Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

Validity and Reliability of the Sustainability Literacy Knowledge Domain among Pre-Service Science Teachers: A Rasch Analysis

Aldeva Ilhami, Topik Hidayat*, Riandi, Siti Sriyati
Universitas Pendidikan Indonesia
Setiabudhi Street No. 229, Bandung, Indonesia
Corresponding author: topikhidayat@upi.edu

Article Information

Article History:
Received: September 23, 2025
Revised: February 03, 2026
Published: April 02, 2026

Keywords:
Sustainability literacy, Validity, Reliability, Rasch analysis

ABSTRACT

Sustainability literacy is increasingly recognized as a critical competence for pre-service science teachers in addressing global sustainability challenges. The availability of psychometrically robust instruments in the Indonesian higher education context remains limited. This study aims to validate a sustainability literacy instrument within the cognitive domain using a Rasch dichotomous model to evaluate reliability, validity, item-fit statistics, and unidimensionality. The assessment instrument was designed to measure the extent of sustainability literacy among pre-service science teachers. A cross-sectional survey design was employed, involving 144 students from universities in Riau, Indonesia. Participants completed a sustainability literacy instrument comprising 30 multiple-choice items, developed based on the Sulitest framework. Data were analyzed using the Rasch measurement model with the assistance of Winsteps version 3.73 software, focusing on item fit indicators (Outfit MNSQ, ZSTD, and PTMEA Corr.), reliability indices, and principal component analysis of residuals. The results demonstrated that the instrument achieved good internal consistency (α = 0.85), high item reliability (0.94), and stable person reliability (0.84). Two items were identified as misfitting and recommended for revision, while the remaining items met acceptable MNSQ and PTMEA criteria. Furthermore, unidimensionality was confirmed, indicating that the instrument measures a single latent construct of cognitive sustainability literacy. The validated instrument can be used by teacher education institutions to diagnose pre-service teachers' sustainability knowledge and to inform curriculum improvement. Its distinctive contribution lies in adapting an internationally recognized framework to the Indonesian higher education context through Rasch-based validation.

How to Cite
Ilhami, A., Hidayat, T., Riandi, Sriyati, S. (2026). Validity and Reliability of the Sustainability Literacy Knowledge Domain among Pre-Service Science Teachers: A Rasch Analysis. *Al Jahiz: Journal of Biology Education Research*. 7(1), 107–121. DOI: https://doi.org/10.32332/al-jahiz.v7i1.11550.

Published by
Al-Jahiz: Journal of Biology Education Research
Website
https://e-journal.metrouniv.ac.id/index.php/Al-Jahiz/index

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

INTRODUCTION

Sustainable development (SD) has become one of the most urgent global agendas of the 21st century, shaping the future of environmental, social, and economic systems worldwide (Hariram et al., 2023). Despite substantial progress across societies, persistent challenges such as poverty, biodiversity loss, and climate change remain unresolved, requiring innovative and systemic responses (Vasiliev, 2021) . Addressing these issues demands new capacities among the current generation, particularly in terms of sustainability literacy, which broadly encompasses knowledge, skills, competencies, and awareness. Sustainability literacy reflects the knowledge base, mindset, and skills that enable individuals to commit to building a sustainable future and to make informed, effective decisions toward this goal (Saarna & Laius, 2025).

Higher education institutions play a pivotal role in advancing global sustainability by fostering pro-environmental knowledge, attitudes, and behaviors. Sustainability literacy has therefore become a core competency for younger generations, especially for pre-service science teachers who are expected to transfer sustainability-related knowledge and skills to future learners (García-González et al., 2020; Nousheen et al., 2024). Beyond environmental awareness, sustainability literacy entails critical thinking, responsible decision-making, and the integration of sustainability values into daily practice (Ekaputri et al., 2025). It is widely conceptualized as a key learning outcome of formal education (Brundiers et al., 2021). Over the past decades, higher education institutions have not only integrated sustainability into their curricula but also developed standardized instruments such as the Assessment of Sustainability Knowledge (Zwickle & Jones, 2018), the Sustainability Attitude Scale (SAS) (Zwickle & Jones, 2018), and the Sulitest (Décamps et al., 2017) to assess sustainability knowledge and attitudes.

