Jurnal Inovasi Pendidikan dan Sains
Vol.6 No.1, April 2025: 89-95 ISSN 2721-9119 (Online)

RASCH MODEL ANALYSIS OF ITEM QUALITY FOR MEASURING CONCEPTUAL UNDERSTANDING OF THE HUMAN BODY SYSTEM

Satwika Trianti Ngandoh[1], Riandi[1*], Adi Rahmat[2], Muslim[3]

[1]Department of Science Education, Faculty of Mathematic and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Department of Biology Education, Faculty of Mathematic and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3]Department of Physics, Faculty of Mathematic and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

*Email: rian@upi.edu

Accepted: February 28th 2025. Approved: April 16th 2025. Published: April 24th 2025

Abstract: This study aimed to evaluate the quality of test items in a science learning assessment instrument designed to measure students' conceptual understanding of the human body system using Rasch model analysis. A quantitative descriptive research design was employed, involving 100 junior high school students as participants. These students completed a conceptual understanding test consisting of 35 multiple-choice items. The Rasch model was applied using Winsteps software to assess item fit, analyze the person-item distribution through the Wright Map, and evaluate the overall validity and reliability of the instrument. The analysis results showed that all test items met the Rasch model criteria, indicating their suitability for assessing students' conceptual understanding. The instrument provided an item reliability score of 0.85, reflecting strong internal consistency, while the item separation index of 2.38 confirmed its capability to distinguish items based on difficulty levels. The Wright Map further illustrated a proper alignment between item difficulty and student ability, ensuring fair measurement across varying levels of proficiency. These findings confirmed that the developed instrument was valid and reliable, making it a valuable tool for both evaluative and diagnostic purposes in science education.

Keywords: rasch analysis, conceptual understanding, reliability, validity, human body system

INTRODUCTION

Science is one of the important fields of study in the secondary education curriculum because it provides a foundation for understanding life and the underlying scientific principles. In science subjects, students learn to understand concepts, think critically, find problem solving and make the right decisions [1]. One of the important topics in science subjects is the human body system, including organ systems and their functions. These systems do not work in isolation; instead, they interact with each other in a complex way to ensure the body functions properly [2]. Therefore, a deep conceptual understanding of the human body system is important for students to be able to apply concepts with real-life contexts [3]. Accurately measuring students' conceptual understanding is a challenge in science learning, especially on this topic. Effective assessment should be able to identify the extent to which students understand concepts, not just memorize facts. Therefore, a valid and reliable assessment instrument is needed to evaluate students' conceptual understanding of the human body system [4].

Various studies show that there are still many assessment instruments used in schools that have not been able to measure students' conceptual understanding in a valid and reliable manner. The assessment method still focuses on testing the memorization aspect so that it does not reflect thinking skills at a higher level [5]. Many educational environments still rely on traditional assessments such as multiple-choice tests that often emphasize memorization and low-level cognitive skills [6]. This approach can lead to a superficial understanding of concepts, as students can remember information but have difficulty applying it in new contexts. So that the assessment process becomes less meaningful and does not provide a complete picture of the student's understanding of concepts. Therefore, it is important to have a strong assessment instrument that accurately measures students' conceptual understanding.

To overcome these problems, a better approach to question item analysis is needed to improve the quality of assessment instruments. One approach that can be used is the Rasch model, which allows for an in-depth analysis of the characteristics of the question items and the ability of students [7]. The Rasch model allows for estimation of the difficulty level of the question item and the student's ability on the same scale, providing a clear picture of how well the student understands the material and how challenging each question item is [8]. The Rasch model is a probabilistic model used in item response theory to analyze data from assessment instruments. Using the Rasch model, researchers can ensure that the assessment instruments have high validity and reliability [9]-[10]. In addition, this model can also be used to identify biases in question items, such as gender bias or educational background.

