JISE 13 (2) (2024): 55-63

Journal of Innovative Science Education

https://journal.unnes.ac.id/journals/jise/

Evaluation of STEM Literacy Test Instrument on Elements, Compounds and Mixtures Using Multidimensional Rasch Analysis

Ikwan Wahyudi ✉, Ani Rusilowati, Sunyoto Eko Nugroho, Wiyanto, Arif Widiyatmoko

Science Education Master's Program, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Semarang, 50299

Article Info

Article History :
December 2023
Accepted
February 2024
Published
August 2024

Keywords:
multidimensional model; Rasch analysis; STEM literacy

Abstract

STEM literacy is crucial in the 21st century, encompassing science, technology, engineering, and mathematics, allowing individuals to understand and apply conceptual knowledge to real-life problems. Evaluation of STEM literacy instruments is essential for improving the quality of STEM learning. This study aims to evaluate the STEM literacy test instrument using multidimensional Rasch analysis. This study employed a quantitative approach, including a description to validate the research instrument. The instrument consists of 22 items measuring science, technology, engineering, and mathematics literacy, focusing on elements, compounds, and mixtures. Participants included 259 eighth-grade students randomly selected from public junior high schools in Balongpanggang and Benjeng, Gresik, East Java. Data analysis employed a Rasch model with ConQuest programs. Results indicated that the multidimensional model had better-fit parameters than the unidimensional model and demonstrated that the item's validity and instrument's reliability meet requirements. The interaction between students' abilities and items revealed that most students have moderate to low STEM literacy abilities. Additionally, the analysis found no differential item functioning based on gender differences. This research indicates that multidimensional Rasch analysis helps evaluate STEM literacy test instruments so that valid and reliable instruments can provide accurate measurements in improving students' STEM literacy skills.

✉ correspondence:
Gedung D5 Lantai 1 Kampus, Jl. Sekaran Raya, Sekaran,
Gunungpati, Semarang City, Central Java 50229, Indonesia
E-mail: ikwanwahyudi@students.unnes.ac.id

p-ISSN 2252-6412
e-ISSN 2502-4523

INTRODUCTION

STEM literacy plays a crucial role in 21st-century learning. STEM literacy is the ability to understand and apply conceptual knowledge to solve problems in science, technology, engineering, and mathematics in everyday life (Wannapiroon et al., 2021). STEM literacy is not limited to disciplinary knowledge; it also shapes the dispositions and skills needed to face a dynamic and complex future (Falloon et al., 2020; Setiawan et al., 2023). In learning, STEM literacy contributes to developing critical thinking and fosters a growth mindset so students succeed in various fields (Wu et al., 2024). Therefore, it is essential to encourage and improve the quality of STEM learning, including in the learning evaluation process.

Evaluation plays a crucial role in enhancing STEM literacy. Practical evaluation assists teachers in identifying the strengths and weaknesses of STEM learning, thus enabling continuous adjustments and improvements (Li et al., 2020). Evaluation can provide valuable feedback to students, including their STEM literacy skills, which can help them correct mistakes, understand STEM concepts more accurately, and apply them in different contexts. In addition, the presence of evaluation can also ensure equal access to learning experiences and create an inclusive learning environment (Jackson et al., 2021). Therefore, there is a need for evaluation instruments to measure valid and reliable STEM literacy and improve STEM learning quality.

STEM literacy involves three domains, namely context, practice, and content (Ardianto et al., 2019). The measurement of STEM literacy in the content domain is related to cognitive abilities in STEM disciplines. STEM literacy measurement can focus on the conceptualization of knowledge acquisition and the application of STEM issues where the dimensions of multidisciplinary STEM literacy become an integrated part of a holistic understanding (Huang et al., 2022; Tang & Williams, 2019). Fallon et al. (2020) reviewed that the existing STEM literacy test instruments can measure conceptual understanding and problem-solving skills. Although Chamrat et al. (2019) recommend that STEM literacy tests focus on STEM integration, several STEM literacy test instruments in Indonesia are still being developed based on the dimensions of multidisciplinary STEM literacy (Nurlaely et al., 2017; Suwana et al., 2023).

