Vol. 11, No. 1, 2026, pp. 657-665
DOI: https://doi.org/ 10.29210/30036872000

Contents lists available at Journal ELORA

JRTI (Jurnal Riset Tindakan Indonesia)
ISSN: 2502-079X (Print) ISSN: 2503-1619 (Electronic)

Journal homepage: https://jrti.eloracenter.org/jrti

Psychometric characteristics of a pisa-based scientific literacy assessment instrument for ecology: a rasch model analysis

Erik Nanda Putra*), Lufri, Syafri Anwar, Yuni Ahda
Universitas Negeri Padang, Indonesia

Article Info

Article history:

Received Jun 4th, 2026
Revised Jul 6th, 2026
Accepted Jul 27th, 2026

Keyword:

Assessment instrument
Ecology
Pisa
Rasch model
Scientific literacy

ABSTRACT

Scientific literacy is an essential competency for the 21st century, yet Indonesian students continue to demonstrate relatively low levels of achievement. This study aimed to analyze the psychometric characteristics of a PISA-based scientific literacy assessment instrument on ecology during the field-testing phase using the Rasch model. The study involved 40 students who completed a 45-item assessment. Data were analyzed using Winsteps based on three Rasch indicators: person–item targeting, item fit, and separation index. The results showed appropriate targeting, with a 0.45-logit difference between the mean student ability and mean item difficulty. A total of 43 of the 45 items (95.6%) met the Rasch model fit criteria. The person separation index of 1.71 and item separation index of 2.18 classified students' abilities and item difficulties into three distinct strata. Overall, the findings indicate that the instrument demonstrates adequate psychometric characteristics for measuring scientific literacy in ecology. However, the relatively modest person separation suggests that additional items targeting lower ability levels may further improve the instrument's measurement precision.

Corresponding Author:

Erik Nanda Putra,
Universitas Negeri Padang, Padang, Indonesia
Email: erikbiologi@gmail.com

Introduction

Scientific literacy is an essential competency that enables students to meet the challenges of the twenty-first century (Turiman et al., 2012). It encompasses the ability to explain scientific phenomena, evaluate evidence-based information, and make informed decisions in various contexts, including environmental, health, and technological issues (Permatasari et al., 2026; Khery et al., 2022). As scientific knowledge continues to expand rapidly, students are increasingly expected to develop scientific reasoning skills that enable them to distinguish scientifically valid information from misinformation (Schmaltz et al., 2017). Consequently, scientific literacy has become one of the core competencies assessed in the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD) since its first implementation in 2000 (Lie & Linnakylä, 2004; Bybee et al., 2009). Beyond assessing scientific knowledge, scientific literacy also plays an important role in fostering critical thinking and evidence-based problem-solving skills.

Despite its growing importance in educational policy, Indonesian students continue to demonstrate relatively low levels of scientific literacy. According to the Programme for International Student Assessment (PISA) 2022, Indonesia achieved a scientific literacy score of 383, which remains below the OECD average of 399. In addition, only approximately 34% of Indonesian students reached the minimum level of scientific proficiency (Ayu et al.,

2025). These findings are consistent with Putra (2019), who reported that junior high school students in Koto XI Tarusan District achieved only 26.81% of the expected scientific literacy competency, indicating a very low level of performance.

The persistently low level of scientific literacy may be associated with assessment practices that primarily emphasize conceptual knowledge and factual recall rather than students' ability to apply scientific knowledge in real-world contexts, as emphasized in the PISA framework, although this relationship was not examined empirically in the present study. Therefore, assessment instruments that align with the characteristics of PISA and provide accurate and consistent measurement are needed.

Several studies have developed PISA-based scientific literacy assessment instruments and reported satisfactory evidence of content validity and practicality (Mulyana & Desnita, 2023; Widiyanti et al., 2017; Alti et al., 2021; Hasnawati et al., 2024). Most of these studies focused on instrument development, expert validation, and item quality evaluation using Classical Test Theory (CTT) or Item Response Theory (IRT). In addition, the Rasch model has been widely applied to evaluate the validity and reliability of educational assessment instruments (Susongko et al., 2019; Susongko et al., 2020; Familia et al., 2026). However, studies examining the psychometric characteristics of PISA-based scientific literacy instruments following field implementation remain limited, particularly regarding person–item targeting, item fit, and the instrument's ability to distinguish different levels of student ability through the separation index. These aspects are essential because they provide empirical evidence that an instrument not only demonstrates acceptable validity and reliability but also functions effectively under actual testing conditions.

