharto, 2024).

The utilization of a scientific literacy framework is a commonly embraced method in the creation of assessment tools. Researchers frequently develop, modify, or enhance such frameworks to correspond with their research aims and educational environments (Istyadji & Sauqina, 2023). Nonetheless, scant research has investigated the incorporation of cultural factors, including indigenous knowledge and region-specific environmental concerns, into the design and evaluation of scientific literacy assessments. Furthermore, limited research has utilized contemporary psychometric methodologies such as Item Response Theory (IRT) to assess the validity and reliability of tools designed for specific local contexts.

Despite the growing recognition of science literacy as a core competency, few assessment tools at the primary level have been systematically examined for their psychometric properties, particularly those incorporating local wisdom as contextual content. Existing studies often emphasize test development, yet provide limited evidence on how such instruments perform in practice when analyzed with advanced psychometric models.

This study seeks to address this gap by examining the psychometric properties of a local wisdom-based science literacy instrument for primary school students through the lens of Item Response Theory (IRT). The analysis focuses on item-level indicators such as difficulty, discrimination, and model fit, thereby providing empirical evidence on the instrument's capacity to measure science literacy reliably across diverse learners. By concentrating on trial outcomes, the study contributes to the refinement of culturally grounded assessments and advances the broader discourse on equitable and inclusive science education.

Method

This study adopted an instrument development and validation design to evaluate the psychometric properties of a local wisdom–based science literacy instrument for primary school students. The instrument was developed based on established science literacy indicators while integrating cultural elements relevant to students' local contexts. The study examined item difficulty, discrimination, and fit statistics within the Item Response Theory (IRT) framework. Validation procedures included expert reviews of content, language, and construct validity, followed by a field trial involving primary school students. Data from the trial were analyzed using the Rasch Model to assess reliability indices, item difficulty, item characteristic, and the overall quality of the instrument.

Research Participants

The participants of this study were 300 Grade IV primary school students drawn from several schools across Jember Regency, Indonesia. The schools were purposively selected to represent diverse geographical and socio-cultural contexts, different levels of institutional readiness, and alignment with the local wisdom elements embedded in the instrument. Each student responded to all 34 items in a paper-based assessment, which was administered in their respective classrooms under standardized testing conditions to ensure fairness and comparability of results.

Instrument

The research instrument consisted of a 34-item multiple-choice test designed to measure students' science literacy in culturally relevant contexts. Each item contained a stimulus, which could take the form of short texts, images, or diagrams, accompanied by four answer choices with only one correct response. Local wisdom was integrated through familiar contexts such as traditional practices, local biodiversity, and community-based problem-solving scenarios. Content validation was conducted by 3 experts in educational assessment, primary school science education, and language, as well as by practicing teachers. The feedback provided by these validators was used to refine the items in terms of

clarity, cultural appropriateness, and curricular alignment, leading to the final version of the instrument employed in this study.

Data Analysis

After validation, the instrument was administered to students for empirical testing, and the responses were analyzed within the framework of Item Response Theory. QUEST software was employed to estimate item and person parameters, examine model–data fit statistics, and calculate reliability indices. In addition, R Studio was utilized to conduct supplementary analyses and generate graphical representations such as Person Item Maps, Item Characteristic Curves, and Item Information Curves.

The analysis covered several dimensions, including item fit, reliability, item difficulty, item characteristic curves, and item information curves. Item fit was assessed using Infit Mean Square (MNSQ) statistics, with values ranging from 0.77 to 1.30 considered acceptable. Reliability was examined using QUEST indices, interpreted based on the classification proposed by Sumintono and Widhiarso (2015), where values greater than 0.94 were categorized as excellent, 0.91–0.94 as very good, 0.81–0.90 as satisfactory, 0.67–0.80 as adequate, and values below 0.67 as weak. Item difficulty was analyzed using threshold maps in QUEST supported by parameter estimates in R Studio to ensure that the distribution of item difficulties adequately covered the student ability spectrum.

Item Characteristic Curves were examined to determine the probability of a correct response across different ability levels, with steeper slopes indicating higher discriminative power, while irregular curves were flagged for potential revision or removal. Furthermore, Item Information Curves were analyzed to evaluate the amount of information provided by each item at varying ability levels, complemented by the Test Information Function, which assessed the overall measurement precision of the instrument across the ability continuum.

Results and Discussion

This study set out to evaluate the psychometric properties of a local wisdom–based science literacy instrument for primary school students using the Rasch Model within the Item Response Theory (IRT) framework. The findings indicate that the majority of items demonstrated a satisfactory fit with the model, the instrument exhibited excellent reliability, and the distribution of item difficulty adequately covered the ability spectrum of the target population. These results suggest that the instrument not only meets statistical requirements for validity and reliability but also holds practical value for assessing science literacy in contexts where cultural relevance is emphasized. The following sections discuss these findings in relation to previous research, theoretical considerations, and practical implications for science education at the primary level.

