JPPIPA 11(3) (2025)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

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

Analysis of the Quality of Sudents' Critical Reasoning Instruments Using Rasch Model

Yokhebed[1*], Rexy Maulana Dwi Karmadi[1], Jayanti Silaen[1], Sherli Meidina[1], Nur Alfiatun Nikmah[1]

[1] Biology Education Study Program, Faculty of Teacher Training and Education, Universitas Tanjungpura, Pontianak City, Indonesia.

Received: October 30, 2024
Revised: January 22, 2025
Accepted: March 25, 2025
Published: March 31, 2025

Corresponding Author:
Yokhebed
yokhebed@fkip.untan.ac.id

DOI: 10.29303/jppipa.v11i3.9584

Abstract: This research aims to analyzes the validity and reliability of a students' critical reasoning questionnaire using the Rasch Model. This quantitative descriptive research employed a non-experimental survey design involving 122 fourth-semester undergraduate students from the College of Teacher Training and Education (STKIP) Melawi. Data were collected through an online questionnaire distributed via Google Forms. Validity was assessed through unidimensionality and item fit analysis, while reliability was measured using person reliability, item reliability, and Cronbach's Alpha. The results indicate that the instrument has not yet achieved good item validity, with only 4 out of 20 items classified as valid. Reliability analysis showed Cronbach's Alpha of 0.55 (bad category), item reliability of 0.98 (special category), and person reliability of 0.44 (weak category). Item difficulty analysis identified four difficulty levels: 2 items were very easy, 9 easy, 6 difficult, and 3 very difficult. The Rasch model's item discrimination analysis produced eight item groups based on an item separation index of H = 9.97. These findings emphasize the need to refine questionnaire items and enhance respondent engagement. A validated instrument will support educational assessments by providing a reliable tool for measuring students' critical reasoning skills in teacher training institutions.

Keywords: Critical reasoning; Instrument; Rasch model; Reliability; Validity

Introduction

Education in the 21st century demands critical reasoning skills to navigate complex global challenges effectively (Asmarita & Prastowo, 2023; Packer et al., 2022). Critical reasoning enables individuals to analyze, evaluate, and formulate arguments systematically, making it a crucial competency for both academic and professional success (Azmi & Prasetya, 2025; Kaur, 2018). In an era dominated by rapid technological advancements and information overload, the ability to critically assess information is more vital than ever. Without adequate reasoning skills, individuals may struggle to distinguish credible information from misinformation, a challenge that is particularly significant in the context of Society 5.0 (Denis, 2022; Gustianingrum et al., 2023).

The Indonesian education system, like many others worldwide, faces difficulties in effectively fostering critical reasoning among students. Traditional assessment methods often emphasize rote memorization rather than analytical thinking, which may hinder students' ability to apply reasoning skills in real-world contexts (Ramdani et al., 2020). Although some previous studies have developed test-based assessments for critical reasoning using socioscientific issues (Yokhebed et al., 2024), these methods may not comprehensively capture the nuances of students' reasoning abilities. A more holistic approach such as the use of questionnaires could provide a broader perspective on students' reasoning capabilities beyond specific test scenarios.

Despite the acknowledged importance of critical reasoning, there is a lack of reliable instruments to measure students' reasoning abilities, particularly in

How to Cite:
Yokhebed, Karmadi, R. M. D., Silaen, J., Meidina, S., & Nikmah, N. A. Analysis of the Quality of Sudents' Critical Reasoning Instruments Using Rasch Model. *Jurnal Penelitian Pendidikan IPA*, *11*(3), 54-60. https://doi.org/10.29303/jppipa.v11i3.9584

Indonesia. Existing studies have primarily focused on test-based assessments, leaving non-test instruments such as questionnaires underexplored. Questionnaires can serve as effective tools for assessing students' self-perceptions of their reasoning abilities and their application of critical reasoning in different contexts. However, the validity and reliability of such instruments must be rigorously tested to ensure their effectiveness. Previous research has not extensively applied the Rasch model to analyze the psychometric properties of critical reasoning instruments, leaving a methodological gap in the field (Anggraini et al., 2023; Sari & Rakhmawati, 2024). Addressing this gap is essential to provide educators and researchers with more robust tools for assessing and enhancing students' reasoning abilities.

