

artikel_Sindi

by suripah suripah

Submission date: 17-Mar-2025 05:30AM (UTC+0700)

Submission ID: 2616324391

File name: Revisi_oke_111859.docx (142.57K)

Word count: 4392

Character count: 25007

OPTIMIZING NUMERACY ASSESSMENTS: A DETAILED ITEM ANALYSIS FOR PHASE D STUDENTS

Received dd Month yy; Received in revised form dd Month yy; Accepted dd Month yy (9pt)

Abstrak

Kemampuan numerasi peserta didik Indonesia yang rendah mestinya menjadi perhatian semua kalangan. Dibandingkan dengan kemampuan literasi, kemampuan numerasi peserta didik Indonesia lebih memprihatinkan. Oleh karena itu, praktik-praktik baik yang dapat mendukung peningkatan kemampuan numerasi sangat dibutuhkan sebagai bentuk usaha untuk meningkatkan prestasi akademik peserta didik. Salah satu praktik baik tersebut adalah pemberian soal numerasi secara rutin untuk melatih peserta didik dalam menghadapi soal-soal berbasis kemampuan numerasi. Penyusunan soal numerasi yang berkualitas perlu melalui tahapan pengembangan yang ilmiah. Pada penelitian sebelumnya, telah dikembangkan instrumen soal numerasi untuk peserta didik fase D yang telah teruji validitas dan kepraktisannya melalui serangkaian kegiatan kualitatif, yaitu *Self-Evaluation*, *Expert Review*, *One-to-one*, dan *Small Group*. Untuk menyempurnakan kualitas soal numerasi bagi peserta didik fase D, penelitian dilanjutkan dengan kegiatan kuantitatif, yakni melalui tahapan *Field Test*. Tujuan khusus dari penelitian ini adalah untuk menganalisis butir soal numerasi untuk peserta didik fase D dengan teknik analisis data menggunakan bantuan program Quest. Penelitian ini menghasilkan 21 butir soal numerasi yang valid dengan tingkat kesukaran yang ideal, yakni 19% soal sulit, 62% soal sedang, dan 19% soal mudah, serta tingkat reliabilitas yang baik sekali (0.92). Butir-butir soal ini dapat digunakan di jenjang SMP atau Fase D sebagai asesmen diagnostik, formatif, maupun sumatif untuk mengukur tingkat kemampuan numerasi peserta didik.

Kata kunci: Kemampuan Numerasi; Peserta Didik Fase D; Program Quest.

Abstract

The low numeracy skills of Indonesian students should be a concern for all stakeholders. Compared to literacy skills, the numeracy skills of Indonesian students are more alarming. Therefore, best practices that support the enhancement of students' numeracy skills are urgently needed as an effort to improve the academic performance of Indonesian students. One effective practice is giving numeracy problems routinely to train students in dealing with numeracy-based problems. Developing high-quality numeracy questions requires a systematic and scientific approach. In previous research, a numeracy instrument for Phase D students was developed and validated through a series of qualitative activities (*Self-Evaluation*, *Expert Review*, *One-to-One*, and *Small Group*). To further improve the quality of the numeracy questions for Phase D students, this study will continue with quantitative activities, specifically through the *Field Test* phase. Aim of this study is to analyze the numeracy questions for Phase D students using data analysis techniques with the aid of the Quest program. This study resulted in 21 valid numeracy questions with an ideal difficulty distribution (19% difficult, 62% moderate, and 19% easy), and a high reliability score (0.92). These questions can be used at the middle school level or Phase D as diagnostic, formative, or summative assessments to measure students' numeracy skills.

Keywords: Numeracy Skill; Quest Program; Student Phase D.



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

DOI: <https://doi.org/10.24127/ajpm>

INTRODUCTION

Numeracy is simply viewed dichotomously as testing whether a person can perform basic arithmetic or not (Parnis & Petocz, 2016). Mastering numeracy skills means having the ability to think critically in processing data, making decisions, and solving problems effectively (Yustitia et al., 2025).

