

Hierarchical Structure and Measurement Precision of Pro-Environmental Behaviour among Indonesian Junior Secondary Students: A Rasch Rating Scale Analysis

Moch. Dwi Irsyad Saputra, Saiful Ridlo*, Fidia Fibrina

State University of Semarang

Sekaran, Gunungpati, Semarang City, Central Java, Indonesia

*Corresponding author : saiful_ridlo@mail.unnes.ac.id

Article Information

Article History:

Received: Januari 05, 2026

Revised: August 11, 2026

Published: August 27, 2026

Keywords:

Pro-environmental behaviour;
Rasch rating scale model;
Behavioural hierarchy; Person-item targeting; Measurement precision

ABSTRACT

Understanding how pro-environmental behaviour (PEB) develops requires measurement approaches that extend beyond initial validation to examine category thresholds and the substantive structure of behaviour. Building upon prior baseline validation work, this study conducts an advanced secondary analysis of a 20-item PEB instrument using the Rasch Rating Scale Model (RSM) to explicitly evaluate rating scale functioning, behavioural difficulty hierarchy, and measurement precision across latent trait levels. Response data from 124 Indonesian male junior secondary students (aged 12–15 years) were re-analysed using Winsteps version 3.73. Category analysis showed ordered thresholds, although overlap among the middle-frequency categories suggests that their descriptors could be clarified. Item calibration revealed a behavioural hierarchy in which routine and rule-supported practices were easier to endorse than behaviours requiring sustained self-regulation, responsibility for shared resources, or peer engagement. The Test Information Function indicated that measurement precision was highest around the middle of the behavioural continuum and declined towards the lower and upper extremes. Rather than re-evaluating general psychometric quality, this study illustrates how Rasch-derived hierarchies map behavioural progression, providing educators with an empirical blueprint for designing progressive environmental education assessments and targeted interventions.

How to Cite

Saputra, M. D. I., Ridlo, S., Fibrina, F. (2026). Hierarchical Structure and Measurement Precision of Pro-Environmental Behaviour among Indonesian Junior Secondary Students: A Rasch Rating Scale Analysis. *Al-Jahiz: Journal of Biology Education Research*. 7(2), 292–314. DOI: <https://doi.org/10.32332/al-jahiz.v7i1.13494>.

Published by

Website

This is an open access article under the CC BY SA license

Al-Jahiz: Journal of Biology Education Research
<https://e-journal.metrouniv.ac.id/index.php/Al-Jahiz/index>

<https://creativecommons.org/licenses/by-sa/4.0/>



INTRODUCTION

Environmental degradation remains one of the most pressing global challenges, requiring educational interventions that extend beyond increasing environmental knowledge to fostering pro-environmental behaviour (PEB) among students (Sierra Barón & Meneses Baez, 2022; Muliadi

et al., 2025). Within biology and environmental education, behavioural change is widely recognised as the ultimate educational outcome because sustainable environmental improvement depends on individuals' willingness to translate ecological knowledge into consistent everyday actions (Ardoin et al., 2020). Consequently, evaluating students' pro-environmental behaviour has become increasingly important for assessing the effectiveness of environmental education programmes and providing evidence for designing educational interventions that promote sustainable lifestyles. This need is particularly relevant in local educational settings, where social norms, school culture, and environmental practices may substantially influence students' behavioural development (Sukri et al., 2022).

Unlike cognitive achievement, pro-environmental behaviour represents a latent construct that cannot be directly observed but must be inferred from observable behavioural indicators. The theoretical foundation of the present study is the Campbell Paradigm, which conceptualises environmental behaviour as a continuum reflecting an individual's underlying environmental commitment (Kaiser et al., 2010; Kaiser & Lange, 2021). According to this perspective, environmental behaviours differ in their behavioural cost or level of difficulty. Individuals possessing stronger environmental commitment are expected to perform behaviours requiring greater effort, resources, or personal sacrifice, whereas less demanding behaviours tend to be adopted by a broader population. Consequently, pro-environmental behaviour should be interpreted as a hierarchically organised behavioural continuum rather than as a collection of independent actions. Understanding this hierarchy is important because it provides insight into how behavioural commitment develops progressively and how environmental education programmes may facilitate movement toward more demanding forms of pro-environmental behaviour.

Accurately measuring this behavioural continuum requires psychometric approaches capable of modelling both respondent characteristics and behavioural difficulty simultaneously. Conventional approaches based on Classical Test Theory (CTT) primarily evaluate total scores, reliability coefficients, and item discrimination, but provide limited information regarding how individual behaviours are hierarchically organised or how response categories function across different levels of behavioural commitment. In contrast, the Rasch Model transforms ordinal questionnaire responses into interval-level measures while simultaneously estimating respondent ability and item difficulty on a common latent scale (Bond et al., 2021). More specifically, the Rasch Rating Scale Model (RSM) enables researchers to evaluate response category functioning, estimate behavioural hierarchy, and examine measurement precision within a unified analytical framework,

thereby providing richer psychometric evidence than conventional validation procedures (Linacre, 2002; Chong et al., 2022; Colledani et al., 2025).

Although Rasch analysis has increasingly been applied in studies of pro-environmental behaviour, previous research has generally investigated different psychometric aspects separately rather than providing an integrated evaluation of the measurement process. For example, Gaborieau and Pronello (2021) primarily examined the hierarchical ordering of environmental behaviours, whereas Kumawat and Pronello (2021) focused on construct validity, model fit, and reliability of the General Ecological Behaviour scale. More recently, Saputra et al. (2026) established the overall psychometric quality of a Pro-Environmental Behaviour instrument through Rasch validation, demonstrating satisfactory reliability, model fit, and unidimensionality. The earlier study reported category and item diagnostics as part of the instrument's initial psychometric validation. However, it did not integrate these findings into a substantive analysis of response-category functioning, behavioural hierarchy, person-item targeting, and measurement precision across the latent continuum. The 20 PEB items analysed in the present study constituted the behavioural component of the environmental literacy instrument reported by Saputra et al. (2026). Furthermore, empirical evidence from Indonesian junior secondary education remains limited, despite the possibility that local educational practices, school culture, and socio-cultural environments shape behavioural hierarchy differently from other contexts (Menardo et al., 2020).

These limitations indicate two important research gaps. First, previous studies have predominantly emphasised the overall psychometric quality of PEB instruments, whereas relatively little attention has been given to evaluating whether the response categories themselves function optimally and discriminate consistently across different levels of behavioural commitment. Second, although behavioural hierarchy represents a fundamental assumption of the Campbell Paradigm, empirical investigations rarely interpret item difficulty as evidence of behavioural progression or examine how measurement precision varies along the latent behavioural continuum. Addressing these methodological gaps is essential because inaccurate response categories or poorly targeted instruments may obscure the substantive structure of pro-environmental behaviour and reduce the interpretability of measurement results.

