Journal of Teaching and Learning Mathematics

Manuscript template of JTLM (pp. 130–140)
https://ejournal.umm.ac.id/index.php/jtlm

Analyzing mathematics formative test questions for higher-order thinking skills utilizing the Rasch model

Clarisa Aurelia Avrinda[1*], Amelline Slazenny EN[2], Fahad Abdul Ghani[3]

1. *SMP Muhammadiyah 9 Bojonegoro, Indonesia*
2. *SMP Muhammadiyah 9 Bojonegoro, Indonesia*
3. *Sindh Madressatul Islam University Karachi Pakistan*

ARTICLE INFO.

Keywords:
Rasch Model;
Hots;
Formative test

Abstract

In the present day, education is anticipated to cultivate individuals who possess the 4C competencies, which encompass creative thinking, critical thinking and problem-solving, communication, and teamwork. In the 21st century, education is of growing significance in equipping students with learning aptitude, creativity, and proficiency in using technology and information media, enabling them to thrive and adapt throughout their lives. This research aims to examine and explain the attributes of questions that assess higher-order thinking skills using the Rasch Item Response Theory (IRT) method. This study is a Descriptive-quantitative-research. This study focused on analyzing the response patterns of 70 class XI students from SMA Negeri 3 Tuban towards a formative test instrument that consisted of five answer choices. The test set is derived from the execution of formative assessments. The Rasch model was utilized for data analysis with the assistance of ConQuest. The analysis results indicate that all items conform to the Rasch model. Based on the difficulty level, 15% of the items fell into the troublesome category, while 85% were classified as moderate. The difficulty level values go from -1,094 to 1,250

To quote this article: *Avrinda et al. (2024). Analyzing mathematics formative test questions for higher-order thinking skills utilizing the Rasch Model. Journal of Teaching and Learning Mathematics, 1(2), 130-140. https://doi.org/10.22219/jtlm.v1i2.31627*

1 Introduction

Education at this time is expected to produce human resources with the 4C abilities (Johansson, 2014; Komboz, 2018; Wood, 1978), namely creative thinking skills (Bond, 2020; Hansen, 2020; Wolfe, 2013), critical thinking and problem-solving, communication (Hansen, 2020; Hee, 2008; Maia, 2013), and collaboration. In the 21st century, education is becoming increasingly important to ensure students have learning skills (Setio & Baiduri, 2023), innovation (Lubis & James, 2024), and skills in using technology and information media and can work and survive using these skills for life (Andrian & Rusman, 2019).

Abdullah Sani, R. (2019) stated that the younger generation must have critical thinking, creative thinking, and problem-solving skills (Abus & Usmiyatun, 2023). These three abilities will help the younger generation face all the more complex challenges of the future (Budiarti et al., 2024; Cholily, 2023; Reihani et al., 2024).

The role of education is significant for the younger generation to gain critical thinking skills, creative thinking, and problem-solving abilities (In'am et al., 2021; Karim & Zoker, 2023; Zahroh et al., 2024). At this time, education has entered the 21st century, where learning no longer focuses on memorizing facts without meaning, but instead knowing that requires students to be skilled in using technology, be creative, innovate, and be intelligent in science (Wijaya et al., 2016). These abilities have been organized into one in the 2013 curriculum, which we currently call higher-order thinking skills or higher-order thinking Skills (HOTS).

HOTS is the ability to combine facts and ideas in analyzing and evaluating up to the creation stage by assessing a fact that has been studied or being able to create from something that has been studied. Investigating, evaluating, and building is part of the cognitive taxonomy created by Benjamin S. Bloom in 1956. In the end, it was refined again by Anderson and Krathwohl (2001) into C1-memory (remembering), C2-understanding (understanding), C3- applying, C4-analysis, C5-evaluating, and C6-creation (Saraswati & Agustika, 2020a).