International assessments such as PISA (Programme for International Student Assessment) have consistently shown that Indonesian students struggle with basic numeracy skills. In line with this, the national assessment, namely Minimum Competency Assessment (AKM), also shows that students' numeracy is still relatively low (Rosnelli & Ristiana, 2023).

One crucial effort to enhance students' numeracy skills is through regular practice with numeracy questions. An activity that can increase numeracy scores is intensive training in answering numeracy-related questions (Ismawati et al., 2023; Kholid et al., 2022). As a tool for practice, teachers need a collection of well-developed numeracy questions. Therefore, the development of high-quality numeracy questions is essential.

The various stages required to produce a high-quality question instrument include expert validation, readability testing (both limited and in small groups), and item analysis. Item analysis aims to assess the validity, reliability, discriminative power, and difficulty level of the questions. In previous research, a set of questions that were valid (based on expert judgment) and practical (through two stages of readability testing with students) was developed. To further test the quality, the instrument must undergo a final stage of quantitative analysis.

Referring to the Minimum Competency Assessment (AKM), which serves as a benchmark for measuring the quality of each school in Indonesia, several forms of numeracy questions are provided: essay, short answer, multiple choice, complex multiple choice, and matching. These forms can be categorized into dichotomous and polytomous questions.

The Quest program, an application used for calculating and analyzing question items, has the advantage of being able to analyze both dichotomous and polytomous questions. Additionally, this program can estimate both item groups and respondent groups, making it the primary choice for researchers to produce truly high-quality numeracy questions. The objective of this study is to determine the quality of numeracy question items (including essay, short answer, multiple choice, complex multiple choice, and matching) for Phase D students through quantitative analysis using the Quest program.

Phase D students are those in grades 7, 8, and 9 in Junior High School. One of the competencies measured in the basic education competency standards (SKL) is the ability to demonstrate numeracy skills by reasoning using mathematical concepts, procedures, facts, and tools to solve problems related to oneself, the immediate environment, and the wider community (Kemendikbudristek RI, 2022).

In relation to numeracy, Phase D students can answer numeracy questions based on domain components aligned with the Learning Outcomes (CP) set by the government through the Kurikulum Merdeka.

DOI: <https://doi.org/10.24127/ajpm>

Table 1 summarizes several numeracy question items studies related to the analysis of

Table 1. State-of-the-Art Analysis of Numeracy Question Items

No	Research Title / Author Name (Year)	Research Design	Result
1	Development and Validation of Diagnostic Assessment Instrument for Numeracy Skills in 7th Grade / (Burgmanis et al., 2021)	Rasch Model	The diagnostic instrument is suitable for evaluating the numeracy skills of 7th-grade students.
2	Validation of a Digital Tool for Diagnosing Mathematical Proficiency / (Junpeng et al., 2020)	Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM)	The instrument is validated based on three arguments: validity, reliability, and item fit, making it suitable for use as a formative test in schools.
3	Analisis Validitas dan Realibilitas Kualitas Soal Pilihan Ganda Asesmen Kompetensi Minimum (AKM) Mata Pelajaran Pendidikan Agama Islam Menggunakan Pendekatan Model Rasch / (Nuri Syifa Azzahra et al., 2024)	Rasch Model with Quest Program	Seven out of ten multiple-choice items are appropriate, and all items are valid.
4	Analisis Soal Literasi Numerasi Menggunakan Pemodelan Rasch Konteks Pemanasan Global Berbasis ESD untuk Sekolah Dasar / (Tiara Dewi Lestari et al., 2023)	Rasch Model	From the multiple-choice questions tested, it was found that 2 questions fall into the difficult category, 8 questions fall into the moderate category, and 1 question falls into the easy category.

From the four aforementioned studies, item analysis tends to focus on the elementary school level. For Phase D, research subjects are only available in grade 7. However, the study's question material covers all levels within Phase D. Furthermore, the item analysis in this study encompasses not only one type of question but includes all question formats present in the Minimum Competency Assessment (AKM). Thus, the purpose of this study

is to examine numeracy questions for Phase D students by applying data analysis techniques with the support of the Quest program.

