**EMTEKA**: Jurnal Pendidikan Matematika
Volume 7, No. 2, 2026. 933-949
DOI: https://doi.org/10.24127/emteka.v7i2.12604

e-ISSN 2746-5594
p-ISSN 2746-5608

# SOLO TAXONOMY AND RASCH-BASED NEEDS ANALYSIS FOR MATHEMATICS LITERACY ASSESSMENT: TEACHERS' AND STUDENTS' PERSPECTIVES

Viona Yuliza[1*], Syamsir Sainuddin[2], Ilham Falani[3]

[1*,2,3] Universitas Jambi, Jambi, Indonesia
** Corresponding author. Jl. Arif Rahman Hakim, Kec. Telanaipura, 36361, Jambi, Indonesia*

*E-mail:* vionayuliza03@gmail.com[1*]
syamsirsainuddin@unja.ac.id [2]
ilhamfalani@unja.ac.id [3]

Received 25 August 2025; Received in revised form 28 June2 026; Accepted 09 September 2026

**ABSTRACT**
Indonesian students' mathematical literacy remains low (PISA 2022 score = 366), partly because school assessments are dominated by procedural items lacking graded cognitive complexity. This study analyzes the need for a SOLO Taxonomy-based mathematics literacy instrument from combined teacher and student perspectives, an empirical foundation not offered by single-stakeholder studies. An Exploratory Sequential Mixed-Methods design used three instruments, a teacher questionnaire (N = 3), a student questionnaire (N = 200), and an 8-item PISA-referenced diagnostic test (N = 200), in Jambi City and Muaro Jambi. Questionnaires were analyzed descriptively; the diagnostic test used the Dichotomous Rasch Model via Joint Maximum Likelihood Estimation. Teachers showed a knowledge-practice gap (highest dimension: need for new instruments, M = 3.24; lowest: item-construction practice, M = 2.92, scale 1-4). Rasch calibration showed strong item separation (Reliability = 0.99, Separation = 9.13) and acceptable person separation for an 8-item instrument (Reliability = 0.52), no item misfit, no significant gender/domicile DIF, and confirmed 87.5% of students (175/200) at PISA Levels 1-3, mean θ = -1.17 logit. Joint Display integration converged on three needs: SOLO-graded items, PISA-aligned real-world contexts, and teacher professional development. Unlike prior PISA- or SOLO-based instruments developed separately, this is the first study to integrate teacher/student needs data with Rasch-calibrated diagnostic evidence into a single empirical foundation for SOLO-based literacy instrument development in Indonesian junior high schools.
**Keywords**: mathematics assessment; mathematical literacy; rasch model; SOLO taxonomy

***ABSTRAK***
*Penelitian Literasi matematika siswa Indonesia masih rendah (skor PISA 2022 = 366), sebagian karena penilaian di sekolah didominasi oleh soal-soal prosedural yang kurang mencerminkan kompleksitas kognitif berjenjang. Penelitian ini menganalisis kebutuhan akan instrumen literasi matematika berbasis Taksonomi SOLO dari perspektif gabungan guru dan siswa, sebuah landasan empiris yang tidak ditawarkan oleh penelitian yang hanya melibatkan satu pemangku kepentingan. Desain Eksploratif Sekuensial Metode Campuran menggunakan tiga instrumen, yaitu kuesioner guru (N = 3), kuesioner siswa (N = 200), dan tes diagnostik berbasis PISA yang terdiri dari 8 soal (N = 200), di Kota Jambi dan Muaro Jambi. Kuesioner dianalisis secara deskriptif; sedangkan tes diagnostik menggunakan Model Rasch Dikotomis melalui Estimasi Maksimum Likelihood Bersama. Guru menunjukkan kesenjangan antara pengetahuan dan praktik (dimensi tertinggi: kebutuhan akan instrumen baru, M = 3,24; terendah: praktik penyusunan item, M = 2,92, skala 1–4). Kalibrasi Rasch menunjukkan pemisahan item yang kuat (Reliabilitas = 0,99, Pemisahan = 9,13) dan pemisahan antarindividu yang dapat diterima untuk instrumen berisikan 8 item (Kekuatan Reliabilitas = 0,52), tanpa misfit butir maupun DIF gender/domisili, serta mengonfirmasi 87,5% siswa (175/200) berada pada Level PISA 1–3, dengan rata-rata θ = –1,17 logit. Integrasi Joint Display menghasilkan tiga kebutuhan konvergen: soal berjenjang berbasis SOLO, konteks nyata sesuai PISA, dan pengembangan professional guru. Berbeda dari instrument berbasis PISA atau SOLO yang dikembangkan terpisah pada penelitian*