This study focuses on STKIP Melawi, where assessing students' critical reasoning skills is particularly relevant due to the institution's role in training future educators. Teachers play a crucial role in shaping students' cognitive abilities, and their ability to foster critical thinking skills in classrooms is directly influenced by their own reasoning capabilities. However, no comprehensive evaluation of critical reasoning instruments has been conducted at STKIP Melawi, highlighting a pressing need for research in this area. Understanding teacher candidates' reasoning skills will provide insights into potential improvements in teacher education programs, ensuring that future educators are better equipped to cultivate critical thinking in their students.

To address these gaps, this study aims to analyze the validity and reliability of a student critical reasoning questionnaire using the Rasch model. The Rasch model is a robust framework that ensures measurement accuracy by assessing item suitability, reliability, and unidimensionality. It provides valuable insights into item difficulty, respondent ability, and the overall effectiveness of an instrument in measuring the intended construct (Alsoudi & Shindi, 2023; Laliyo et al., 2022). By applying this model, this study will contribute to the development of a well-validated questionnaire that can be used as a standardized tool for assessing students' critical reasoning skills in teacher training institutions.

The findings of this research are expected to have significant implications for educational assessment and curriculum development. A validated questionnaire will enable educators to better understand students' reasoning strengths and weaknesses, informing instructional strategies that promote deeper analytical thinking. Additionally, this study will provide empirical evidence for integrating more structured reasoning assessments into teacher education curricula, ultimately fostering a generation of educators who can effectively cultivate critical thinking in their students. By bridging the research gap in non-test assessments of critical reasoning and applying a rigorous analytical framework, this study contributes to the advancement of educational practices in Indonesia and beyond.

Method

The type of research used is quantitative descriptive with a non-experiment/survey research design. The respondents in this study were 122 students in the 4th (fourth) semester of the undergraduate program of the Melawi College of Teacher Training and Education (STKIP), with 13 males and 109 females.

Data collection was conducted online using a questionnaire distributed via Google Forms. The questionnaire employed a four-point Likert scale for scoring responses. For positive statements, the response options were: strongly agree (4), agree (3), disagree (2), and strongly disagree (1). Conversely, for negative statements, the scoring was reversed: strongly disagree (4), disagree (3), agree (2), and strongly agree (1). This scoring approach ensures consistency in measuring the strength of agreement across all items (Allen & Seaman, 2007; Mawardi, 2019).

To calculate the total score, each respondent's responses were first converted into their respective numerical values based on the scoring scheme. The scores for all items were then summed to generate a total score for each respondent, representing their overall position on the measured construct. Additionally, scores were grouped according to item categories to facilitate subscale analysis where applicable. The dataset was reviewed for completeness and consistency before proceeding to the next step (Boone, 2016).

Once collected, the responses were compiled into a structured dataset where each row represented a respondent, and each column corresponded to an item. The assigned numerical scores were arranged in a tabular format and saved as a spreadsheet, ensuring compatibility with the Winsteps 3.75 software. The software requires data to be formatted appropriately, ensuring that each response is mapped correctly to the corresponding item and respondent. Prior to analysis, the dataset was examined for missing values or inconsistencies, and necessary adjustments were made to maintain data integrity (Mignogna et al., 2023).

After the initial processing, the data were analyzed using the Winsteps 3.75 application to assess the validity and reliability of the instrument. The analysis included evaluating item fit, unidimensionality, and person-item reliability based on the Rasch model. The responses were converted into logits, a logarithmic scale representing the probability of respondents endorsing an item relative to their ability level. This transformation allowed for an in-depth examination of item difficulty, respondent ability, and overall measurement accuracy (Balbuena, 2023; Linacre, 2012). The output of the analysis provided key indicators such as item difficulty indices, person reliability scores, and fit statistics, which were then interpreted to determine the quality and effectiveness of the measurement instrument (Madzlan et al., 2022). Testing the validity and reliability of instruments includes, namely:

Validity analysis is carried out through the unidimensional approach, meaning that the instrument measures only one dimension. The validity test based on item unidimensionality is determined through the raw variance value explained by measures. If the raw variance value explained is more than 20%, the minimum criterion is met, while a value above 40% is considered good, and more than 60% falls into the special category. In addition, to identify problematic and unsuitable question items, an eigenvalue analysis is conducted on the unexplained variance 1st contrast. An item is categorized as problematic if the eigenvalue is greater than 3, while the observed variance must be less than 15% to ensure the item is appropriate (Balbuena, 2023; Sumintono, 2016; Tennant & Küçükdeveci, 2023).