METHOD

In general, this research constitutes a series of item development studies utilizing a formative evaluation design. This study is at the Field Test stage, where in previous research, a

DOI: <https://doi.org/10.24127/ajpm>

numeracy item instrument was obtained, which was both valid (92%) and practical (88%) (Amelia et al.,

2023) through the stages of self-evaluation, expert review, one-to-one, and small group, as illustrated in Fig 1.

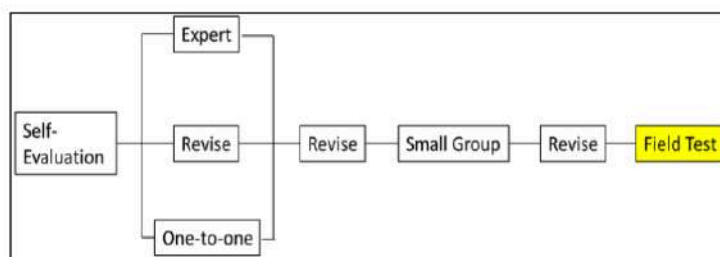


Figure 1. Desain Formative Evaluation

Specifically, this research is evaluative in nature, employing a quantitative descriptive approach. The evaluation focuses on numeracy items for Phase D students, assessing their quality through quantitative item analysis. The Quest program is used as a tool for this analysis. Items are considered to be of good quality if they meet the established criteria for item evaluation.

In analyzing items, computer programs are commonly used to facilitate the calculation process. One such program is Quest. By utilizing the Quest program, users can effectively and quickly analyze items (Ofianto Ofianto, 2021).

The advantage of this program is its capability to analyze both dichotomous and polytomous data. The program's output allows for the analysis of items from various perspectives within classical theory, such as reliability, difficulty level, discrimination, and distractor items. Additionally, it can analyze the difficulty level of the Rasch model (Fine Reffiane et al., 2021).

The numeracy questions tested quantitatively consist of 30 items, and the percentage distribution of questions based on their components can be seen in Table 2.

Table 2. The Proportion of Numeracy Items Before Analysis

Components	Subcomponents	Items	Proportion
Domain	Number	11, 12, 15, 16, 17, 18	20%
	Measurement and Geometry	4, 5, 6, 7, 26, 27	20%
	Data and Uncertainty	19, 20, 21, 24, 25, 28, 29, 30	27%
	Algebra	1, 2, 3, 8, 9, 10, 13, 14, 22, 23	33%
Context	Personal	1, 2, 3, 8, 9, 10, 11, 12, 24, 25	33%
	Socio-Cultural	13, 14, 15, 16, 17, 18	20%
	Scientific	4, 5, 6, 7, 19, 20, 21, 22, 23,	47%

Components	Subcomponents	Items	Proportion
Cognitive Level	Understanding	26, 27, 28, 29, 30 8, 9, 10, 11, 12, 19, 20, 21, 25, 28, 29, 30	40%
	Application	1, 2, 3, 4, 5, 13, 14, 22, 23	30%
	Reasoning	6, 7, 15, 16, 17, 18, 24, 26, 27	30%
Question Format	Essay	4, 5, 6, 7, 12, 26, 27	23%
	Short Answer	1, 2	7%
	Multiple Choice	11, 13, 14, 20, 21, 22	20%
	Complex Multiple Choice	3, 8, 9, 15, 16, 17, 18, 19, 24, 28, 29, 30	40%
	Matching	10, 23, 25	10%

From the Table, 30 numeracy items are presented, distributed across each component and subcomponent of numeracy questions. The proportion of the distribution of numeracy questions is aimed to approximate the proportion of AKM numeracy questions (Pusat Asesmen dan Pembelajaran, 2020).

These questions were administered to 32 Phase D students at

SMPN 34 Pekanbaru, then assessed based on their answer alternatives for subsequent analysis using the Quest program.

The output of the Quest program includes item validity estimates, difficulty level estimates, item passing estimates, and reliability estimates (Rizbudiani et al., 2021) (see Fig. 2).