In Indonesia, the validation of sustainability literacy instruments for pre-service teachers has largely relied on classical test theory (CTT). Fildzah Sani (2025) and (Putri et al., 2025) employed pearson correlation to assess validity and reliability among high school students, while Himmah et al., (2025) applied the same approach in junior high schools. However, sustainability literacy instruments specifically designed for pre-service science teachers remain limited, and existing tools have not sufficiently addressed item-level diagnostics or construct validity in teacher education contexts. Although widely used, CTT suffers from significant limitations, including sample dependency and restricted diagnostic capacity. Reliability and validity parameters often vary across populations, and CTT does not provide detailed insights into item-level performance, respondent

ability, or structural validity such as unidimensionality (Bond et al., 2021; Markus & Borsboom, 2024). CTT is unable to identify item bias or to separate reliability estimates for items and respondents (Sumintono & Widhiarso, 2015). These constraints undermine the accuracy of literacy assessments and reduce the effectiveness of educational interventions for pre-service teachers.

To address these limitations, the rasch measurement model offers a more robust and diagnostic alternative. Rooted in Item Response Theory (IRT), rasch analysis allows researchers to evaluate the validity and reliability of instruments while ensuring interval-level scaling (Boone, 2016). It further provides tools for identifying misfitting items, analyzing person ability, and assessing unidimensionality (Yu, 2020). Rasch analysis enables the development of instruments that are both conceptually valid and empirically sound. In this study, Rasch analysis is used primarily to strengthen measurement precision and construct validation, while broader methodological reflections are reserved for the Discussion section.

Empirical evidence demonstrates that rasch analysis yields more accurate and meaningful results compared to classical approaches. It detects biased items, offers clear guidance for revisions (Müller, 2020; Prasetya & Pratama, 2023), and enhances measurement precision by aligning item difficulty with respondent ability (Metsämuuronen, 2023). Susbiyanto et al. (2025) research utilized the Rasch model to examine problem-solving skills, thereby allowing students to be systematically classified according to their ability levels.In this study, ethnoscience is understood as the integration of local and indigenous knowledge systems into science education, emphasizing culturally grounded understandings of environmental interactions (Aikenhead & Ogawa, 2007; Snively & Corsiglia, 2001). Ethnoscience is relevant to sustainability literacy because it contextualizes global sustainability concepts within local socio-ecological practices, which are reflected in several item contexts of the instrument.

Despite the growing emphasis on sustainability education, no prior study has developed and validated a Rasch-based sustainability literacy knowledge instrument specifically for pre-service science teachers in the Riau context. This study addresses this gap by adapting the Sulitest cognitive framework and embedding ethnoscience perspectives into item design, followed by rigorous Rasch-based validation. The scope includes examining item fit, unidimensionality, and both item and person reliability. The findings are expected to inform future research and educational practice, supporting the achievement of the Sustainable Development Goals (SDGs) through strengthening the quality of pre-service science teachers. Based on this rationale, the objective of the study is to

validate a cognitive-domain sustainability literacy instrument for pre-service science teachers in Indonesia using Rasch analysis

RESEARCH METHODS

Research design

This study employed a cross-sectional survey design, focusing on the collection of information from participants through a structured questionnaire. Data collection was conducted in an offline classroom setting during regular lecture hours at participating universities in Riau Province. The test administration lasted approximately 50 minutes and was supervised by the researchers. Prior to testing, participants received standardized instructions regarding the purpose of the study and how to complete the instrument. Ethical approval for this study was obtained from the institutional research ethics committee of the authors' affiliated university. Participation was voluntary, and all students provided informed consent prior to data collection. Participants were informed that their responses would be used solely for research purposes, anonymity would be maintained, and withdrawal was possible at any time without academic consequences.

Instruments

The instrument used in this study measured sustainability literacy within the cognitive domain, following the framework proposed by Décamps et al. (2017). It consisted of four main constructs: (1) Sustainable humanity and ecosystems on planet Earth, (2) Global and local human-constructed systems to answer people's needs, (3) Transitions towards sustainability, and (4) Roles to play to create and maintain individual and systemic changes. The instrument comprised 30 multiple-choice items, each with four response options. Items were scored dichotomously (1 = correct, 0 = incorrect) and analyzed using the Rasch dichotomous model. Content validity was established through expert judgment involving three experts in science education and sustainability. The review process was conducted in two rounds, focusing on conceptual relevance, clarity of wording, and alignment with indicators. Experts provided qualitative feedback, which resulted in minor revisions to item phrasing and contextual clarity. No items were removed at this stage. A pilot test was subsequently conducted to ensure readability and comprehensibility. Table 1 summarizes the indicators and sub-indicators used to measure cognitive aspects of sustainability literacy.