A good instrument for measuring STEM literacy must meet the criteria of validity and reliability. Validity refers to the extent to which an instrument can accurately measure what it intends to measure, while reliability refers to the consistency of the measurement (Ghazali, 2016). In the context of STEM literacy, ensuring the validity and reliability of the testing instrument is crucial to accurately and consistently measure STEM skills and knowledge. The test instrument must also be free from bias and be effectively applicable in various contexts and populations. Additionally, Arikan et al. (2020) revealed that the test instrument should measure various skills in the STEM field, such as problem-solving, critical thinking, and reasoning. Therefore, evaluating test instruments, including STEM literacy test instruments, is essential in education and research.

Rasch analysis is a statistical approach used to evaluate STEM literacy test instruments. Rasch analysis operates within the Item Response Theory (IRT) framework, known for its robust approach to measuring latent traits. This allows for a more precise and objective measurement of students' abilities compared to classical test theory (CTT) (Bashooir & Supahar, 2018). Rasch analysis applies statistical methods to predict the mathematical relationship between items and the measured latent trait. This analysis involves logistic and probabilistic models, which suggest that responses to test items are influenced by the individual's and item's quality (Medvedev & Krageloh, 2022). Rasch analysis provides detailed diagnostic information about individual items and person abilities, enabling educators to identify specific strengths and weaknesses in student performance (Murti & Sunarti, 2021). This is crucial for targeted interventions in STEM education, especially evaluating the validity and reliability of test instruments.

Rasch analysis has been applied in various studies, including those related to STEM literacy, such as the development and validation of the Scientific Inquiry Literacy Instrument (Darman et al., 2024) and the STEM-Science Achievement Test (Jamaludin et al., 2021). However, the Rasch model is still limited to a unidimensional model and has not been applied to element, compound, and mixture subjects. The unidimensional model assumes that all items in the instrument measure a single latent trait (Combrinck, 2020; Farris et al., 2023). However, in reality, many constructs consist of several dimensions that are relevant to each other (Sigfrid, 2024). This indicates that the unidimensional model is unsuitable for more than one dimension variables. Therefore, a multidimensional Rasch model can be selected to overcome these limitations. Multidimensional Rasch models facilitate a better understanding of the dimensional structure of a variable. Additionally, they can enhance subtest reliability and serve as a confirmatory approach (Baghael, 2017).

In response to this challenge, this research offers a solution by evaluating STEM literacy test instruments on elements, compounds, and mixtures using multidimensional Rasch analysis. This research is essential because it will provide a more comprehensive and accurate understanding of the complexity of the dimensions of STEM literacy. This study aims to evaluate STEM literacy test instruments using multidimensional Rasch analysis. This research's benefits are not limited to increasing our understanding of STEM literacy instrument evaluation. However, they may also assist in developing more effective curricula and teaching strategies in the future, including in STEM fields.

METHODS

This study employed a quantitative approach with a cross-sectional research design, including a description for constructing and validating the research instrument (Bungin, 2021). The flow of research methods is presented in Figure 1. This study involved 259 eighth-grade students randomly selected from public junior high schools in Balongpanggang and Benjeng, Gresik, East Java. All participants received preliminary learning on elements, compounds, and mixtures at their respective schools. The STEM literacy test instrument was constructed around four dimensions: scientific literacy, technological literacy, engineering literacy, and mathematical literacy (Ardianto et al., 2019; Harpian et al., 2023; Suwana et al., 2023; Tang & William, 2019). This instrument comprised 22 multiple-choice questions and underwent rigorous validation by science experts and teachers before use. However, this paper focuses solely on analyzing the results of testing the STEM literacy measurement instrument using the Rasch model.