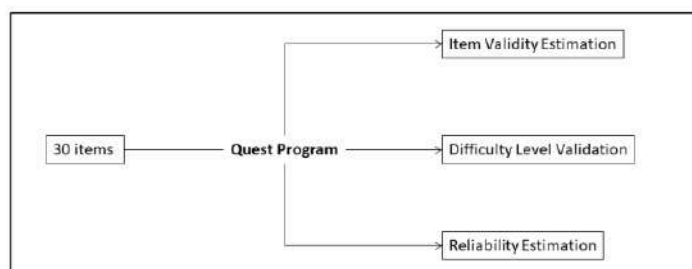


Figure 2. Scheme of the Item Analysis Process Using the Quest Program

In Item Validity Estimation, based on the Rasch Model, the validity of the analyzed items can be assessed using the INFIT MNSQ and OUTFIT t output values (Saryanto et al., 2020). An item is considered valid if the INFIT MNSQ value falls within the range of 0.5 – 1.5 (Aryadoust et al., 2021) and the OUTFIT t value is less than 2.0 (Abu

Bakar et al., 2023; Guo et al., 2020; Muslih et al., 2022).

The item estimate (Threshold) analysis can also be used to determine the difficulty level of the item. The difficulty levels are categorized as follows: 1) $b > 2$ (very difficult); 2) $1 < b \leq 2$ (difficult); 3) $-1 < b \leq 1$ (moderate); 4) $-2 < b \leq -1$

DOI: <https://doi.org/10.24127/ajpm>

(easy); and 5) $b < -2$ (very easy) (Dewi et al., 2023).

The criteria for Rasch model reliability values are as follows: 1) < 0.67 (weak); 2) $0.67 - 0.80$ (sufficient); 3) $0.81 - 0.90$ (good); 4) $0.91 - 0.94$ (very good); and > 0.94 (excellent) (Bambang Sumintono & Wahyu Widhiarso, 2015).

RESULT AND DISCUSSION

1. Item Validity Estimation

The validity results of the numeracy items based on the INFIT MNSQ and OUTFIT t values are shown in Table 3.

Table 3. Recapitulation of Numeracy Item Validity

Item	INFIT MNSQ Value	OUTFIT t Value	Interpretation	Item	INFIT MNSQ Value	OUTFIT t Value	Interpretation
1	1,33	1,2	Valid	16	0,73	-1,1	Valid
2	0,85	-0,5	Valid	17	0,76	-0,9	Valid
3	0,67	-0,8	Valid	18	-	-	Not Valid
4	0,79	-0,7	Valid	19	1,12	1,0	Valid
5	0,86	-0,7	Valid	20	1,61	3,2	Not Valid
6	0,85	-0,8	Valid	21	1,42	3,0	Not Valid
7	0,78	-0,4	Valid	22	0,97	-0,6	Valid
8	1,19	0,4	Valid	23	0,48	-2,3	Valid*
9	1,19	0,4	Valid	24	1,05	1,0	Valid
10	1,58	1,1	Valid*	25	-	-	Not Valid
11	0,81	-0,2	Valid	26	-	-	Not Valid
12	0,82	-0,7	Valid	27	-	-	Not Valid
13	-	-	Not Valid	28	1,14	1,2	Valid
14	1,32	3,1	Not Valid	29	0,60	-1,5	Valid
15	0,72	-1,1	Valid	30	-	-	Not Valid

Note: *: valid with consideration

The table above provides information about the validity of each numeracy item. The INFIT MNSQ values of the 30 numeracy items above range from $0.48 - 1.58$ and the OUTFIT t values range from -2.3 to 3.2 . This means that there are numeracy items that are not valid according to the Rasch model, namely items 10, 14, 20, 21, and 23. In the Rasch model, items that cannot be answered by all respondents are not counted and are automatically considered invalid, such as items 13, 18, 25, 26, 27, and 30.

In this analysis, it was found that some invalid items (10, 23, and 25) are matching type questions, which means

that if all of these questions are eliminated, there will be no matching type questions in this numeracy item instrument. However, to familiarize students with practicing numeracy questions in AKM format, all question components must be fulfilled.