Item validity is tested through item fit analysis using several statistical indicators, namely Outfit Mean Square (MNSQ) with an acceptable range of 0.5 < MNSQ < 1.5, Outfit Z-Standard (ZSTD) with a range of -2 < ZSTD < +2, and Point Measure Correlation (Pt Measure Corr) with a value of 0.4 < Pt Measure Corr < 0.85 (Sumintono, 2016). An item is considered valid if it meets these three criteria. This test ensures that the items follow a consistent pattern in measuring the intended construct (Sumintono & Widhiarso, 2015).

Testing the difficulty analysis of items for approval by respondents using a logical value (measure). When the standard deviation value is combined with the average value of the logit, the difficulty level of the item can be classified into several levels, which consist of a very difficult level (more than + 1 SD), difficult level (0.0 logit to + 1 SD), easy level (0.0 logit to – 1 SD), and very easy level (less than -1 SD) (Natsir et al., 2022; Salsabila, 2023; Sumintono, 2016). Based on the standard deviation measure value, the difficulty level of the items approved by the respondents is grouped based on the following Table 1.

Table 1. Categories of Interpretation of Measure Values

Measure values	Interpretation of the difficulty of question items
Measure logit < -1.25	Very Easy
-1.25 ≤ measure logit ≤ 0	Easy
0 ≤ measure logit ≤ 1.25	Difficult
Measure logit > 1.25	Very Difficult

Instrument reliability testing is carried out by reviewing the person and item reliability values, as well as Cronbach's alpha coefficient. The interpretation of person and item reliability values is categorized as special if greater than 0.94, very good in the range of 0.91-0.94, good in 0.81-0.90, sufficient in 0.67-0.80, and weak if less than 0.67. Meanwhile, Cronbach's alpha reliability is categorized as poor if less than 0.5, bad in the range of 0.5-0.6, good in 0.8-0.90, very good in 0.91-0.94, and special if greater than 0.94. Furthermore, the separation of persons and items is also conducted by calculating the separation value (H), which is determined using the formula 1 (Salsabila, 2023; Sumintono, 2016)

$$H = \frac{(4 \times \text{separation}) + 1}{3} \tag{1}$$

Result and Discussion

Unidimensional Instrument Validity

Unidimensional analysis identifies several aspects that are measured through instruments by paying attention to the value of raw variance explained by measures (Ridzuan et al., 2020). The results of the analysis of the validity of the instrument through unidimensional can be seen in the following Table 2.

Table 2. Results of Unidimensional Analysis

Item information units	Eigenvalue	Observed
Raw variance explained by measures	12.9736	39.3%
Unexplned variance in 1st contrast	4.7998	14.6%

Based on Table 2, the raw variance explained by measure has a score of 39.3% so that the minimum criterion of validity of instrument through unidimensional can be said to be met (>20%). Unexplained variance in 1st contrast has an observed score below 15% (14.6%), but the eigenvalue score is more than 3, so it is necessary to conduct further analysis with the analysis of fit order items to determine whether an item can be maintained or must be replaced. Eigenvalue score helps determine whether the instrument is unidimensional. A score above 3 indicates that there are several constructions that can jeopardize the validity of the instrument, so it is necessary to review each item (Darmana et al., 2021).

Several studies Putra et al. (2020), Tarigan et al. (2022), and Wahyudi et al. (2020) point out that although the minimum criteria for unidimensionality have been met based on raw variance explained by measure, further interpretation is needed to ensure that the instrument actually measures one major construct. Further evaluation steps are essential to ensure the overall validity and reliability of the instrument.

Item Validity through Item Fit

The matching of items with the Rasch model was reviewed from the value of the outfit mean square (MNSQ), standard Z outfit (ZSTD), and the value of Pt measure correlation. Based on the analysis using the Winstep application, the fit item score was obtained as follows.