Mathematics subjects are one area of knowledge central to developing the competencies needed to face the 21st-century environment (Hamdi et al., 2018). As a basis for the development of modern knowledge and technology, mathematics also plays a role in advancing human thinking power. Mathematics subjects provide the ability to think logically, analytically, systematically, critically, innovatively, and collaborate. These subjects must be taught from the basic level of education (Darmayanti, Arif, et al., 2023; Gunawan et al., 2023; Nisa et al., 2023). Through mathematics subjects, students are expected to be able to apply their uses to everyday life (Saraswati & Agustika, 2020b). Everyone needs mathematics subjects to solve various problems through calculating and thinking (Hasanah et al., 2023; Laila et al., 2023; Rofiah et al., 2023). Being able to solve problems means being able to examine a problem and being able to apply that knowledge to new situations. This ability is usually called higher-order thinking Skills (Dinni, 2018).

Ideally, learning-oriented towards Higher Order Thinking Skills (HOTS) starts from designing, implementing learning in class, and evaluating learning (Aji et al., 2023; Astuti et al., 2023; Setiawati et al., 2023). The learning evaluation process is essential to measure the extent of each student's abilities. Suppose implementing teaching and learning activities is oriented towards higher-order thinking Skills (HOTS). In that case, the learning evaluation must also be oriented toward High Order Thinking Skills (HOTS). Student learning can be evaluated in various ways, including tests (Legi Aspriyanti et al., 2022).

Research on students' ability test questions based on higher-order thinking Skills (HOTS) has been increasingly carried out in recent years. For example, in 2020, Yuli Setiawan and Siti Fatonah (2020) conducted analytical research on thematic mid-semester assessment questions at MI Sleman Regency, and the result was that the percentage of HOTS questions was 18.6%. Analysis of question items helps improve the quality through revision or removing ineffective questions (Darmayanti, Hidayat, et al., 2023; Iwuanyanwu, 2021; Shania et al., 2024). Besides that, it can be used as diagnostic information for students, whether they understand the material that has been taught. Analysis of questions in education can be done using two approaches, namely the classical and modern approaches. Modern analysis of question items is the study of items using Item Response Theory (IRT) or the theory of question item answers (Fauziana & Dessy Wulansari, 2021).

Rasch modeling provides a different approach to using test scores or raw data in educational Assessment (Cahyadi et al., 2023; Lestari et al., 2023; Setiawan et al., 2023). The main goal is to produce a measurement scale with equal intervals that will later provide accurate information about test takers and the quality of the questions they work on (Perdana, 2018). The Rasch model is based on two principles. The first principle is the subject's ability, in this case, the student's ability to answer a question, which can be predicted using traits (Wahyuningsih, 2020). Mardapi (2012) said that measurement in the Rasch model directly compares individuals and items. In this case, the individual is the test taker's ability, and the items are the difficulty level parameters. So that under certain conditions, for example, the test taker's ability increases, the chance of correctly answering the test item becomes greater. Therefore, the opportunity to answer a test item correctly refers to two things, namely, the test taker's ability and the difficulty level of the item.

Regarding the basic concept of the Rash model, Sumintono (2014) also stated the same thing, that the Rasch model is a measurement model that determines the relationship between the test taker's ability and the level of difficulty of the test items. Furthermore, Sumintono (2014) illustrates this relationship; for example, if a test taker can answer 80% of the questions correctly, he has higher abilities than a test taker who can only answer 60% of the questions correctly. Based on this, researchers are interested in analyzing hots formative questions using the Rasch model

2 Theoretical Review

Higher Order Thinking Skills

Sani (2019) defines higher-order thinking Skills (HOTS) as the capacity to engage in strategic thinking, utilizing knowledge to resolve problems, examine arguments, negotiate difficulties, or make predictions. Stein & Lane argue that higher-order thinking skills are intricate mental processes that lack a systematic solution, cannot be

anticipated, and can only be addressed by employing a novel approach to established problems or tasks, distinct from the provided examples (Ayuningtyas & Rahaju, 2017).

According to Lewis & Smith, high-level thinking results from combining existing knowledge with new information and then organizing and developing that information to achieve a specific goal or find answers and solutions to complex problems. This process requires advanced cognitive skills. The Higher Order Thinking Skills (HOTS) encompass essential cognitive abilities such as critical thinking, creative thinking, problem-solving, and decision-making (Sani, 2019). "Higher Order Thinking Skill" refers to the capacity to engage in complex cognitive processes such as describing, concluding, analyzing, and employing other advanced thinking skills to solve problems that lack a predetermined algorithm, cannot be predicted, and require diverse approaches.

