Integrated Science Education Journal
Vol. 6, No. 3, September 2025, pp. 153~162
ISSN: 2716-3725, DOI: 10.37251/isej.v6i3.2025

Measuring Energy Literacy: Validation of Knowledge, Attitude, and Behavior Instruments Using the Rasch Model

Inda Hasanah Bimastari[1], Muhamad Yusup[1,*], Kistiono[1]
[1]Master Program of Physics Education, Universitas Sriwijaya, Sumatera Selatan, Indonesia

Article Info

Article history:

Received Jul 17, 2025
Revised Jul 27, 2025
Accepted Jul 30, 2025
OnlineFirst Aug 08, 2025

Keywords:

Attitude
Behavior
Energy literacy
Knowledge
Rasch model

ABSTRACT

Purpose of the study: This study aims to develop and validate a Rasch Model-based energy knowledge, attitude, and behavior assessment instrument to measure energy literacy among high school students.

Methodology: The research design followed the stages of theory-based instrument development and the Rasch modeling approach. The quantitative approach and instrument development process were carried out through the stages of item preparation based on literature review and theoretical indicators, validation by experts, and field testing of 50 participants.

Main Findings: The knowledge instrument has high quality, obtained as much as 9 data showing a fit for the Rasch model, such as MNSQ infit and outfit values ranging from 0.84 to 1.25. and positive point measure correlations. Instrument reliability shows infit and outfit results that are still within the ideal tolerance range, obtained sequentially the person reliability values of knowledge, attitude, and behavior of 0.50, 0.76 and 0.75. So that the instrument is suitable for accurate and consistent measurement.

Novelty/Originality of this study: this study provides innovation in the energy literacy measurement literature (knowledge, attitude, and behavior) with an in-depth evaluation of item performance and respondent ability distribution using empirical and applicable methods for the development of evidence-based and Rasch model-based assessment instruments.

This is an open access article under the CC BY license

Corresponding Author:
Muhamad Yusup,
Master Program of Physics Education, Universitas Sriwijaya,
Padang Selasa Road, No. 524, Palembang, Sumatera Selatan, 30139, Indonesia
Email: rohyanafelisa@students.unnes.ac.id

1. INTRODUCTION

Energy-related global challenges, such as climate change and the need for energy transition Sustainable energy issues, further highlight the urgency of energy literacy. Individuals are required to have cognitive and demonstrate responsible behavior regarding energy consumption [1]. Energy literacy has a multidimensional construct that includes cognitive, affective, and behavior towards energy [2]-[4]. The knowledge dimension relates to cognitive understanding of energy concepts, energy efficiency, renewable and non-renewable resources, and the impact of energy use on the environment, and the attitude dimension reflects an individual's views or values on energy and environmental issues [5], [6]. The behavioral dimension reflects a person's tendencies and habits in using energy efficiently and responsibly [7]. Meanwhile, research results show that the level of energy literacy of students, especially knowledge, attitudes, and behavior, is still in the low category [8]. Increasing energy awareness and literacy among the public is considered a key pillar in the transition to a more efficient and environmentally friendly energy system [9], [10].

Low energy literacy is not only a problem in Indonesia, but also a global concern that affects efforts toward sustainable energy transition [11]. Several studies indicate that secondary school curricula have not fully integrated energy competencies in an explicit and balanced manner, leaving students with limited understanding of energy issues [12], [13]. Low critical and analytical thinking skills in the context of energy pose a challenge in fostering sustainable behavior [14]. However, energy literacy has been proven to have a strong correlation with pro-environmental attitudes and energy-saving habits [15], [16]. Therefore, an instrument capable of measuring the integration of cognitive, affective, and behavioral dimensions is essential as a basis for evaluating energy education programs [17]. The Rasch model offers high psychometric validity in assessing participants' abilities and responses in an invariant and objective manner [18], [19]. Rasch-based instruments can be flexibly developed for various student backgrounds, ensuring measurement accuracy and freedom from contextual bias [20]. Therefore, it is important to develop instruments that are not only statistically valid and reliable but also relevant to the current energy education context. Various initiatives and educational programs are promoted to foster energy literacy among the public, starting from the primary education level to the general public. Energy literacy, which includes knowledge, attitudes and behaviors related to energy use, is considered an important pillar in realizing a just and sustainable energy transition [21]. However, to measure the effectiveness of these programs, valid and reliable measurement instruments are needed.

The Rasch model, one of the approaches in Item Response Theory (IRT), has been the choice for generating interval scales and ensuring the invariant nature of measurement policies. With Rasch, items are thoroughly evaluated based on item difficulty and person ability, and tested for the extent to which the data fulfill the expected probabilistic structure. As a more sophisticated alternative, Item Response Theory (IRT) has emerged as a more robust framework for psychometric analysis. One of the most frequently used IRT models that has the advantage of easy interpretation is the Rasch model [22]. The Rasch model offers a different approach from CTT by modeling the probability of individual responses to items based on the interaction between respondent ability and item difficulty. The advantage of the Rasch Model lies in its ability to identify misfit items, estimate item difficulty, and measure respondent ability on an interval scale, which classical test theory cannot do directly [19], [23]. The application of the Rasch Model has been widely used in various fields, including education, psychology, and health, to ensure the quality of the instruments used [24]. The application of the Rasch Model in various fields of learning has shown encouraging results. For example, [25] successfully designed and validated an energy literacy instrument for prospective physics teachers using a four-tier approach and Rasch analysis, showing that the instrument was feasible and able to measure energy literacy accurately. In addition, the digital literacy study developed a digital literacy measurement tool with high validity and reliability through Rasch verification, ensuring the data is unbiased and meets the assumptions of a single-dimensional structure [26]. The Rasch model not only produces consistent measures, but is also able to detect items that do not converge ("misfit") through statistical analysis, as well as validate the equality of item functions (invariance) between subgroups. A valid measurement instrument is one that actually measures what it is supposed to measure. This method is used to generate construct validity (through unidimensionality analysis, infit/outfit, PTMEA) and adequate reliability [27]. Therefore, it is important to conduct an in-depth analysis of the instruments used to measure energy knowledge and behavior to ensure the validity and quality of the data. Instrument validation is an important process to ensure that the instrument measures the intended aspects precisely and consistently [28].