**EMTEKA**: Jurnal Pendidikan Matematika
Volume 7, No. 2, 2026. 933-949
DOI: https://doi.org/10.24127/emteka.v7i2.12604

e-ISSN 2746-5594
p-ISSN 2746-5608

*terdahulu, studi ini pertama kali mengintegrasikan data kebutuhan guru-siswa dengan bukti diagnostic terkalibrasi Rasch sebagai landasan empiris tunggal pengembangan instrument literasi matematika berbasis SOLO di SMP Indonesia.*
***Kata kunci**:* asesmen matematika; literasi matematika; rasch model; taksonomi SOLO

This is an open access article under the Creative Commons Attribution 4.0 International License

## Introduction

Mathematical literacy is a fundamental ability that enables individuals to formulate, apply, and interpret mathematics in various real-life contexts, as well as to make rational and constructive decisions (OECD, 2023a). Within the framework of the Programme for International Student Assessment (PISA), mathematical literacy is defined as an individual's capacity to reason mathematically and use mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena (OECD, 2023). This ability is not only relevant for academic purposes but is also essential for preparing the younger generation to face the increasingly complex challenges of the 21st century.

However, results from international assessments indicate that Indonesian students' mathematical literacy remains at an alarming level. According to the 2022 PISA results released in December 2023, Indonesia achieved an average score of 366 points in the mathematics domain, compared to the OECD average of 472 points. This score of 366 places Indonesia 67th out of 81 participating countries (OECD, 2023b). This is the lowest score since 2006 and reflects a 13-point decline compared to the 2018 PISA results (OECD, 2023b). Even more concerning is that only 18% of Indonesian students were able to achieve at least Level 2 proficiency in mathematics, far below the OECD average of 69% (OECD, 2023b). Almost no Indonesian students have managed to reach Level 5 or Level 6, which require the ability to mathematically model complex situations and evaluate problem-solving strategies (OECD, 2023b).

The low level of mathematical literacy among Indonesian students cannot be separated from the quality and characteristics of the assessment instruments used in schools. Several studies indicate that mathematics assessment practices in Junior High Schools (SMP) in Indonesia are still dominated by procedural and routine questions that require the memorization of algorithms, without integrating demands for higher-order thinking skills (HOTS) and more relevant real-life contexts (Stacey, 2011; Nusantara et al., 2021). This situation creates a significant gap between what is measured through classroom assessments and what is actually required by the mathematics literacy framework.

This gap indicates an urgent need to develop valid and reliable mathematics literacy assessment instruments capable of measuring students' abilities at different levels. One relevant taxonomic framework that has yet to be widely integrated into the development of mathematics literacy assessment instruments in Indonesia is the SOLO Taxonomy (Structure of Observed Learning Outcomes). The SOLO Taxonomy provides a hierarchy of learning responses ranging from the Pre-structural (L1) to Extended Abstract (L5) levels, which conceptually aligns with the hierarchy of abilities within the PISA framework from Level 1 to Level 6 (Biggs & Collis, 1982; Dong & Zhang, 2024).

e-ISSN 2746-5594
p-ISSN 2746-5608

Previous research on the development of mathematics literacy assessment instruments has been conducted by several researchers. Aisyah & Juandi (2022) developed PISA-based mathematics items aimed at measuring the mathematics literacy skills of junior high school students. However, the instrument produced by Aisyah & Juandi (2022) has not integrated the SOLO Taxonomy framework as the basis for the hierarchy of skill measurement. Similarly, Nusantara et al. (2021) examined PISA-based contextual mathematics questions at the secondary school level and found that most of the questions used by teachers were only at the low to intermediate levels of the PISA framework, without systematic mapping to cognitive taxonomy.

In the field of SOLO Taxonomy-based instrument development, Yuanita et al. (2023) successfully developed a SOLO-based assessment instrument aimed at measuring students' mathematical thinking skills. However, this instrument has not yet been validated using the Rasch Model, which provides stricter psychometric measurement standards and can produce a true interval scale. Research using the Rasch Model as a method for validating mathematics instruments generally focuses only on questions based on the conventional curriculum, not on SOLO Taxonomy-based mathematical literacy assessment instruments (Sumintono & Widhiarso, 2015).

Thus, based on the above discussion, three research gaps were identified as the foundation for this study. These three gaps are: (1) The lack of a mathematics literacy assessment instrument that explicitly integrates the PISA framework with the SOLO Taxonomy; (2) The lack of instrument validation using the Rasch Model to produce stable item and person parameter estimates; and (3) No needs analysis simultaneously integrates teacher and student perspecties within one framework. This study's novelty lies in closing all three at once: to the authors' knowledge, it is the first Indonesian needs-analysis study to triangulate teacher assessment-practice data, student self-report data, and a Rasch-calibrated PISA-referenced diagnostic test into a single Joint Display, producing an empirical rather than purely theoretical foundation for SOLO-based instrument development.