Several previous studies that applied the Rasch model to analyze the quality of the assessment instruments stated that the Rasch model was effective in identifying inappropriate and biased question items in the research instrument [11]. The Rasch model provides complete diagnostic information that helps in improving the measurement properties of scale. This model can identify non-compliant items and suggest modifications to improve the accuracy of the instrument [12]. The analysis of the Rasch model can explain the quality of critical thinking questions based on the level of difficulty and suitability and can categorize students' abilities and their suitability with the learning carried out [13]. For question items that are identified as inappropriate, the Rasch model provides information that helps researchers determine whether they need to be revised, replaced, or maintained [14]. Rasch's analysis model produces more stable and accurate measurements than the classical analysis approach [15].

Although the Rasch model has been widely used in various studies to analyze the quality of assessment instruments in education, its specific application to evaluate students' conceptual understanding of the material of the human body system is still relatively limited. Existing research generally focuses more on other topics such as ecosystems, genetics, or on measuring critical thinking skills and science literacy in general. In fact, the human body system is one of the core materials in the biology curriculum which has high complexity and requires a deep understanding from students [16]. Instruments designed to measure understanding of the topic of human body systems need to be carefully evaluated in order to truly reflect students' abilities objectively. The integration of Rasch analysis in the development and evaluation of instruments is specifically aimed at measuring students' conceptual understanding of human body systems. Therefore, this study applies the Rasch model as the main tool in analyzing the quality of question items, so that it can provide a valid and comprehensive empirical picture of the effectiveness of the instruments used in measuring student competence in this material.

The application of the Rasch model in analyzing the quality of question items on the topic of human body systems is expected to make an important contribution to the development of more valid and reliable assessment instruments in the context of biology learning. Through the Rasch approach, this research offers an objective quantitative perspective and creates a new way of evaluating the appropriateness and accuracy of question items to measure students' conceptual understanding. Rasch's analysis allows the identification of problem items that are not functioning optimally [17]. Problematic question items can be revised or even eliminated immediately, so that the overall quality of the instrument improves. The results of the assessment become more representative of the actual competencies of students, and support efforts to improve the quality of biology learning in a sustainable manner.

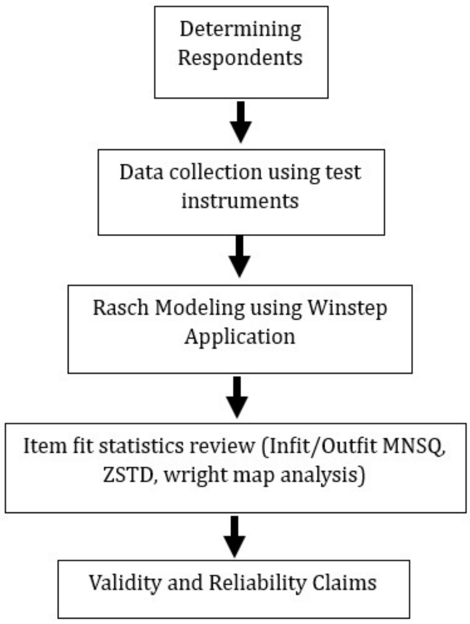
The main problem that is the focus of this study is how to evaluate the quality of question items in assessment instruments specifically designed to measure students' conceptual understanding of the material of the human body system, using the Rasch model analysis approach. This study aims to assess whether the assessment instrument can accurately measure student competence related to the understanding of the concept of human body system material. By applying the Rasch method, it is hoped that it can provide a strong empirical basis for the development and refinement of these assessment instruments so as to support a more objective, diagnostic, and oriented evaluation process towards improving the quality of learning.

RESEARCH METHODS

This study is quantitative descriptive research that focuses on the analysis of the Rasch model for the evaluation of the quality of students' concept understanding test questions about the material of the human body system. The stages in this study include testing test instruments to respondents, conducting Rasch modeling using the Winstep application, reviewing item match statistics and wright maps, and determining the validity and reliability of concept understanding test instruments. The research procedure is shown in Figure 1.