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

Tabel 1. Indicator of sustainability literacy

Indicator	Sub-Indicator	Item
Sustainable humanity and ecosystems on planet Earth	Ecosystem	7
	Humanity	2
	Sustainability concepts	2
Global and local human-constructed systems to answer people's needs	Social structure and governance	3
	Education and culture	2
	Water issues	4
Transitions towards sustainability	Acceleration of system change	2
	Implementation of sustainability-based concepts, instruments, and frameworks	2
	Examples and ideas to learn from	2
Roles to play to create and maintain individual & systemic changes	Awareness of contribution and impact	2
	Ways to act effectively and efficiently	2

Participants

The participants were 144 pre-service science teachers enrolled in science education programs at higher education institutions in Riau Province, Indonesia. Inclusion criteria were active enrollment in a science education program and completion of at least four semesters of study. Quota sampling was applied to ensure proportional representation across universities and academic semesters. The sample size was considered adequate for Rasch analysis. Table 2 presents the demographic characteristics of the participants. The sample consisted of 22 males (15%) and 122 females (85%), with 58 students in the fourth semester (40%) and 86 students in the sixth semester (60%). In terms of institutional distribution, 118 students were from UIN Sultan Syarif Kasim Riau (82%), while 26 students were from Universitas Muhammadiyah Riau (18%).

Data analysis

The data collected from student responses were first tabulated in Microsoft Excel, coded, and then converted into.prn format for analysis using WINSTEPS version 3.73, a software package for Rasch measurement modeling. The Rasch dichotomous model was applied for right–wrong item responses (Ilhami et al., 2025). In this study, validity was conceptualized through item fit statistics and unidimensionality analysis. Person misfit was examined using Outfit MNSQ and ZSTD criteria; no respondents were removed, as all met acceptable thresholds. Analytical parameters were employed to evaluate the quality of the instrument: validity, reliability, unidimensionality, item fit analysis, person fit, separation index, and Wright Map.

Item fit analysis was conducted using three indices: Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point-Measure Correlation (PTMEA Corr.). An item was retained if it met at least one of the three criteria, while items that failed two or more criteria were recommended

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

for removal or revision. The cut-off values for item fit were: MNSQ between 0.5 and 1.5, ZSTD between -2.0 and +2.0, and PTMEA Corr. between 0.40 and 0.85 (Bond et al., 2021).

Instrument reliability was evaluated using item reliability, person reliability, and Cronbach's alpha. The criteria for interpretation followed (Sumintono & Widhiarso, 2015), where reliability values below 0.67 were categorized as weak, 0.67–0.80 as sufficient, 0.80–0.90 as good, 0.91–0.94 as very good, and above 0.94 as excellent. These parameters were employed to ensure that the instrument consistently measured sustainability literacy across different respondent profiles and item distributions.

RESEARCH RESULT

The analysis of students' sustainability literacy in the cognitive domain was conducted using the Rasch measurement model to obtain both technical and substantive information about instrument performance. The results of the analysis addressed five main aspects: (a) item fit based on Mean Square (MNSQ), Z-Standard (ZSTD), and Point-Measure Correlation (PTMEA Corr.); (b) unidimensionality; (c) item reliability; (d) person reliability; and (e) person ability distribution visualized through the Wright Map. The statistical outcomes for MNSQ, ZSTD, and PTMEA Corr. are summarized in Table 5.