Therefore, the present study investigates the psychometric characteristics of a Pro-Environmental Behaviour (PEB) instrument using the Rasch Rating Scale Model within the context of Indonesian junior secondary education. Specifically, this study aims to (1) evaluate the functioning of the five-category response scale, (2) identify the hierarchical structure of pro-environmental behaviour based on item difficulty, and (3) examine measurement precision across

different levels of respondent ability. By integrating these three complementary perspectives within a single Rasch measurement framework, this study extends previous research beyond conventional instrument validation and provides both methodological and substantive contributions to environmental behaviour measurement. The findings are expected to strengthen the application of Rasch measurement in environmental education while providing empirical guidance for designing progressive environmental education programmes that align with students' behavioural development.

RESEARCH METHODS

Research Design

This study conducted an advanced psychometric evaluation of a Pro-Environmental Behaviour (PEB) instrument using the Rasch Rating Scale Model (RSM). The instrument was originally constructed following the ADDIE framework (Analysis, Design, Development, Implementation, and Evaluation), which provided a systematic process for building educational assessment instruments through iterative design, expert review, and field testing (Nesri & Kristanto, 2020; Latip, 2022; Saputra et al., 2026). While the initial development phase established the preliminary overall validation of the instrument (Saputra et al., 2026), the present study specifically investigates its fine-grained measurement characteristics through Rasch analysis, including response category functioning, behavioural hierarchy under the Campbell Paradigm, and measurement precision.

Participants

The study involved 124 Indonesian male junior secondary students aged 12–15 from MTs Tahfidz Yanbu'ul Qur'an Kudus, Indonesia. Participants were selected using convenience sampling because data collection was conducted within a single educational institution during the regular academic schedule, making probability sampling impractical (Etikan et al., 2016). Students were eligible to participate if they were actively enrolled during the data collection period, completed all questionnaire items, and voluntarily agreed to participate.

The sample size exceeded the minimum recommendation for stable estimation of person and item parameters in Rasch analysis of polytomous data (Linacre, 2002). Nevertheless, because the participants were recruited from a single institution using convenience sampling, the findings should be interpreted within the characteristics of the study sample rather than generalised to the broader population of Indonesian junior secondary students.

The dataset analysed in the present study was obtained during the field-testing phase of a previously developed PEB instrument (Saputra et al., 2026). While the earlier publication reported

the overall psychometric validation of the instrument, the present study addresses different research questions by examining rating scale functioning, behavioural hierarchy, and measurement precision using the Rasch Rating Scale Model.

Instrument Development

The instrument was constructed based on four dimensions of pro-environmental behaviour: waste management, participation in environmental conservation, environmentally friendly lifestyle and conservation of natural resources (McBeth & Volk, 2009). The Analysis stage involved reviewing literature on environmental behaviour and environmental education to identify the conceptual dimensions of the construct. During the Design stage, the dimensions were translated into behavioural indicators and an instrument blueprint. In the Development stage, 20 positively and negatively worded items were constructed using a five-point Likert response format to capture variations in students' environmental behaviour while reducing acquiescence bias (Podsakoff et al., 2003). A complete blueprint of the instrument, including item codes, dimensions, behavioural indicators, item wording, and item direction (positive/negative), is provided in Appendix A

Content Validity

Content validity was established through expert judgement involving four specialists, comprising two experts in environmental education and two experts in educational measurement and evaluation. Each expert independently evaluated the relevance, clarity, and representativeness of every item using a four-point relevance scale. Content validity was quantified using Aiken's V (Aiken, 1985), adopting $V \geq 0.62$ as the minimum acceptable criterion (Azwar, 2012). The obtained Aiken's V values ranged from 0.63 to 0.94, with an average coefficient of 0.82, indicating satisfactory content validity.

Following expert review, several items were revised to improve wording clarity, eliminate ambiguity, and strengthen alignment with the intended behavioural indicators before field implementation.

Data Collection Procedure

Data were collected through direct questionnaire administration during regular school hours. Prior to completing the questionnaire, participants received information regarding the objectives of the study, administration procedures, and their rights as voluntary respondents. The questionnaire consisted of a five-point Likert scale 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often and 5 = Always. To ensure consistent score interpretation, all negatively worded items were reverse-coded before Rasch analysis so that higher scores consistently represented higher levels of

pro-environmental behaviour. Completed questionnaires were checked for completeness before analysis. No missing responses were identified in the final dataset.

Ethical Considerations

Participation was voluntary. Respondents were informed about the objectives of the study and assured that all responses would remain anonymous and confidential. The collected data were used exclusively for research purposes. Permission to conduct the study was obtained from the school administration before data collection.

Rasch Analysis Procedure

Data were analysed using Winsteps version 3.73 based on the Rasch Rating Scale Model (RSM). The RSM was selected because all questionnaire items used the same five-category frequency response scale and were designed to share a common category structure across items (Linacre, 2002). Parameter estimation was performed using Joint Maximum Likelihood Estimation (JMLE). The analyses were organised according to the three research objectives. First, rating scale functioning was evaluated using category frequencies, observed average measures, Andrich thresholds, category fit statistics, and category probability curves (Linacre, 2002; Chong et al., 2022; Colledani et al., 2025). Second, the hierarchical structure of pro-environmental behaviour and person–item targeting were examined using item difficulty estimates expressed in logits and visualised through the Wright Map. Item-level and category-level Infit and Outfit MNSQ values between 0.50 and 1.50 were interpreted as productive for measurement (Bond et al., 2021). Third, measurement precision across the latent continuum was examined using the Test Information Function and interpreted in relation to person–item targeting. Overall evidence concerning person and item reliability, separation, and residual dimensionality was reported in the initial validation study (Saputra et al., 2026) and was therefore not repeated as a primary outcome in the present study.

RESEARCH RESULT

The overall psychometric validation of the instrument, including person and item reliability, separation, and residual dimensionality, was reported in the initial validation study (Saputra et al., 2026). Therefore, the present Results section focuses on the three analyses corresponding to the research questions: response category functioning, the hierarchical structure of pro-environmental behaviours and person–item targeting, and measurement precision across the latent continuum.

Response Category Functioning

The functioning of the five-category response scale was evaluated using category frequency, observed average measures, category fit statistics, Andrich thresholds, and category probability curves. Table 1 presents the category structure obtained from the Rasch Rating Scale Model.

Table 1. Functioning of the Five Response Categories

Category	Response label	Observed count	Percentage	Observed average	Infit MNSQ	Outfit MNSQ	Andrich threshold
1	Never	188	8%	−1.51	1.08	1.05	—
2	Rarely	435	18%	−0.89	0.90	0.92	−2.05
3	Sometimes	621	25%	0.07	0.86	0.83	−0.66
4	Often	790	32%	1.39	0.86	1.03	0.43
5	Always	446	18%	2.06	1.08	1.07	2.28

All five response categories were utilised by the respondents, with category frequencies ranging from 188 responses (8%) for the lowest category to 790 responses (32%) for the fourth category. The observed average measures increased monotonically from −1.51 logits in Category 1 to 2.06 logits in Category 5, indicating that higher response categories consistently represented higher levels of pro-environmental behaviour. Category fit statistics were acceptable across all response categories, with Infit Mean Square values ranging from 0.86 to 1.08 and Outfit Mean Square values ranging from 0.83 to 1.07. The Andrich thresholds progressed in the expected order (−2.05, −0.66, 0.43, and 2.28 logits), indicating that respondents generally distinguished the five response categories sequentially.