Statistical validation methods play a major role in this context, and the Rasch Model approach is increasingly being used in assessment instrument development due to its ability to determine item fit and measure respondent characteristics simultaneously [29]. Recent research shows that the use of Rasch models is able to improve measurement accuracy, identify problematic items, and ensure invariance of item functions across different cultural and demographic populations [30]. In addition, validation through statistical analysis such as fit statistics, unidimensionality, and item bias analysis are key things that must be done to ensure the instrument can be used widely and sustainably [31]. However, many instruments in circulation still rely on traditional approaches such as classical test theory (CT) item analysis, which has limitations, especially in terms of the mismatch between item characteristics and respondent abilities [32]. Research shows the potential of the Rasch model in analyzing instruments in the field of education. For example, in a validation study of science literacy instruments, the Rasch Model was shown to be able to detect biased items and provide more stable parameter estimates than CTT [33]. Similarly, in the context of energy literacy, the application of the Rasch model can provide a more accurate picture of the level of respondents' energy knowledge and attitudes, as well as identify problematic items [34]. Based on this urgency, this study aims to comprehensively analyze the validation of energy knowledge, attitude, and behavior question instruments using the Rasch Model. This analysis will include testing the fit of the data with the model (fit analysis), unidimensionality, and identification of misfit items. Thus, the results of this study are expected to provide strong empirical evidence of the quality of the instrument used, as well as provide recommendations for instrument improvement to more accurately measure energy literacy competencies.

2. RESEARCH METHOD

This instrument was developed through several stages starting from developing an assessment framework: analyzing the competencies and components of energy literacy and organizing them. The items developed followed the framework of [35]. The research design followed the stages of theory-based instrument development and the Rasch modeling approach, which is known to produce interval-scale measurements and maintain the principle of invariance in measurement [36]. The instrument development procedure was carried out through several stages, starting from the preparation of question items based on literature review and theoretical indicators, which were then validated by experts (expert judgment) to ensure content validity. The revised question items were then tested on an initial sample of 50 high school students using purposive sampling technique, in order to obtain a diversity of responses and backgrounds.

The results of data collection were analyzed using the Rasch Model with the help of software such as Winsteps to evaluate various aspects of item quality. The analysis included item fit statistics, consisting of infit and outfit Mean Square (MnSq) values with ideal criteria ranging from 0.5 to 1.5; values over 2.0 indicate the presence of item misfit [37]. In addition, item-total correlations (PTMEA Corr) were also analyzed, with a minimum cutoff of 0.20 for an item to be considered as having a good contribution to the construct being measured To ensure that the instrument measures one consistent construct (unidimensionality), Principal Component Analysis (PCA) analysis is carried out on the residuals, with the criteria that at least 40-50% of the variance is explained by the main dimensions [38].

The quality of the instrument is also determined through reliability analysis and separation index. A good instrument is expected to have person and item reliabilities above 0.80, as well as a stratum value of at least 2.0, which indicates the instrument's ability to distinguish student ability levels well Visualization of the distribution of respondent ability and item difficulty is displayed through the Wright map (Item-Person Map), to see whether the items are evenly distributed across student ability levels [39]. After all analyses were conducted, items that were inappropriate or showed bias were revised or deleted. The revised instrument is then retested on a larger sample group (field test II) to confirm final validity and reliability. The final result is an instrument that is valid, reliable, unidimensional, bias-free, and ready to be used for research and learning evaluation. This approach has been widely used in the validation of science and energy literacy instruments.

3. RESULTS AND DISCUSSION

The assessment instrument consists of 10 knowledge items, 10 attitude items, and 9 behavior items. The average person size is.

3.1. Evidence of validity argument

3.1.1 Knowledge instrument items

It shows that most items fall within the ideal MNSQ range, which is between 0.84 and 1.25. This range indicates that most items have a good level of fit to the model. However, there is one item, item number 4, which has an *MNSQ outfit* value of 2.13, which exceeds the upper tolerance limit (1.5). This indicates that the item shows a mismatch to the model and most likely contains *noise* or measurement inaccuracy, which could be due to student guessing or the mismatch of the item context to the ability being measured.

Furthermore, the ZSTD values for almost all items were within ±2.0, indicating that none of the items statistically deviated too far from the model expectations, although the ZSTD value for item 4 was also high (2.06), corroborating the previous finding that this item needed to be re-examined. In addition, the *point measure* correlation values for all items are positive, which means that all items make a positive contribution to the measurement of student ability. However, it should be noted that the correlation value of item 4 is only 0.23, which is lower than the other items which are generally above 0.30. This low correlation suggests that students' responses to item 4 are not consistent with their general ability.