This study employed the Rasch measurement model using the Partial Credit Model (PCM), which estimates student ability and item difficulty on a common interval scale (logits) independently of the sample and test items (Sumintono & Widhiarso, 2015). Unlike previous studies that primarily focused on instrument development and expert validation, the present study specifically examined the field-testing phase to evaluate the psychometric characteristics of an instrument that had previously demonstrated acceptable content validity and practicality. The sample size of 40 students was determined based on Linacre's (1994) recommendation that 27–61 respondents are sufficient to obtain stable item calibration within ±1 logit, provided that the distribution of respondent ability adequately matches the distribution of item difficulty.

The Rasch model provides a robust psychometric framework for measuring student ability and item difficulty on the same logit scale (Sumintono & Widhiarso, 2015). This approach enables independent estimation of person ability and item difficulty, thereby producing more objective and accurate measurements (Ayub et al., 2020). In addition, the Rasch model provides several important psychometric indicators, including item fit, person fit, targeting, reliability, and separation index, allowing researchers to evaluate both individual item performance and the overall measurement quality of an assessment instrument (Sumintono & Widhiarso, 2014; Krishnan & Idris, 2018). Furthermore, the analysis of item and respondent characteristics offers comprehensive evidence regarding the psychometric quality of the instrument (Ibrahim, 2025; Ngandoh et al., 2025).

Based on these considerations, this study aimed to analyze the psychometric characteristics of a PISA-based scientific literacy assessment instrument on ecology during the field-testing phase using the Rasch model. The analysis focused on person–item targeting, item fit, and the instrument's ability to distinguish different levels of student ability and item difficulty through the separation index. The findings are expected to provide empirical evidence of the instrument's psychometric performance under actual testing conditions, complementing the evidence of content validity and practicality established during the previous development phase.

Method

This study was part of a Research and Development (R&D) project based on the Plomp development model, which consists of three phases: preliminary research, development or prototyping, and assessment. This article specifically reports the findings from the assessment phase, focusing on the psychometric characteristics of a Programme for International Student Assessment (PISA)-based scientific literacy assessment instrument. The instrument was developed during the preliminary research and development phases and had previously undergone expert validation and practicality testing, demonstrating its suitability for field implementation. Detailed procedures and findings of the expert validation and practicality testing have been reported separately and are therefore beyond the scope of this article, which focuses exclusively on the instrument's psychometric characteristics during the field-testing phase.

The field test involved 40 students who completed a scientific literacy assessment on ecology consisting of 45 items, including 34 multiple-choice items scored dichotomously and 11 constructed-response items scored polytomously using an analytic scoring rubric. This sample size falls within the range recommended by Linacre

(1994) for achieving stable item calibration. The instrument was developed based on the PISA scientific literacy framework, encompassing the dimensions of context, competencies, and scientific knowledge. Data consisted of students' responses to all assessment items collected during a single administration.

The data were analyzed using the Rasch model with the Partial Credit Model (PCM) (Masters, 1982), which was selected because the instrument comprised both dichotomous and polytomous items. The analyses were conducted using Winsteps software.

The instrument's psychometric characteristics were evaluated using three complementary Rasch indicators: person–item targeting, item fit, and the separation index (Bond & Fox, 2015). These indicators were selected because they collectively represent three essential aspects of measurement quality within the Rasch framework. Person–item targeting evaluates the alignment between the distribution of student ability and item difficulty, item fit provides evidence of construct validity at the individual item level, and the separation index reflects the instrument's precision in distinguishing different levels of student ability and item difficulty. Targeting was examined using the Wright Map by comparing the distributions of person ability and item difficulty. Item fit was evaluated using Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point-Measure Correlation (Pt-Measure Corr). An item was considered to fit the Rasch model when it satisfied the following criteria: $0.5 \leq$ Outfit MNSQ ≤ 1.5, $-2.0 \leq$ Outfit ZSTD $\leq +2.0$, and a positive Point-Measure Correlation between 0.40 and 0.85 (Bond & Fox, 2015). The separation index was used to evaluate the instrument's ability to distinguish different levels of student ability and item difficulty. Collectively, these three indicators served as the basis for evaluating the psychometric characteristics of the PISA-based scientific literacy assessment instrument on ecology.