Empirical Analysis of the Instrument

The trialed instrument was subsequently subjected to empirical analysis using the QUEST program, based on the Item Response Theory (IRT) framework, specifically employing the Rasch model (1-Parameter Logistic/1PL). The analysis aimed to evaluate the quality of each test item in terms of model fit (goodness of fit), instrument reliability, and item difficulty level. The results of the analysis are presented as follows:

Model Fit (Goodness of Fit)

The trial data were examined using the Rasch Dichotomous Model, suitable for multiple-choice items evaluated dichotomously (0 for incorrect responses and 1 for correct responses). This model was utilized to assess the degree of alignment between the empirical data and the theoretical Rasch model, focusing on item parameters within a unidimensional framework. Figure 1 presents the results, summarizing item estimates and fit statistics for the science literacy instrument.

```
Literasi Sains (34 ITEM PG dengan 5 alternatif)
------------------------------------------------------------------------------
-------------------------------
Item Estimates (Thresholds)
12/ 7/25 11: 5
all on all (N = 300 L = 34 Probability Level= .50)
------------------------------------------------------------------------------
-------------------------------

Summary of item Estimates
-------------------------

Mean                        .00
SD                          .71
SD (adjusted)               .70
Reliability of estimate     .97


 Fit Statistics
 ---------------

 Infit Mean Square        Outfit Mean Square

    Mean     .99            Mean    1.03
    SD       .16            SD       .27


      Infit t                  Outfit t

    Mean    -.28            Mean     .03
    SD      2.78            SD      2.30

   0 items with zero scores
   0 items with perfect scores
------------------------------------------------------------------------------
```

Figure 1. Item estimates and fit statistics of the science literacy instrument

Figure 1 illustrates that the alignment of the test items with the Rasch model can be assessed by various indicators derived from the fit statistics, namely Infit Mean Square, Outfit Mean Square, Infit t, and Outfit t.

The analysis results indicate that the Infit Mean Square value is 0.99 and the Outfit Mean Square value is 1.03, both of which are within the permitted range. Items conforming to the Rasch model generally exhibit Infit MNSQ values ranging from 0.77 to 1.30, with Outfit t < 2 (Rosana & Sukardiyono, 2015). The presence of both Infit and Outfit MNSQ values inside the optimal range signifies that the items exhibit a strong alignment with the Rasch model. The standard deviations (SD) for Infit and Outfit Mean Square are 0.16 and 0.27, respectively, indicating an acceptable degree of variability among the items. Meanwhile, the average Infit t value of -0.28 and Outfit t value of 0.03, with standard deviations of 2.78 and 2.30, respectively, further support the assumption that there are no significant deviations from the model. All t values fall within an acceptable range, indicating that no items exhibit extreme anomalies in the respondents' answers. This strengthens the conclusion that the test items fit well with the Rasch model.

Based on the analysis results, no items were found to have zero scores or perfect scores. This indicates that student response distribution is balanced, and there are no items that are either too easy or too difficult for all participants. Overall, this suggests that the Rasch model is appropriate for analyzing the developed local wisdom-based science literacy test. Furthermore, these results also demonstrate that the student response data meet the assumptions of the Rasch model, making the instrument suitable for further in-depth item characteristic analysis. The overall fit statistics findings provide strong support for the construct validity of the locally developed wisdom-based science literacy instrument.

Instrument Reliability

The reliability of the instrument in the Rasch Model is indicated by the Reliability of Estimate value, which describes the consistency of the instrument items in measuring respondents' abilities. Based on the analysis results using the Quest application, a reliability value of 0.97 was obtained. This value is considered very high or exceptional (Sumintono & Widhiarso, 2015) and is above the commonly used minimum threshold of 0.70 in psychometric measurement (Bond & Fox, 2015). This result indicates that the instrument has an excellent ability to consistently distinguish respondents with different ability levels.

According to Sumintono et al. (2015), reliability in the Rasch Model reflects internal consistency and indicates the extent to which the items can separate individuals based on their ability levels. The higher the reliability value, the greater the confidence that the instrument can measure participants' abilities accurately and stably. This high reliability also reflects that the items in the instrument have internal stability and collectively contribute to measuring the construct of science literacy based on local wisdom as intended. Therefore, the developed instrument can be considered to have an excellent level of reliability and is suitable for use in the context of educational assessment in elementary schools.