Table 1. Misfit order of knowledge

TABLE 10.1 pengetahuan 1.xlsx ZOU843WS.TXT Jul 02 2025 10:56
INPUT: 50 Person 10 Item REPORTED: 50 Person 10 Item 2 CATS MINISTEP 5.7.0.0

Person: REAL SEP.: 1.01 REL.: .50 ... Item: REAL SEP.: 2.44 REL.: .86

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
4	10	50	1.69	.38	1.05	.31	2.13	2.06	A .23	.34	80.0	79.9	item4
7	27	50	-.20	.32	1.16	1.18	1.25	1.29	B .34	.47	70.0	71.0	item7
10	25	50	.00	.32	1.14	1.12	1.22	1.15	C .35	.47	66.0	70.0	item10
6	20	50	.51	.32	1.10	.84	1.13	.64	D .35	.44	72.0	69.3	item6
1	29	50	-.41	.33	.95	-.28	.94	-.27	E .51	.48	74.0	71.8	item1
5	37	50	-1.35	.37	.95	-.18	.82	-.53	e .51	.46	78.0	79.1	item5
8	17	50	.83	.33	.93	-.50	.89	-.32	d .47	.42	78.0	71.9	item8
9	20	50	.51	.32	.93	-.52	.84	-.66	c .50	.44	68.0	69.3	item9
3	26	50	-.10	.32	.86	-1.05	.84	-.84	b .57	.47	76.0	70.6	item3
2	38	50	-1.48	.37	.81	-.94	.57	-1.41	a .63	.46	80.0	80.2	item2
MEAN	24.9	50.0	.00	.34	.99	.00	1.06	.11			74.2	73.3	
P.SD	8.2	.0	.91	.02	.11	.78	.41	1.05			4.8	4.3	

Overall, eight out of ten items were classified as excellent and valid, namely items 1, 3, 5, 6, 7, 8, 9, and 10, as they met the fit criteria and adequate correlation. One item (item 4) is recommended for revision as it shows significant misfit, while item 2 can still be used but needs further monitoring as its correlation value is at the lower limit. Thus, this instrument generally has good validity, but there is a need for improvement or refinement on one item in order to measure students' abilities more accurately and representatively.

Attitude instrument item

Table 2. Misfit order of attitude

TABLE 10.1 Data sikap 1.xlsx ZOU391WS.TXT Jul 02 2025 11:38
INPUT: 50 PERSON 10 ITEM REPORTED: 50 PERSON 10 ITEM 5 CATS MINISTEP 5.7.0.0

PERSON: REAL SEP.: 1.94 REL.: .79 ... ITEM: REAL SEP.: 2.75 REL.: .88

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
10	115	50	1.21	.14	2.95	6.82	4.32	7.43	A .03	.52	12.2	35.7	Item 10
1	211	50	-.66	.17	1.35	1.35	1.17	.62	B .67	.68	46.9	50.2	Item1
3	183	50	.00	.14	.92	-.36	.86	-.51	C .67	.63	36.7	35.3	Item3
2	206	50	-.52	.16	.90	-.35	.82	-.54	D .74	.67	51.0	48.5	Item2
4	182	50	.02	.14	.78	-1.15	.89	-.39	E .67	.63	44.9	35.9	Item4
5	172	50	.21	.13	.89	-.57	.85	-.64	e .62	.62	57.1	33.4	Item5
7	188	50	-.10	.14	.60	-2.27	.77	-.93	d .71	.64	59.2	36.2	Item7
8	190	50	-.14	.14	.67	-1.77	.72	-1.14	c .74	.64	44.9	36.2	Item8
6	183	50	.00	.14	.60	-2.35	.57	-2.07	b .75	.63	49.0	35.3	Item6
9	184	50	-.02	.14	.42	-3.79	.39	-3.22	a .79	.63	59.2	35.4	Item9
MEAN	181.4	50.0	.00	.15	1.01	-.44	1.14	-.14			46.1	38.2	
P.SD	24.7	.0	.47	.01	.69	2.77	1.08	2.71			13.2	5.6	

Based on the table, the average *measure* value of the items is 0.00 logit, indicating that the difficulty level of the items is around the average ability of the respondents. However, there are significant findings in the *fit statistics* values, especially in the *infit* and *outfit mean square* (MNSQ). The average infit MNSQ of 1.01 and outfit MNSQ of 1.14 is still within the tolerance range (0.5-1.5), but there are items that deviate greatly, namely Item10 with an infit MNSQ of 2.95 and outfit MNSQ of 4.32, accompanied by high ZSTD values (6.82 and 7.43). This indicates that participants' responses to this item were highly at odds with the model predictions, which could be due to guessing, ambiguity in the statement, or unclear context of the item.

The *point measure* correlation value (PTMEASUR-AL) for Item10 is also very low, only 0.03, indicating that this item barely contributes to measuring the intended attitude construct. In contrast, other items such as Item9, Item6, and Item7 show high correlations (above 0.70), as well as more ideal fit values, indicating that they consistently measure aspects of attitude according to the Rasch model. Items that have low correlation and high deviant fit such as Item10 need to be evaluated and revised immediately.

Almost all items were categorized as good. However, the presence of one highly misfit item (Item10) needs serious attention. According to [40] items like this risk lowering the overall validity of the instrument and

should be considered for deletion or substantial revision. This item may have a language structure or context that is incompatible with respondents' understanding, thus not measuring the attitude construct validly.