Table 3. Item Validity Test Results Based on Fit Items

Item	Measure	Outfit		Point measure correlation	Valid
		MNSQ	ZSTD		
1	-0.73	1.32	2.03	0.12	X
2	1.16	1.12	1.05	0.47	✓
3	-0.79	1.32	2.00	0.31	X
4	0.97	0.98	-0.11	0.52	✓
5	-0.91	0.84	-1.12	0.38	X
6	-1.24	1.23	1.56	0.23	X
7	2.11	0.94	-0.45	-0.06	X
8	-0.76	0.62	-2.93	0.24	X
9	1.18	0.86	-1.22	0.34	X
10	-0.88	0.75	-1.83	0.23	X
11	1.38	1.06	0.56	0.35	X
12	-1.45	0.51	-4.28	0.50	X
13	0.60	1.00	0.06	0.57	✓
14	-1.24	0.83	-1.20	0.24	X
15	0.77	1.11	0.91	0.49	✓
16	2.68	1.15	1.34	-0.25	X
17	-1.45	-0.81	-1.43	0.34	X
18	-1.09	1.28	1.83	0.35	X
19	-0.94	0.76	-1.71	0.35	X
20	0.62	1.33	2.46	0.55	X

Based on Table 3, there are 4 items (2, 4, 13, 15) that are valid while the other 16 items (1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20) are invalid. The number of invalid items was due to most of the items not meeting (red) the point measure correlation criteria (14 items), followed by 5 items that did not meet the ZSTD fit criteria, and 1 item that did not meet the MNSQ fit.

Point-measure correlation is analyzed to ensure that each item has a positive correlation with the respondent's ability, therefore if the value is not within the specified criteria, even if it is negative, then the item does not support the measurement of the skill in question. Items with MNSQ values outside the set range are considered not fit, as well as ZSTD values as a support in evaluating fit, values outside the specified range indicate possible response patterns that do not conform to the model. Items with values that do not meet all of these validity criteria need to be revised or deleted (Loh et al., 2022; Sumintono & Widhiarso, 2014).

The number of items that have not reached the value of the point measure correlation criteria is suspected to be due to the construction of these items does not lead to the skills to be measured or the statements on the items are too difficult for respondents to understand (Yunus et al., 2017). The number of invalid items also shows how accurate the analysis of question items is with the Rasch modeling approach, because when the question item meets 3 (three) criteria, namely the MNSQ Outfit value, the ZSTD Outfit value, and the Point Measure Correlation value, the question can only be considered valid (Erfan et al., 2020). Therefore, it is necessary to conduct a difficulty level analysis to narrow down the alleged causes of invalid items.

Difficulty of Items to Approve for Respondents

Analyzing the difficulty of the item is essential to ensure that the survey instrument is valid and effectively measures the construction in question, as indicated by the fit criteria in the study (Lailiyah et al., 2018). Based on Table 3, by looking at the measure value, each item is then grouped in different levels of difficulty related to respondents' approval. The difficulty level of each item for respondents to approve can be seen in Table 4 below.

Table 4. Difficulty of Items for Respondents to Agree on

Category	Item Number
Very difficult for respondents to agree	16, 7, 11
Difficult for respondents to agree	9, 2, 4, 15, 20, 13
Easy for respondents to agree	1, 8, 3, 5, 10, 19, 18, 6, 14
Very easy for respondents to agree	12, 17

Table 4 shows the diversity of item difficulty levels for respondents to agree on. This illustrates the variation in the difficulty level of each statement on the student's critical reasoning questionnaire. There are 3 items that are included in the very difficult category, 6 difficult items, 9 easy items, and 2 very easy items. The varying levels of difficulty in assessment items improve the accuracy and validity of measurements, allow for better differentiation among test takers and reduce the risk of differential item function across sub-populations (Wilson & Dulhunty, 2019).

The balance of the number of questions with varying levels of difficulty shows that the distribution of questions has been designed to cover different levels of test takers' abilities. However, if there are still invalid question items, it is likely that the problem lies in the formulation of the question item itself. Factors such as the clarity of the question construction, conformity with the measured indicators, and the potential for bias in the preparation of questions can be the main causes of inconsistencies in the measurement results (Jumadi et al., 2023; Laliyo et al., 2022). Table 4 show the number of items with difficult-very difficult and easy-very easy difficulty levels tends to be balanced, therefore, the suspicion related to the cause of the number of invalid items can be narrowed down to the construction of these items that have not yet led to the skill to be measured. Further analysis of each question item is needed to ensure that all questions used actually measure the skills in question accurately and consistently.