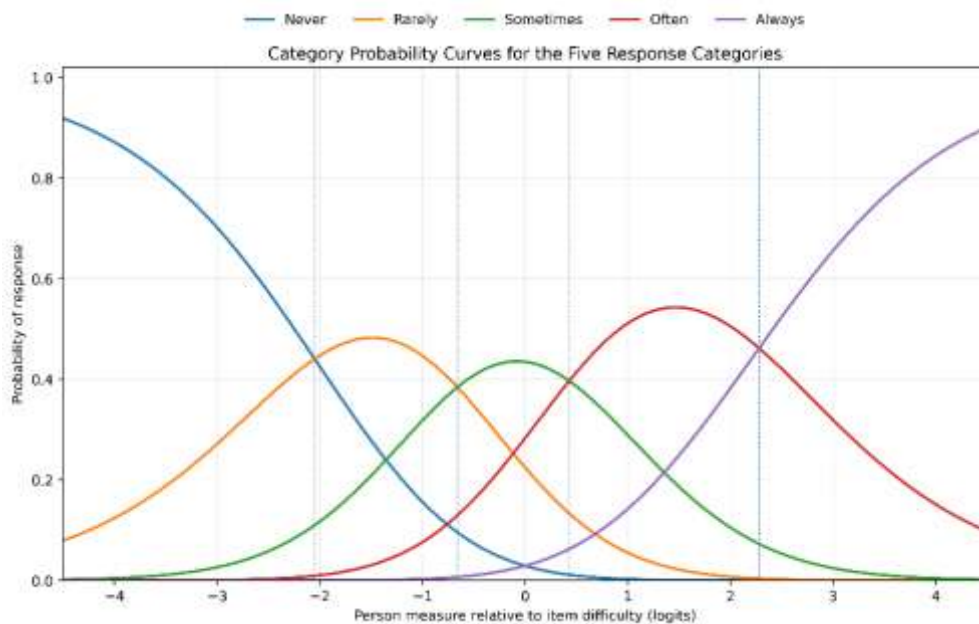


Figure 1. Category Probability Curves for the Five Response Categories

The category probability curves showed that each response category occupied a distinct region along the latent continuum. Slight overlap was observed among the middle categories,

whereas the lowest and highest response categories demonstrated clearer dominance across the lower and upper ranges of respondent ability.

Hierarchical Structure of Pro-Environmental Behaviours and Person–Item Targeting

The hierarchical structure of pro-environmental behaviour was examined using item difficulty estimates expressed in logits. Negatively worded items are presented according to their meaning after reverse scoring so that higher scores consistently represent stronger pro-environmental behaviour. Table 2 presents the hierarchy of pro-environmental behaviours estimated from item difficulty measures.

Table 2. Hierarchy of Pro-Environmental Behaviours

Item	Behaviour represented after scoring	Dimension	Logit	SE	Infit MNSQ	Outfit MNSQ	Relative position
I19	Avoiding leaving water running when it is not being used	Natural-resource conservation	1.53	0.12	1.11	1.11	Uppermost
I18	Turning off unused lights and electrical equipment	Natural-resource conservation	1.30	0.11	0.81	0.81	Uppermost
I10	Encouraging peers to participate in environmental conservation	Environmental participation	1.15	0.11	0.86	0.86	Uppermost
I16	Walking or cycling for short-distance travel	Environmentally friendly lifestyle	0.89	0.11	1.00	1.01	Uppermost
I9	Participating in environmental activities despite perceiving them as tiring	Environmental participation	0.75	0.11	1.08	1.07	Upper
I8	Volunteering in school or community greening activities	Environmental participation	0.62	0.11	0.89	0.86	Upper
I15	Limiting the use of tissue and paper	Environmentally friendly lifestyle	0.54	0.11	0.86	0.89	Upper
I14	Considering environmental impacts when purchasing products	Environmentally friendly lifestyle	0.40	0.11	0.94	1.00	Upper
I7	Participating in school or neighbourhood cleaning activities	Environmental participation	0.28	0.11	0.95	0.93	Near mean
I13	Choosing environmentally friendly products despite higher prices	Environmentally friendly lifestyle	−0.09	0.12	1.04	1.01	Near mean
I6	Avoiding littering even when waste bins are difficult to access	Waste management	−0.18	0.12	0.91	0.91	Near mean
I5	Reusing items that remain suitable for use	Waste management	−0.24	0.12	0.92	0.92	Near mean
I17	Saving water during everyday activities	Natural-resource conservation	−0.29	0.12	0.96	0.94	Near mean
I4	Carrying a reusable drinking bottle	Waste management	−0.35	0.12	0.87	0.85	Lower
I12	Carrying a reusable shopping bag	Environmentally friendly lifestyle	−0.66	0.12	0.99	1.02	Lower
I11	Participating actively in school environmental programmes	Environmental participation	−0.77	0.12	1.01	1.08	Lower
I2	Separating organic and inorganic waste before disposal	Waste management	−1.04	0.13	0.88	0.98	Lowermost
I3	Avoiding the burning of plastic waste	Waste management	−1.12	0.13	1.00	1.05	Lowermost

Item	Behaviour represented after scoring	Dimension	Logit	SE	Infit MNSQ	Outfit MNSQ	Relative position
I20	Planting and caring for plants at home	Natural-resource conservation	-1.13	0.13	1.12	1.13	Lowermost
I1	Disposing of waste in designated bins	Waste management	-1.58	0.14	0.89	1.16	Lowermost

Note. Relative-position labels are descriptive groupings based on locations within the observed item hierarchy and should not be interpreted as universal Rasch cut-offs. Negatively worded items are presented according to their pro-environmental meaning after reverse scoring.

Item difficulty estimates ranged from -1.58 to 1.53 logits. The behaviour with the highest difficulty estimate was avoiding unnecessary water flow (I19; 1.53 logits), followed by turning off unused electrical equipment (I18; 1.30 logits) and encouraging peers to participate in conservation activities (I10; 1.15 logits).

At the lower end of the continuum, disposing of waste in designated bins (I1; -1.58 logits) had the lowest difficulty estimate. This was followed by planting and caring for plants at home (I20; -1.13 logits) and avoiding plastic-waste burning (I3; -1.12 logits). The item measures therefore formed an ordered continuum spanning routine behaviours and less frequently endorsed behaviours.