Behavioral instrument item

Table 3. Misfit order of behavior

TABLE 10.1 Data Perilaku 1.xlsx ZOU234WS.TXT Jul 02 2025 11:31
INPUT: 50 PERSON 9 ITEM REPORTED: 50 PERSON 9 ITEM 5 CATS MINISTEP 5.7.0.0

PERSON: REAL SEP.: 1.59 REL.: .72 ... ITEM: REAL SEP.: 2.52 REL.: .86

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
1	115	50	.90	.13	2.23	4.61	4.94	5.97	A .28	.63	21.3	38.5	Item1
2	189	50	-.24	.13	1.32	1.46	1.33	1.06	B .57	.63	40.4	38.8	Item2
4	186	50	-.19	.13	1.01	.13	.79	-.63	C .66	.63	36.2	36.3	Item4
3	166	50	.13	.12	.88	-.59	.83	-.53	D .66	.63	40.4	30.3	Item3
5	205	50	-.56	.15	.87	-.45	.59	-1.21	E .69	.63	51.1	50.2	Item5
8	161	50	.21	.12	.84	-.84	.77	-.78	d .67	.63	46.8	30.2	Item8
9	175	50	-.01	.13	.73	-1.45	.70	-1.04	c .70	.63	34.0	31.4	Item9
7	178	50	-.05	.13	.69	-1.71	.61	-1.46	b .70	.63	40.4	31.6	Item7
6	187	50	-.20	.13	.51	-2.79	.44	-2.22	a .75	.63	55.3	38.3	Item6
MEAN	173.6	50.0	.00	.13	1.01	-.18	1.22	-.09			40.7	36.2	
P.SD	24.1	.0	.38	.01	.48	2.03	1.34	2.30			9.4	6.0	

Based on the results of the analysis using the Rasch Model on 9 behavioral items with 50 respondents, it was found that there were several items that had *fit statistics* values outside the ideal limits. The average *infit mean square* (MNSQ) value is 1.01 and *the outfit MNSQ* is 1.22. Although on average it is still within the tolerance range of 0.5-1.5, there are some items that deviate quite significantly, especially in the MNSQ outfit value. The first item (Item1) shows an outfit MNSQ value of 5.97 and a ZSTD of 4.61, which indicates an *overfit* or response pattern that is very inconsistent with the model predictions. Item2 and Item4 also showed outfit values above 2.0, indicating the possibility that the items were not working as intended in measuring student behavior, which could be due to guessing, miscoding, or ambiguous item interpretation.

Overall, the behavioral question instrument has good measurement quality based on the Rasch Model. However, there are some items that show high statistical misfit and potentially interfere with the overall validity of the instrument. Therefore, revision and revalidation of these items are needed, especially those with outfit MNSQ > 2.0 and low correlation, in order to improve the accuracy and suitability of the instrument in measuring student behavior effectively and objectively.

3.2. Evidence for reliability argument

3.2.1 Reliability of knowledge instruments

Table 4. Summary statistics of knowledge

TABLE 3.1 pengetahuan 1.xlsx ZOU843WS.TXT Jul 02 2025 10:56
INPUT: 50 Person 10 Item REPORTED: 50 Person 10 Item 2 CATS MINISTEP 5.7.0.0

SUMMARY OF 50 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	5.0	10.0	-.05	.77	1.00	.00	1.06	.01
SEM	.3	.0	.17	.02	.04	.13	.13	.13
P.SD	2.1	.0	1.17	.11	.28	.92	.90	.93
S.SD	2.1	.0	1.18	.11	.28	.93	.91	.94
MAX.	8.0	10.0	1.61	1.09	1.72	2.29	6.75	2.52
MIN.	1.0	10.0	-2.50	.69	.59	-1.78	.36	-1.43

REAL RMSE .82 TRUE SD .83 SEPARATION 1.01 Person RELIABILITY .50
MODEL RMSE .78 TRUE SD .87 SEPARATION 1.12 Person RELIABILITY .56
S.E. OF Person MEAN = .17

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .56 SEM = 1.41
STANDARDIZED (50 ITEM) RELIABILITY = .86

The *person reliability* value based on real data of 0.50 indicates that the instrument's ability to distinguish between high and low ability students is limited. The separation index value of only 1.12 further emphasizes that this instrument is only able to group students into about two ability strata. This finding is in line with the opinion of [41] which states that a separation index below 2.0 indicates the limitations of the instrument in detecting variations in the ability of participants, and indicates a mismatch between the items and the construct being measured, so it is necessary to review the content and construct validity of the questions. Therefore, it is necessary to increase both the number of items and the variation in the level of difficulty of the questions so that the instrument can measure students' abilities more sharply.

Overall, this knowledge assessment instrument is good enough to be used in the context of Rasch-based measurement, especially because the students' conformity value to the model is ideal and the distribution of students' abilities is normal. However, to improve the quality of measurement, further development is needed, especially on the reliability and differentiation aspects of the instrument, for example through increasing the number of items and a wider variety of question difficulties in order to cover a more comprehensive and representative spectrum of student abilities.

3.2.2 Reliability of attitude instruments

Table 5. Summary statistics of attitude

TABLE 3.1 Data sikap 1.xlsx ZOU391WS.TXT Jul 02 2025 11:38
INPUT: 50 PERSON 10 ITEM REPORTED: 50 PERSON 10 ITEM 5 CATS MINISTEP 5.7.0.0

SUMMARY OF 49 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	36.8	10.0	.60	.35	1.04	-.08	1.14	.02
SEM	1.1	.0	.12	.02	.09	.21	.15	.25
P.SD	7.8	.0	.84	.12	.63	1.47	1.03	1.72
S.SD	7.9	.0	.85	.12	.63	1.49	1.04	1.74
MAX.	48.0	10.0	2.25	.95	3.00	2.51	4.51	4.52
MIN.	11.0	10.0	-2.84	.27	.24	-3.02	.26	-2.76

REAL RMSE	.41	TRUE SD	.73	SEPARATION 1.78	PERSON RELIABILITY .76
MODEL RMSE	.37	TRUE SD	.75	SEPARATION 2.04	PERSON RELIABILITY .81

S.E. OF PERSON MEAN = .12

MINIMUM EXTREME SCORE: 1 PERSON 2.0%

Based on the reliability aspect. The person reliability value of 0.76 (real) indicates that the instrument is reliable enough to distinguish between individuals with different attitudes. The *person separation* values of 1.78 (real) and 2.04 (model) indicate that the instrument can distinguish respondents into at least two different ability groups. These results indicate that although still in the moderate category, this attitude instrument has the potential to be further developed to have higher classification acuity.