Item Difficulty Level

The assessment of item difficulty is essential for evaluating test quality, as it reveals the distribution and ratio of easy, moderate, and difficult items within the instrument. Figure 2 illustrates the distribution of item thresholds according to the Item Response Theory (IRT) framework.

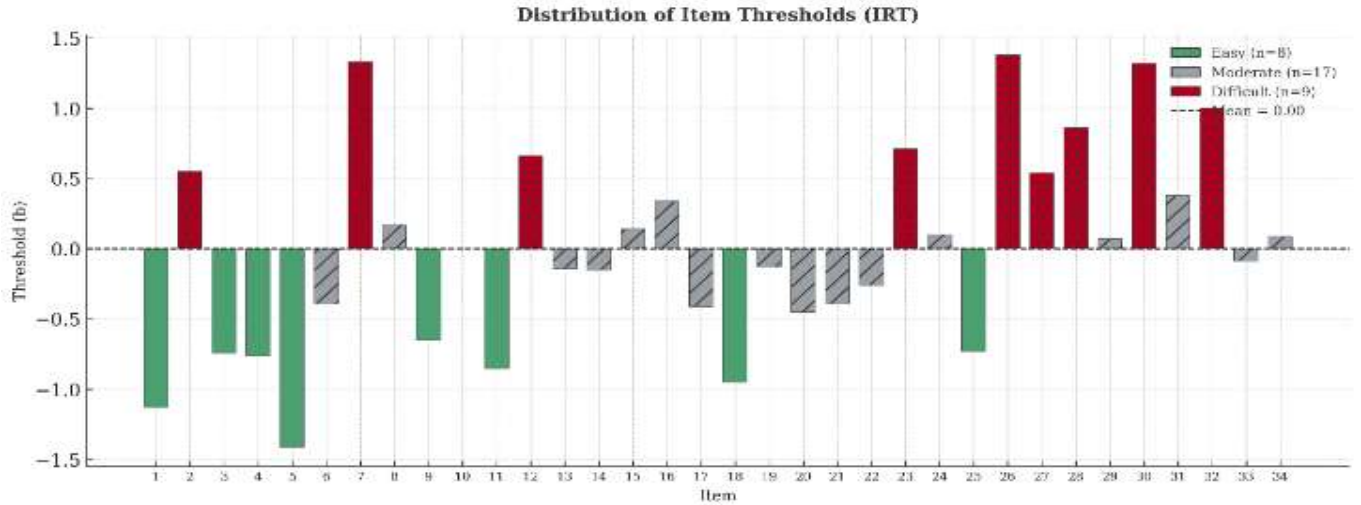

Figure 2. Distribution of item thresholds (IRT) for the Science Literacy Instrument

The analysis conducted with the Quest program revealed that the threshold (b) or difficulty level values for the items varied from −1.41 to 1.38. The values denote the comparative complexity of the items, with higher b-values signifying greater difficulty. Out of the 34 items examined, 8 were classified as easy, 17 as moderate, and 9 as difficult. This distribution indicates that the developed instrument exhibits a relatively balanced variation in difficulty levels, with the majority concentrated in the moderate group. Test construction recommendations generally stipulate that around 25% of items should be classified as easy, 50% as intermediate, and 25% as difficult (Hambleton, 1991). This distribution guarantees item equilibrium and optimizes the information yielded by the instrument within the intermediate range of ability levels (Boone et al., 2014; Stemler, 2021).

The findings of this study indicate that the obtained distribution roughly corresponds with these guidelines, especially with the majority of items categorized as moderate. This finding suggests that the instrument can successfully differentiate students' abilities at the median level. Furthermore, the incorporation of both easy and difficult items enhances the instrument's measurement efficacy by catering to students with diverse ability levels, as highlighted by Rasch methodologists who contend that a range of easy, moderate, and difficult items is essential for thorough measurement (Kean et al., 2018; Sumintono & Widhiarso, 2015).

Analysis of Person–Item Mapping

The analysis employing Item Response Theory (IRT) produced a person–item map, illustrated in Figure 3. This map depicts the distribution of student skills (person parameters) along the latent trait continuum, alongside the difficulty levels of the test items (item parameters). These visualizations are crucial in IRT-based measurement, as they enable researchers to assess the alignment of test questions with the ability levels of respondents (Bond, 2015; Boone et al., 2014).

The examination of the person–item map indicated that the distribution of respondents' abilities and the difficulty levels of the local wisdom-based science literacy instrument items spanned from −2 to +2 logits. The distribution of the person parameter predominantly clustered within the −0.5 to +1.0 logit range, indicating that most students involved in this study exhibited moderate abilities. The test items exhibited a wide range of difficulty levels with a fairly equitable distribution. Multiple items (e.g., I2, I4, I6) were in the negative logit region, signifying that these items were comparatively easy for students. In contrast, items I28, I30, I32, and I34 were in the positive logit region, indicating that they required elevated levels of scientific literacy proficiency.