Formative Test Questions

Formative Assessment is a method used to gather information regarding the progress of the learning process. The primary objective of formative Assessment is to systematically track the progress of students' learning and offer continuous feedback to instructors to enhance their learning experience. In addition, the utilization of Formative Assessment in education has the advantage of enabling pupils to discern their aptitudes and deficiencies in specific subjects inside particular classes. Formative Assessment also aids in identifying student challenges, allowing for prompt intervention to support their learning progress (Sinaga et al., 2016; Susanti, 2018).

Rasch Model

The Rasch model is a modern evaluation theory that can classify item and human data on a distribution map. The Rasch model is a theoretical framework that outlines the anticipated data organization to obtain precise measurements. It functions as a benchmark for effective quantification. Rasch's equations provide factual information and encapsulate the anticipated correlations we expect to observe in the real world. Education aims to provide students with the essential skills and information needed to address a diverse range of life's difficulties, going beyond the material taught in textbooks or assessed on tests. Rasch models enable the testing of the hypothesis that the specific challenges presented in a curriculum and on a test accurately represent the entire population of possible challenges in that domain. This is achieved by ensuring that the measures remain consistent across different tests measuring the same thing. A Rasch model is a theoretical framework that serves as a conceptual standard, providing a useful way to organize and understand data, even though it may not be directly observed in real-world practice (Asrijanty, 2014; Sabekti & Khoirunnisa, 2018)

3 Method

This quantitative descriptive study aims to obtain an overview of the quality of higher-order thinking skills (HOTS) formative tests through the characteristics of the Rasch model test. The subjects of this research were 70 response patterns of class. The ice instrument is teacher-made and taken from a formative test's results in multiple choices with five alternative answers. Quantitative data analysis was carried out using the Item Response Theory (IRT) approach with the Rasch model with the help of the ConQuest program. The quality of questions in the Rasch model can be determined by estimating parameters such as the validity of the difficulty level and item fit with the Rasch model. In the validity estimation test, items can be compared using the criteria below (Pratama, 2020).

Table 1. INFIT MNSQ Score Criteria

INFIT MNSQ value	Information
> 1.33	It does not match the model.
$0.77 - 1.33$	Matches the model
< 0.77	It does not match the model.

In estimating an item's difficulty level, the criteria for determining the level of difficulty of an item revolve around the value in the calculated passing item -2.0 to 2.0. Figure 1 illustrates the sequential phases of the research procedure to the Rasch Model.

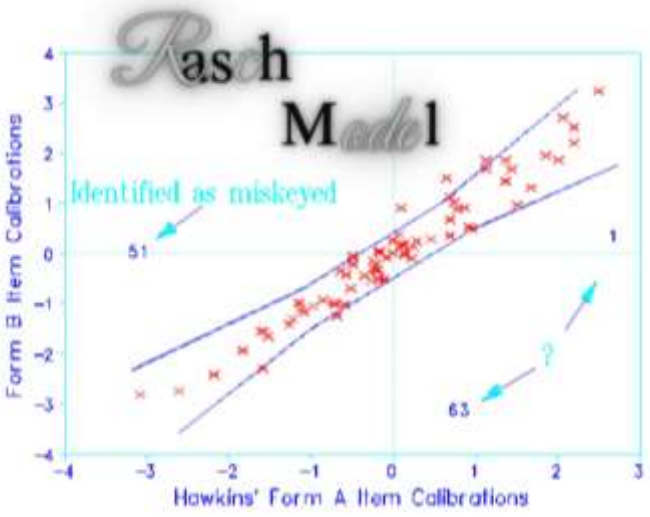

Figure 1 depicts the sequential process of the Rasch Model analysis procedure (Yolcu, 2019; Yuwono et al., 2021).

4 Results and Discussion

The formative test instrument for arithmetic sequences and series material has 20 items with five answer choices. Respondents' answer patterns were analyzed using the Rasch model via ConQuest software.