There is one item that deviates significantly from the model, Item10, which has an *infit MNSQ* value of 2.95 and *an outfit MNSQ* of 4.32. These values are well above the tolerance limit of 2.0 and are accompanied by high *ZSTDs* (6.82 and 7.43), indicating extreme misfit. This item is suspected of not measuring the attitude construct consistently, which could be due to ambiguity, multiple interpretations, or inappropriate context. [42] 4) recommend that items with outfit MNSQ > 2.0 should be revised or deleted to improve the accuracy of the model.

3.2.3 Behavioral Reliability

Table 6. Summary statistics of behavior

TABLE 3.1 Data Perilaku 1.xlsx ZOU234WS.TXT Jul 02 2025 11:31
INPUT: 50 PERSON 9 ITEM REPORTED: 50 PERSON 9 ITEM 5 CATS MINISTEP 5.7.0.0

SUMMARY OF 47 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	31.1	9.0	.29	.34	.99	-.10	1.22	.05
SEM	1.3	.0	.12	.02	.10	.21	.20	.24
P.SD	8.5	.0	.80	.15	.65	1.45	1.38	1.60
S.SD	8.6	.0	.81	.16	.66	1.46	1.39	1.62
MAX.	44.0	9.0	2.24	.89	3.10	3.10	6.40	4.89
MIN.	10.0	9.0	-2.03	.25	.13	-3.19	.17	-2.49

REAL RMSE	.40	TRUE SD	.69	SEPARATION 1.71	PERSON RELIABILITY .75
MODEL RMSE	.37	TRUE SD	.71	SEPARATION 1.88	PERSON RELIABILITY .78
S.E. OF PERSON MEAN = .12					

MAXIMUM EXTREME SCORE: 2 PERSON 4.0%
MINIMUM EXTREME SCORE: 1 PERSON 2.0%

The *person reliability* value of 0.75 and the *person separation of* 1.71 indicate that the instrument is good at differentiating students' abilities, although not enough to group participants into more than two ability levels in a stable manner. In the context of Rasch modeling, a separation value of at least 2.0 is recommended to classify students into three different ability levels [23] Although not ideal, these results suggest that with item improvements or an increase in the number of questions, the reliability and distinctiveness of the instrument can be further improved. This finding is in line with the research by [28]. who emphasized the importance of considering the separation value in evaluating the quality of Rasch-based educational measurement tools.

It was concluded that the instrument was sufficiently reliable and feasible to be used to measure behavioral differences between respondents in the research sample. Participants in this study had enough behavioral variation to be mapped by the instrument, so its use in behavioral evaluation or mapping was considered quite effective. Thus, it can be concluded that the instrument has good internal validity in distinguishing behavioral levels between individuals.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	14.5832	100.0%	100.0%
Raw variance explained by measures =	4.5832	31.4%	29.5%
Raw variance explained by persons =	2.1645	14.8%	13.9%
Raw Variance explained by items =	2.4186	16.6%	15.6%
Raw unexplained variance (total) =	10.0000	68.6% 100.0%	70.5%
Unexpended variance in 1st contrast =	1.8272	12.5% 18.3%	
Unexpended variance in 2nd contrast =	1.6394	11.2% 16.4%	
Unexpended variance in 3rd contrast =	1.3281	9.1% 13.3%	
Unexpended variance in 4th contrast =	1.1540	7.9% 11.5%	
Unexpended variance in 5th contrast =	1.1328	7.8% 11.3%	

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	18.9655	100.0%	100.0%
Raw variance explained by measures =	8.9655	47.3%	47.6%
Raw variance explained by persons =	5.7845	30.5%	30.7%
Raw Variance explained by items =	3.1811	16.8%	16.9%
Raw unexplained variance (total) =	10.0000	52.7% 100.0%	52.4%
Unexplained variance in 1st contrast =	2.4013	12.7% 24.0%	
Unexplned variance in 2nd contrast =	1.6442	8.7% 16.4%	
Unexplned variance in 3rd contrast =	1.5073	7.9% 15.1%	
Unexplned variance in 4th contrast =	1.1159	5.9% 11.2%	

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	17.6826	100.0%	100.0%
Raw variance explained by measures =	8.6826	49.1%	48.7%
Raw variance explained by persons =	5.6891	32.2%	31.9%
Raw Variance explained by items =	2.9935	16.9%	16.8%
Raw unexplained variance (total) =	9.0000	50.9% 100.0%	51.3%
Unexplned variance in 1st contrast =	2.3051	13.0% 25.6%	
Unexplned variance in 2nd contrast =	1.8632	10.5% 20.7%	
Unexplned variance in 3rd contrast =	1.1877	6.7% 13.2%	
Unexplned variance in 4th contrast =	1.0282	5.8% 11.4%	
Unexplned variance in 5th contrast =	.8503	4.8% 9.4%	

Figure 1. Standardized residual variance in eigenvalue units.