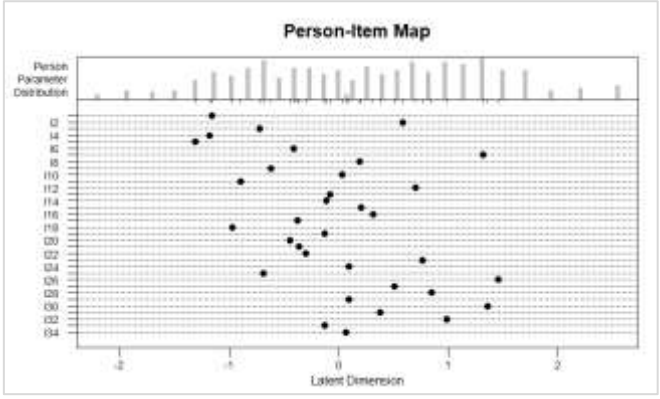

Figure 3. Person–item map of the instrument for science literacy based on local wisdom

This distribution illustrates that the test instrument effectively assesses student capabilities across a broad spectrum, aligning with the unidimensionality concept in Item Response Theory (Embretson & Reise, 2013; Hambleton, 1991). The equilibrium between student ability distribution and item difficulty levels serves as a crucial indicator of instrument quality. An excellent tool must reliably assess student abilities across the spectrum and furnish significant information for both low- and high-ability pupils (DeMars, 2010).

The incorporation of local wisdom into the science literacy items facilitated sufficient variance in difficulty levels, enabling the instrument to evaluate students' competencies more authentically. Incorporating local contexts in assessment design enhances both validity and engagement by rendering items more relevant to learners' daily experiences (Kean et al., 2018; Sumintono & Widhiarso, 2015). This enhances the instrument's capacity to differentiate among varying levels of competence while also incorporating cultural significance.

In conclusion, the person–item map offers empirical evidence that the locally designed wisdom-based science literacy instrument satisfies the standards of a high-quality assessment tool. It exhibits a uniform distribution of item difficulty, the ability to assess a broad spectrum of student competencies, and significant contextual relevance to the daily experiences of primary school students.

Item Characteristic Curves (ICC)

Figure 4 displays the item characteristic curves (ICCs) for the 34 science literacy items created in this study. These curves depict the correlation between students' proficiency levels (θ) and the likelihood of accurately responding to each item. Consistent with the Item Response Theory (IRT) framework, the Item Characteristic Curves (ICCs) typically exhibit a sigmoid (S-shaped) pattern, demonstrating that the items operated in accordance with theoretical assumptions

(Embretson & Reise, 2013; Hambleton, 1991). The S-shaped curve indicates that as ability rises, the likelihood of correct responses similarly escalates, illustrating that the items can be differentiated among students with varying proficiency levels.

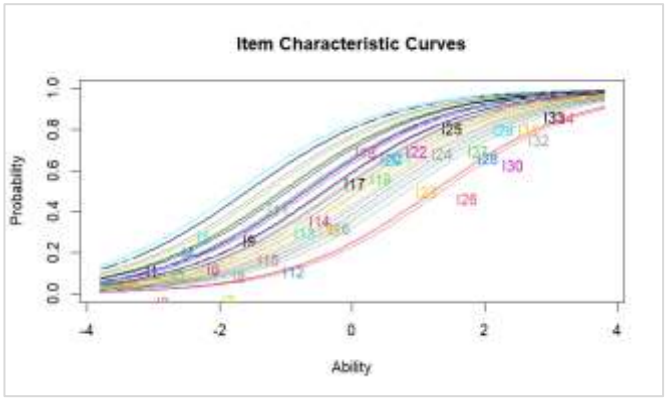

Figure 4. Item characteristic curves (ICCs) of the science literacy instrument based on local wisdom

Most items fell within the θ range of −2 to +2, signifying that the instrument is very adept at evaluating students with low to moderate ability levels. Multiple questions (e.g., I23, I26, I30) were displaced towards higher θ values, indicating that these items were comparatively challenging and could only be accurately answered by students with advanced science literacy. In contrast, items I1, I5, and I9 exhibited relative ease, since their probability curves reached elevated values even at $\theta < 0$.

The gradient of the Item Characteristic Curves (ICCs) provides insight into how sensitively items respond to changes in student ability. In the Rasch framework, the discrimination parameter is held constant; however, visual differences in curve steepness can still reflect how effectively an item contributes to measuring ability along the continuum. Items with steeper ICCs (e.g., I17, I25) were observed to provide more information for distinguishing among students of different ability levels, whereas items with flatter ICCs (e.g., I12, I23) contributed less information. This pattern aligns with the principle that well-functioning items should vary in difficulty and provide adequate information across the ability spectrum (Baker, 2001; Wyse, 2010).