Item Fit

Item fit analysis was performed to ensure that each item functioned appropriately in measuring the intended construct and did not cause confusion or misinterpretation among respondents. The criteria used included Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point-Measure Correlation (PTMEA Corr.). An item was considered acceptable if at least one of these criteria was met (Sumintono & Widhiarso, 2015). Beyond statistical fit, substantive interpretation was also considered to support meaningful instrument refinement

Table 5. Item fit

No	MNSQ	ZSTD	PTMEASUR-AL	Item
5	1.96	4.78	-0.06	A5
26	1.84	2.88	-0.08	C6
20	1.44	2.61	0.33	B9
16	1.2	1.2	0.27	B5
12	1.16	1.15	0.44	B1
6	1.05	0.44	0.38	A6
8	1.1	0.77	0.44	A8
21	0.99	-0.01	0.42	C1
2	1.03	0.27	0.41	A2

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

No	MNSQ	ZSTD	PTMEASUR-AL	Item
4	0.96	-0.23	0.44	A4
7	0.92	-0.57	0.45	A7
9	0.86	-0.48	0.4	A9
30	0.89	-0.79	0.46	D4
3	0.94	-0.12	0.41	A3
10	0.9	-0.76	0.5	A10
13	0.84	-1.24	0.51	B2
15	0.87	-1.04	0.51	B4
17	0.82	-0.74	0.45	B6
22	0.79	-1.44	0.51	C2
1	0.77	-1.02	0.48	A1
18	0.8	-1.31	0.52	B7
24	0.85	-1.13	0.54	C4
28	0.76	-1.12	0.49	D2
27	0.8	-1.15	0.52	D1
19	0.83	-1.37	0.55	B8
29	0.73	-1.89	0.55	D3
14	0.65	-1.61	0.53	B3
23	0.79	-1.43	0.57	C3
25	0.68	-1.69	0.61	C5
11	0.55	-1.01	0.49	A11

Table 5 shows that most items demonstrated acceptable fit within the expected ranges: MNSQ values between 0.5 and 1.5, ZSTD between -2.0 and +2.0, and PTMEA Corr. between 0.40 and 0.85 (Boone, 2016). Only two items (Items 5 and 26) did not meet the fit requirements across multiple criteria and should therefore be revised. Items 20 and 16, although not meeting ZSTD thresholds, were retained because they still satisfied the MNSQ criterion. These findings confirm that the majority of items contributed meaningfully to the measurement of the intended construct.

Instrument reliability

Reliability testing was conducted to verify the consistency and stability of the instrument in differentiating both items and respondents along the sustainability literacy continuum. Within Rasch modeling, reliability is assessed through three indicators: person reliability, item reliability, and Cronbach's alpha (Sumintono & Widhiarso, 2015).

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

Table 6. Validity and reliability measurement

Measurement	Score
Pearson Measurement	
Person reliability	0.84
Separation	2,27
MNSQ	0,96
ZSTD	0,06
Item Measurument	
Item reliability	0.94
separation	4,09
Cronbach alpha	0,85

As presented in Table 6, the instrument demonstrated excellent reliability indices. Item reliability reached 0.94 with a separation index of 4.09, indicating a highly stable scale capable of distinguishing items across a wide spectrum of difficulty levels. Person reliability was reported at 0.84 with a separation index of 2.27, suggesting that the instrument was capable of distinguishing participants into at least two to three distinct ability levels. Cronbach's alpha reached 0.85, confirming high internal consistency between items and respondents.

Person ability

The Wright Map (Figure 1) generated through Rasch analysis provides a comprehensive visualization of the interaction between item difficulty and person ability in the context of sustainability literacy assessment. On the left-hand side of the map, the distribution of respondents is presented according to their estimated ability levels, while on the right-hand side, the relative difficulty of items is displayed. This representation allows for an immediate comparison of whether the test items align appropriately with the range of student abilities, ensuring that the instrument is both challenging and fair across different levels of competence.