Results and Discussions

Psychometric Characteristics of the Instrument Based on Person–Item Targeting

Table 1 <Summary of the Rasch Model Targeting Analysis>

Parameter	Value
Mean Item Measure	0.00 logit
Item Standard Deviation	0.83 logit
Mean Person Measure	−0.45 logit
Person Standard Deviation	0.61 logit

The Rasch analysis indicated that the PISA-based scientific literacy assessment instrument on ecology demonstrated adequate person–item targeting between students' ability and item difficulty. As presented in Table 1, the mean person measure was −0.45 logits, whereas the mean item measure was fixed at 0.00 logits. The difference of 0.45 logits indicates that students' average ability was slightly lower than the average item difficulty, but the magnitude of this difference remained within a range that suggests an acceptable alignment between respondent ability and the characteristics of the instrument.

The distribution of student ability and item difficulty is illustrated in the Wright Map (Figure 1). Overall, the assessment items covered a broad range of student abilities, from low to high. However, several items were relatively difficult for most respondents, and only a limited number of items were available at the lower end of the ability continuum. These findings indicate that the measurement distribution was not yet fully balanced across all ability levels.

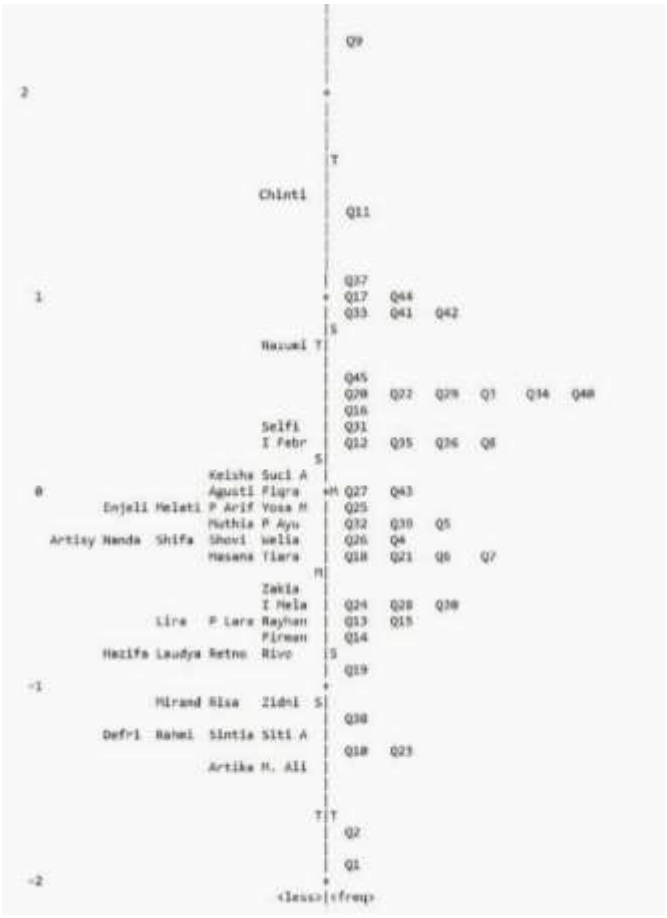

Figure 1 <Wright Map of the PISA-Based Scientific Literacy Assessment Instrument on Ecology>

The findings indicate that the instrument was able to measure the abilities of most students, although there remains room for improvement, particularly at the lower end of the ability continuum where item coverage was limited. This finding is consistent with Neumann et al. (2010), who reported that the Wright Map is a useful tool for evaluating the alignment between student ability and item difficulty while identifying ability ranges that are insufficiently represented by an instrument. A similar pattern was reported by Darman et al. (2024), who found that gaps at one end of the ability continuum are a common challenge in the development of Rasch-based scientific literacy instruments and can be addressed by developing additional items targeting the underrepresented ability range. Therefore, adding more items for students with lower ability levels should be considered to achieve a more balanced measurement distribution.