The ICC analysis confirms that the developed instrument comprises items with a wide range of difficulty levels and satisfactory fit to the Rasch model. This variation enhances the instrument's effectiveness as a reliable measure of primary students' science literacy. Moreover, these findings highlight the advantages of applying Item Response Theory–based methods in educational assessment, as they yield more nuanced and precise insights into item functioning compared to classical test theory (van der Linden & Hambleton, 1997; Sinharay, 2015).

Item Information Curves (IIC)

Figure 5 displays the Item Information Curves (IICs) for the 34 items in the local wisdom-based science literacy assessment. The IICs denote the contribution of each item to the assessment of students' ability levels (θ). An IIC denotes the proficiency level at which an item yields optimal information, enabling researchers to ascertain the spectrum of abilities most precisely assessed by the item (Embretson & Reise, 2013; Hambleton, 1991).

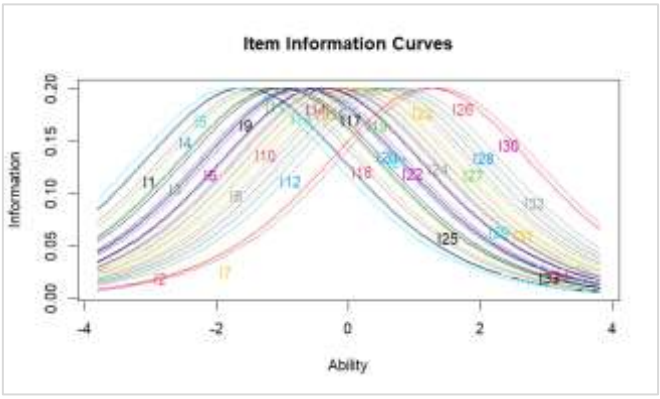

Figure 5. Item information curves (IICs) for the local wisdom-based science literacy instrument

The research indicated that the majority of items yield optimal information within the interval of $\theta = -2$ to +2. This outcome suggests that the instrument is especially adept at assessing students across a spectrum of skills, with the greatest density of information centered around $\theta = 0$. This distribution indicates that the instrument is most accurate in differentiating students at the average ability level, aligning with the principle that optimal test information should ideally be centered around the mean ability of the target population (Baker, 2001; Boone et al., 2014).

Several items, including I23, I26, and I30, reached $\theta > +1$, indicating their utility in evaluating higher-ability students. The incorporation of these items improves the instrument's overall quality, guaranteeing its capacity to assess a wide spectrum of abilities instead of being limited to average populations. The convergence of numerous Item Information Curves (IICs) around $\theta = 0$ indicates substantial reliability for most students, as overlapping peaks imply consistent measurement within a common range of skill levels (Sinharay, 2015; Wyse, 2010).

The correlation between the Item Information Curve (IIC) and the item discrimination parameter (a) is also apparent. Items with steeper slopes in their Item Characteristic Curves (ICCs) correspond to more

pronounced and elevated IIC peaks, indicating enhanced discriminative ability. These items exhibit more sensitivity in distinguishing between students with similar ability levels. In contrast, items exhibiting flatter ICCs demonstrate fewer peaks in the IIC, hence supplying diminished information to the measurement process (Chalmers, 2012; Linden & Hambleton, 1997).

Notwithstanding these features, the analysis revealed comparatively poor information at $\theta < -2$. This indicates that the instrument is less successful in evaluating students with significantly low levels of science literacy. Consequently, additional refining is required by incorporating items particularly intended to address the lower end of the ability distribution. These enhancements would expand the measurement parameters and guarantee that all students, including those with the lowest proficiency levels, are evaluated appropriately.

In conclusion, the IIC analysis indicates that the proposed instrument offers adequate information across a broad spectrum of abilities, particularly excelling in assessing average to high ability levels. This data substantiates the validity and reliability of the local wisdom-based science literacy instrument for application in elementary schools and emphasizes the significance of IRT-based methodologies for creating adaptive and equitable educational assessments.

Consequences, Constraints, and Suggestions

The results of this investigation yield several important consequences. A science literacy test grounded in local wisdom for primary school pupils shows that contextual integration can yield items with diverse difficulty levels and sufficient discriminative ability. This renders the instrument not only valid and reliable but also more authentic in reflecting students' real-life experiences, a conclusion that aligns with prior research indicating that instruments incorporating cultural or ethnoscience components typically attain high validity, reliability, and authentic contextual representation (Amalia et al., 2024; Muniroh et al., 2022). International research has similarly indicated that incorporating pertinent settings into assessment items enhances validity and increases their discriminative capacity (Follette et al., 2015; Phadke et al., 2024). The instrument can function as a diagnostic tool for educators to discern students' strengths and weaknesses in science literacy, thereby facilitating differentiated instruction and remedial programs in primary schools, as previously documented in Rasch-based evaluations of science literacy instruments (Sihombing et al., 2019).