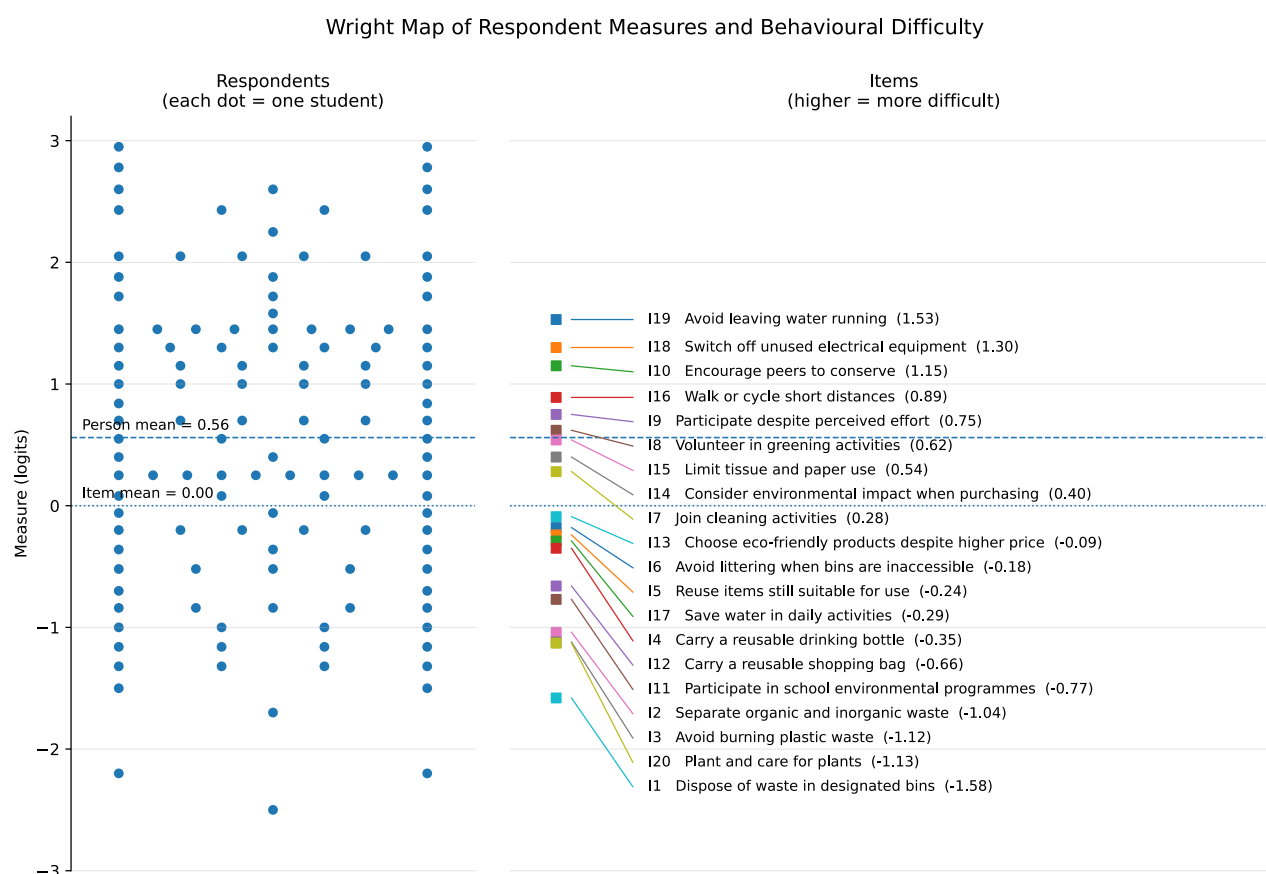


Figure 2. Wright Map of Respondent Measures and Behavioural Difficulty

Respondent measures ranged from -2.55 to 2.92 logits, whereas item difficulty estimates ranged from -1.58 to 1.53 logits. The mean person measure was 0.56 logits, while the item mean was fixed at 0.00 logits. The average respondent measure was therefore 0.56 logits higher than the average item difficulty. Most respondents were located within the central portion of the continuum. However, several respondent measures extended above the most difficult item and below the easiest item. The item distribution therefore covered the middle range of respondent measures more fully than the extreme upper and lower ranges.

Measurement Precision Across the Latent Continuum

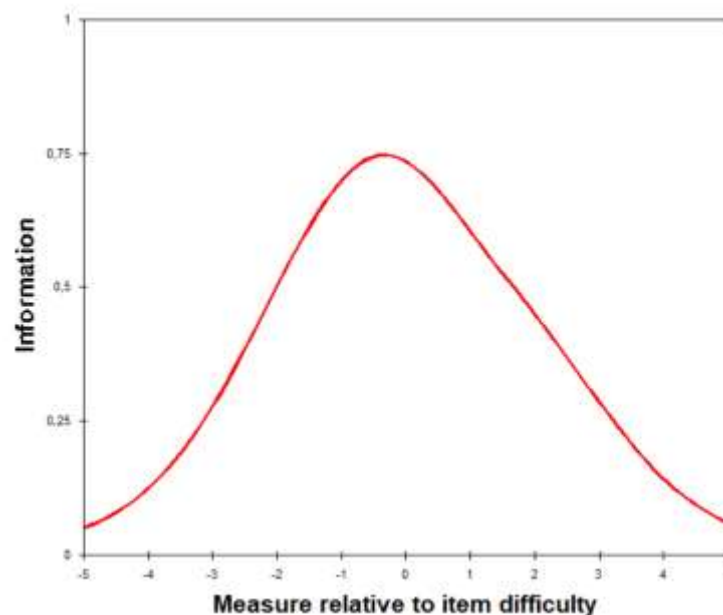


Figure 3. Information Function Across the Pro-Environmental Behaviour Continuum

The information function indicated that information was concentrated around the middle of the continuum and declined towards both extremes. The information curve reached its maximum around 0 logits, indicating that the instrument provided its greatest measurement precision for respondents located near the middle of the continuum. Measurement information decreased progressively towards the lower and upper ends, particularly below approximately -3 logits and above $+3$ logits. Thus, the instrument provided greater precision for respondents with moderate behavioural measures than for respondents at the extreme ends of the continuum. The reduction in information at the upper end was consistent with the upper-end targeting gap identified in the Wright Map, although this pattern alone was not interpreted as definitive evidence of a ceiling effect.

DISCUSSION

The present study examined the functioning of the response scale, the substantive hierarchy of pro-environmental behaviours, and the targeting and precision of the instrument among

Indonesian junior secondary students. These analyses extend the initial psychometric validation of the instrument by focusing on what the Rasch measures reveal about students' interpretation of response categories, the relative difficulty of different environmental actions, and the extent to which the available items cover the behavioural levels of the respondents. The discussion therefore emphasises the substantive and educational meaning of the measurement results rather than repeating the reliability, item-fit, and dimensionality findings reported in the earlier validation study (Saputra et al., 2026).

Functioning of the Response Categories

The overall ordering of the five response categories indicates that the scale was capable of representing progressively higher frequencies of pro-environmental behaviour. In a Rasch rating scale, the meaningfulness of category scores depends not merely on the number of response options but on whether respondents use them in an ordered and interpretable manner. Ordered average measures and thresholds indicate that the transition from Never to Always generally corresponded to movement towards higher levels of the latent behavioural construct (Chong et al., 2022; Colledani et al., 2025).