The respondents in this study were 100 junior high school students on the condition that they had received lessons about the human body system. Respondents were then asked to solve 35 concepts understanding problems. The test questions are multiple-choice questions and are made according to the indicators of concept understanding by Mayer. These indicators include: interpreting, exemplifying, classifying, summarizing, summarizing and comparing, and explaining.

After the data collection is complete, data analysis is carried out using the Winstep Rasch Model. Rasch analysis offers an objective and comprehensive approach to evaluating the quality of test instruments.



Figur 1. Research Procedures

Instrument tests were carried out to see the level of conformity of the item (fit-statistic) with the Rasch measurement model on the quality of the question item. To see the conformity of an item, Output Tables with Item Fit Order are used.

The measurement of the validity of the instrument was also based on the logit values on the criteria of outfit means-square (MNSQ), outfit z-standard (ZSTD), and point-measure correlation (PTMEASUR-AL CORR). The range of values in the three criteria is shown in Table 1.

Table 1. Criteria for accepting question items

Criteria	Value Range
MNSQ	0,50 < MNSQ < 1,50
ZSTD	-2,0 < ZSTD < +2,0
PTMEASURE-AL CORR	0,32 < PTMEASURE AL CORR < 0,85

Each of the values of the three criteria is interpreted to find out the extent to which the question items on the test instrument meet the criteria that have been set. There are four categories of interpretation of the suitability of question items based on the number of criteria met. If all three criteria (MNSQ, ZSTD, PTMEASURE-AL CORR) are met, the question item is considered highly appropriate, indicating that it meets high quality standards and is reliable to measure the ability or knowledge in question. If two of the three criteria are met, the question item is considered appropriate, although it may still need a slight revision to the question item. If only one of the three criteria is met, the question item is considered inappropriate, which means that the item requires further revision or refinement. Finally, if all criteria are not met, the question item is considered inappropriate and should be deleted or replaced with a better question item [18].

RESULT AND DISCUSSION

The results of the test analysis of the conceptual understanding test instrument using the Rasch analysis model have a question item reliability value of 0.85 which is shown in Table 2. This means that the reliability of the question items is in the very good category which shows that the instrument has high consistency in measuring students' understanding of concepts. Conceptual understanding instruments are able to distinguish precisely between students with different levels of conceptual understanding in the material of the human body system. This shows that students with high abilities tend to answer questions correctly, while students with low abilities tend to answer incorrectly, according to the Rasch model's prediction. The high reliability value indicates that the test question items provide stable and reliable results for repeated measurements. Reliability refers to the consistency of the test in measuring what it is intended to measure, free from measurement errors [19]. Varying scores are more due to differences in students' abilities than measurement error factors. Test instruments that have high reliability and are able to accurately differentiate student ability levels reduce the possibility of measurement errors as the main cause of score variation [20].

Table 2. Summary Statistics of Rasch Model

Measure		MNSQ		Separation	S.E. Item Mean	Reliabilitas Item
Mean	S.D	Infit	Outfit			
0,00	0,57	1,00	1,00	2,38	0,22	0,85

Based on the results of the analysis presented in Table 2, it is known that the reliability value of the item reaches 0.85. According to Bond [21], the reliability value above 0.80 is relatively high, which indicates that the instrument has excellent internal consistency. This means that most of the variability in a student's score is due to a noticeable difference in their ability level, not due to measurement errors or noise in the data. This is important because high reliability strengthens the validity of decision-making based on test results. For example, Saputra et al. [22] showed that a reliable scientific literacy test could effectively reflect real differences in student ability, reinforcing the importance of reliability in fair educational assessment.

In addition to reliability, an Item Separation value of 2.38 indicates that the instrument is able to distinguish items based on their difficulty level well. A separation score above 2.0 indicates that items are evenly distributed from easy to difficult, thus being able to accurately measure students with different skill levels. Thus, the difficulty level of the questions in this test is distributed quite widely and is able to accommodate the spectrum of students' abilities effectively [23]. The standard error (S.E) on the mean item was 0.22, relatively low, reinforcing the finding that the estimation of the item parameters was carried out with adequate precision. Meanwhile, the Mean Square (MNSQ) value for Infit and Outfit is both at 1.00, which is the ideal value in Rasch's analysis. This shows that the student's response to the question item is according to the model's expectations, without any systematic deviant answer patterns [24].