The INFIT MNSQ scores for item 10 (1.58) and item 23 (0.48) indicate that they are less productive as measurement instruments, but they do not decrease data quality, although item 23 specifically can affect reliability scores (Boone et al., 2014). The INFIT MNSQ scores for these two items also have a small difference from the validity category threshold ($0.5 - 1.5$).

DOI: <https://doi.org/10.24127/ajpm>

Meanwhile, in terms of OUTFIT t scores, both of these items fall into the fit category.

Considering these factors, items number 10 and 23 need to be reviewed (Erfan et al., 2020) or not discarded. However, to improve the quality of these two items, minor revisions are needed, such as improving the wording

of the questions to facilitate students' understanding.

2. Difficulty Level Estimation

The analysis results of the difficulty level of numeracy questions for Phase D students can be seen in Table 4.

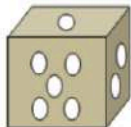
Table 4. Recapitulation of Numeracy Question Item Difficulty Levels

Item	Threshold Value	Interpretation	Item	Threshold Value	Interpretation
1	-0,74	Moderate	16	-0,65	Moderate
2	-0,20	Moderate	17	-0,84	Moderate
3	0,34	Moderate	18	-	-
4	0,44	Moderate	19	1,93	Difficult
5	0,03	Moderate	20	-0,46	Moderate *
6	2,26	Very Difficult	21	-0,46	Moderate *
7	1,33	Difficult	22	0,63	Moderate
8	-1,34	Easy	23	-0,11	Moderate
9	-1,34	Easy	24	-0,65	Moderate
10	-0,55	Moderate	25	-	-
11	-1,74	Easy	26	-	-
12	-1,07	Easy	27	-	-
13	-	-	28	1,93	Difficult
14	1,56	Difficult*	29	0,34	Moderate
15	-0,65	Moderate	30	-	-

Note: *: not valid

The table above presents information on the difficulty levels of numeracy question items. After excluding invalid items (13, 14, 18, 20, 21, 25, 26, 27, and 30), it was found that 4 (19%) items are categorized as very difficult and difficult, 13 (62%) items are moderately difficult, and 4 (19%) items fall into the easy category. This proportion of difficulty levels is considered ideal, with the number of difficult and easy questions together accounting for 19% and the remaining 62% falling into the moderate category.

Item number 6 is an essay question with a maximum score of 2. This item is categorized as very difficult, as only 3 out of 34 respondents scored 1, even though none achieved the perfect score. The numeracy component of question number 6 is an essay format, with reasoning as the cognitive level, falling under the measurement and geometry domain, and using a scientific context, as shown in Fig 3.



DICE PAINTING

The picture beside shows a cube-shaped dice with an edge length of 1 cm. Each side of the dice is painted white and features a circle with a diameter of 4 mm.

Question 6: Dice Painting
Find the surface area of the dice that is not white on the 1 and 6-edged sides!

Figure 3. Question Item Number 6 (Very Difficult Category)

This question item involves reasoning, and during the validation stage with experts, the validators recommended increasing its cognitive level compared to the previous question design. Therefore, at this estimation stage, item 6 has not been removed.

3. Reliability Estimation

The reliability of item estimate for multiple-choice questions is 0.92 (very good), and the respondent reliability is 0.81 (good). This means that the respondent reliability is lower than the

item reliability. This can occur for several reasons, including respondents answering questions carelessly and the sample size being less than 100 respondents, specifically 32 respondents.

From the results of the three estimations above, 21 valid items were obtained with varying difficulty levels and very good item reliability. The distribution of these 21 numeracy items for Phase D students is presented in the Table 5.