This result is methodologically important because ordinal labels cannot automatically be assumed to function as intended. A five-point scale may appear logically organised during instrument development while producing indistinguishable, rarely used, or disordered categories after administration. The empirical examination of category functioning is therefore essential before total scores or person measures are interpreted as representing meaningful behavioural differences (Hergesell, 2022; Jin & Eckes, 2024).

Despite the overall ordering, the overlap among Rarely, Sometimes, and Often suggests that respondents distinguished the middle levels of behavioural frequency less clearly than the extreme categories. Verbal frequency labels do not always have identical meanings across respondents; one student may understand Often as several times per week, whereas another may use the same label for a behaviour performed only when circumstances permit. Closely spaced verbal descriptors can consequently result in similar response probabilities within the middle of the latent continuum even when the thresholds remain ordered (Chong et al., 2022; Jin & Eckes, 2024).

The overlap does not, by itself, provide sufficient justification for collapsing the middle categories. Category collapsing is more defensible when adjacent categories show disordered thresholds, extremely low frequencies, severe misfit, or a consistent inability to occupy distinguishable regions of the continuum. Because the categories in the present study retained their intended ordering and were all used by the respondents, revising the descriptors is preferable to

immediately reducing the response format from five categories to four or three categories (Colledani et al., 2025; Tsai et al., 2025).

Future revisions could define the middle categories using clearer behavioural anchors. For example, the instructions could specify approximate weekly or monthly frequencies associated with Rarely, Sometimes, and Often, while allowing minor adjustments for behaviours that do not occur daily. More explicit anchors could reduce subjective differences in category interpretation without sacrificing the additional information potentially provided by a five-category scale (Chong et al., 2022; Van Zile-Tamsen, 2017).

The comparatively frequent selection of Often also requires a cautious interpretation. It may reflect genuinely established environmental routines, but it may also be influenced by the socially desirable character of actions such as conserving resources, participating in cleaning activities, and disposing of waste correctly. Self-reported environmental behaviour can be affected by respondents' desire to present themselves as environmentally responsible, especially when the measured behaviours are closely connected to school rules and socially approved conduct (Deltomme et al., 2023; Lange & Dewitte, 2019; Mónus, 2021).

The school and boarding context may also have shaped how the categories were used. Daily routines, adult supervision, collective responsibilities, and institutional rules can make certain environmental practices more frequent than they would be in less structured settings. The category distribution should therefore be understood as representing behaviour within a particular educational and social environment rather than as a context-free indication of students' environmental commitment (Menardo et al., 2020; Mónus, 2021).

Hierarchical Structure of Pro-Environmental Behaviours

The item hierarchy demonstrates that pro-environmental behaviour is not composed of actions that are equally easy to perform. Instead, the behaviours occupy different positions on a latent continuum, with some practices being readily incorporated into students' routines and others requiring greater persistence, initiative, or personal responsibility. Similar Rasch-based research has shown that ecological behaviours can be ordered according to relative difficulty, thereby providing more substantive information than a simple count of environmentally friendly actions (Gaborieau & Pronello, 2021; Kumawat & Pronello, 2021; Muliadi et al., 2025).

This interpretation is consistent with the Campbell Paradigm, in which the likelihood of performing an environmental behaviour depends partly on the cost or demand associated with that behaviour. Behavioural cost does not refer only to financial expense; it can include effort, inconvenience, time, self-regulation, disruption of established habits, and the social consequences

of influencing other people. Individuals with stronger environmental commitment are expected to perform not only low-cost behaviours but also actions that remain demanding under less supportive conditions (Kaiser & Lange, 2021).

Avoiding unnecessary water flow and consistently switching off unused electrical equipment occupied the difficult end of the hierarchy. Although these actions may appear simple, consistent performance requires continuous attention rather than a single decision. Students must recognise resource waste, interrupt their current activity, and accept personal responsibility for resources that are often used collectively. These requirements make the behaviours more dependent on sustained self-regulation than visible actions performed only at particular times (Kaiser & Lange, 2021; Steg & Vlek, 2009).

The shared-resource setting provides a plausible contextual explanation for their difficulty. In communal school or boarding facilities, water taps, lights, and electrical equipment may be used by many students, making individual responsibility less clearly defined. A student may notice wastage but assume that another person, a class officer, or school staff member is responsible for addressing it. This explanation remains interpretive because the present study did not directly investigate students' reasons for performing or neglecting each behaviour, but it illustrates how infrastructure and social context can influence the location of an item within a behavioural hierarchy (Menardo et al., 2020; Mónus, 2021).

Encouraging peers to participate in conservation activities was also relatively difficult because it extends beyond private environmental practice. Students must communicate their environmental expectations, influence their peers, and potentially respond to resistance or indifference. Peer encouragement therefore combines environmental commitment with communication, social confidence, and perceived collective efficacy, making it more demanding than silently performing an individual routine (De Leeuw et al., 2015; Stevenson et al., 2013).

By contrast, disposing of waste in designated bins, caring for plants, and avoiding the burning of plastic waste were located at the easier end of the continuum. These actions are comparatively visible, familiar, and more likely to be reinforced through school regulations, environmental campaigns, facilities, and repeated instructions. Their lower difficulty may therefore reflect the availability of supportive cues and routines rather than the behaviours being intrinsically unimportant (Mónus, 2021; Steg & Vlek, 2009).

This pattern is relevant to the Indonesian context. A recent Rasch evaluation of environmentally friendly behaviour among Indonesian university students also found that item hierarchy and sample characteristics influenced the interpretation of environmental behaviour,

while gaps between item locations indicated the need for additional, carefully formulated behaviours. Although the populations and instruments differ, both studies demonstrate that environmental behaviour scales should be interpreted through the specific content and contextual demand of their items rather than through total scores alone (Muliadi et al., 2025).

The hierarchy also indicates that behaviour may develop progressively. Performing accessible environmental actions can support a sense of competence that subsequently facilitates engagement in more demanding behaviours. Previous research has shown that experience with relatively easy pro-environmental actions can contribute to self-efficacy, which may encourage intentions and subsequent engagement in more difficult environmental behaviour (Lauren et al., 2016).

Nevertheless, progression should not be interpreted as an automatic sequence in which every student moves through exactly the same stages. Behavioural performance remains affected by opportunity, infrastructure, social support, personal control, and the relevance of each behaviour to the respondent's daily life. The hierarchy identifies the relative empirical difficulty of actions in the present sample, not a universal developmental order applicable to every educational context (Lange & Dewitte, 2019; Mateer et al., 2022; Menardo et al., 2020).