Item Validity Estimates

The results of the INFIT MNSQ value analysis in the QUEST program can be seen in the image below:

ConQuest: Generalised Item Response Modelling Software Wed Jan 11 19:37 2023
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item

	VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
	item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	1	-0.000	0.186	0.88	(0.67, 1.33)	-0.7	0.92	(0.71, 1.29)	-0.5
2	2	0.507	0.177	0.97	(0.67, 1.33)	-0.1	0.97	(0.82, 1.18)	-0.4
3	3	-0.258	0.191	0.95	(0.67, 1.33)	-0.3	0.97	(0.65, 1.35)	-0.1
4	4	1.250	0.173	1.00	(0.67, 1.33)	0.1	1.01	(0.90, 1.10)	0.1
5	5	-0.167	0.189	1.19	(0.67, 1.33)	1.1	1.11	(0.67, 1.33)	0.7
6	6	1.191	0.173	1.07	(0.67, 1.33)	0.5	1.06	(0.90, 1.10)	1.2
7	7	-0.165	0.189	0.95	(0.67, 1.33)	-0.2	0.97	(0.67, 1.33)	-0.1
8	8	-0.672	0.200	1.11	(0.67, 1.33)	0.7	1.05	(0.53, 1.47)	0.3
9	9	-0.349	0.193	0.92	(0.67, 1.33)	-0.4	0.95	(0.63, 1.37)	-0.2
10	10	0.004	0.185	0.92	(0.67, 1.33)	-0.4	0.95	(0.72, 1.28)	-0.3
11	11	-0.672	0.200	0.90	(0.67, 1.33)	-0.5	0.95	(0.53, 1.47)	-0.2
12	12	0.159	0.183	1.09	(0.67, 1.33)	0.6	1.04	(0.75, 1.25)	0.4
13	13	0.305	0.180	0.95	(0.67, 1.33)	-0.3	0.95	(0.78, 1.22)	-0.4
14	14	0.082	0.184	1.06	(0.67, 1.33)	0.4	1.02	(0.73, 1.27)	0.2
15	15	-0.938	0.206	1.16	(0.67, 1.33)	1.0	1.05	(0.44, 1.56)	0.3
16	16	0.303	0.180	1.00	(0.67, 1.33)	0.0	0.99	(0.78, 1.22)	-0.1
17	17	-0.258	0.191	1.08	(0.67, 1.33)	0.5	1.01	(0.65, 1.35)	0.1
18	18	-0.354	0.193	1.15	(0.67, 1.33)	0.9	1.09	(0.62, 1.38)	0.5
19	19	-1.095	0.209	0.98	(0.67, 1.33)	-0.1	1.01	(0.38, 1.62)	0.1
20	20	1.129*	0.823	1.06	(0.67, 1.33)	0.4	1.05	(0.90, 1.10)	1.0

Figure 1. Recapitulation of Item Validity

Figure 1 above provides information regarding item validity where all items fit or match the Rasch model with a range of INFIT MNSQ values between 0.88–1.19.

Estimated Level of Difficulty

To determine an item's difficulty level through the ConQuest program, you can find out by looking at the item estimate (Threshold) analysis results. The criteria for determining the item difficulty level range from -2.0 – 2.0. The items are included in the easy category if the range or distribution of items or test participants is extensive. Meanwhile, the items are in the difficult category if the range or distribution of items or test participants is extensive. The results of the item estimate (Threshold) analysis in the ConQuest program can be seen in the image below:$<-2.0 > 2.0$.