The test instrument used in this study has a strong performance in measuring the construct of concept understanding. The high reliability value, supported by a balanced distribution of items (adequate separation) and a low error standard, indicates that the instrument can be used reliably and fairly in the context of biology learning, both for evaluative and diagnostic purposes [25]. With the absence of an indicator of the item that deviating from the model, as well as the stability of the data in terms of measurement consistency, it can be concluded that this instrument is suitable for widespread use to measure students' conceptual understanding of the material of human body systems.

In addition, the high reliability value suggests that test results tend to be stable even if they are used at different times or applied to other groups of students who come from similar populations. In biology learning, this condition is important because it shows that the instrument can be relied upon to be used repeatedly, both as an evaluation and diagnostic tool, without compromising the accuracy in measuring the understanding of the targeted concept.

For the validity of the concept comprehension test instrument, it is seen from the values of MNSQ, Z-STD, and Point Measure Correlation shown in Table 3. Based on the results of the analysis, it can be seen that most of the question items meet two of the three criteria set so that they fall into the appropriate category and can be used. Overall, the MNSQ outfit value is in the range of 0.91 to 1.07 which indicates that the question items correspond to the Rasch model. There are no question items that are overfit (score too low) or underfit (score too high). The overall ZSTD value of the question item is in the ideal range, which is -2.0 to +2.0. Although there are some question items that have ZSTD values close to the lower and upper limits, they are still within the tolerance limit, so the question items are considered appropriate and can be used. The PT Measure-Al Corr score shows the correlation value between the score of the question item and the total score which ranges from 0.05 to 0.36. Question items with high correlation values mean that the items are very appropriate and able to distinguish students' abilities well. Question items with lower correlation values are still considered appropriate even though they need a slight revision to improve their quality [26].

Table 3. Results of Rasch Test Instrument Analysis

Question Item	**Outfit**	**PT Measure-Al Corr**	**Decision**	**Question Item**
	MNSQ	**ZSTD**		
1	1,03	0,68	0,14	Accepted
2	1,01	0,16	0,14	Accepted
3	1,00	0,02	0,19	Accepted
4	0,98	-0,10	0,18	Accepted
5	1,01	0,11	0,13	Accepted
6	0,95	-0,74	0,29	Accepted
7	0,96	-0,80	0,29	Accepted
8	0,97	-0,15	0,19	Accepted
9	1,00	0,09	0,19	Accepted
10	0,99	-0,06	0,19	Accepted
11	1,01	0,16	0,18	Accepted
12	0,98	-0,38	0,23	Accepted
13	0,95	-0,22	0,29	Accepted
14	1,00	0,07	0,19	Accepted
15	0,93	-1,23	0,35	Accepted
16	1,07	1,00	0,08	Accepted
17	0,98	-0,18	0,22	Accepted
18	0,91	-1,00	0,36	Accepted
19	1,05	0,51	0,07	Accepted
20	1,01	0,15	0,19	Accepted
21	1,02	0,15	0,15	Accepted
22	0,95	-0,29	0,24	Accepted
23	1,02	0,21	0,19	Accepted
24	1,04	0,47	0,19	Accepted
25	0,99	-0,11	0,19	Accepted
26	0,94	-0,63	0,28	Accepted
27	1,07	1,10	0,05	Accepted
28	1,07	0,84	0,08	Accepted
29	1,04	0,55	0,09	Accepted
30	0,96	-0,35	0,27	Accepted
31	1,06	0,95	0,09	Accepted
32	0,95	-0,47	0,27	Accepted
33	1,04	0,51	0,09	Accepted
34	0,98	-0,19	0,22	Accepted
35	1,04	0,37	0,13	Accepted
36	1,00	0,07	0,20	Accepted

Validity is an important thing in evaluating assessment instruments, because it shows the extent to which a test is able to measure the construct of the instrument that was developed appropriately. The Rasch approach provides a strong analytical framework in assessing construct validity, through examining the level of agreement between student responses and the predictions of the measurement model used. This finding indicates that all test items are acceptable and in accordance with the assumptions of the Rasch model. Each test item contributes meaningfully to measuring students' conceptual understanding of the human body system and does not cause measurement error or bias.