Table 5. Proportion of Numeracy Question Items After Analysing

Components	Subcomponents	Items	Proportion
Domain	Number	11, 12, 15, 16, 17	24%
	Measurement and Geometry	4, 5, 6, 7	19%
	Data dan Uncertainty	19, 24, 28, 29	19%
	Algebra	1, 2, 3, 8, 9, 10, 22, 23	38%
Context	Personal	1, 2, 3, 8, 9, 10, 11, 12, 24	43%
	Socio-Cultural	15, 16, 17	14%
	Scientific	4, 5, 6, 7, 19, 22, 23, 28, 29	43%
Cognitive Level	Understanding	8, 9, 10, 11, 12, 19, 28, 29	38%
	Application	1, 2, 3, 4, 5, 22, 23	33%
	Reasoning	6, 7, 15, 16, 17, 24	29%
Question Format	Essay	4, 5, 6, 7, 12	24%
	Short Answer	1, 2	9,3%
	Multiple Choice	11, 22	9,3%
	Complex Multiple Choice	3, 8, 9, 15, 16, 17, 19, 24, 28, 29	48%
	Matching	10, 23	9,3%

DOI: <https://doi.org/10.24127/ajpm>

8
From the table above, it can be seen that the distribution of numeracy questions for Phase D students changed after several items were dropped following analysis with the Quest program.

This research found that nine discarded consisted of Essay (28%), Multiple Choice (66,7%), Complex Multiple Choice (16,7%), and Matching (33,3%) question types. Multiple-choice questions were the most frequently rejected due to their lack of validity.

In this study, questions that students were unable to answer could not be classified as difficult, moderate, or easy. Additionally, item validity, whether based on INFIT MNSQ or OUTFIT t, was not interconnected with difficulty levels. This finding aligns with previous research (Nurhalimah et al., 2022; Van Vo & Csapó, 2021). In other words, invalid questions cannot be categorized as difficult, moderate, or easy.

The use of the Quest Program is relatively simple, as it provides readily available command templates. Users only need to input data into the lightweight application. However, Quest has a limitation in that its reliability calculations apply only to multiple-choice questions. This opens opportunities for future researchers to combine Quest with other formulas to obtain reliability values for all question types.

If the development of numeracy questions has the benefit of improving their quality in terms of fulfilling curriculum standards and assessing the validity and practicality of the questions, then the analysis of numeracy items is a follow-up activity to evaluate the potential effects through the validity and reliability of the questions. Item analysis is a simple yet valuable activity

for teachers in providing questions as an evaluation tool to their students (Kumar et al., 2021). It also strengthens scientific decisions based on the quantitative analysis of the questions' level of difficulty. Thus, integrating both qualitative and quantitative approaches, educators can create a collection of numeracy questions that are truly valid, practical, and effective for classroom use. These process of item quality control is essential for test development (Quagrains & Arhin, 2017).

This numeracy test instrument can be further developed to assess the numeracy skills of students in Phase D and analyze their difficulties in solving numeracy questions.

CONCLUSIONS

7 The Quest program, which is part of the Rasch model, helps in analyzing the validity and reliability of numeracy questions. Additionally, the Quest program provides an overview of the difficulty level of the questions, making it easier to make scientific decisions to produce good numeracy questions for Phase D learners.

After being analyzed using the Quest program, this study produced 21 valid numeracy items with an ideal level of difficulty (19% difficult questions, 62% medium questions, and 19% easy questions), as well as a very good level of reliability (0.92). These 21 items not only fulfill curriculum standards but can also be used accurately at the junior high school level or Phase D as diagnostic, formative, and summative assessments to measure students' numeracy skills.

ACKNOWLEDGEMENTS

Thanks to DRTPM Dikti for improving this article, and DPPM UIR

DOI: <https://doi.org/10.24127/ajpm>

as research sponsor with contract number: 486/KONTRAK/P-PT/DPPM-UIR/06-2023.