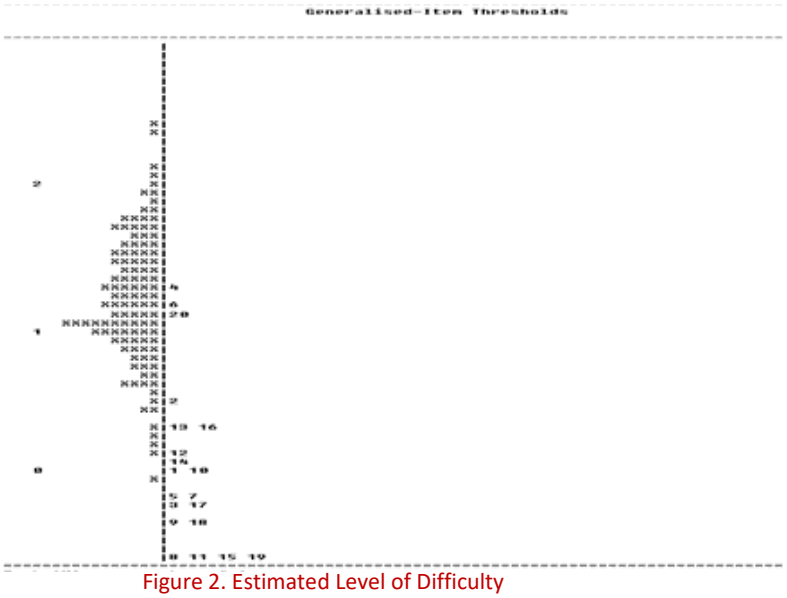

Figure 2. Estimated Level of Difficulty

From picture 2 above, you can see that question item 4 is the most challenging item. Even if compared with the test taker's ability, the possibility of the test taker answering item number 4 correctly is minimal, or one could say impossible. Apart from that, items number 8, 11, 15, and 19 are the most straightforward questions among the other questions. The threshold value for all items is in the range, which means all items meet the item difficulty criteria.$-1.094 - 1.250$.

Estimated Item Passed$(\boldsymbol{Fit})$

Finding which items have failed or passed is based on the OUTFIT t value in the ConQuest program. If the OUTFIT t value ≤ 2.00, the item passes; if the OUTFIT t value ≥ 2.00, the item fails.

Table 4. Recapitulation of Items$\boldsymbol{Fit}$

Items	Outfit Value t	Information	Items	Outfit Value t	Information
1	-0.5	*Getaway*	*11*	-0.2	*Getaway*
2	-0.4	*Getaway*	*12*	0.4	*Getaway*
3	-0.1	*Getaway*	*13*	-0.4	*Getaway*
4	0.1	*Getaway*	*14*	0.2	*Getaway*
5	0.7	*Getaway*	*15*	0.3	*Getaway*
6	1.2	*Getaway*	*16*	-0.1	*Getaway*
7	-0.1	*Getaway*	*17*	0.1	*Getaway*
8	0.3	*Getaway*	*18*	0.5	*Getaway*
9	-0.2	*Getaway*	*19*	0.1	*Getaway*
10	-0.3	*Getaway*	*20*	1.0	*Getaway*

Based on Table 4 above, it is known that all items passed, so it can be concluded that all items can be used.

The results of this study show that all items are valid with a value INFIT MNSQ between 0.88 and 1.19. In the research conducted, The INFIT MNSQ value is in the range, which means all items are right; the INFIT MNSQ value is not much different from the INFIT MNSQ value in this study. The item difficulty level in this study, the Threshold value for all items, is in the range, which means all items meet the item difficulty criteria. In the research carried out$0.25 - 1.24 - 1.094 - 1.50$ (Elisa et al., 2020), Based on the Threshold value, 1 out of 10 items have a Threshold value, so the item does not meet the item difficulty criteria. The OUTFI t value in this study is OUTFIT t ≤ 2.00 so that all items pass and can be used; this is different from research conducted by< -2.0. (Yulianto & Widodo, 2020)If several items do not meet the requirements, you can be sure that the item is not good and needs to be repaired or replaced.

5 Conclusion

Based on the results of the analysis of the HOTS test questions with material on sequences and arithmetic series, several characteristics of the test and test participants can be described as follows: 1) estimate the validity of item fit or fit with the Rasch model for 20 items with a range of INFIT MNSQ values between 0.88 – 1.19. 2) The estimated difficulty level is in the range, meaning all items meet the criteria. 3) All items in the test can be used based on the estimated value of OUTFIT t ≤ 2.00-1.094 - 1.250.

The researcher suggests that future researchers conduct research with a more significant number of student response patterns. They can also use summative test questions to contain varied material.

6 Reference

Abus, O., & Usmiyatun, U. (2023). TAYO Cards in Understanding Numbers 1-10 for Early Childhood, Improve? *Journal of Teaching and Learning Mathematics*, *1*(1), 13–24.