Based on the results of the analysis, no misfit or deviation items were found, which means that the instrument is valid to measure the competence of students' understanding of concepts towards the material of the human body system.

In addition to reliability and validity, the Wright map is important to explain the relationship between the difficulty level of the question and the ability of students to understand the concept of the material of the human body system. So that overall, the quality of the concept understanding instruments developed can be ensured to be accurate and consistent. The use of the Wright Map in the development and evaluation of concept comprehension test instruments not only improves the accuracy and consistency of measurements, but also ensures that the instrument is in accordance with the characteristics and needs of learners. This is in line with efforts to improve the quality of learning and assessment in science education, especially in the material of the human body system. Wright's map is shown in figure 2.

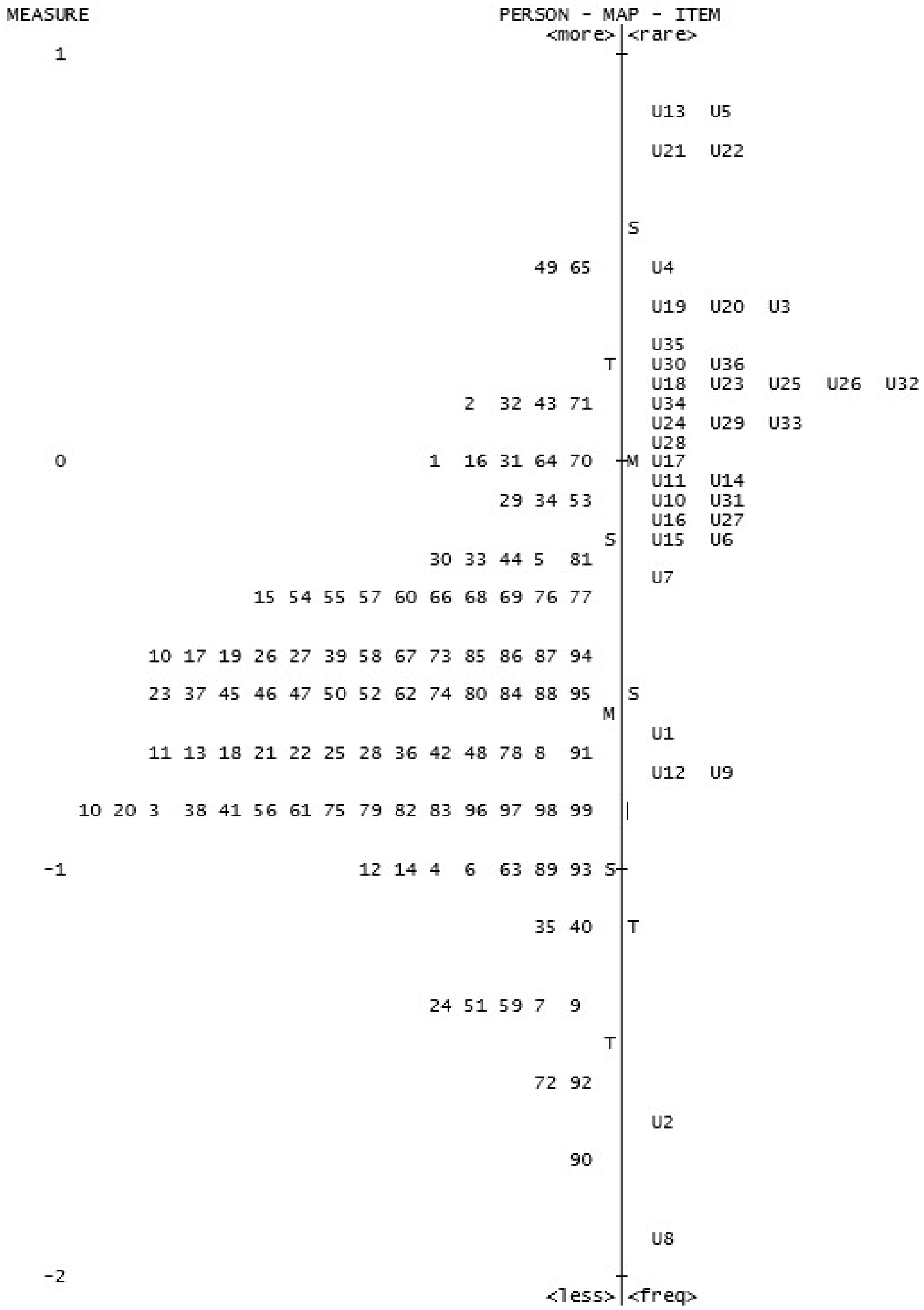


Figure 2. Item person map (Wright map) for conceptual understanding

From the Wright map, it is explained that the distribution of students' abilities is relatively even and forms a normal curve that is quite ideal. Students have an adequate variety of abilities, so that the instrument can detect differences in students' levels of understanding. Most students are in the middle of the scale, but there are also students with very high and very low abilities, which is characterized by the logical position at the top and bottom of the map.

On the other hand, the distribution of question items covers a proportional range of difficulty. There are question items located in high logit (U5, U13), indicating that the question is in the difficult category, because it can only be answered by students with high ability. In contrast, some items such as U7, U1, and U9 are at the bottom of the scale, which indicates that the questions are more accessible to students with low abilities. This distribution is important because it allows the instrument to detect the entire spectrum of capabilities, not just at one specific level [27]. Recent studies have highlighted the importance of analyzing item difficulty to determine test quality and provide feedback for learning. For example, Dewi and Putrayasa (2020) emphasized that analyzing item difficulty is crucial for increasing test validity and reliability, adjusting to students' abilities, and improving test quality for future assessments [28].