REFERENCES

- Abu Bakar, S. N., Mahmood, N., Ismail, A., Hashim, A., Hamsan, N. Y., Nor Hamid, N. I., & Che Yakzam, R. (2023). Validity and Reliability of Competency Analysis Instrument for Cooperative Board Members using Rasch Measurement Model Approach. *International Journal of Academic Research in Business and Social Sciences*, 13(6), Pages 2408-2419. <https://doi.org/10.6007/IJARBS/v13-i6/17543>
- Amelia, S., Widiati, I., & Yadraka, G. (2023). PENGEMBANGAN SOAL NUMERASI UNTUK PESERTA DIDIK FASE D. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(3), 3048. <https://doi.org/10.24127/ajpm.v12i3.7236>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6–40. <https://doi.org/10.1177/0265532220927487>
- Bambang Sumintono & Wahyu Widhiarso. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.
- Boone, W. J., Staver, J. R., & Yalc, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands.
- <https://doi.org/10.1007/978-94-007-6857-4>
- Burgmanis, G., France, I., Namsone, D., & Čakāne, L. (2021). *DEVELOPMENT AND VALIDATION OF DIAGNOSTIC ASSESSMENT INSTRUMENT FOR NUMERACY SKILLS IN 7TH GRADE*. 7781–7791. <https://doi.org/10.21125/iceri.2021.1747>
- Dewi, H. H., Damio, S. M., & Sukarno, S. (2023). Item analysis of reading comprehension questions for English proficiency test using Rasch model. *REID (Research and Evaluation in Education)*, 9(1), 24–36. <https://doi.org/10.21831/reid.v9i1.53514>
- Erfan, M., Maulida, 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. <https://doi.org/10.33292/petier.v3i1.51>
- Fine Reffiane, Sudarmin, Wiyanto, & Sigit Saptono. (2021). The instrument analysis of students' problem-solving ability on hybrid learning model using ETNO-STEM Approach through Quest Program in COVID-19 Pandemic. *Pegem Journal of Education and Instruction*, 11(4), 1–8. <https://doi.org/10.47750/pegegog.11.04.01>
- Guo, C., Liu, Y., Hao, S., Xie, L., Xiang, G., Wu, Y., & Li, S. (2020). The Reliability and

DOI: <https://doi.org/10.24127/ajpm>

- Validity of the “Activity and Participation” Component in the Brief ICF Core Set for Chronic Obstructive Pulmonary Diseases Based on Rasch Analysis. *International Journal of Chronic Obstructive Pulmonary Disease*, Volume 15, 1191–1198. <https://doi.org/10.2147/COPD.S249704>.
- Ismawati, E., Hersulastuti, H., Amertawengrum, I. P., & Anindita, K. A. (2023). Portrait of Education in Indonesia: Learning from PISA Results 2015 to Present. *International Journal of Learning, Teaching and Educational Research*, 22(1), 321–340. <https://doi.org/10.26803/ijlter.22.1.18>.
- Junpeng, P., Marwiang, M., Chinjunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., & Wilson, M. (2020). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education (IJERE)*, 9(3), 665. <https://doi.org/10.11591/ijere.v9i3.20503>.
- Kemendikbudristek RI. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 5 Tahun 2022 tentang Standar Kompetensi Lulusan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*.
- Kholid, M. N., Rofi'ah, F., Ishartono, N., Waluyo, M., Maharani, S., Swastika, A., Faiziyah, N., & Sari, C. K. (2022). What Are Students' Difficulties in Implementing Mathematical Literacy Skills for Solving PISA-Like Problem? *Journal of Higher Education Theory and Practice*, 22(2), 181–200.
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85–S89. <https://doi.org/10.1016/j.mjafi.2020.11.007>.
- Muslih, H. Y., Suryana, D., Universitas Pendidikan Indonesia, Indonesia, dodisuryana@upi.edu, Ahman, A., Universitas Pendidikan Indonesia, Indonesia, ahman@upi.edu, Suherman, U., Universitas Pendidikan Indonesia, Indonesia, umans@upi.edu, Dahlan, T. H., & Universitas Pendidikan Indonesia, Indonesia, tinadahlan_psi@upi.edu. (2022). Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using Rasch Model. *International Journal of Instruction*, 15(2), 207–222. <https://doi.org/10.29333/iji.2022.15212a>.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). HUBUNGAN ANTARA VALIDITAS ITEM DENGAN DAYA PEMBEDA DAN TINGKAT KESUKARAN SOAL PILIHAN GANDA PAS. *Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>.
- Nuri Syifa Azzahra, Sri Sumarni, & Himawan Putranta. (2024). Analisis Validitas dan Realibilitas