Unidimensionality is a fundamental right of the Rasch model. Unidimensionality analysis is conducted to ensure the measurement scale accurately measures one domain of energy literacy. Based on the three residual variance analysis results shown, it can be concluded that the knowledge instrument consistently shows a tendency to fulfill the assumption of unidimensionality according to the Rasch model. The three tables show that the proportion of *raw variance explained by measures* is 31.4%, 47.3% and 49.1% respectively, all of which are above the minimum threshold of 30% as recommended by [43] thus indicating that one main dimension in the instrument is strong enough to explain the data structure. Meanwhile, the value of *unexplained variance in 1st contrast in* two of the three analyses had an eigenvalue of (1.8272 2.4013 and 2.3051 respectively, indicating that the items have a secondary dimension equivalent to three items. The instrument continues to measure predominantly the same construct. However, the examination of the items of the three aspects of the instrument can be declared to have strong unidimensional properties and valid and the items belong to the same construct, namely the energy literacy domain.

The results of this instrument validation indicate that the Rasch Model is effective in identifying misfit items, while also providing guidance for revisions based on statistical analyses such as infit, outfit, ZSTD, and PTMEA Corr. A study by Linacre (2012) shows that item evaluation based on MNSQ and ZSTD can detect items with extreme misfit. These results align with previous research in the context of validating science and environmental literacy instruments [44], [45]. Additionally, a study by [46] shows that Rasch analysis on pro-environmental behavior instruments enables the detection of cross-group bias. Positive correlations between items are also an important indicator of construct validity, widely used in the development of attitude and value measurement tools [47]. Reliability findings with separation values < 2.0 also suggest expanding the range of question difficulty levels, as demonstrated by [48]. in their development of a climate attitude instrument among students. Considering these validation results, further revisions and testing are needed to obtain an instrument that can comprehensively and deeply describe energy literacy while being applicable to broader populations and regions [49], [50].

This study has a unique feature in its application of the Rasch Model as a statistical approach in developing and validating an energy literacy measurement instrument that comprehensively covers the dimensions of knowledge, attitude, and behavior. Unlike most previous studies that used a classical approach (CTT), this study adopted the Rasch approach to produce objective, unidimensional, and evidence-based measurements, as well as to detect statistically misfit items. Additionally, this study specifically designed the instrument for the context of high school students in Indonesia, which has rarely been the focus of modern psychometric-based energy literacy measurement tool development. This study's main uniqueness lies in the application of the Rasch Model as a statistical approach in developing and validating an energy literacy measurement instrument that comprehensively covers the dimensions of knowledge, attitude, and behavior. Unlike most previous studies that used a classical approach (CTT), this study adopted the Rasch approach to produce objective, unidimensional, and evidence-based measurements, as well as to detect statistically misfit items. This study specifically designed an instrument for the context of high school students in Indonesia, which has rarely been the focus of modern psychometric-based energy literacy measurement tool development.

This study opens opportunities for the development of similar instruments at different educational levels or in other literacy domains, such as digital and scientific literacy. Additionally, the developed instrument can be used in longitudinal studies to track the development of students' energy literacy over time.

4. CONCLUSION

The final results of the assessment instrument consisting of 10 knowledge items, 10 attitude items, and 9 behavior items have been tested using the Rasch model. The test results showed that almost all items met the criteria for fit with the Rasch model and had evidence for the validity and reliability arguments of the instrument. The use of the Rasch Model proved to be very effective in measuring participants' abilities equitably, as well as providing statistical parameters that are able to identify inappropriate or problematic items. In general, this instrument provides a strong basis for evaluation and research activities in the field of energy literacy, and can be adopted for a wider range of populations and educational contexts. In the future, the development of the instrument should continue to be carried out continuously through revision and retesting on a larger and more diverse sample, so that its validity and reliability are increasingly tested and guaranteed.

ACKNOWLEDGEMENTS

Thank you to all respondent and stakeholder for the permission and opportunity given to carry out this research. Then, the researcher also thanked all parties who had contributed to the success of this research.

REFERENCES

[1] US EIA, "World Energy Outlook 2023 - DOE/EIA (2023)," *World Energy Outlook, 2023*, 2023.
[2] K. L. Chen, S. Y. Liu, and P. H. Chen, "Assessing multidimensional energy literacy of secondary students using