Overall, the results from the Wright Map support previous findings that this instrument has good measurement quality. An even distribution of items from easy to difficult allows measurements to be made fairly against students of different skill levels. The absence of a stack of question items in one particular area also indicates that there is no "over-represented" ability level, which can lead to bias in measurement.

The balance between the difficulty of the question and the ability of the students strengthens the argument that the conceptual understanding instruments that have been developed are constructively and operationally appropriate for evaluative and diagnostic purposes**.** Teachers can use this instrument not only to assess overall comprehension, but also to map students at a particular level of mastery of a particular concept. Thus, the Wright Map provides visual and statistical confirmation that the instrument has met the principles of validity and measurability in the context of Rasch**,** and can be used as a credible measuring tool in biology learning at the secondary school level.

CONCLUSION

Based on the results of the research and analysis that has been conducted, it can be concluded that the assessment instrument developed to measure students' conceptual understanding of the human body system has met the characteristics as a valid and reliable measuring tool. The application of the Rasch model allows for comprehensive instrument testing, both from the aspect of item function, conformity with the measurement model, to mapping students' abilities to the level of difficulty of the questions. This research contributes to the development of assessment instruments based on modern measurement theory, which emphasizes accuracy, fairness, and measurability. In the future, similar models can be adapted and expanded to other materials or levels of education to support more meaningful assessments and are oriented towards strengthening students' conceptual understanding.

ACKNOWLEDGMENTS

We would like to thank the Ministry of Education and Culture of the Republic of Indonesia through DRTPM (Direktorat Riset, Teknologi dan Pengabdian Masyarakat), Universitas Pendidikan Indonesia. BPPT (Balai Pembiayaan Pendidikan Tinggi) Ministry of Education and Culture and LPDP (Lembaga Pengelola Dana Pendidikan) Ministry of Finance of the Republic of Indonesia.