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

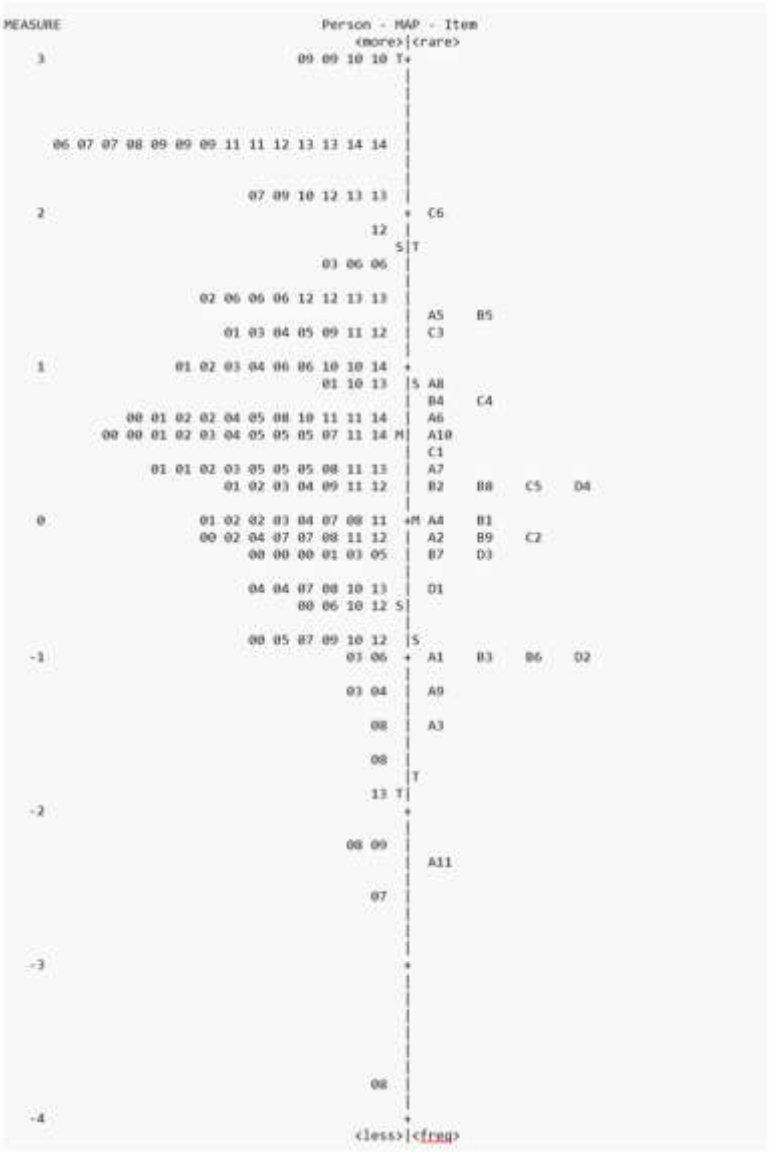

Figure 1. Wright map analysis of students' sustainability literacy

The Wright Map results indicate that student ability ranged from approximately -2.0 to +3.0 logits, while item difficulty spanned a comparable range. No strong floor or ceiling effects were observed, as most students clustered around the mid-range of ability with items available to differentiate higher- and lower-performing respondents. A brief inspection suggested similar ability distributions across semesters and institutions, although subgroup comparisons were not formally tested.

The instrument functions well in differentiating levels of sustainability literacy among pre-service science teachers. Most students clustered around the mid-range of ability, indicating that the test was appropriately targeted for the sample. A few items were positioned at higher levels of

Al-Jahiz: Journal of Biology Education Research
ISSN 2722-5070 (Print) ISSN 2722-5275 (Online)
Vol. 7 No. 1 January-June 2026
Available online at:
http://e-journal.metrouniv.ac.id/index.php/Al-Jahiz

difficulty, such as C6, which suggests that these questions were particularly challenging and therefore useful for identifying students with stronger sustainability literacy. Certain items, such as A11, appeared at the lower end of the difficulty scale, meaning that they were relatively easy and accessible to students with lower ability levels. This spread of item difficulty is advantageous because it allows the instrument to discriminate effectively across a wide spectrum of respondent abilities.

Unidimensionality

Unidimensionality analysis was conducted to confirm whether the instrument measured a single latent construct using principal component analysis (PCA) of residuals. It was employed, with two criteria used as benchmarks: (a) raw variance explained by the measures should exceed 20%, and (b) unexplained variance in the first contrast should not exceed 15% (Sumintono & Widhiarso, 2015).

Table 7. Unidimensionality measurement

Measurement	Score
Raw variance	27.5%
Unexplained variance 11st contrast	6.1%

The results (Table 7) showed that the raw variance explained by the measures was 27.5%, surpassing the minimum threshold, while the unexplained variance in the first contrast was 6.1%, well below the 15% limit. These outcomes confirm that the instrument is unidimensional, meaning that all items coherently measure the cognitive domain of sustainability literacy.