Notwithstanding these strengths, certain limits must be recognized. The instrument exhibited diminished sensitivity in assessing pupils with extremely low science literacy skills, as seen by the little data in the $\theta < -2$ range. Comparable results have been documented in research indicating that instruments grounded in local wisdom demonstrated modest efficacy but exhibited reduced sensitivity towards pupils with significantly poor abilities (Amalia et al., 2024). This result indicates that the instrument may inadequately reflect the performance of students with the lowest proficiency levels. Moreover, contextualized instruments created in many worldwide studies have encountered difficulties in accurately assessing the capabilities of low-performing pupils, underscoring the necessity of broadening coverage at the lower spectrum of the ability continuum (Phadke, et al., 2024). The study was confined to a particular regional setting, potentially affecting the generalizability of the findings to wider populations, as observed in similar region-specific instrument development studies (Kirana, 2024).

Future study should focus on creating supplementary items tailored for students with extremely low literacy skills to enhance the breadth of the ability continuum. Furthermore, extensive validation across various educational institutions and cultural contexts would strengthen the instrument's reliability, consistent with previous studies advocating for multi-contextual validation of assessment instruments grounded in local wisdom (Kirana, 2024; Muniroh et al., 2022). In alignment with worldwide assessments that authenticate contextualized science literacy tools across diverse communities (Follette et al., 2015), subsequent efforts should prioritize inclusivity and generalizability. The study underscores the potential of incorporating local knowledge into science literacy evaluations and offers guidance for further refinement to enhance inclusion and broader application.

Conclusion

This study evaluated the psychometric properties of a local wisdom–based science literacy instrument for primary school students using the Rasch model within the Item Response Theory framework. The results demonstrated excellent reliability (0.97), satisfactory model fit (Infit MNSQ = 0.99; Outfit MNSQ = 1.03), and an appropriate range of item difficulty (−1.41 to 1.38). These findings confirm that the instrument performs effectively across various ability levels and provides stable measurement results. Integrating local wisdom contexts strengthened the cultural authenticity and relevance of the instrument, allowing it to represent students' lived experiences in science learning more accurately. The instrument thus meets both psychometric and contextual criteria, offering a reliable tool for evaluating students' science literacy and supporting teachers in implementing culturally

responsive and differentiated instruction. Nevertheless, the instrument showed limited sensitivity for students with very low ability levels and was tested only within the Jember region, which may limit the generalizability of the findings. Future research should expand the item pool to better capture lower proficiency ranges and validate the instrument across more diverse educational and cultural contexts. Overall, this study demonstrates the value of integrating local cultural perspectives with modern psychometric approaches to promote more inclusive and context-sensitive science literacy assessment in primary education.

Acknowledgments
The University of Jember, through the Research and Community Service Institute (LPPM), supported this research with an internal grant. The authors thank the experts who provided input, the participating elementary school students and teachers, and the elementary schools in Jember that assisted with data collection. We greatly appreciate all parties' support and cooperation for the success of this research.

Author Contributions
Conceptualization, N. and T.A.; methodology, N.; formal analysis, N.; investigation, N. and T.A.; resources, E.C. and N.; data curation, N.; writing original draft preparation, N. and T.A.; writing, review and editing, N., T.A., E.C. and N.; visualization, N. and E.C. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by the Beginner Lecturer Research Grant, internal funding source of the University of Jember, grant number 2751/UN25.3.1/LT/2025.

Conflicts of Interest
The authors declare no conflicts of interest.

References

Amalia, D. V., Ilhami, A., Fuadiyah, S., & Kusumanegara, A. (2024). Development of a scientific literacy instrument based on Riau Malay ethnoscience in science subjects. *Pedagogik: Jurnal Pendidikan*, *11*(1), 1–18. https://doi.org/10.33650/pjp.v11i1.6382

Anau, N., Hakim, A., Lekson, A. S., & Setyowati, E. (2019). Local Wisdom Practices of Dayak Indigenous People in the Management of Tana' Ulen in the Kayan Mentarang National Park of Malinau Regency, North Kalimantan Province, Indonesia. *Russian Journal of Agricultural and Socio-Economic Sciences*, *91*(7), 156–167. https://doi.org/10.18551/rjoas.2019-07.16

Arjaya, I. B. A., Suastra, I. W., & Redhana, I. W. (2024). Global Trends in Local Wisdom Integration in Education: A Comprehensive Bibliometric Mapping Analysis from 2020 to 2024. *International Journal of Learning, Teaching and Educational Research*, *23*(7), 120–140. https://doi.org/10.26803/ijlter.23.7.7

Asmayawati, Yufiarti, & Yetti, E. (2024). Pedagogical innovation and curricular adaptation in enhancing digital literacy: A local wisdom approach for sustainable development in Indonesia context. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), 100233. https://doi.org/10.1016/j.joitmc.2024.100233.