DOI: <https://doi.org/10.24127/ajpm>

- Kualitas Soal Pilihan Ganda Asesmen Kompetensi Minimum (AKM) Mata Pelajaran Pendidikan Agama Islam Menggunakan Pendekatan Model Rasch. *QuranicEdu: Journal of Islamic Education*, 4(1), 94. <https://doi.org/10.37252/quranicedu.v4i1.681>
- Ofianto Ofianto. (2021). Analysis Of Instrument Test Of Historical Thinking Skills In Senior High School History Learning With Quest Programs. *Indonesian Journal of History Education*, 6(2), 184–192.
- Parnis, A. J., & Petocz, P. (2016). Secondary school students' attitudes towards numeracy: An Australian investigation based on the National Assessment Program—Literacy and Numeracy (NAPLAN). *The Australian Educational Researcher*, 43(5), 551–566. <https://doi.org/10.1007/s13384-016-0218-3>
- Pusat Asesmen dan Pembelajaran. (2020). *Desain Pengembangan Soal AKM*. Jakarta: Balitbang Kemdikbud.
- Quagrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1301013. <https://doi.org/10.1080/2331186X.2017.1301013>
- Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar: Jurnal Pendidikan Matematika*, 12(2), 399–412. <https://doi.org/10.24042/ajpm.v12i2.9939>
- Rosnelli, R., & Ristiana, P. A. (2023). Independent Curriculum Learning Management to Improve Students' Literacy and Numerical Competence in Schools. *International Journal of Education in Mathematics, Science and Technology*, 11(4), 946–963. <https://doi.org/10.46328/ijemst.3513>
- Saryanto, S., Sumiharsono, R., Ramadhan, S., & Suprpto, E. (2020). The Analysis of Instrument Quality to Measure the Students Higher Order Thinking Skill in Physics Learning. *Turkish Journal of Science Education*, 17(4), 520–527. <https://doi.org/10.36681/tused.2020.42>
- Tiara Dewi Lestari, Ghullam Hamdu, & Erwin Rahayu Saputra. (2023). Analisis Soal Literasi Numerasi Menggunakan Pemodelan Rasch Konteks Pemanasan Global Berbasis ESD untuk Sekolah Dasar. *Pendas: Jurnal Ilmu Pendidikan Dasar*, 8(2), 2489–2503. <https://doi.org/10.23969/jp.v8i2.9270>
- Van Vo, D., & Csapó, B. (2021). Development of scientific reasoning test measuring control of variables strategy in physics for high school students: Evidence of validity and latent predictors of item difficulty. *International Journal of Science Education*,

DOI: <https://doi.org/10.24127/ajpm>

43(13), 2185–2205.
<https://doi.org/10.1080/09500693.2021.1957515>

Yustitia, V., Kusmaharti, D., & Wardani, I. S. (2025). Students' critical thinking in numeracy problem-solving through moderate self-Efficacy: A mixed-methods study. *Multidisciplinary Science Journal*, 7(8), 2025410. <https://doi.org/10.31893/multiscience.2025410>

ORIGINALITY REPORT

12%	10%	11%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to UW, Stevens Point Student Paper	8%
2	Sindi Amelia, Indah Widiati, Gusri Yadrika. "PENGEMBANGAN SOAL NUMERASI UNTUK PESERTA DIDIK FASE D", AKSIOMA: Jurnal Program Studi Pendidikan Matematika, 2023 Publication	1%
3	doaj.org Internet Source	1%
4	Submitted to Universitas Pendidikan Indonesia Student Paper	1%
5	Eny Cahyaningsih, Iva Sarifah, Riyadi Riyadi. "Harmonizing technology and tradition: Analysis of grade VI mathematics midterm exam questions", Jurnal Gantang, 2024 Publication	1%
6	repository.uhamka.ac.id Internet Source	1%
7	Nabila Lutfiati, Himawan Putranta. "Analysis of the Quality of the Faraid Material AKM Instrument to Measure Students' Reasoning Ability with Assistance Ministep", ej, 2025 Publication	1%
8	mafiadoc.com	

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		