Separation Index and Instrument Reliability

Based on the results of the analysis using the Winstep application, the separation index and reliability of the instrument can be seen in the following Table 5.

Table 5. Separation Index and Instrument Reliability

	Mean logit (SD)	Separation	Reliability	Alpha cronbach	H
Person	0.41	0.89	0.44	0.55	1.52
Item	1.24	7.23	0.98		9.97

Rasch modeling can identify groups of respondents based on the respondent's separation index. The greater the item separation value, the better the quality of the instrument in terms of the overall respondents and item items, because it can identify the respondent group and the item group (Sumintono & Widhiarso, 2015). Based on Table 5, it is known that the item item separation value is 7.23, then the H = 9.97 value is rounded to 10, so there are ten groups of item items, and for the respondents a separation value of 0.89 with H = 1.52 rounded to 2, indicating that the respondent group can be distinguished into two groups based on the respondent's separation value. This is in line with the results of the analysis of the difficulty of the questions in Table 4, where the number of respondents who are difficult-very difficult to understand and answer items and the number of respondents who are easy-very easy to understand and answer items tend to be balanced.

The greater the price of person separation, the better the test is compiled because the items in it are able to reach responden with high to low levels of ability. Item separation indicates how much sample subjected to measurement is spread along a linear interval scale. The higher the item separation value, the better the measurement is made (Sumintono, 2016).

When viewed from Table 5, item reliability of 0.98 is in the Special category, while person reliability of 0.44 is in the weak category. This shows that the consistency of the answers from the respondents is weak, but the quality of the items in the instrument is good. This is also related to the alpha cronbach value (interaction between person and item) of 0.55 with the bad category. Cronbach's alpha value with a bad criterion shows a lack of interaction between the person and the item as a whole (Purwana et al., 2020).

Conclusion

The results showed that the students' critical reasoning skills through Rasch modeling analysis in general had not achieved a good quality of item validity (only 4 items were categorized as valid out of a total of 20 items). The reliability analysis of the question items obtained an Alpha Cronbach value for a reliability value of 0.55 with the bad category and from Item Reliability obtained a reliability value of 0.98 with the special category, while the person reliability value was 0.44 with the weak category. The results of the difficulty level analysis showed that there were four categories of item difficulty, namely 2 items which included very easy, 9 easy items, 6 difficult items, and 3 very difficult items. The item differentiation power of Rasch modeling obtained ten groups of question items based on the question item separation index H = 9.97. The quality of the instrument that has not met expectations is suspected to be due to the construction of the items that have not led to the skills to be measured, therefore it is necessary to revise or replace these items so that the quality of the instruments becomes better.

Acknowledgments

The author(s) would like to express sincere gratitude to all individuals who contributed to this research, including colleagues, faculty members, and participants. Their valuable insights and support were instrumental in completing this research.

Author Contributions

Research design, Y., R.M.D.K.; Conducting article, Y.; Writing article, Y., R.M.D.K., J.S., S.M., N.A.N.; Collecting and analyzing data, Y., J.S., S.M., N.A.N. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

Allen, E., & Seaman, C. (2007). Likert scales and data analysis. *Quality Progress*, *40*(7), 64–65. Retrieved from https://www.bayviewanalytics.com/reports/asq/likert-scales-and-data-analyses.pdf

Alsoudi, S., & Shindi, Y. A. (2023). Rasch rating scale modelling of the Arabic version of the critical thinking disposition scale. *Journal of Educational and Psychological Studies [JEPS]*, *17*(4), 359–369. https://doi.org/10.53543/jeps.vol17iss4pp359-369

Anggraini, N., Nabillah, C., Lonan, H. D., & Sain, H. (2023). Application of the Rasch model to test tools in the analysis of survey design. *Parameter: Journal of Statistics*, *3*(1), 19–27. https://doi.org/10.22487/27765660.2023.v3.i1.16412

Asmarita, D., & Prastowo, A. (2023). The learning process of Islamic education to foster students' critical reasoning at Madrasah Ibtidaiyah Teluk Panji, North Sumatra. *DAYAH: Journal of Islamic Education*, *6*(2), 233. https://doi.org/10.22373/jie.v6i2.18508