Aji, F. B., Sugianto, R., & Usmiyatun, U. (2023). Analysis Think Mathematical comedian with Method Inductive, Deductive, and Analogy in Make Material Stand Up Comedy. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*.

Andrian, Y., & Rusman, R. (2019). Implementasi pembelajaran abad 21. *Jurnal Penelitian Ilmu Pendidikan*, *12*(1), 14–23.

Asrijanty. (2014). Model Rasch sebagai Kerangka Acuan Penyusunan Alat Ukur. *Jurnal Pendidikan Dan Kebudayaan*, *20*(1), 109–123. https://doi.org/10.24832/jpnk.v20i1.130

Astuti, P., Anwar, M. S., & Juarlan, A. E. (2023). Pengaruh Kecerdasan Logis Matematis Terhadap Kemampuan Pemecahan Masalah Pada Penyelesaian Soal Cerita. *Delta-Phi: Jurnal Pendidikan Matematika*, *2*, 156–164.

Bond, T. G. (2020). Applying the rasch model: Fundamental measurement in the human sciences. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 1–348. https://doi.org/10.4324/9780429030499

Budiarti, E., Lestari, J. T., & Hagenimana, E. (2024). The educational game" MARIO" for early childhood number recognition to improve cognitive abilities:

Attempts and Problems. *Journal of Teaching and Learning Mathematics*, *2*.

Cahyadi, M. R., Ariansyah, F., & da Silva Santiago, P. V. (2023). Analysis of Using Pattern Finding Strategies Skills in Mathematical Problem Solving Viewed from Gender Differences. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*, 6–10.

Cholily, Y. M. (2023). Analysis of Mathematical Communication Ability on Problem-Solving Pythagoras Theorem Viewed From Students' Cognitive Style. *Journal of Teaching and Learning Mathematics*, *1*(1), 42–52.

Darmayanti, R., Arif, V. R., & Usmiyatun, U. (2023). My HAIR SALON: Collaboration of Right Brain and Left Brain Students in the Mind Mapping Model for Statistical Material. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*(2).

Darmayanti, R., Hidayat, A., da Silva Santiago, P. V, Gunawan, I. I., & Dhakal, A. (2023). Post-Math: An innovative math approach to engage children (Case Studies). *Journal of Teaching and Learning Mathematics*, *1*(1).

Dinni, H. N. (2018). HOTS (High Order Thinking Skills) dan Kaitannya dengan Kemampuan Literasi Matematika. *Prisma*, *1*, 170–176.

Elisa, R. S., Sutisnawati, A., & Nurasiah, I. (2020). Jurnal perseda. *Jurnal Persada*, *III*(3), 115–120.

Fauziana, A., & Dessy Wulansari, A. (2021). Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch. *Ibriez : Jurnal Kependidikan Dasar Islam Berbasis Sains*, *6*, 10–19. https://doi.org/10.21154//ibriez.v6i1.112

Gunawan, I. I., Darmayanti, R., In'am, A., Vedianty, A. S. A., & Vereshchaha, V. (2023). Clap-Breathe-Count: Using Ice-Breaking Ma-Te-Ma-Ti-Ka to Increase High School Students' Learning Motivation. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*(1), 51–57.

Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *Research and Evaluation in Education*, *4*(2), 126–135. https://doi.org/10.21831/reid.v4i2.22089

Hansen, T. (2020). Item analysis of the Eating Assessment Tool (EAT-10) by the Rasch model: A secondary analysis of cross-sectional survey data obtained among community-dwelling elders. *Health and Quality of Life Outcomes*, *18*(1). https://doi.org/10.1186/s12955-020-01384-2

Hasanah, N., Laila, A. R. N., & Nurmalitasari, D. (2023). Development of Audiovisual Ethnomathematics Teaching Materials Assisted by Kinemaster Applications Algebraic Function Derivative Application Materials for Class XI SMA. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*.

Hee, S. H. (2008). Validation of the Korean version of the pediatric quality of life inventory™ 4.0 (PedsQL™) generic core scales in school children and adolescents using the rasch model. *Health and Quality of Life Outcomes*, *6*. https://doi.org/10.1186/1477-7525-6-41

In'am, A., Maghfiroh, W., & Rachmawati, N. I. (2021). Designing math learning through teaching the material" Number Patterns" by developing means to train students' reasoning skills. *AMCA Journal of Education and Behavioral Change*, *1*.