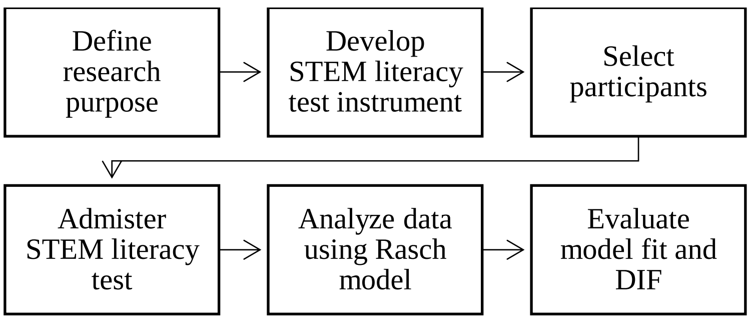

Figure 1. Steps evaluation of STEM literacy test instrument

The data analysis utilizes a Rasch model approach with the ConQuest program. The Rasch model is chosen based on comparing these three parameter values (final deviance, AIC, and BIC), with a smaller value indicating a more appropriate model. The validity of item fit is assessed by reviewing the MNSQ value. Items meet the validity criteria if the MNSQ value falls within the 95% confidence interval. The item-person map will explore the relationship between participant responses and item difficulty (Yudiana et al., 2023). The impact of existing DIF is gauged by its magnitude, determined through the difference between two estimated values based on gender. The corresponding chi-square test is also derived from DIF and standard estimate errors. An item is flagged for DIF if the chi-square is significant at 0.01 (Danuwijaya & Roebiyanto, 2020) and the absolute DIF value exceeds 0.43. DIF is categorized into three sizes: moderate to large (DIF $\geq$ 0.64 logits), slight to moderate (DIF $\geq$ 0.43 logits), and negligible (DIF < 0.43 logits) (Hairida et al., 2023).

RESULTS AND DISCUSSION

We need to compare unidimensional and multidimensional models to investigate the dimensional structure in the STEM literacy test instrument. The statistical procedure used to identify model fit is Rasch analysis, which involves comparing the final deviance, AIC, and BIC as model parameters. Table 1. shows the model comparison.

Table 1. Model comparison for investigating dimension structure in STEM Literacy

Model	The final deviance	AIC	BIC
Unidimensional	7226.32	7272.32	7354.13
Multidimensional	7131.11	7195.11	7308.93

Based on Table 1. shows that the multidimensional model exhibits the smallest parameter values for the final deviance, AIC, and BIC. Therefore, the multidimensional model statistically outperforms the unidimensional model in terms of fit. In line with Sigfrid (2024), the unidimensional model has limitations as it assumes that all items in the test measure a single dimension despite the reality of numerous complex and multidimensional constructs, such as STEM literacy. Multidimensional models accurately depict the STEM literacy complexity construct and provide more comprehensive insights into students' abilities. This multidimensional model of STEM literacy includes four dimensions, namely scientific literacy (SL), technological literacy (TL), engineering literacy (EL), and mathematical literacy (ML).

Table 2. Psychometric properties of the Rasch analysis with the multidimensional model

Dimension	Number of Items	Estimate	SE	Weighted Fit		Differences (M – L)
				MNSQ	CI	
SL	1	-0.05	0.10	1.01	(0.92. 1.08)	0.296
	2	-0.17	0.10	0.98	(0.92. 1.08)	-0.380
	5	0.01	0.10	1.05	(0.92. 1.08)	-0.088
	6	-0.88	0.10	1.06	(0.92. 1.08)	0.006
	8	0.37	0.10	0.95	(0.89. 1.11)	0.400
	9	-0.45	0.10	0.97	(0.93. 1.07)	0.082
	12	0.32	0.10	1.00	(0.90. 1.10)	0.046
	13	-0.10	0.10	0.97	(0.92. 1.08)	0.114
	15	0.95	0.29	1.02	(0.84. 1.16)	-0.418
TL	3	-0.05	0.10	1.04	(0.90. 1.10)	0.044
	4	-0.02	0.10	1.01	(0.90. 1.10)	-0.028
	16	0.16	0.10	1.04	(0.88. 1.12)	-0.132
	17	-0.09	0.17	0.97	(0.90. 1.10)	0.116
EL	7	-0.13	0.10	0.96	(0.92. 1.08)	-0.010
	14	-0.04	0.10	1.02	(0.92. 1.08)	-0.248
	18	-0.16	0.10	1.01	(0.92. 1.08)	0.404
	19	0.10	0.10	1.01	(0.91. 1.09)	-0.170
	20	0.23	0.19	1.00	(0.89. 1.11)	0.026
ML	10	0.15	0.10	1.00	(0.87. 1.13)	-0.260
	11	-0.25	0.10	1.02	(0.90. 1.10)	0.282
	21	0.19	0.10	0.99	(0.87. 1.13)	-0.086
	22	-0.09	0.17	0.97	(0.89. 1.11)	0.064