Person–Item Targeting and Measurement Precision

The person–item relationship indicates that the existing items provided their strongest coverage around the middle of the behavioural continuum. This targeting is suitable for distinguishing students with moderate levels of pro-environmental behaviour, but it provides less information about respondents whose behavioural measures are located beyond the available item range. Rasch-based scale evaluation should therefore consider not only whether the items fit the model but also whether their difficulties are appropriately distributed for the population and intended use of the instrument (Hergesell, 2022; Muliadi et al., 2025).

The location of several respondents above the most difficult item indicates an upper-end targeting gap. The scale contains too few behaviours that challenge students who already report consistently performing routine and moderately demanding environmental actions. Consequently, students with comparatively high environmental commitment may receive similar estimated measures even though they differ in their capacity to initiate, maintain, and lead more complex environmental practices (Muliadi et al., 2025; Wind & Jones, 2019).

This pattern should be described as a potential upper-end targeting gap rather than definitive evidence of a ceiling effect. A formal ceiling effect would require supporting evidence such as a substantial concentration of maximum raw scores or a considerable number of extreme person

estimates. The available evidence more directly demonstrates that the upper part of the person distribution is less well covered by item difficulty than the centre of the continuum (Colledani et al., 2025; Wind & Jones, 2019).

The Test Information Function reinforces the person–item targeting results by demonstrating that measurement precision was not uniform across the latent continuum. Information was concentrated around the middle range, where the available item difficulties most closely matched the respondent measures. In contrast, the decline in information at the upper and lower ends indicates reduced differentiation among respondents whose behavioural levels extended beyond the central item range. Therefore, the instrument should not be interpreted as providing equal measurement precision at all behavioural levels (Hergesell, 2022; Wind & Jones, 2019).

The reduced information at the upper end is consistent with a potential upper-end targeting gap rather than definitive evidence of a ceiling effect. The scale contained relatively few items representing highly demanding environmental behaviours. Future development should therefore include items involving sustained commitment, environmental leadership, monitoring of communal resources, and environmentally responsible decisions despite additional effort or inconvenience (Kaiser & Lange, 2021; Muliadi et al., 2025).

Improving upper-end targeting requires more difficult items, but difficulty should not be produced merely by making statements linguistically complicated. New items should represent substantively more demanding forms of environmental action involving sustained commitment, reduced convenience, leadership, monitoring, or responsibility for collective outcomes. Behavioural cost should arise from the action being measured rather than from unclear wording or excessive reading difficulty (Kaiser & Lange, 2021; Mateer et al., 2022).

Potential upper-level items could examine whether students initiate environmental activities without waiting for teacher instructions, coordinate peer conservation programmes, monitor communal water and electricity use over time, establish measurable waste-reduction targets, or evaluate whether environmental programmes achieve their intended outcomes. Other items could assess whether students maintain environmentally responsible choices when those alternatives require additional time, effort, or inconvenience (Kaiser & Lange, 2021; Lauren et al., 2016).

The item pool could also include behaviours that distinguish individual compliance from environmental leadership. For instance, participating in a teacher-organised cleaning activity may be less demanding than planning the activity, allocating responsibilities, persuading reluctant peers, and evaluating its environmental impact. Including both levels would extend the continuum from

routine participation to autonomous and collective environmental agency (Ardoin et al., 2020; De Leeuw et al., 2015).

A smaller targeting gap was also present at the lower end, indicating that very low behavioural levels may not be optimally differentiated. Future development could therefore include a limited number of basic diagnostic items, such as performing an environmental action after being reminded, recognising the appropriate waste-disposal facility, or following a simple conservation instruction under supervision. These items should be included only when the scale is intended to identify students who have not yet established foundational environmental routines (Muliadi et al., 2025; Sukri et al., 2022).

Additional items should be tested in a new and more diverse sample rather than simply appended to the existing scale. Content validation, cognitive interviews, category-functioning analysis, item calibration, person–item targeting, and local-dependence checks would be needed to determine whether the new behaviours genuinely extend measurement coverage without introducing redundancy or a secondary construct (Colledani et al., 2025; Hergesell, 2022; Muliadi et al., 2025).

Implications for Environmental Education

The behavioural hierarchy provides a practical basis for designing environmental education progressively. Educational programmes should not assume that knowledge about environmental problems will immediately produce consistent and demanding environmental action. Instruction should instead create opportunities for students to establish routine behaviours, practise self-regulation, and gradually assume responsibility for collective environmental outcomes (Ardoin et al., 2020; Steg & Vlek, 2009).

At the foundational level, schools can reinforce behaviours located at the easier end of the continuum, including disposing of waste in designated bins, separating waste, caring for plants, and using reusable containers. These behaviours can be embedded in daily routines, supported by adequate facilities, and monitored consistently until they become less dependent on direct reminders. Environmental education is more likely to influence practice when students are provided with concrete opportunities to act rather than only receiving information about environmental problems (Ardoin et al., 2020; Mónus, 2021).

At the intermediate level, activities can focus on self-regulation and responsibility for resource use. Students could maintain weekly records of water or electricity use, inspect whether lights and equipment have been switched off, calculate paper or plastic consumption, and reflect on barriers that prevented consistent conservation. Such activities connect environmental

knowledge with measurable daily behaviour and help students recognise that apparently minor actions accumulate into collective environmental consequences (Steg & Vlek, 2009; Sukri et al., 2022).

At the advanced level, students can be assigned leadership and inquiry-oriented responsibilities. Examples include conducting a school waste audit, coordinating peer campaigns, establishing class-level reduction targets, designing conservation projects, and presenting evidence of environmental improvement. These activities reflect the more demanding end of the hierarchy because they require initiative, collaboration, communication, and responsibility for outcomes beyond the individual student (Ardoin et al., 2020; De Leeuw et al., 2015).

The hierarchy can therefore be translated into a three-stage educational sequence: routine compliance, self-regulated conservation, and environmental leadership. This sequence should remain flexible because students may demonstrate advanced behaviour in one domain while requiring support in another. A student may independently conserve electricity but remain reluctant to influence peers, illustrating why environmental behaviour should be examined at the item level rather than represented only by a global score (Mateer et al., 2022; Mónus, 2021).

The results also have implications for programme assessment. An intervention focused on foundational environmental routines should be evaluated using items that represent those routines, whereas programmes designed to develop leadership and collective action require more difficult items. Using a scale dominated by easy behaviours to assess an advanced environmental programme may produce apparently high scores while failing to detect meaningful differences among highly engaged students (Muliadi et al., 2025; Sukri et al., 2022).

Schools can additionally combine questionnaire measures with observable indicators, such as reductions in waste volume, electricity consumption, water use, or the frequency of student-led initiatives. Behavioural records would complement self-reports and help determine whether reported frequency corresponds to environmental practice. Systematic reviews of environmental education have similarly emphasised the importance of linking educational activities with direct, observable, and locally relevant environmental outcomes (Ardoin et al., 2020).