Iwuanyanwu, P. N. (2021). Contemporary problems of teaching and learning in mathematics education. *Mathematics Teaching-Research Journal*, *13*(2), 23–35.

Johansson, S. (2014). Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure - a comparative study using a Rasch model approach. *Health and Quality of Life Outcomes*, *12*(1). https://doi.org/10.1186/1477-7525-12-24

Karim, S., & Zoker, E. M. (2023). Technology in Mathematics Teaching and Learning: An Impact Evaluation in Selected Senior Schools in Masingbi Town. *Assyfa Learning Journal*, *2*, 60–72.

Komboz, B. (2018). Tree-Based Global Model Tests for Polytomous Rasch Models. *Educational and Psychological Measurement*, *78*(1), 128–166. https://doi.org/10.1177/0013164416664394

Laila, A. R. N., Cholily, Y. M., Syaifuddin, M., & Kurniawati, W. (2023). Functional Inverse Material Description Test: Development of Islamic Integrated Communication Ability Instruments. *Delta-Phi: Jurnal Pendidikan Matematika*, *2*, 107–124.

Legi Aspriyanti, Rustam, & Setyonegoro, A. (2022). Analisis Soal Tes Formatif Bahasa Indonesia Berorientasi pada Higher Order Thinking Skills (HOTS). *Lintang Aksara: Jurnal Pendidikan Bahasa Dan Sastra*, *1*(2), 88–100.

Lestari, A. S. B. L., Wahyono, A., AkkuÅŸci, Y. E., Purwanto, P., Anas, K., Nurmalasari, Y., Bibi, R., & Yunus, M. (2023). PLANâ€"DOâ€"SEE: Lesson Study-Based Differentiated Learning in Middle Schools. *Delta-Phi: Jurnal Pendidikan Matematika*, *1*, 41–45.

Lubis, M., & James, M. L. (2024). Compared concrete and digital mathematics learning media: a systematic literature review in the age of Industrial Revolution 5.0. *Journal of Teaching and Learning Mathematics*, *2*.

Maia, A. (2013). Cross-cultural adaptation and analysis of the psychometric properties of the balance evaluation systems test and MiniBESTest in the elderly and individuals with Parkinson's disease: Application of the Rasch model. *Brazilian Journal of Physical Therapy*, *17*(3), 195–217. https://doi.org/10.1590/S1413-35552012005000085

Nisa, H., Choirudin, C., Anwar, M. S., & Wardana, M. R. F. (2023). Implementasi Etnomatemtika Berbasis Alat Kesenian Rebana Dalam Pembelajaran Bangun Ruang. *Delta-Phi: Jurnal Pendidikan Matematika*, *3*, 205–210.

Perdana, S. A. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, *6*(1), 41–48. https://doi.org/10.31629/kiprah.v6i1.574

Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy : Jurnal Pendidikan Islam*, *7*(1), 61–70. https://doi.org/10.32923/tarbawy.v7i1.1187

Reihani, M., Arifin, S., Afriani, A., & Choirudin, C. (2024). How valid is AR-enabled e-module development in a Palembang wardrobe context? *Journal of Teaching and Learning Mathematics*, *2*.

Rofiah, N., Anwar, M. S., Choirudin, C., & Ridho'i, A. V. (2023). Analisis Kesulitan Belajar Matematika ditinjau dari Motivasi Belajar Siswa. *Delta-Phi: Jurnal Pendidikan Matematika*, *3*.

Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model Untuk Mengembangkan Instrumen Pengukuran Kemampuan Berikir Kritis Siswa Pada Topik Ikatan Kimia. *Jurnal Zarah*, *6*(2), 68–75. https://doi.org/10.31629/zarah.v6i2.724

Saraswati, P. M. S., & Agustika, G. N. S. (2020a). Kemampuan Berpikir Tingkat Tinggi Dalam Menyelesaikan Soal HOTS Mata Pelajaran Matematika. *Jurnal Ilmiah Sekolah Dasar*, *4*(2), 257. https://doi.org/10.23887/jisd.v4i2.25336

Saraswati, P. M. S., & Agustika, G. N. S. (2020b). Kemampuan Berpikir Tingkat Tinggi Dalam Menyelesaikan Soal HOTS Mata Pelajaran Matematika. *Jurnal Ilmiah Sekolah Dasar*, *4*(2), 257. https://doi.org/10.23887/jisd.v4i2.25336

Setiawan, A., Anwar, S., & Ariansyah, F. (2023). Pengaruh Penggunaan Metode Team Quiz Terhadap Hasil Belajar Siswa Pada Mata Pelajaran Matematika. *Delta-Phi: Jurnal Pendidikan Matematika*, *2*, 147–155.

Setiawati, D., Setiawan, A., Choirudin, C., & Muniri, M. (2023). Pengaruh Pembelajaran dengan Pendekatan Interkoneksi Matematika Al-Qurâ€™an terhadap Hasil Belajar Siswa. *Delta-Phi: Jurnal Pendidikan Matematika*, *2*, 183–190.

Setio, A., & Baiduri, B. (2023). Statistical Literacy: A Preliminary Research to Identify Student's Level in Solving AKM Problems Based on Watson Category. *Journal of Teaching and Learning Mathematics*, *1*(1), 33–41.

Shania, F., Inganah, S., Rosyadi, A. A. P., & Afolaranmi, A. O. (2024). What are junior high maths students' Edmodo interests and learning outcomes? *Journal of Teaching and Learning Mathematics*, *2*.

Sinaga, B., Harahap, M., Sinambela, P., & Sinaga, L. (2016). *… of Mathematics Instructional and Authentic Assessment Model Based on Curriculum 2013 to Improve the Attitude Quality, Creative Thinking Ability and Mathematics …*. digilib.unimed.ac.id.

Susanti, R. D. (2018). Penerapan Edmodo Pada Pemberian Tes Formatif Mahasiswa Pendidikan Profesi Guru Matematika. *Aksioma Jurnal*.

Wahyuningsih, S. (2020). *PENGGUNAAN MODEL RASCH UNTUK ANALISIS INSTRUMEN TES PADA MATA KULIAH MATEMATIKA AKTUARIA*. *3*, 45–50.

Wijaya, E. Y., Sudjimat, D. A., & Nyoto, A. (2016). Transformasi Pendidikan Abad 21 sebagai Tuntutan Pengembangan Sumber Daya Manusia di Era Global. *Prosiding Seminar Nasional Pendidikan Matematika*, *1*, 263–278.

Wolfe, E. (2013). A bootstrap approach to evaluating person and item fit to the Rasch model. *Journal of Applied Measurement*, *14*(1), 1–9.

Wood, R. (1978). Fitting the Rasch model—A heady tale. *British Journal of Mathematical and Statistical Psychology*, *31*(1), 27–32. https://doi.org/10.1111/j.2044-8317.1978.tb00569.x

Yolcu, H. H. (2019). Analysis and evaluation of 3. and 4. Grade science course learning outcomes according to revised bloom taxonomy. *Elementary Education Online*, *18*(1), 253–262. https://doi.org/10.17051/ilkonline.2019.527214

Yulianto, A., & Widodo, A. (2020). Disclosure of Difficulty Distribution of HOTS-Based Test Questions through Rasch Modeling. *Indonesian Journal of Primary Education*, *4*(2), 197–203. https://doi.org/10.17509/ijpe.v4i2.29318

Yuwono, T. A., Sulistiadi, S., & Atmiasih, D. (2021). Pengaruh Teknologi Ramah Lingkungan Sonic Bloom Menggunakan Musik Hard Rock dan Asmaul Husna Terhadap Pertumbuhan Kangkung (Ipomoea Aquatic). *MEKANIKA*. https://www.ejournal.unugha.ac.id/index.php/me/article/view/411

Zahroh, U., Ariansyah, F., & Farooq, S. M. Y. (2024). Video Blog (Vlog): Media Innovation in Mathematics Learning explores creative mathematics learning. *Journal of Teaching and Learning Mathematics*, *1*(2).