Note: SE = standard error, MNSQ = mean square, CI = confidence interval, M = male, F = female

STEM literacy entails comprehending and utilizing conceptual knowledge to address science, technology, engineering, and mathematics challenges. Assessing STEM literacy may concentrate on conceptualizations linked to STEM knowledge acquisition within core disciplinary practices, such as scientific inquiry, technological literacy, engineering design, and mathematical thinking (Huang et al., 2022; Kelley & Knowles, 2016). Hence, multidimensional Rasch models prove more adept at capturing variations in the measured constructs due to the complex nature of STEM disciplines. In line with Wannapiroon et al. (2021), STEM literacy includes the ability to identify, apply, and integrate concepts in STEM fields of science to understand complex problems and solve them innovatively. This multidimensional model is substantiated by lower parameter values (final deviance, AIC, and BIC), indicating decreased measurement estimate deviation. Table 2. shows the item fit validity, the item difficulty, and the DIF analysis.

Based on Table 2, the MNSQ value for each item falls within the 95% confidence interval, neither below the lower limit nor exceeding the upper limit. This MNSQ value indicates that all items in the STEM literacy test instrument meet the validity criteria for fit and demonstrate reliable constructs. In Rasch analysis, an item with an MNSQ value closer to 1.0 is considered fitting, indicating adequate construct homogeneity with other items (Kaharu et al., 2024). An item with an MNSQ value outside the 95% confidence interval is identified as a possible misfitting item, suggesting that further examination may be warranted. Additionally, the precision of all items meets satisfactory, with a standard error value ranging from 0.10 to 0.29 (less than 0.5). Rasch's analysis also provides insight into the difficulty level of STEM literacy. According to the estimated values, item difficulty is classified into two groups: easy (13 items) and complex (9 items).

The reliability of the STEM literacy test instrument is assessed by examining its reliability both at the latent and dimensional levels. Overall, the multidimensional model reliability yielded a Cronbach's alpha coefficient of 0.790 and a separation reliability of 0.880. The EAP/PV reliability for each dimension is as follows: 0.846 for SL, 0.796 for TL, 0.795 for EL, and 0.813 for ML. These findings suggest that the instrument meets acceptable reliability criteria, surpassing the threshold of 0.60 (Sarstedt et al., 2017). The estimated reliability coefficient indicates a high level of consistency in students' responses. Moreover, the separation reliability meets the required criteria, indicating that the STEM literacy test instrument is adequately sensitive to distinguish between students with varying ability levels.

The STEM literacy test instrument is deemed appropriate based on the results of the validity and reliability analysis. Valid items demonstrate their ability to measure STEM literacy in each dimension accurately. Reliable instruments indicate consistent measurement results that can be trusted, reflecting students' abilities (Bashooir & Supahar, 2018). Therefore, the strong validity and reliability of the STEM literacy test instrument ensure its accuracy, enabling measurement results to inform further decision-making regarding students' abilities. In line with Erfan et al. (2020), assessment instruments for learning outcomes should possess good validity and reliability, instilling confidence in the measured results throughout the learning evaluation process.

The DIF analysis was conducted to ascertain whether the STEM literacy test instrument exhibited gender bias at the item level. Based on Table 1, none of the items in the STEM literacy test exhibit DIF or gender bias issues. The absolute difference in estimated values does not exceed 0.5, as affirmed by the non-significant chi-square test results (21.42, df = 18), indicating an absence of DIF problems. A negative estimated value signifies that male students' logit value is lower than female students and vice versa. Moreover, the DIF analysis results validate that the STEM literacy test demonstrates gender-based invariance. DIF analysis in Rasch measurements serves a similar purpose to testing measurement invariance in different groups, akin to structural equation modeling analysis (Hairida et al., 2023).