From a theoretical perspective, the alignment between student ability and item difficulty represents an important indicator of instrument quality (Bond & Fox, 2015). The smaller the difference between the mean person measure and the mean item measure, the better the instrument is targeted to the ability distribution of the respondents at the aggregate level. In the present study, the 0.45-logit difference was relatively small, indicating acceptable overall targeting. However, this average alignment does not necessarily imply that all ability levels were equally represented. As shown in the Wright Map, several items remained relatively difficult for most respondents, while the number of items targeting lower ability levels was still limited. Therefore, the instrument can be considered to demonstrate adequate person–item targeting at the aggregate level, although broader coverage of lower ability levels should be considered in future instrument development.

Psychometric Characteristics of the Instrument Based on Item Fit

The item fit analysis indicated that most items in the PISA-based scientific literacy assessment instrument functioned consistently with the Rasch model.

Table 2 <Summary of Item Fit Analysis Based on the Rasch Model>

Item Category	Number of Items	Percentage (%)
Fit	43	95.6
Misfit	2	4.4
Total	**45**	**100**

As presented in Table 2, 43 of the 45 items (95.6%) met the Rasch model fit criteria, whereas 2 items (4.4%), namely Q17 and Q18, were identified as misfit. These findings indicate that the majority of the items functioned consistently in measuring the intended construct of scientific literacy.

Table 3 <Statistics of Items Requiring Revision Based on the Rasch Model Analysis>

Item	Measure (logits)	Outfit MNSQ	Outfit ZSTD	Pt-Measure Corr	Status
Q16	0.45	1.61	1.46	0.26	Marginal
Q17	1.04	1.30	1.02	–0.05	Misfit
Q18	–0.29	1.70	2.72	0.32	Misfit

As shown in Table 3, item Q18 was classified as misfit because its Outfit Mean Square (MNSQ = 1.70) and Outfit Z-Standard (ZSTD = 2.72) exceeded the acceptable fit criteria. In contrast, item Q17 showed a negative Point-Measure Correlation (Pt-Measure Corr = −0.05), indicating that the item did not measure the scientific literacy construct in the same direction as the remaining items. In addition, item Q16 had an Outfit MNSQ value of 1.61, and was therefore classified as a marginal item requiring further review.

The high proportion of fitting items indicates that the measurement structure of the instrument is largely consistent with the theoretical expectations of the Rasch model. This finding is consistent with Susongko et al. (2019), who reported that 16 of 17 items in a PISA-based scientific literacy instrument satisfied the Rasch model fit criteria and were considered suitable for measuring scientific literacy. Similar findings were reported by Suryadi et al. (2025), who found that most items met the Rasch fit criteria and demonstrated good psychometric quality, although several items required revision to better represent the intended construct. Likewise, Bashooir and Supahar (2018) reported that the majority of items in a STEM-based scientific literacy performance assessment satisfied the Rasch model fit criteria following field testing. The consistency of these findings across different subject areas and educational contexts provides further evidence that the Rasch model is effective in identifying both well-functioning items and those requiring revision in scientific literacy assessments.

From a theoretical perspective, items that satisfy the Rasch fit criteria indicate that students' responses conform to the expectations of the measurement model, suggesting that the items consistently measure the same underlying construct (Bond & Fox, 2015). Conversely, misfitting items, such as Q17 and Q18 in the present study, indicate unexpected response patterns that may result from ambiguous item wording, incorrect answer keys, or the possibility that the items measure constructs beyond the intended domain of scientific literacy. Consequently, these items should be carefully reviewed and revised before the instrument is used more broadly. Overall, the high proportion of fitting items demonstrates that the instrument possesses good psychometric quality, although the two misfitting items (Q17 and Q18) require revision to improve their alignment with the intended construct.

Psychometric Characteristics of the Instrument Based on the Separation Index

The separation index analysis indicated that the instrument demonstrated a good ability to distinguish both student ability and item difficulty. A summary of the Rasch descriptive statistics for persons and items is presented in Table 4.