Theoretical and Methodological Contribution

The principal theoretical contribution lies in demonstrating how the concept of behavioural cost can be interpreted within a junior secondary educational context. The hierarchy did not merely separate environmentally friendly from environmentally harmful behaviour; it distinguished routine actions from behaviours involving sustained self-regulation, shared-resource responsibility, and social initiative. This supports the conceptualisation of pro-environmental behaviour as a

continuum of increasingly demanding actions rather than a homogeneous collection of practices (Kaiser & Lange, 2021; Lauren et al., 2016).

Methodologically, the study shows why scale evaluation should move beyond global reliability and validity coefficients. Rating category functioning determines whether responses carry their intended meaning, item calibration reveals the substantive hierarchy of behaviours, and person–item targeting identifies where the scale provides adequate or inadequate coverage. Considering these components together produces a more informative interpretation than relying exclusively on a total score or internal-consistency estimate (Colledani et al., 2025; Gaborieau & Pronello, 2021; Hergesell, 2022).

Previous Rasch studies of environmental behaviour have provided important evidence concerning unidimensional measurement, item calibration, reliability, differential item functioning, and ecological behaviour scales among travellers or adult populations. The present study complements that literature by focusing more explicitly on rating-scale operation, item-content interpretation, and targeting among Indonesian junior secondary students (Gaborieau & Pronello, 2021; Kumawat & Pronello, 2021; Muliadi et al., 2025).

The findings also reinforce the importance of context in environmental measurement. An item is not inherently easy or difficult independently of the population, infrastructure, social expectations, and behavioural opportunities surrounding it. Accordingly, the hierarchy observed in this study should be treated as an empirical representation of the participating students rather than as a universally fixed ordering of environmental behaviour (Lange & Dewitte, 2019; Menardo et al., 2020; Mónus, 2021).

Limitations and Directions for Future Research

The findings should be interpreted in relation to the characteristics of the sample. Participants were selected through convenience sampling from one private Islamic junior secondary and boarding-school context, and all participants were male. Convenience sampling was appropriate for access to the field-test population, but it limits the extent to which the item hierarchy and targeting pattern can be generalised to other schools, female students, non-boarding settings, or different regions of Indonesia (Etikan et al., 2016).

The use of self-reported behavioural frequency represents another limitation. Respondents may differ in recall accuracy, interpretation of frequency labels, or willingness to report environmentally undesirable behaviour. The findings therefore represent students' reported behavioural frequency rather than direct evidence that each behaviour was consistently performed in daily practice (Deltomme et al., 2023; Lange & Dewitte, 2019).

The data were also obtained from the field-testing dataset used in the initial instrument-development study. The earlier publication reported the instrument's general psychometric properties, whereas the present study addressed distinct questions concerning category functioning, substantive behavioural hierarchy, targeting, and measurement precision. Nevertheless, independent replication is required to determine whether the category structure and item hierarchy remain stable in a new sample (Saputra et al., 2026).

Future studies should involve students from different school types, regions, genders, and socio-economic settings. Multi-group Rasch analysis and differential item functioning could then examine whether behaviours have comparable difficulties across participant groups. Such analyses are particularly important because apparently identical environmental behaviours may involve different levels of opportunity, cost, and social support across educational settings (Ivanova et al., 2023; Mateer et al., 2022; Muliadi et al., 2025).

Longitudinal research could also investigate whether students actually move from routine practices towards more demanding environmental behaviours after participating in progressive environmental education. This would allow the proposed hierarchy to be evaluated not only as a cross-sectional ordering of item difficulty but also as a potential framework for monitoring behavioural development over time (Ardoin et al., 2020; Lauren et al., 2016).

Overall, the findings suggest that the five-category scale retains a meaningful ordering but would benefit from clearer middle-category descriptors. The behavioural hierarchy distinguishes routine practices from actions requiring sustained self-regulation and social initiative, while the targeting analysis indicates that additional difficult items are needed to differentiate students with advanced levels of environmental engagement. These findings provide a basis for refining the instrument and designing environmental education that progresses from supported daily routines towards autonomous and collective environmental responsibility

CONCLUSION

This study examined the functioning of the response scale, the hierarchical structure of pro-environmental behaviours, and measurement precision among Indonesian junior secondary students. The five response categories generally operated in the intended order, although overlap among the middle categories indicates that the descriptors for Rarely, Sometimes, and Often may require further clarification. The behavioural hierarchy showed that routine and rule-supported actions were more readily endorsed than behaviours requiring sustained self-regulation, responsibility for shared resources, and peer engagement.

Person-item targeting indicated that the instrument provided its strongest measurement coverage for students located around the middle of the pro-environmental behaviour continuum.

Reduced information at the upper and lower ends suggests that the instrument was less capable of differentiating students with extreme behavioural levels, particularly those whose measures exceeded the difficulty of the available items. Future instrument development should therefore include more demanding behaviours involving environmental leadership, long-term commitment, communal resource monitoring, and environmentally responsible choices despite additional inconvenience or effort.

These findings provide context-specific evidence regarding the functioning and substantive interpretation of the instrument within the participating sample. They should not be interpreted as demonstrating universally accurate measurement of students' actual environmental behaviour. The data were self-reported and obtained through convenience sampling from one private Islamic junior secondary and boarding-school context, and responses may have been influenced by social desirability and institutional routines. Further studies should replicate the analysis using more diverse and independent samples, incorporate observational or behavioural evidence, and examine whether the category structure and behavioural hierarchy remain stable across gender, school type, and socio-cultural context.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the management, teachers, and students of the participating MTs Tahfidz Yanbu'ul Qur'an Kudus for their cooperation and support during the data collection process. The authors also thank the experts in environmental education and educational measurement who provided valuable feedback during the content-validation stage of the instrument. Their contributions supported the refinement and implementation of this study.

FUNDING STATEMENT

This research received no external funding and was financially supported by the authors.

CONFLICTS OF INTEREST

The authors declare that they have no competing financial, professional, or personal interests that could have influenced the research reported in this article.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Author 1: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, and Writing – review and editing. **Author 2:** Conceptualization, Methodology, Supervision, and Validation. **Author 3:** Methodology, Supervision, and Validation.

DECLARATION OF THE USE OF GENERATIVE AI

The authors used OpenAI's ChatGPT solely for language refinement and grammatical correction. All AI-assisted content was reviewed and revised by the authors, who take full responsibility for the final manuscript.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available because they contain information collected from minor participants and are subject to institutional and ethical confidentiality considerations.