Even though there was no potential for DIF in the STEM literacy test instrument, what is interesting from Table 1 shows that most of the items in the scientific literacy dimension have differences in logit values that are positive. The scientific literacy items are mostly easier for male students than female students, while they are pretty balanced on other dimensions. Overall, the difference in logit scores for male and female students is 0.008 (chi-square = 0.02, df = 1), which means that the STEM literacy abilities of male and female students are not significantly different. However, further examination to obtain substantive reasons as a basis for consideration for making decisions about item improvements needs to be carried out (Danuwijaya & Roebiyanto, 2020). Wu et al. (2016) suggested using statistical analysis to identify DIF items and examining item content to investigate theoretical explanations.

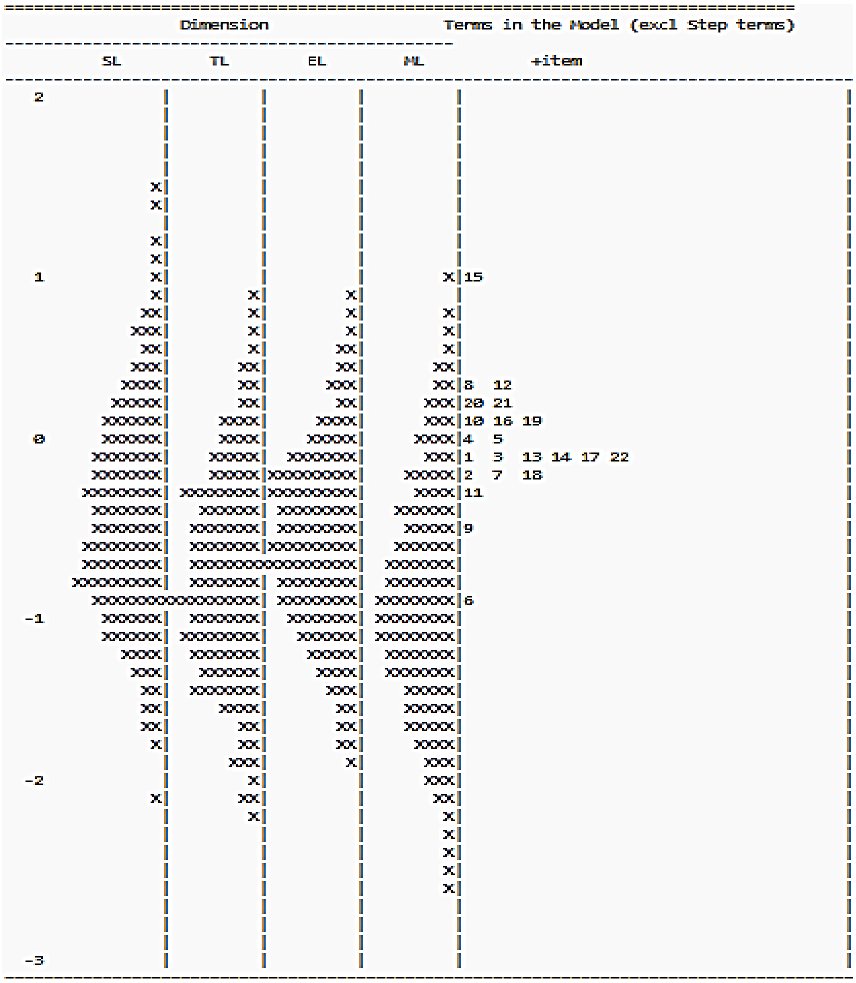

Figure 2. Item-person map with the multidimensional model

Figure 2 is an item-person map illustrating the relationship between a student's abilities in STEM literacy and the estimated difficulty level of corresponding items along the same latent dimension. Elevated student logit scores signify heightened STEM literacy ability, while a high logit value for an item indicates increased difficulty. According to the item-person map, students' abilities in STEM literacy are primarily classified as medium to low, attributed to the trend of students' logit scores being lower than the item logit scores. This implies that despite the perceived ease of item difficulty, most students encounter challenges in answering questions across SL, TL, EL, and ML dimensions. Interestingly, several previous studies have also revealed that students' literacy abilities in STEM disciplines are low (Adnan et al., 2021; Imamuddin et al., 2022; Nurlaely et al., 2017).