Table 4 <Summary of Rasch Descriptive Statistics for Persons and Items>

Statistical Parameter	Persons	Items
Number (N)	40	45
Mean Measure (logits)	−0.45	0.00
Standard Deviation (SD)	0.61	0.83
Maximum Measure	1.49	2.23
Minimum Measure	−1.41	−1.94

Statistical Parameter	Persons	Items
Mean Infit MNSQ	1.03	1.00
Mean Infit ZSTD	0.11	−0.01
Mean Outfit MNSQ	1.00	1.00
Mean Outfit ZSTD	0.01	0.01
Separation Index	1.71	2.18
Reliability	0.75	0.83
Cronbach's Alpha (α)	0.78	–

As presented in Table 4, the person separation value of 1.71 indicates that the instrument was able to classify students into three ability strata: high, moderate, and low. According to the Rasch instrument quality criteria proposed by Fisher (2007), separation values of 2.0 or higher and reliability values of 0.80 or higher are considered good, whereas values below these thresholds are classified as marginal. Based on these criteria, the person separation value of 1.71 and the person reliability value of 0.75 in the present study fall into the marginal category. Nevertheless, these values remain sufficient in practical terms to distinguish students into three ability strata. These findings indicate that the instrument has an adequate, although not yet optimal, ability to classify students according to their level of scientific literacy. Therefore, increasing both the number and diversity of items in future instrument development may further improve the precision of person separation.

The item separation value of 2.18 also produced three levels of item difficulty: easy, moderate, and difficult. Unlike the person separation, the item separation value exceeded the recommended threshold of 2.0 and was therefore classified as good according to Fisher's (2007) criteria. This finding indicates that the assessment items provide an adequate range of difficulty, enabling the instrument to measure students across different ability levels.

These findings are consistent with those of Arabbani et al. (2019), who reported that the item separation index can be used to evaluate an instrument's ability to distinguish different levels of item difficulty within the Rasch measurement framework. Similar findings were reported by Mellyzar et al. (2024), who demonstrated that person separation and item separation analyses provide useful information about an instrument's ability to classify students according to their ability levels while differentiating item difficulty. Likewise, Familia et al. (2026), in developing a Rasch-based scientific literacy assessment instrument for elementary school students, reported a similar separation pattern and further examined the potential for measurement bias across groups using Differential Item Functioning (DIF). Because DIF analysis was beyond the scope of the present study, potential measurement bias across groups could not be evaluated and should therefore be considered a limitation to be addressed in future research.

From a theoretical perspective, higher separation values indicate that an instrument has greater discriminating power to distinguish both respondents and items, resulting in more stable and consistent measurement (Bond & Fox, 2015). Based on the findings of this study, the item separation value of 2.18 met Fisher's (2007) criterion for good instrument quality, whereas the person separation value of 1.71 was classified as marginal, although it remained within an acceptable range for distinguishing students into three ability strata. Therefore, the instrument demonstrated good measurement characteristics in differentiating item difficulty, although its ability to distinguish students with different levels of scientific literacy could be further improved in future instrument development.

Overall Psychometric Characteristics of the Instrument

Based on the results of the three main Rasch indicators, person, item targeting, item fit, and the separation index, the PISA-based scientific literacy assessment instrument on ecology demonstrated adequate psychometric characteristics for measuring students' scientific literacy. The instrument showed acceptable person–item targeting at the aggregate level, although item coverage for students with lower ability remained limited. In addition, most items (95.6%) satisfied the Rasch model fit criteria, and the instrument was able to distinguish both student ability and item difficulty into three categories. However, the discriminating power for student ability, as indicated by the person separation index, remained in the marginal category.

Although two items did not satisfy the Rasch model fit criteria, these findings provide valuable diagnostic information for further instrument refinement. The presence of a small number of misfitting items does not substantially affect the overall quality of the instrument, as the vast majority of items functioned consistently with the Rasch measurement model.