DISCUSSION

This study aimed to validate a sustainability literacy instrument in the cognitive domain for pre-service science teachers in Indonesia using the Rasch measurement model. The findings revealed that the instrument fulfilled the essential assumptions of Rasch analysis, including item fit, reliability, and unidimensionality, thereby providing strong evidence of its validity and reliability. The item fit analysis indicated that the majority of the items were within the acceptable range for Outfit Mean Square (MNSQ), Z-Standard (ZSTD), and Point-Measure Correlation (PTMEA Corr.), with only two items requiring revision. Rasch analysis has been widely acknowledged for its ability to identify misfitting items and to ensure that each item contributes meaningfully to the underlying construct being measured (Boone, 2016; Müller, 2020). The identification of such items should not be viewed as a weakness but rather as evidence of the diagnostic power of Rasch analysis, which allows researchers to refine instruments and improve

construct validity. This aligns with the view of Sumintono and Widhiarso (2015), who argued that Rasch analysis goes beyond psychometric evaluation and serves as a guide for enhancing the overall quality of measurement tools. The rasch model emphasizes construct validity by ensuring that each item contributes consistently to the latent trait being measured. Item fit provides evidence of how consistently respondents' patterns of responses align with expected measurement outcomes. Identifying misfitting items at this stage is crucial, as it allows researchers to refine instruments and prevent measurement errors in subsequent applications (Müller, 2020).

The reliability indices obtained in this study further affirm the robustness of the instrument. Item reliability was found to be 0.94, categorized as very high, while person reliability was reported at 0.84, indicating consistent responses among participants. These findings are in line with previous studies that employed rasch analysis for evaluating educational instruments and similarly reported high reliability levels (Prasetya & Pratama, 2023; Metsämuuronen, 2023). The Cronbach's alpha of 0.85 provides further support for internal consistency. As Bond et al. (2021) highlighted, reliability values above 0.80 signify measurement stability and minimize the risk of error, thereby enhancing the interpretability and trustworthiness of the results.

One of the most significant contributions of this study lies in its confirmation of the instrument's unidimensionality. The Principal Component Analysis (PCA) of residuals demonstrated that the variance explained by measures reached 27.5%, surpassing the minimum threshold of 20%, while the unexplained variance in the first contrast was only 6.1%, far below the 15% benchmark. These findings confirm that the instrument is unidimensional, meaning that all items coherently measure the same latent construct. Bond et al. (2021) noted that unidimensionality is often assumed but rarely tested empirically in social science research, which can weaken the credibility of findings. By empirically validating this assumption, the present study provides strong methodological legitimacy and enhances the credibility of the instrument for assessing sustainability literacy. In social science research, unidimensionality is often assumed rather than tested; however, in new contexts such as sustainability education, empirical verification is essential to ensure validity and precision (Bond et al., 2021).

The development of valid and reliable instruments for measuring sustainability literacy is of critical importance. Sustainability education has been recognized as a key driver in addressing global challenges such as climate change, biodiversity loss, and social inequality (Hariram et al., 2023; Vasiliev, 2021). García-González et al. (2020) and Nousheen et al. (2024) emphasize that

pre-service teachers play a crucial role in embedding sustainability values and knowledge into formal education systems. The validated instrument in this study represents an important contribution for higher education institutions, as it provides a reliable tool to assess and monitor the cognitive dimension of sustainability literacy among pre-service science teachers.

The findings of this research are also relevant in the broader debate between classical test theory (CTT) and modern approaches such as Rasch. Previous studies in Indonesia have predominantly used CTT methods, particularly Pearson correlation, to validate sustainability literacy instruments (Fildzah, 2025; Putri et al., 2025; Himmah et al., 2025). However, CTT is limited by its dependence on sample characteristics and its inability to provide detailed diagnostic information about item quality and respondent ability (Markus & Borsboom, 2024). In contrast, Rasch analysis allows for interval scaling, separate evaluation of item and person reliability, and the detection of item bias (Yu, 2020; Sumintono & Widhiarso, 2015). This study thus provides empirical support for the superiority of Rasch analysis in validating sustainability literacy instruments.