Astari, T., Purwanti, K. Y., Arditama, A. Y., Subhananto, A., & Nuryanti, M. S. dkk. (2024). *Ekologi sosialisasi anak: Perspektif keluarga, sekolah dan komunitas*. Majalengka: Cv. Edupedia Publisher.

Ayu, G. N., Putri, C. A., Riyanto, A. R., & Koto, I. (2025). The Scientific Literacy Competence of Students in Indonesia and Mexico Based on PISA 2022: An International Comparative Study. *TOFEDU: The Future of Education Journal*, *4*(5), 1033-1038. https://doi.org/10.61445/tofedu.v4i5.525

Baker, F. B. (2001). *The basics of item response theory*. US: ERIC Clearinghouse on Assessment and Evaluation.

Bimastari, I. H., Yusup, M., & Kistiono, K. (2025). The Measuring Energy Literacy: Validation of Knowledge, Attitude, and Behavior Instruments Using the Rasch Model. *Integrated Science Education Journal*, *6*(3), 153–162. https://doi.org/10.37251/isej.v6i3.2025

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd edition). Routledge, Taylor & Francis Group.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4

Bürkner, P.-C. (2021). Bayesian Item Response Modeling in *R* with brms and *Stan*. *Journal of Statistical Software*, *100*(5). https://doi.org/10.18637/jss.v100.i05

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6). https://doi.org/10.18637/jss.v048.i06

Coppi, M., Fialho, I., & Cid, M. (2023). Scientific Literacy Assessment Instruments: A Systematic Literature Review. *Educação Em Revista*, *39*, e37523. https://doi.org/10.1590/0102-4698237523-t

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and Validation of Scientific Inquiry Literacy Instrument (SILI) Using Rasch Measurement Model. *Education Sciences*, *14*(3), 322. https://doi.org/10.3390/educsci14030322

DeMars, C. (2010). *Item Response Theory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory*. Psychology Press. https://doi.org/10.4324/9781410605269

Faizin, A., Susantini, E., & Raharjo. (2024). Application of A Guided Inquiry Learning Model to Improve Students' Scientific Literacy Skills. *IJORER : International Journal of Recent Educational Research*, *5*(2), 490–503. https://doi.org/10.46245/ijorer.v5i2.573

Follette, K., McCarthy, D., Dokter, E., Buxner, S., & Prather, E. (2015). The quantitative reasoning for college science (QuaRCS) assessment, 1: Development and validation. *Numeracy*, *8*(2). https://doi.org/10.5038/1936-4660.8.2.2

Hambleton, R. K., S. H. and R. H. J. (1991). *Fundamentals of item response theory*. SAGE Publications.

Istyadji, M. & Sauqina. (2023). Conception of scientific literacy in the development of scientific literacy assessment tools: A systematic theoretical review. *Journal of Turkish Science Education*, *20*(2), 281–308. https://doi.org/10.36681/tused.2023.016

Jafar, L., Riyadi, R., & Ridwan, A. (2024). Penggunaan Teori Tes Klasik untuk Analisis Butir Soal Pada Asesmen. *IJSH: Indonesian Journal of Social and Humanities*, *2*(3), 1–10. Retrieved from https://jurnal.academiacenter.org/index.php/IJSH

Kean, J., Bisson, E. F., Brodke, D. S., Biber, J., & Gross, P. H. (2018). An introduction to item response theory and Rasch analysis: Application using the Eating Assessment Tool (EAT-10). *Brain Impairment*, *19*(1), 91–102. https://doi.org/10.1017/BrImp.2017.31

Kelp, N. C., McCartney, M., Sarvary, M. A., Shaffer, J. F., & Wolyniak, M. J. (2023). Developing Science Literacy in Students and Society: Theory, Research, and Practice. *Journal of Microbiology & Biology Education*, *24*(2), e00058-23. https://doi.org/10.1128/jmbe.00058-23

Kirana, R. N. (2024). Development of a mathematical literacy test instrument based on local wisdom in Blitar Raya for fifth-grade elementary students. *Journal of Development Research*, *8*(2), Process. https://doi.org/10.28926/jdr.v8i2.386