REFERENCES

[1] E. Twizeyimana, T. Shyiramunda, B. Dufitumukiza, and G. Niyitegeka, "Teaching and learning science as inquiry: an outlook of teachers in science education," SN Soc. Sci., vol. 4, no. 2, 2024, doi: 10.1007/s43545-024-00846-4.

[2] R. Tacutu, D. Toren, E. Ursu, G. Bunu, and T. B. Mracica, "Healthy Biological Systems," Springer, Cham, 2020, pp. 53–78. doi: 10.1007/978-3-030-52663-4_5.

[3] M. A. Mustofa, "Model-model Pembelajaran Sains di MTs Negeri Semarang," JISIP (Jurnal Ilmu Sosial dan Pendidikan), vol. 6, no. 2, Mar. 2022, doi:10.58258/jisip.v6i2.3179.

[4] V. Devaki, "Exploring the Impact of Innovative Assessment Methods on Learning Outcomes," Advances in educational technologies and instructional design book series, pp. 343–374, Aug. 2024, doi: 10.4018/979-8-3693-3645-8.ch015.

[5] [D. P. Collins, D. Rasco, and V. A. Benassi, "Test-Enhanced Learning: Does Deeper Processing on Quizzes Benefit Exam Performance?," Teach. Psychol., vol. 45, no. 3, pp. 235–238, 2018, doi: https://doi.org/10.1177/0098628318779262.

[6] S. Ghosh, B. Brooks, D. Ranmuthugala, and M. Bowles, "Authentic Versus Traditional Assessment: An Empirical Study Investigating the Difference in Seafarer Students' Academic Achievement," J. Navi, vol. 73, no. 4, pp. 797–812, 2020. doi:10.1017/S0373463319000894.

[7] X. Liu and W. J. Boone, "Introduction to Advanced in Applications of Rasch Measurement in Science Education," Contemp. Trends Issues Sci. Educ., vol. 57, pp. 1–18, 2023, doi: https://doi.org/10.1007/978-3-031-28776-3_1.

[8] H. Nitta and T. Aiba, "An alternative learning gain based on the rasch model," Phys. Educ., vol. 1, no. 1, pp. 1–7, 2019, doi: https://doi.org/10.1142/S2661339519500057.

[9] E. Avinç and F. Doğan, "Digital literacy scale: Validity and reliability study with the rasch model," Educ. Inf. Technol., vol. 29, no. 17, pp. 22895–22941, 2024, doi: 10.1007/s10639-024- 12662-7.

[10] A. Winarti, Almubarak, and S. Annurc, "How does the rasch model justify multiple choice question items as a measure of student understanding of acid-base material at the sub-microscopic level?," Int. J. Innov. Creat. Chang., vol. 7, no. 11, pp. 344–360, 2019.

[11] Y. Ö. Özkan and M. A. Güvendir, "Differential item functioning analysis of a high stake test in terms of statistical regions of Turkey," J. Pedagog. Res., vol. 5, no. 3, pp. 122–134, 2021, doi: 10.33902/JPR.2021371303.

[12] M. Farozin, B. Astuti, S. Sanyata, D. Titi, E. Nurmalasari, and N. Sutanti, "Rasch model for the need assessment instrument of academic guidance and counseling program in junior high school," J.

Penelit. dan Eval. Pendidik., vol. 26, no. 2, pp. 177–185, 2022, doi: 10.21831/pep.v26i2.51816.

[13] G. Hamdu, F. N. Fuadi, A. Yulianto, and Y. S. Akhirani, "Items Quality Analysis Using Rasch Model To Measure Elementary School Students' Critical Thinking Skill On Stem Learning," JPI (Jurnal Pendidik. Indones., vol. 9, no. 1, p. 61, 2020, doi: 10.23887/jpi-undiksha.v9i1.20884.

[14] Marantika Lia Kristyasari and Sri Yamtinah, "Validation of assessment instruments for integrated science learning on the ability of student using rasch model," Edusains, vol. 14, no. 1, pp. 24–33, 2022. http://doi.org/10.15408/es.v13i2.22468.