Azmi, I., & Prasetya, D. S. B. (2025). Profil berpikir kritis siswa SMP pada mata pelajaran IPA. *Journal of Classroom Action Research*, *7*(1), 163–175. https://doi.org/10.21449/ijate.1210479

Balbuena, S. (2023). Application of the Rasch model in streamlining an instrument measuring depression among college students. *International Journal of Assessment Tools in Education*, *10*(2), 257–278. https://doi.org/10.21449/ijate.1210479

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of Rasch model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 329–345. https://doi.org/10.24815/jpsi.v9i3.19618

Denis, B. (2022). Critical thinking technology as a tool for innovative activity in mathematics lessons. *Baltic Journal of Legal and Social Sciences*, *2*(2), 23–30. https://doi.org/10.30525/2592-8813-2022-2-4

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19. https://doi.org/10.23887/ijerr.v3i1.24080

Gustianingrum, R. A., Murni, A., & Maimunah. (2023). Analisis kemampuan berpikir kritis peserta didik dalam menunjang penguatan profil pelajar Pancasila. *PRISMA, Prosiding Seminar Nasional Matematika*, *6*(2), 465–471. Retrieved from https://journal.unnes.ac.id/sju/prisma/article/view/66908

Jumadi, J., Sukarelawan, M. I., & Kuswanto, H. (2023). An investigation of item bias in the four-tier diagnostic test using Rasch model. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(2), 622–629. https://doi.org/10.11591/ijere.v12i2.22845

Kaur, M. K. (2018). Critical thinking for global peace: A key for sustainable development. *International Journal of Trend in Scientific Research and Development*, *2*(3), 1498–1501. https://doi.org/10.31142/ijtsrd11439

Lailiyah, L., Supriyati, Y., & Komarudin, K. (2018). Analysis of measures items in development of instruments self-assessment (Rasch modeling application). *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *4*(1), 1–9. https://doi.org/10.21009/jisae.v4i1.7819

Laliyo, L. A. R., Kilo, A. L., Paputungan, M., Kunusa, W. R., Dama, L., & Panigoro, C. (2022). Rasch modelling to evaluate reasoning difficulties, changes of responses, and item misconception pattern of hydrolysis. *Journal of Baltic Science Education*, *21*(5), 817–835. https://doi.org/10.33225/jbse/22.21.817

Linacre, J. M. (2012). *Winsteps Rasch tutorial 2*. Retrieved from https://www.winsteps.com/a/winsteps-tutorial-2.pdf

Loh, S. L., Jamil, N. A., & Ng, K. T. (2022). Establishing the validity and reliability of a program evaluation questionnaire using Rasch measurement model. *DINAMIKA: Jurnal Ilmiah Pendidikan Dasar*, *14*(2), 150–157. https://doi.org/10.30595/dinamika.v14i2.14786

Madzlan, N. A., Rus, R. C., Mustafa, M. C., & Yassin, S. M. (2022). Validity and reliability of survey items in employer perspective construct on the quality of ECCE: Rasch measurement model analysis. *World Journal of English Language*, *12*(2), 288–293. https://doi.org/10.5430/wjel.v12n2p288

Mawardi. (2019). Rambu-rambu penyusunan skala sikap model Likert untuk mengukur sikap siswa. *Scholaria: Jurnal Pendidikan Dan Kebudayaan*, *9*(3), 292–304. https://doi.org/10.24246/j.js.2019.v9.i3.p292-304

Mignogna, G., Carey, C. E., Wedow, R., Baya, N., Cordioli, M., Pirastu, N., Bellocco, R., Malerbi, K. F., Nivard, M. G., Neale, B. M., Walters, R. K., & Ganna, A. (2023). Patterns of item nonresponse behaviour to survey questionnaires are systematic and associated with genetic loci. *Nature Human Behaviour*, *7*(8), 1371–1387. https://doi.org/10.1038/s41562-023-01632-7

Natsir, I., Mayasari, D., & Munfarikhatin, A. (2022). Rasch model: Quality of test instruments to measure students' ability in algebra. *SHS Web of Conferences*, *149*, 1–5. https://doi.org/10.1051/shsconf/202214901033