contextualized assessment," *Int. J. Environ. Sci. Educ.*, 2015, doi: 10.12973/ijese.2015.241a.
[3] S. Wu, Y. Li, C. Fang, and P. Ju, "Energy Literacy of Residents and Sustainable Tourism Interaction in Ethnic Tourism: A Study of the Longji Terraces in Guilin, China," *Energies*, 2023, doi: 10.3390/en16010259.
[4] Y. F. Lee, H. B. N. Nguyen, and H. Te Sung, "Energy literacy of high school students in Vietnam and determinants of their energy-saving behavior," *Environ. Educ. Res.*, 2022, doi: 10.1080/13504622.2022.2034752.
[5] M. Maurer, P. Koulouris, and F. X. Bogner, "Green awareness in action-how energy conservation action forces on environmental knowledge, values and behaviour in adolescents' school life," *Sustain.*, 2020, doi: 10.3390/su12030955.
[6] M. Yusup, A. Setiawan, N. Y. Rustaman, and I. Kaniawati, "Pre-service Physics Teachers' Knowledge, Decision Making, and Self-system Toward Energy Conservation," *J. Pendidik. Fis. Indones.*, 2018, doi: 10.15294/jpfi.v14i2.16638.
[7] K. A. Babatunde, N. F. Wakimin, S. H. B. Zakaria, M. Soman, and K. Charman, "Determinants of Sustainable Energy Saving Behavior among University Students in Malaysia: A Case Study of Management & Science University (MSU)," *Inf. Manag. Bus. Rev.*, 2023, doi: 10.22610/imbr.v15i3(i).3534.
[8] OECD, "Revenue Statistics in Asia and the Pacific: Key findings for Indonesia," *OECD Tax Stat.*, 2022.
[9] R. Kc, J. Föhr, and T. Ranta, "Public Perception on the Sustainable Energy Transition in Rural Finland: A Multi-criteria Approach," *Circ. Econ. Sustain.*, 2023, doi: 10.1007/s43615-022-00206-5.
[10] P. Papon, "AIE (Agence internationale de l'énergie). World Energy Outlook 2023, Paris : AIE octobre 2023, 353 p. URL : https://iea.blob.core.windows.net/assets/42b23c45-78bc-4482-b0f9-eb826ae2da3d/WorldEnergyOutlook2023.pdf . Consulté le 10 janvier 2024," *Futuribles*, 2024, doi: 10.3917/futur.459.0114.
[11] K. S. Ermenc and E. K. Niemczyk, "Education for Sustainable Development," *Sodobna Pedagogika/Journal of Contemporary Educational Studies* . 2022. doi: 10.21300/21.4.2020.2.
[12] A. Lange Salvia, L. Londero Brandli, W. Leal Filho, B. Gasparetto Rebelatto, and G. Reginatto, "Energy sustainability in teaching and outreach initiatives and the contribution to the 2030 Agenda," *Int. J. Sustain. High. Educ.*, 2020, doi: 10.1108/IJSHE-05-2020-0180.
[13] H. Husamah, H. Suwono, H. Nur, and A. Dharmawan, "Sustainable development research in Eurasia Journal of Mathematics, Science and Technology Education: A systematic literature review," *Eurasia J. Math. Sci. Technol. Educ.*, 2022, doi: 10.29333/ejmste/11965.
[14] D. R. E. Cotton, W. Miller, J. Winter, I. Bailey, and S. Sterling, "Developing students' energy literacy in higher education," *Int. J. Sustain. High. Educ.*, 2015, doi: 10.1108/IJSHE-12-2013-0166.
[15] J. DeWaters and S. Powers, "Establishing measurement criteria for an energy literacy questionnaire," *J. Environ. Educ.*, 2013, doi: 10.1080/00958964.2012.711378.
[16] R. Rohmatulloh, A. Hasanah, L. Sahlani, M. T. Zuhri, N. Kholifah, and M. Nurtanto, "A Systematic Review of Energy Literacy Programs at Primary and Middle Schools," *Pegem Egit. ve Ogr. Derg.*, 2022, doi: 10.47750/pegegog.13.01.17.
[17] A. Reid, "Climate change education and research: possibilities and potentials versus problems and perils?," *Environmental Education Research*. 2019. doi: 10.1080/13504622.2019.1664075.
[18] M. Wilson, *CONSTRUCTING MEASURES: An Item Response Modeling Approach, Second Edition*. 2023. doi: 10.4324/9781003286929.
[19] W. J. Boone and J. R. Staver, "Correction to: Advances in Rasch Analyses in the Human Sciences," in *Advances in Rasch Analyses in the Human Sciences*, 2020. doi: 10.1007/978-3-030-43420-5_21.
[20] A. A. Latif, M. Najib, W. H. Libunao, N. M. Amin, and C. J. Peter, "Enhancing student ' s akhlaq behaviour through holistic assessment," *Int. Conf. Soc. Sci. Res. ICSSR 2013*, 2013.
[21] B. D. Solomon, "United Nations Environment Programme (UNEP)," in *Dictionary of Ecological Economics: Terms for the New Millennium*, 2023. doi: 10.4337/9781788974912.U.7.
[22] R. García-Sierra, M. I. Fernández-Cano, M. Jiménez-Pera, M. Feijoo-Cid, and A. Arreciado Marañón, "Knowledge about the best practice guidelines in the nursing degree: A non-randomized post-test design," *Nurs. Open*, 2024, doi: 10.1002/nop2.2074.
[23] T. G. Bond, Z. Yan, and M. Heene, *Applying the rasch model: Fundamental measurement in the human sciences*. 2020. doi: 10.4324/9780429030499.
[24] L. Tesio, A. Caronni, D. Kumbhare, and S. Scarano, "Interpreting results from Rasch analysis 1. The 'most likely' measures coming from the model," *Disabil. Rehabil.*, 2024, doi: 10.1080/09638288.2023.2169771.
[25] T. H. Fiskawarni, R. Rahmawati, W. Widiasih, and R. Saad, "The Design and Validation of the Four Tier Test Instrument for Energy Literacy Using the Rasch Model Analysis," *JIPF (Jurnal Ilmu Pendidik. Fis.*, vol. 9, no. 1, p. 74, 2024, doi: 10.26737/jipf.v9i1.4717.
[26] E. Avinç and F. Doğan, *Digital literacy scale: Validity and reliability study with the rasch model*, vol. 29, no. 17. Springer US, 2024. doi: 10.1007/s10639-024-12662-7.
[27] P. Susongko, "Construct Validation with Rasch Modeling of Pedagogical Literacy Measurement Instrument for Teachers in Junior High School Environment," vol. 10, no. 16, pp. 312–322, 2025.
[28] J. Yang, Z. Reheman, Y. Liu, and S. Zhao, "Applying Rasch analysis in refinement and validation of interpersonal skills measure for gifted children," *Front. Psychol.*, 2023, doi: 10.3389/fpsyg.2023.1236640.
[29] G. H. Alnahdi, J. Goldan, and S. Schwab, "Psychometric Properties and Rasch Validation of the Teachers' Version of the Perception of Resources Questionnaire," *Front. Psychol.*, 2021, doi: 10.3389/fpsyg.2021.633801.
[30] A. Rahim and H. Haryanto, "Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics," *J. Educ. Res. Eval.*, 2021, doi: 10.15294/jere.v10i2.51802.
[31] O. Aman, T. Z. Ningsih, and N. F. Abidin, "The Development of Historical Thinking Assessment to Examine Students' Skills in Analyzing the Causality of Historical Events," *Eur. J. Educ. Res.*, 2022, doi: 10.12973/eu-jer.11.2.609.
[32] M. Bond, S. Bedenlier, V. I. Marín, and M. Händel, "Emergency remote teaching in higher education: mapping the first