Item number 6 in the scientific literacy dimension has the lowest difficulty level. The indicator of this item is interpreting scientific evidence that supports alleged causes in cases of exploding gas balloons. In this item, students are presented with the statement that the cause of gas balloons exploding is due to the use of hydrogen gas, and they are asked to scientifically interpret the evidence available in the answer choices according to the context of the question. Thus, this item belongs to cognitive level C2 (Mita et al., 2021). Conversely, item 15 in the scientific literacy dimension has the highest difficulty level. The indicator for this item is to conclude the type of mixture (heterogeneous or homogeneous) based on its characteristics. In this item, students are presented with a discourse containing three cases of healthy water pollution from the mass media, and they are asked to conclude what type of mixture was in one of the well water cases in question. Thus, this item is included in the C5 cognitive level (Laila & Fitriyah, 2022).

Items in other STEM literacy dimensions tend to have logit values close to zero, unlike items in the science literacy dimension, which exhibit significant differences. In the technological literacy dimension, students are generally tasked with identifying appropriate technological principles, such as techniques for separating mixtures or altering the properties of elements. In the engineering literacy dimension, students are prompted to devise solutions, such as proposing alternative methods or designing schemes for separating mixtures. Lastly, items in the mathematics dimension emphasize problem formulation in mathematical terms, often requiring students to express them as simple ratios. While, in general, the items in STEM literacy are relatively straightforward, there is a need for comprehensive improvement in the future, aligning them more closely with students' ability levels to yield more precise insights into their STEM literacy ability.

The findings of this study could significantly enhance understanding of how to assess STEM literacy effectively. By using multidimensional Rasch analysis, the study aims to capture the multidimensional nature of STEM literacy, which includes knowledge and understanding of STEM concepts and the ability to use this knowledge in various contexts. The use of multidimensional Rasch analysis allows for a more nuanced and comprehensive evaluation of students' abilities, considering the complexity and multidimensionality of STEM literacy. Furthermore, a reasonable assessment that can maximize students' STEM literacy should be able to measure students' understanding of STEM concepts accurately, their ability to apply this knowledge, and their problem-solving skills in STEM dimensions. Based on the findings, this study can help educators and researchers develop more effective STEM education strategies and curricula. It can also provide valuable insights for educational policymakers, primarily developing and evaluating STEM literacy test instruments.

CONCLUSION

In conclusion, the results of this study have contributed to evaluating the STEM literacy test instrument, mainly through multidimensional Rasch analysis, and have shed light on the interaction between items and students' abilities in STEM literacy. The STEM literacy test instrument focusing on element, compound, and mixture subjects has demonstrated validity and reliability based on multidimensional Rasch analysis. The findings affirm that the multidimensional model is more suitable for assessing STEM literacy, as indicated by lower values of the final deviance, AIC, and BIC, than the unidimensional model. The item-person map reveals that most students have moderate to low abilities in STEM literacy. Finally, DIF analysis based on gender has confirmed that all items are classified as negligible, indicating the absence of DIF problems.

ACKNOWLEDGMENT

Thanks to the expert validators (Wiwin Puspita Hadi and Dwi Bagus Rendy Astid Putera from Science Education, Faculty of Education, Universitas Trunojoyo Madura) and science teachers (Faridah UPT SMPN 14 Gresik) who contributed through face validation and their suggestions for improving the STEM literacy test instrument before testing.

REFERENCES

Adnan, A., Mulbar, U., Sugiarti, S., & Bahri, A. (2021). Scientific literacy skills of students: Problem of biology teaching in junior high school in South Sulawesi, Indonesia. In *International Journal of Instruction*, *14*(3), 847–860. https://doi.org/10.29333/iji.2021.14349a

Ardianto, D., Firman, H., Permanasari, A., & Ramalis, T. R. (2019). What is science, technology, engineering, mathematics (STEM) literacy. *Advances in Social Science, Education and Humanities Research*, *253*, 381–384.