[15] R. M. Lia, A. Rusilowati, and W. Isnaeni, "NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model," Res. Eval. Educ., vol. 6, no. 1, pp. 41–50, 2020, doi: 10.21831/reid.v6i1.30112.

[16] S. Huerta, "Methodological Strategies for Teaching Human Anatomy: A Systematic Review," Int. J. Morphol., vol. 42, no. 5, pp. 1355–1360, 2024. http://dx.doi.org/10.4067/S0717-95022024000501355.

[17] Z. Lu, J. I. Vincent, and J. C. MacDermid, "Evaluation of the Structural Validity of the Work Instability Scale Using the Rasch Model," Arch. Rehabil. Res. Clin. Transl., vol. 3, no. 1, p. 100-103, 2021, doi: 10.1016/j.arrct.2021.100103.

[18] B. Sumintono, Aplikasi pemodelan rasch pada assessment pendidikan. Penerbit Trim Komunikata, 2015.

[19] H. L. Yang, K. R. Chou, S. C. Lee, P. H. Lin, and H. Y. Chiang, "Test–Retest Reliability and Random Measurement Error of the Multifactorial Memory Questionnaire in Older Adults With Subjective Memory Complaints," Gerontol. Geriatr. Med., vol. 9, no. 510, pp. 1–7, 2023, doi: 10.1177/23337214231171981.

[20] K. A. Zainina, M. Z. Mufiqoh, N. Aprilia, and B. Isnaeni, "Rasch Model : Analysis of Biology Question Item in the Indonesia Independent Curriculum," J. Penelit. Pendidik. IPA, vol. 10, no. 12, pp. 10990–10998, 2025, doi: 10.29303/jppipa.v10i12.7661.

[21] Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences. Routledge

[22] W. T. Saputra, N. Y. Rustaman, and L. Rusyati, "Evaluating Validity and Reliability of Scientific Literacy Among Indonesian Secondary Students in PISA 2015 Context : Rasch Model Analysis," J. Res. Educ. Res. Eval., vol. 13, no. 2, pp. 169–179, 2024.

[23] N. Nordin, M. Ghazali, N. F. J. Radzali, N. H. Hashim, and S. L. S. Chear, "Validity and Reliability of Pre-Service Teacher Competency Level in Teaching Early Mathematics Instrument Using the Rasch Model," Cakrawala Dini J. Pendidik. Anak Usia Dini, vol. 14, no. 2, pp. 139–148, 2023, [Online]. Available: http://dx.doi.org/10.17509/cd.v14i2.63683.

[24] W. A. Prasetya and A. T. Pratama, "Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school," J. Penelit. dan Eval. Pendidik., vol. 27, no. 1, pp. 76–91, 2023, doi: 10.21831/pep.v27i1.58873.

[25] A. Jowinis and N. M. Siew, "Validity and Reliability of Multiple-Choice Biology Test: a Rasch Analysis Perspective," Int. J. Mod. Educ., vol. 6, no. 21, pp. 131–139, 2024, doi: 10.35631/ijmoe.621010.

[26] N. R. Akbarini and A. Anggrawal, "Validity and Reliability of Test Questions To Measure the Information Literacy Skills of Prospective Teacher Students," J. Eduscience, vol. 11, no. 1, pp. 163–177, 2024, doi: 10.36987/jes.v11i1.5640.

[27] A. Yulianto and A. Widodo, "Disclosure of Difficulty Distribution of HOTS-Based Test Questions through Rasch Modeling," Indones. J. Prim. Educ., vol. 4, no. 2, pp. 197–203, 2020, doi: 10.17509/ijpe.v4i2.29318.

[28] A. H. Abad Badruzaman. Hosaini2, "An Application of Difficulty Level Analysis of Question Items in Language Learning Evaluation," Bull. Sci. Educ., vol. 1, no. 1, pp. 60–67, 2021.