Several limitations should be considered when interpreting the findings of this study. First, the data were collected through a single administration involving 40 students from a single school; therefore, caution is needed when generalizing the findings to broader populations or different educational contexts. Second, this study did not examine the instrument's convergent validity against an established scientific literacy assessment, nor did it evaluate its ecological validity in authentic classroom assessment settings. Consequently, the validity evidence presented in this study is limited to Rasch-based indicators, including person–item targeting, item fit, and the separation index, and does not encompass other sources of validity evidence. Third, as noted previously, Differential Item Functioning (DIF) analysis was not conducted; therefore, potential measurement bias across student groups could not be evaluated. Finally, evidence from expert validation and practicality testing was reported separately during the earlier stages of instrument development and is therefore beyond the scope of the present article. These limitations provide opportunities for future studies to establish more comprehensive evidence of the instrument's validity.

Considering both the findings and the limitations of the study, the PISA-based scientific literacy assessment instrument on ecology demonstrates adequate psychometric characteristics based on Rasch model analysis. The instrument has the potential to serve as an alternative assessment tool for measuring students' scientific literacy in biology learning, particularly on the topic of ecology. Nevertheless, further refinement of item coverage and the accumulation of additional validity evidence are recommended to strengthen the instrument's measurement quality.

Conclusions

The PISA-based scientific literacy assessment instrument on ecology demonstrated adequate psychometric characteristics for measuring students' scientific literacy based on Rasch model analysis. These characteristics were reflected in acceptable person–item targeting at the aggregate level, with a mean difference of 0.45 logits between student ability and item difficulty; 43 of the 45 items (95.6%) satisfying the Rasch model fit criteria; and an item separation value of 2.18, which was classified as good according to Fisher's (2007) criteria and successfully distinguished item difficulty into three strata. Although the person separation value of 1.71 was sufficient to classify students into three ability strata, it was categorized as marginal according to Fisher's (2007) criteria, indicating that the instrument's ability to differentiate student ability could still be improved.

The conclusions of this study are limited to Rasch-based psychometric evidence obtained from a single field test involving 40 students from one school. Convergent validity with other established scientific literacy assessments, ecological validity in authentic classroom assessment contexts, and potential measurement bias across groups through Differential Item Functioning (DIF) were not examined and should therefore be addressed in future research.

Within these limitations, the instrument has the potential to serve as an alternative assessment for measuring scientific literacy in biology education, particularly on the topic of ecology. Future studies are recommended to develop additional items targeting lower ability levels to achieve a more balanced measurement distribution and to strengthen the instrument by providing more comprehensive evidence of its validity and reliability.

Acknowledgments

The authors would like to express their sincere gratitude to the Department of Biology Education, Universitas Negeri Padang, for providing academic support during this study. The authors also thank the teachers and students who participated in the field testing of the instrument. Their cooperation and valuable contributions made this research possible.

References

Alti, R. P., Lufri, L., Helendra, H., & Yogica, R. (2021). Pengembangan instrumen asesmen berbasis literasi sains tentang materi keanekaragaman hayati kelas X. *Journal for Lesson and Learning Studies, 4*(1). https://doi.org/10.23887/jlls.v4i1.34270

Arabbani, F. K., Mulyani, S., Mahardiani, L., & Ariani, S. R. D. (2019). Analysis the quality of instrument for measuring chemical literacy abilities of high school student using Rasch model. *AIP Conference Proceedings, 2194*, 020007. https://doi.org/10.1063/1.5139739

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

Ayub, M. R. S. S. N., Istiyono, E., Munadi, S., Permadi, C., Pattiserlihun, A., & Sudjito, D. N. (2020). Analisa penilaian soal fisika menggunakan model Rasch dengan program R. *Jurnal Sains dan Edukasi Sains, 3*(2), 46–52. https://doi.org/10.24246/juses.v3i2p46-52

Bashooir, K., & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan, 22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.19590

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

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching, 46*(8), 865–883. https://doi.org/10.1002/tea.20333

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

Familia, M. N., Haerani, R. P. R., & Suhartini, E. (2026). Development of scientific literacy skills test instruments in elementary school: Analysis of the Rasch model. *Integrated Science Education Journal, 7*(2), 255–264. https://doi.org/10.37251/isej.v7i2.2383

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21*(1), 1095.