Arikan, S., Erktin, E., & Pesen, M. (2020). Development and validation of a STEM competencies assessment framework. *International Journal of Science and Mathematics Education*, *20*, 1-24. https://doi.org/10.1007/s10763-020-10132-3

Baghael, P. (2017). The application of multidimensional rasch model in large scale assessment and validation: an empirical example. *Electronic Journal of Research in Educational Psychology*, *10*(26), 233-252. https://doi.org/10.25115/ejrep.v10i26.1493

Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan*, *22*(2), 219-230. https://doi.org/10.21831/PEP.V22I2.19590

Bungin, B. (2021). *Metode Penelitian Sosial*. Universitas Airlangga Pres.

Chamrat, S., Manokarn, M., & Thammaprateep, J. (2019). STEM literacy as an instrument for STEM education research field: development, implementation, and utility. *AIP Conference Proceeding*, *2081*, 030013. https://doi.org/10.1063/1.5094011

Combrinck, C. (2020). Is this a useful instrument? An introduction to Rasch measurement models. In *Research Methods*. Psychological Society of South Africa. https://doi.org/10.17605/OSF.IO/BNPFS

Danuwijaya, A., & Roebiyanto, A. (2020). Performance differences by gender in English reading test. *Jurnal Penelitian dan Evaluasi Pendidikan*, *24*(2), 190-197. https://doi.org/10.21831/pep.v24i2.34344

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and validation of scientific inquiry literacy instrument using rasch measurement model. *Education Sciences*, *14*(3), 322. https://doi.org/10.3390/educsci14030322

Erfan, M., Maulyda, M. A., Ermiana, I., Hidayati, V. R., & Widodo, A. (2020). Validity and reliability of cognitive tests study and development of elementary curriculum using Rasch model. *Psychology, Evaluation, and Technology in Educational Research*, *3*(1), 26-33. http://doi.org/10.33292/petier.v3i1.51

Fallon, G., Hatzigianni, M., Bower, M., Forbes, A., & Stevenson, M. (2020). Understanding K-12 STEM education: a framework for developing STEM literacy. *Journal of Science Education and Technology*, *29*, 369-385. https://doi.org/10.1007/s10956-020-09823-x

Fariss, C. J., Kenwick, M. R., & Reuning, K. (2020). Measurement models. In *The SAGE handbook of research methods in political science and international relations*, 353-370.

Ghazali, N. H. M. (2016). A reliability and validity of an instrument to evalute the school-based assessment system: a pilot study. *Internasional Journal of Evaluation and Research in Education*, *5*(2), 148-157.

Hairida, H., Beno, C., Soeharto, S., Charalambos, C., Rasmawan, R., Martono, M., Arifiyanti, F., Winarti, A., & Enawaty, E. (2023). Evaluating digital literacy of pre-service chemistry teachers: Multidimensional rasch analysis. *Journal of Science Education and Technology*, *32*, 643-654. https://doi.org/10.1007/s10956-023-10070-z

Harpian, H., Suwarna, I. R., & Setiawan, A. (2023). The application of STEM learning to improve students' STEM literacy in the knowledge aspect. *Journal of Innovation in Educational and Cultural Research*, *4*(3), 450–457. https://doi.org/10.46843/jiecr.v4i3.676

Huang, X., Erduran, S., Luo, K., Zhang, P., & Zheng, M. (2022). Investigating in-service teachers' STEM literacy: the role of subject background and gender. *Research in Science and Technological Education, 1-21*. https://doi.org/10.1080/02635143.2022.2153243

Imamuddin, M., Zulmuqin, Z., & Sepriyanti, N. (2022). Exploring madrasah students mathematics literacy ability. *International Journal Of Humanities Education And Social Sciences (IJHESS)*, *1*(4), 357–361.