global online semester," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, 2021, doi: 10.1186/s41239-021-00282-x.
[33] R. Sari and M. N. I. Saleh, "Rasch Model Application: Instrument Development of Readiness to Conduct Inclusive Learning," *Psympathic J. Ilm. Psikol.*, 2023, doi: 10.15575/psy.v9i2.10058.
[34] N. V. Morgacheva, S. V. Shcherbatykh, and E. B. Sotnikova, "Assessment and analysis of the level of natural science literacy of students," *Perspekt. Nauk. i Obraz.*, 2023, doi: 10.32744/pse.2023.2.4.
[35] M. Yusup, A. Setiawan, N. Y. Rustaman, and I. Kaniawati, "Developing a Framework for the Assessment of Pre-Service Physics Teachers' Energy Literacy," *J. Phys. Conf. Ser.*, vol. 877, no. 1, 2017, doi: 10.1088/1742-6596/877/1/012014.
[36] M. Yusup, "Using Rasch model for the development and validation of energy literacy assessment instrument for prospective physics teachers," in *Journal of Physics: Conference Series*, 2021. doi: 10.1088/1742-6596/1876/1/012056.
[37] S. A. Alatas, M. Jahja, and M. Yusuf, "Assessment of students' science literacy skills using diagnostic items on the concept of energy with Rasch modeling," *World J. Adv. Res. Rev.*, 2024, doi: 10.30574/wjarr.2024.21.1.2697.
[38] M. Yusup, A. Setiawan, N. Y. Rustaman, and I. Kaniawati, "Assessing Pre-Service Physics Teachers' Energy Literacy: An Application of Rasch measurement," *J. Phys. Conf. Ser.*, vol. 895, no. 1, 2017, doi: 10.1088/1742-6596/895/1/012161.
[39] K. Thayaseelan, Y. Zhai, S. Li, and X. Liu, "Revalidating a measurement instrument of spatial thinking ability for junior and high school students," *Discip. Interdiscip. Sci. Educ. Res.*, vol. 6, no. 1, 2024, doi: 10.1186/s43031-024-00095-8.
[40] Y. Oktalia, A. B. Susila, and B. H. Iswanto, "Virtual test instruments to measure scientific literacy of high school students on work and energy," in *Journal of Physics: Conference Series*, 2021. doi: 10.1088/1742-6596/2019/1/012009.
[41] E. Bartholomew, N. Iqbal, and O. Medvedev, "Enhancing the Assessment of Gratitude in Mindfulness Research: a Rasch Analysis of the 6-Item Gratitude Questionnaire," *Mindfulness (N. Y).*, 2022, doi: 10.1007/s12671-022-02006-2.
[42] L. Tesio, A. Caronni, A. Simone, D. Kumbhare, and S. Scarano, "Interpreting results from Rasch analysis 2. Advanced model applications and the data-model fit assessment," *Disabil. Rehabil.*, 2024, doi: 10.1080/09638288.2023.2169772.
[43] J. M. Linacre, "Rasch dichotomous model vs. one-parameter logistic model (1PL 1-PL)," *Rasch Meas. Trans.*, 2005.
[44] A. Sukri, M. A. Rizka, E. Purwanti, S. Ramdiah, and M. Lukitasari, "Validating Student's Green Character Instrument Using Factor and Rasch Model," *Eur. J. Educ. Res.*, 2022, doi: 10.12973/eu-jer.11.2.859.
[45] M. N. Mamat and F. Mokhtar, "Reliable assessment model for affective environmental education: Rasch model made simple," in *2012 International Conference on E-Learning and E-Technologies in Education, ICEEE 2012*, 2012. doi: 10.1109/ICeLeTE.2012.6333387.
[46] H. Lahza, H. Khosravi, and G. Demartini, "Analytics of learning tactics and strategies in an online learnersourcing environment," *J. Comput. Assist. Learn.*, 2023, doi: 10.1111/jcal.12729.
[47] N. F. Ramli, O. Talib, S. A. Hassan, and U. K. A. Manaf, "Development and Validation of an Instrument to Measure STEM Teachers' Instructional Preparedness," *Asian J. Univ. Educ.*, 2020, doi: 10.24191/ajue.v16i3.11084.
[48] N. E. M. Razali, R. Z. Ramli, H. Mohamed, N. A. Mat Zin, F. Rosdi, and N. Mat Diah, "Identifying and validating game design elements in serious game guideline for climate change," *Heliyon*, 2022, doi: 10.1016/j.heliyon.2022.e08773.
[49] T. LeVasseur and C. Ciarcia, "Sustainability literacy in a time of socio-ecological crisis: Using reaccreditation as a leverage point for innovation in higher education," *Sustain.*, 2019, doi: 10.3390/su11185104.
[50] L. Wang, Y. Yuan, and G. Wang, "The Construction of Civil Scientific Literacy in China from the Perspective of Science Education," *Sci. Educ.*, 2024, doi: 10.1007/s11191-022-00367-7.