Hasnawati, Syazali, M., & Putra, G. P. (2023). Pengembangan asesmen literasi sains berbasis PISA untuk siswa sekolah dasar. *BADA'A: Jurnal Ilmiah Pendidikan Dasar, 5*(2), 240–250. https://doi.org/10.37216/badaa.v5i1.1213

Ibrahim, N. L. (2025). Analisis literatur penggunaan model Rasch untuk validasi instrumen asesmen pendidikan: Tren, temuan, dan implikasi. *Lencana: Jurnal Inovasi Ilmu Pendidikan, 3*(2), 46–60. https://doi.org/10.55606/lencana.v3i2.5012

Khery, Y., Sarjan, M., Ahzan, S., & Efendi, I. (2022). Konseptualisasi literasi sains mengacu pada kerangka sains PISA sejak tahun 2000. *Educatoria: Jurnal Ilmiah Ilmu Pendidikan, 2*(4), 200–231. https://doi.org/10.36312/ejiip.v2i4.117

Krishnan, S., & Idris, N. (2018). Using partial credit model to improve the quality of an instrument. *International Journal of Evaluation and Research in Education, 7*(4), 313–321. https://doi.org/10.11591/ijere.v7i4.15146

Lie, S., & Linnakylä, P. (2004). Nordic PISA 2000 in a sociocultural perspective. *Scandinavian Journal of Educational Research, 48*(3), 227–230. https://doi.org/10.1080/00313830410001695709

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Mellyzar, M., Andriani, R., Lukman, I. R., Muttakin, M., Wati, E., Pasaribu, A. I., & Fadli, M. R. (2024). Digital assessment tools as a media for assessing high school science learning. *Jurnal Penelitian Pendidikan IPA, 10*(3), 1362–1374. https://doi.org/10.29303/jppipa.v10i3.6416

Mulyana, V., & Desnita, D. (2023). Empirical validity and reliability of the scientific literacy assessment instrument based on the tornado physics enrichment book. *Jurnal Penelitian Pendidikan IPA, 9*(5), 3961–3967. https://doi.org/10.29303/jppipa.v9i5.3290

Neumann, I., Neumann, K., & Nehm, R. (2010). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education, 33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Ngandoh, S. T., Riandi, Rahmat, A., & Muslim. (2025). Rasch model analysis of item quality for measuring conceptual understanding of the human body system. *Jurnal Inovasi Pendidikan dan Sains, 6*(1), 89–95.

Permatasari, A. I., Maftukh, F. D., As'ari, & Muda, R. C. (2026). Insights from a teaching assistance program in Thailand: A scientific literacy analysis of a lower secondary science textbook. *Science Education and Application Journal, 8*(1), 76–87. https://doi.org/10.30736/seaj.v8i1.1346

Putra, E. N. (2019). *Analisis kemampuan literasi sains biologi siswa SMP se-Kecamatan Koto XI Tarusan Kabupaten Pesisir Selatan* (Bachelor's thesis). STKIP PGRI Sumatera Barat. https://repo.upgrisba.ac.id/id/eprint/6964/

Schmaltz, R. M., Jansen, E., & Wenckowski, N. (2017). Redefining critical thinking: Teaching students to think like scientists. *Frontiers in Psychology, 8*, 459. https://doi.org/10.3389/fpsyg.2017.00459

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial* (Edisi revisi). Trim Komunikata.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Institut Teknologi Sepuluh November.

Suryadi, N. R. S., Nurmegawati, L., Mu'aziyah, S. E. S., & Perdani, A. S. (2025). Rasch model for analysis of scientific attitude instruments in the context of secondary school science education. *Equator Science Journal, 3*(2), 98–106. https://doi.org/10.61142/esj.v3i2.234

Turiman, P., Omar, J., Daud, A. M., & Osman, K. (2012). Fostering the 21st century skills through scientific literacy and science process skills. *Procedia – Social and Behavioral Sciences, 59*, 110–116. https://doi.org/10.1016/j.sbspro.2012.09.253

Widiyanti, I. S. R., Winata, A., & Cacik, S. (2017). Pengembangan perangkat penilaian literasi sains berorientasi *Programme for International Student Assessment* (PISA). *Education and Human Development Journal, 2*(1). https://doi.org/10.33086/ehdj.v2i1.377