The ability of the instrument to differentiate levels of respondent ability, as demonstrated by the person reliability and Wright Map, offers substantial value for educational practice. By identifying students with different levels of sustainability literacy, educators can design targeted interventions to address specific gaps in knowledge and competencies. Saarna and Laius (2025) argued that sustainability literacy encompasses not only knowledge but also attitudes and competencies that enable individuals to make responsible decisions. Instruments capable of accurately differentiating these competencies are essential for improving the quality of sustainability education.

The presence of misfitting items (Items 5 and 26) in this study indicates the need for refinement. While the overall instrument demonstrated strong reliability and validity, the existence of such items suggests that revisions are necessary to further optimize measurement quality. Müller (2020) and Prasetya & Pratama (2023) emphasized the importance of iterative refinement in instrument development to achieve precision and accuracy. Accordingly, revising these items will strengthen the robustness of the instrument and ensure that it continues to provide valid assessments across diverse populations of pre-service teachers.

Beyond methodological contributions, the validation of this sustainability literacy instrument carries broader implications for educational practice and policy. At the institutional level,

universities can employ the instrument as a diagnostic tool to evaluate the cognitive dimension of sustainability literacy among pre-service science teachers. By identifying strengths and weaknesses in students' knowledge, educators can design targeted interventions to foster critical thinking, responsible decision-making, and the integration of sustainability principles into professional practice. This aligns with Brundiers et al. (2021), who highlighted the need for competency-based frameworks that not only impart sustainability knowledge but also prepare learners to act effectively in addressing global challenges. From a pedagogical perspective, the validated instrument supports curriculum development and instructional design. García-González et al. (2020) emphasized that sustainability education should go beyond awareness and aim to develop higher-order competencies such as problem solving and systems thinking. The instrument validated in this study provides reliable feedback for curriculum designers to evaluate the extent to which these competencies are being achieved while enabling continuous monitoring of students' progress. At the policy level, the findings suggest that validated instruments should be integrated into national education quality assurance systems. Policymakers and accreditation bodies can adopt such tools to evaluate whether teacher education programs align with national commitments to the Sustainable Development Goals (SDGs), particularly SDG 4 on quality education and SDG 13 on climate action. As Saarna and Laius (2025) argued, reliable assessments of sustainability literacy are essential for ensuring that education systems contribute meaningfully to national and global sustainability agendas.

CONCLUSION

This study confirms that the sustainability literacy instrument satisfies key Rasch measurement assumptions, as evidenced by acceptable item fit, high reliability, and confirmed unidimensionality. Item reliability (0.94) and person reliability (0.84) indicate that the instrument consistently differentiates item difficulty and respondent ability, while the variance explained by the measures (27.5%) supports internal construct validity. Only two items were identified as misfitting and require minor revision. Beyond its psychometric strength, this study offers a clear practical contribution. The validated instrument can be used by teacher education institutions to assess and monitor pre-service science teachers' sustainability knowledge, diagnose learning needs, and inform curriculum improvement, particularly within sustainability-oriented science education programs. From a policy and practice perspective, the instrument supports efforts to align teacher education with sustainability competencies and national commitments to the Sustainable

Development Goals, especially SDG 4 (quality education). Future research should extend validation to broader geographic contexts, incorporate affective and behavioral dimensions of sustainability literacy, and employ longitudinal designs to examine changes in sustainability competencies over time.

REFERENCE

Bond, T. G., Yan, Z., & Heene, M. (2021). Applying the Rasch Model:Fundamental Measurement in The Human Science. In *Applying the Rasch Model* (Fouth Edit). Routledge. https://doi.org/10.4324/9781410614575

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE—Life Sciences Education*, *15*(4), rm4.