Jackson, C., Schroeder, M. J. M., Bush, S. B., Maiorca, C., Roberts, T., Yost, C., & Fowler, A. (2021). Equity-oriented conceptual framework for K-12 STEM literacy. *Internasional Journal of STEM Education*, *8*, 38. https://doi.org/10.1186/s40594-021-00294-z

Jamaludin, J., Lay, Y. F., Khoo, C. H., & Leong, A. S. Y. (2021). Examining the STEM-science achievement test using rasch dichotomous measurement model. *Jurnal Pendidikan IPA*

Indonesia, *10*(4), 571-584. https://doi.org/jpii.v10i4.29405

Kaharu, S. N., Gagaramusu, Y., Azizah, A., Kamisani, N., Tadeko, N., & Mansyur, J. (2024). Development of an online two-tier test on objects in static fluid. *Pegem Journal of Education and Instruction*, *14*(3), 363-375. https://doi.org/10.47750/pegegog.14.03.34

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Laila, I., & Fitriyah, I. (2022). An analysis of reading comprehension questions in English textbook based on revised bloom's taxonomy. *JET (Journal of English Teaching)*, *8*(1), 71-83. https://doi.org/10.33541/jet.v8i1.3394

Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: a systematic review of journal publications. *Internasional Journal of STEM Education*, *7*, 11. https://doi.org/10.1186/s40594-020-00207-6

Medvedev, O. N., & Krageloh, C. U. (2022). Rasch measurement model. In *Handbook of Assessment in Mindfulness Research*, 1-18. Springer. https://doi.org/10.1007/978-3-030-77644-2_4-1

Mita, D. O., Agustinsa, R., & Susanto, E. (2021). Cognitive level analysis of problems in mathematics textbook class XII revision 2018 materials of congress and construction based on the revised bloom taxonomy. *Journal of Education and Learning Mathematics Research (JELMaR)*, *2*(2), 14-25. https://doi.org/10.37303/jelmar.v2i2.56

Murti, W. W., & Sunarti, T. (2021). Pengembangan instrumen tes literasi sains berbasis kearifan lokal di Trenggalek. *Orbita*, *7*(1), 33-43. https://doi.org/10.31764/orbita.v7i1.4386

Nurlaely, N., Permanasari, A., & Riandi, R. (2017). Student's STEM literacy in biotechnology learning at junior high school. *Journal of Physics: Conference Series*, *895*, 012155. https://doi.org/10.1088/1742-6596/895/1/012155

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. In *Handbook of Markert Research*, 1-40. http://doi.org/10.1007/978-3-319-05542-8_15-1

Setiawan, I., Sudarmin, S., & Partaya, P. (2023). Development of project based ethno-STEM online learning module to increase interpersonal literacy and learning outcome. *Journal of Innovative Science Education*, *12*(2), 192-200.

Sigfrid, K. (2024). IRT for voting advice applications: a muti-dimensional test that is adaptive and interpretable. *Quality & Quantiy*. https://doi.org/10.1007/s11135-024-01845-6

Suwarna, I. R., Riandi, R., Kumano, Y., Permanasari, A., Sudarmin, S., & Widyatmoko, A. (2023). Science teacher experiences in developing STEM literacy assessment. In *Education – Annual Volume 2023*, Intechopen. https://doi.org/10.5772/intechopen.112308

Tang, K., & Williams, P. J. (2019). STEM literacy or literacies? Examining the empirical basis of these constructs. *Review of Education*, *7*(3), 675–697. https://doi.org/10.1002/rev3.3162

Wannapiroon, P., Nilsook, P., Techakosit, S., & Kamkhuntod, S. (2021). STEM literacy of students in vocational education. *International Journal of Technology in Education and Science*, *5*(4), 527–549.

Wu, M., Tam, H. P., & Jen, T. H. (2016). *Educational measurement for applied researchers: Theory into practice.* Springer.

Wu, Z., Huang, L., Liu, Y.K., & Chiang, F. K. (2024). Developing a framework of STEM literacy for kindergarten children. *Research in Science Education, 1-23*. https://doi.org/10.1007/s11165-024-10157-6

Yudiana, W., Triwahyuni, A., & Susanto, H. (2023). Multidimensional rasch analysis of gender differences in tes inelegensi kolektif Indonesia-tinggi (TIKI-T). *JP3I: Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, *12*(1), 1-16. http://doi.org/10.15408/jp3i.v12i1.20417