Packer, J., Croker, H., Goddings, A.-L., Boyland, E. J., Stansfield, C., Russell, S. J., & Viner, R. M. (2022). Advertising and young people's critical reasoning abilities: Systematic review and meta-analysis. *Pediatrics*, *150*(6). https://doi.org/10.1542/peds.2022-057780

Purwana, U., Rusdiana, D., & Liliawati, W. (2020). Pengujian kemampuan menginterpretasikan grafik kinematika calon guru fisika: The polytomous Rasch analysis. *ORBITA: Jurnal Kajian, Inovasi Dan Aplikasi Pendidikan Fisika*, *6*(2), 259–266. https://doi.org/10.31764/orbita.v6i2.3264

Putra, M. D. K., & Retnawati, H. (2020). Analisis Rasch terhadap instrumen general self-efficacy scale-12 versi bahasa Indonesia: Studi komparasi rating scale model (RSM) dan partial credit model (PCM). *Indonesian Journal of Educational Assesment*, *3*(1), 52–60. https://doi.org/10.26499/ijea.v3i1.55

Ramdani, A., Jufri, A. W., Jamaluddin, & Setiadi, D. (2020). Kemampuan berpikir kritis dan penguasaan konsep dasar IPA peserta didik. *Jurnal Penelitian Pendidikan IPA*, *6*(1), 119–124. https://doi.org/10.29303/jppipa.v6i1.388

Ridzuan, M. F., Lim, L. H., Fozee, F. A. A., & Nasser, S. N. A. M. (2020). Rasch analysis model: Reliability and validity of superitem test instrument. *International Journal of Academic Research in Progressive Education and Development*, *9*(4), 1–11. https://doi.org/10.6007/IJARPED/v9-i4/8166

Salsabila, F. (2023). Pengujian validitas dan reliabilitas instrumen manajemen diri remaja: Rasch model analysis. *Jurnal Bimbingan Dan Konseling Terapan*, *7*(1), 15–25. https://doi.org/10.30598/jbkt.v7i1.1741

Sari, T. N. I., & Rakhmawati, A. (2024). Analysis the quality of critical thinking and creativity questions in high school biology subjects with the Rasch model. *Research and Development in Education (RaDEn)*, *4*(1), 602–616. https://doi.org/10.22219/raden.v4i1.32758

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk penelitian ilmu-ilmu sosial*. Bandung: Tim Komunika Publishing House

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

Sumintono, B. (2016). *Aplikasi Pemodelan Rasch pada asesmen pendidikan: Implementasi penilaian formatif (assessment for learning)*. Retrieved from https://shorturl.at/uKbbw

Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis instrumen tes menggunakan Rasch model dan software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, *16*(2), 92–96. https://doi.org/10.15294/jipk.v16i2.30530

Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: Early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences*, *4*, 1–17. https://doi.org/10.3389/fresc.2023.1208670

Wahyudi, A., Mahyuddin, M. J., Irawan, A. W., Silondae, D. P., Lestari, M., Basco, F. H., & Kurniawan, S. J. (2020). Model Rasch: Analisis skala resiliensi Connor-Davidson versi Bahasa Indonesia. *Advice: Jurnal Bimbingan Dan Konseling*, *2*(1), 28–35. https://doi.org/10.32585/advice.v2i1.701

Wilson, M., & Dulhunty, M. (2019). Interpreting the relationship between item difficulty and DIF: Examples from educational testing. *Journal of Physics: Conference Series*, *1379*(1). https://doi.org/10.1088/1742-6596/1379/1/012026

Yokhebed, Sutarno, Masykuri, M., Prayitno, B. A., & Sudirman. (2024). Validity and reliability of online socio-scientific issue-based critical reasoning instrument. *Jurnal Pendidikan IPA Indonesia*, *13*(2), 229–238. https://doi.org/10.15294/2hnw4674

Yunus, F. A. N., Suki, N. A., Baser, J. A., Masran, S. H., Marian, M. F., & Rahim, M. B. (2017). Kesahan dan kebolehpercayaan instrumen kompetensi pengajar TVET terhadap pengajaran teknikal berdasarkan pendekatan model Rasch. *Online Journal for TVET Practitioners*, *2*(2). Retrieved from https://publisher.uthm.edu.my/ojs/index.php/oj-tp/article/view/4780