Lestari, N., Paidi, P., & Suyanto, S. (2024). A systematic literature review about local wisdom and sustainability: Contribution and recommendation to science education. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(2), em2394. https://doi.org/10.29333/ejmste/14152

Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-2691-6

Muniroh, N., Rusilowati, A., & Isnaeni, W. (2022). Instrument development of science literacy assessment with socio-sciences contains in natural science learning for elementary school. *Journal of Educational Research and Evaluation*, *11*(1), 15–22. https://doi.org/10.15294/jere.v11i1.55421

Nisfatulsanah, A., & Sugiharto, B. (2024). Scientific Literacy Assessment Instruments: A Systematic Literature Review. *Pionir: Jurnal Pendidikan*, *13*(3), 70. https://doi.org/10.22373/pjp.v13i3.25435

Nurhasanah, Hidayatullah, Z., & Arif, M. B. S. (2024). Karakteristik Instrumen Tes Literasi Digital Ditinjau dari Validitas Isi dan Validitas Empiris (Kecocokan Butir dengan Model, Reliabilitas, serta Tingkat Kesukaran Butir). *Journal of Classroom Action Research*, *6*(4), 916–923. https://doi.org/10.29303/jcar.v6i4.9650

Nurhasanah, N., Jumadi, J., Herliandry, L. D., Zahra, M., & Suban, M. E. (2020). Perkembangan penelitian literasi sains dalam pembelajaran fisika di Indonesia. *Edusains*, *12*(1), 38-46. https://doi.org/10.15408/es.v12i1.14148

OECD. (2023). *PISA 2022 results: What students know and can do*. Paris: OECD Publishing. https://doi.org/10.1787/53f23881-en

Phadke, S., Beckman, M., & Morgan, K. L. (2024). Examining the role of context in statistical literacy outcomes using an isomorphic assessment instrument. *Statistics Education Research Journal*, *23*(1). https://doi.org/10.52041/serj.v23i1.529

Roy, G., Sikder, S., & Danaia, L. (2025). Adopting scientific literacy in early years from empirical studies on formal education: A systematic review of the literature. *International Journal of STEM Education*, *12*(1), 26. https://doi.org/10.1186/s40594-025-00547-1

Setiawati, I., Wardani, S., & Lestari, W. (2024). Development of Wordwall-based Indonesian Geographical Condition Assessment Instrument in Modipaskogo E-Book for Elementary School Students. *Riwayat: Educational Journal of History and Humanities*, *7*(1), 48–65. https://doi.org/10.24815/jr.v7i1.36597

Shofiyah, N., Afrilia, I., & Wulandari, F. E. (2020). Scientific Approach and The Effect on Students Scientific Literacy. *Journal of Physics: Conference Series*, *1594*(1), 012015. https://doi.org/10.1088/1742-6596/1594/1/012015

Sihombing, R. U., Naga, D. S., & Rahayu, W. (2019). A rasch model measurement analysis on Indonesian science literacy test: Smart way to improve the learning assessment. *Indonesian Journal of*

Educational Review (IJER), *6*(2), 42–53. Retrieved from https://journal.unj.ac.id/unj/index.php/ijer/article/view/14071

Sinharay, S. (2015). Book review: Handbook of item response theory modeling: Applications to typical performance assessment. *Applied Psychological Measurement*, *39*(6), 499–502. https://doi.org/10.1177/0146621615590600

Stemler, S. E. & N. A. (2021). Rasch measurement v. item response theory: Knowing when to cross the line. *Practical Assessment, Research, and Evaluation*, *26*(1), 1–16. https://doi.org/10.7275/v2gd-4441

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Trim Komunikata.

Suprapto, N., Prahani, B. K., & Cheng, T. H. (2021). Indonesian Curriculum Reform in Policy and Local Wisdom: Perspectives from Science Education. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 69–80. https://doi.org/10.15294/jpii.v10i1.28438

Suryanti, S., Widodo, W., & Yermiandhoko, Y. (2021). Gadget-Based Interactive Multimedia on Socio-Scientific Issues to Improve Elementary Students' Science Literacy. *International Journal of Interactive Mobile Technologies (iJIM)*, *15*(01), 56. https://doi.org/10.3991/ijim.v15i01.13675

Wahyuni, E., & Tandon, M. (2024). Leveraging Local Wisdom in Curriculum Design to Promote Sustainable Development in Rural Schools. *Journal of Social Science Utilizing Technology*, *2*(3), 446–459. https://doi.org/10.70177/jssut.v2i3.1347

Wyse, A. E. (2010). *The theory and practice of item response theory*. New York: The Guilford Press. https://doi.org/10.1007/s11336-010-9179-z