REFERENCES

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Ardoin, N. M., Bowers, A. W., & Gaillard, E. (2020). Environmental education outcomes for conservation: A systematic review. *Biological Conservation*, 241, 108224. <https://doi.org/10.1016/j.biocon.2019.108224>
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Pustaka Pelajar.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. <https://doi.org/10.4324/9780429030499>
- Chong, J., Mokshein, S. E., & Mustapha, R. (2022). Applying the Rasch Rating Scale Model (RSM) to investigate the rating scales function in survey research instrument. *Jurnal Cakrawala Pendidikan*, 41(1), 97–111. <https://doi.org/10.21831/cp.v41i1.39130>
- Colledani, D., González Pizzio, A. P., Devita, M., & Anselmi, P. (2025). Investigating the functioning of rating scales with Rasch models. *Assessment*, 32(3), 434–446. <https://doi.org/10.1177/10731911241245792>
- De Leeuw, A., Valois, P., Ajzen, I., & Schmidt, P. (2015). Using the theory of planned behavior to identify key beliefs underlying pro-environmental behavior. *Journal of Environmental Psychology*, 42, 128–138. <https://doi.org/10.1016/j.jenvp.2015.03.005>
- Deltomme, B., Gorissen, K., & Weijters, B. (2023). Measuring pro-environmental behavior: Convergent validity, internal consistency, and respondent experience of existing instruments. *Sustainability*, 15(19), 14484. <https://doi.org/10.3390/su151914484>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Gaborieau, J.-B., & Pronello, C. (2021). Validation of a unidimensional and probabilistic measurement scale for pro-environmental behaviour by travellers. *Transportation*, 48(2), 555–593. <https://doi.org/10.1007/s11116-019-10068-w>
- Hergesell, A. (2022). Using Rasch analysis for scale development and refinement in tourism: Theory and illustration. *Journal of Business Research*, 142, 551–561. <https://doi.org/10.1016/j.jbusres.2021.12.063>

- Ivanova, A. A., Agissova, F. B., Sautkina, E. V., Kabanova, V. S., Patrakova, N. A., & Ivande, K. S. (2023). Russian scale of pro-environmental behavior: Development and psychometric assessment. *Experimental Psychology (Russia)*, 16(2), 218–234. <https://doi.org/10.17759/exppsy.2023160213>
- Jin, K., & Eckes, T. (2024). The impact of insufficient effort responses on the order of category thresholds in the polytomous Rasch model. *Educational and Psychological Measurement*, 84(6), 1203–1231. <https://doi.org/10.1177/00131644241242806>
- Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review*, 14(4), 351–367. <https://doi.org/10.1177/1088868310366452>
- Kaiser, F. G., & Lange, F. (2021). Offsetting behavioral costs with personal attitude: Identifying the psychological essence of an environmental attitude measure. *Journal of Environmental Psychology*, 75, 101619. <https://doi.org/10.1016/j.jenvp.2021.101619>
- Kumawat, P., & Pronello, C. (2021). Validating Italian General Ecological Behaviour Questionnaire of travellers using dichotomous Rasch model. *Sustainability*, 13(21), 11976. <https://doi.org/10.3390/su132111976>
- Lange, F., & Dewitte, S. (2019). Measuring pro-environmental behavior: Review and recommendations. *Journal of Environmental Psychology*, 63, 92–100. <https://doi.org/10.1016/j.jenvp.2019.04.009>
- Latip, A. (2022). Penerapan model ADDIE dalam pengembangan multimedia pembelajaran berbasis literasi sains. *DIKSAINS: Jurnal Ilmiah Pendidikan Sains*, 2(2), 102–108. <https://doi.org/10.33369/diksains.2.2.102-108>
- Lauren, N., Fielding, K. S., Smith, L., & Louis, W. R. (2016). You did, so you can and you will: Self-efficacy as a mediator of spillover from easy to more difficult pro-environmental behaviour. *Journal of Environmental Psychology*, 48, 191–199. <https://doi.org/10.1016/j.jenvp.2016.10.004>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Mateer, T. J., Melton, T. N., Miller, Z. D., Lawhon, B., Agans, J. P., & Taff, B. D. (2022). A multi-dimensional measure of pro-environmental behavior for use across populations with varying levels of environmental involvement in the United States. *PLOS ONE*, 17(10), e0274083. <https://doi.org/10.1371/journal.pone.0274083>
- McBeth, W., & Volk, T. L. (2009). The national environmental literacy project: A baseline study of middle grade students in the United States. *The Journal of Environmental Education*, 41(1), 55–67. <https://doi.org/10.1080/00958960903210031>
- Menardo, E., Brondino, M., & Pasini, M. (2020). Adaptation and psychometric properties of the Italian version of the Pro-Environmental Behaviours Scale (PEBS). *Environment, Development and Sustainability*, 22(7), 6907–6930. <https://doi.org/10.1007/s10668-019-00520-3>
- Mónus, F. (2021). Environmental perceptions and pro-environmental behavior: Comparing different measuring approaches. *Environmental Education Research*, 27(1), 132–156. <https://doi.org/10.1080/13504622.2020.1842332>

- Muliadi, R., Jatnika, R., Hanami, Y., & Adishesa, M. S. (2025). Evaluation of an environmentally friendly behavior scale among Indonesian university students using the Rasch model. *Acta Psychologica*, 257, 105087. <https://doi.org/10.1016/j.actpsy.2025.105087>
- Nesri, F. D. P., & Kristanto, Y. D. (2020). Pengembangan modul ajar berbantuan teknologi untuk mengembangkan kecakapan abad 21 siswa. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(3), 480–492. <https://doi.org/10.24127/ajpm.v9i3.2925>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Saputra, M. D. I., Susilaningsih, E., Rusilowati, A., & Lathif, Y. F. (2026). Rasch model-based development of an environmental literacy instrument. *Jurnal Biologi-Inovasi Pendidikan*, 8(1), 112–122. <https://doi.org/10.20527/bino.v8i1.24953>
- Sierra Barón, W., & Meneses Baez, A. L. (2022). Psychometric indicators of the pro-environmental attitudes' questionnaire: Colombian version. *Frontiers in Psychology*, 13, 886769. <https://doi.org/10.3389/fpsyg.2022.886769>
- Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, 29(3), 309–317. <https://doi.org/10.1016/j.jenvp.2008.10.004>
- Stevenson, K. T., Peterson, M. N., & Bondell, H. D. (2013). The influence of personal beliefs, friends, and family in building climate change concern among adolescents. *Environmental Education Research*, 19(5), 654–669. <https://doi.org/10.1080/13504622.2012.714851>
- Sukri, A., Rizka, M. A., Purwanti, E., Ramdiah, S., & Lukitasari, M. (2022). Validating student's green character instrument using factor and Rasch model. *European Journal of Educational Research*, 11(2), 859–872. <https://doi.org/10.12973/eu-jer.11.2.859>
- Tsai, C.-L., Wind, S. A., & Estrada, S. (2025). Exploring the effects of collapsing rating scale categories in polytomous item response theory analyses: An illustration and simulation study. *Measurement: Interdisciplinary Research and Perspectives*, 23(1), 66–89. <https://doi.org/10.1080/15366367.2023.2288791>
- Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, 58(8), 922–933. <https://doi.org/10.1007/s11162-017-9448-0>
- Wind, S. A., & Jones, E. (2019). The many interpretations of Rasch measurement. *Educational Measurement: Issues and Practice*, 38(2), 39–49. <https://doi.org/10.1111/emip.12264>

Copyright Holder:

© Saputra, M. D. I. et al. (2026)

First Publication Right:

© Al-Jahiz: Journal of Biology Education Research

This article is under:

CC BY SA