Brundiers, K., Barth, M., Cebrián, G., Cohen, M., Diaz, L., Doucette-Remington, S., Dripps, W., Habron, G., Harré, N., Jarchow, M., Losch, K., Michel, J., Mochizuki, Y., Rieckmann, M., Parnell, R., Walker, P., & Zint, M. (2021). Key competencies in sustainability in higher education—toward an agreed-upon reference framework. *Sustainability Science*, *16*(1), 13–29. https://doi.org/10.1007/s11625-020-00838-2

Décamps, A., Barbat, G., Carteron, J. C., Hands, V., & Parkes, C. (2017). Sulitest: A collaborative initiative to support and assess sustainability literacy in higher education. *International Journal of Management Education*, *15*(2), 138–152. https://doi.org/10.1016/j.ijme.2017.02.006

Ekaputri, R. Z., Hidayat, T., Surtikanti, H. K., & Surakusumah, W. (2025). Prospective science teachers in the conservation education study. *AIP Conference Proceedings*, *3206*(1), 080033. https://doi.org/10.1063/5.0259754

Fildzah, S. A., Fajrina, S., Ardi, & Selaras, G. H. (2025). Analisis Literasi Keberlanjutan Peserta Didik Fase E SMA Negeri 2 Padang pada Materi Keanekaragaman Hayati. *Jurnal Jendela Pendidikan*, *5*.

García-González, E., Jiménez-Fontana, R., & Azcárate, P. (2020). Education for sustainability and the sustainable development goals: Pre-service teachers' perceptions and knowledge. *Sustainability*, *12*(18), 7741.

Hariram, N. P., Mekha, K. B., Suganthan, V., & Sudhakar, K. (2023). Sustainalism: An integrated socio-economic-environmental model to address sustainable development and sustainability. *Sustainability*, *15*(13), 10682.

Himmah, R. H., Jatmiko, B., & Prahani, B. K. (2025). Validity and Reliability of ADI-ESD Based Teaching Materials for Scientific Literacy in Junior High School Students. *SAR Journal - Science and Research*, *9*(8), 84–92.

Ilhami, A., Hidayat, T., Riandi, & Sriyati, S. (2025). Psychometric validation of a research skills assessment for pre-service science teachers using the rasch model. *THABIEA: Journal of Natural Science Teaching*, 8(2), 149–15 http://dx.doi.org/10.21043/thabiea.v8i2.343599.

Markus, K. A., & Borsboom, D. (2024). *Frontiers of test validity theory: Measurement, causation, and meaning*. Routledge.

Metsämuuronen, J. (2023). Seeking the real item difficulty: bias-corrected item difficulty and some consequences in Rasch and IRT modeling. *Behaviormetrika*, *50*(1), 121–154.

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, *7*(1), 5.

Nousheen, A., Zia, M. A., & Waseem, M. (2024). Exploring pre-service teachers' self-efficacy, content knowledge, and pedagogical knowledge concerning education for sustainable development. *Environmental Education Research*, *30*(3), 321–333.

Prasetya, W. A., & Pratama, A. T. (2023). Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(1), 76–91.

Putri, D. D. N., Rahmawati, Fadilah, M., & Rahmi, F. O. (2025). Analisis Sustainability Literacy (Literasi Keberlanjutan) Peserta Didik Fase E pada Isu Bioteknologi di SMAN 12 Padang. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, *4*(1), 2559–2565. https://doi.org/10.31004/jerkin.v4i1.1832

Saarna, R., & Laius, A. (2025). The sustainability literacy assessment instruments: a systematic literature review. *International Journal of Sustainability in Higher Education*. https://doi.org/10.1108/IJSHE-01-2025-0048

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan* (1st ed.). Trim Komunikata Publishing House.

Susbiyanto, S., Hidayat, T., Koosbandiah Surtikanti, H., & Riandi, R. (2025). *Measuring changes of students' problem-solving skills for citizen science projects that integrated in ecology courses: Stacking analysis techniques in Rasch model*. https://doi.org/10.20527/bino.v7i1.19574

Vasiliev, D. (2021). Combating global challenges–climate change, biodiversity loss and poverty simultaneously: Unachievable goal or the only way forward? *SHS Web of Conferences*, *129*, 09022.

Yu, H. C. (2020). Objective Measurement: How Rasch Modeling Can Simplify and Enhance Your Assessment. In M. S. Khine (Ed.), *Rasch measurement: Application in Quantitative Educational Research* (1st ed.). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-1800-3

Zwickle, A., & Jones, K. (2018). Sustainability Knowledge and Attitudes—Assessing Latent Constructs. In W. Leal Filho, R. W. Marans, & J. Callewaert (Eds.), *Handbook of Sustainability and Social Science Research* (pp. 435–451). Springer International Publishing. https://doi.org/10.1007/978-3-319-67122-2_25