Based on the background and previous research studies presented, this study aims to analyze the need for the development of a mathematics literacy assessment instrument based on the SOLO Taxonomy through a study of the perspectives of junior high school teachers and students in the cities of Jambi and Muaro Jambi. Specifically, this study aims to: (1) Describe teachers' mathematics assessment practices in terms of item characteristics based on the SOLO Taxonomy and the dimensions of the mathematics literacy process; (2) Identify teachers' perceptions and needs regarding the development of mathematics literacy assessment instruments; (3) Map students' mathematical literacy ability profiles based on PISA-based diagnostic tests analyzed using the Rasch Model; and (4) Integrate qualitative and quantitative findings as an empirical foundation for the development of SOLO Taxonomy-based mathematical literacy assessment instruments.

## Research Methods

*Research Design*

This study is a mixed-methods study employing an exploratory sequential design (Creswell & Clark, 2018). This design was chosen because of the primary phenomenon under investigation: the need to develop a mathematics literacy assessment instrument based on the SOLO Taxonomy. This phenomenon requires in-depth qualitative exploration of teachers' practices and perspectives before quantitative patterns can be

e-ISSN 2746-5594
p-ISSN 2746-5608

meaningfully measured and interpreted. The research proceeds through six stages, as ahown in Figure 1 below:

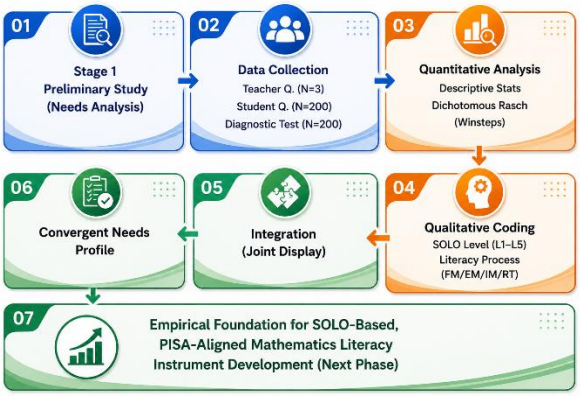



Figure 1. Exploratory Sequential Mixed-Methods Research Design Flowchart

Based on Figure 1 above, the research proceeds from an initial needs analysis through parallel data collection, quantitative and qualitative analysis, integration via Joint Display, and finally the formulation of a convergent needs profile that serves as the empirical foundation for the next instrument-development phase.

*Participants*

This study was conducted at public junior high schools (SMP) in the cities of Jambi and Muaro Jambi during the second semester of the 2025/2026 academic year. The selection of these locations was based on preliminary observations that identified that mathematics assessment practices in these schools were still dominated by the use of procedural and routine problems. This finding was further supported by the limited integration of higher-order thinking skills and real-life contexts aligned with the mathematical literacy framework. The subjects in this study were selected using purposive sampling and consisted of two main groups, namely: (1) three mathematics teachers actively teaching at SMPN Kota Jambi and/or Muaro Jambi who were willing to participate in the study; (2) 200 students in grades VII (36.5%), VIII (22.0%), and IX (41.5%). The determination of the student sample size was based on the minimum threshold recommended for Rasch Model analysis to produce stable item and person parameter estimates (Linacre, 1994).

*Instruments*

Data collection in this study was conducted using five instruments that had been theoretically validated by experts and implemented in a coordinated manner, as presented in Table 1:

Table 1. Data Collection Instruments

| Instrument | Target | Number of Items | Technique | Analysis |
|---|---|---|---|---|
| **Teacher Questionnaire** | Math Teachers | 29 Closed-ended (scale 1-4) + 4 Open-ended | Google Form | Descriptive Statistics |
| **Semi-Structured Interview** | Math Teachers | 15 Key Questions | Face-to-Face/Voice Recording | Content Analysis |

e-ISSN 2746-5594
p-ISSN 2746-5608

| Instrument | Target | Number of Items | Technique | Analysis |
|---|---|---|---|---|
| **Test Item Analysis** | Teacher-Created Midterm/Final Exam Questions | All Available Test Items | Documentation | Content Analysis |
| **Student Questionnaire** | 7th, 8th, and 9th Grade Students | 25 Closed-Ended (scale 1-4) + 3 Open-Ended | Google Form | Descriptive Statistics |
| **Diagnostic Test** | Grade 7, 8, and 9 Students | 8 items, PISA Level 1c–6 Questions | Written Test (60 Minutes) | Descriptive Statistics |

Table 1 presents details of the data collection instruments used in this study. Based on Table 1, it is evident that several data analyses will be conducted, encompassing both qualitative and quantitative data. Data collection techniques are also categorized into several criteria, including the use of Google Forms, voice recording, and written tests to assess students' mathematical literacy outcomes.

In this research, validity evidence differs by instrument. For the diagnostic test, content and construct validity are inherited from PISA's own official item development and international field-trial process (OECD, 2023a). Since all eight items are levels-preserving translations of released PISA items rather than newly authored questions; no additional local index is required. For the teacher and student questionnaires (D1-D4), authored specificall for this study, content validity was established through qualitative expert judgment by two validators, who assessed item relevance and clarity regarding teachers' instructional practices and needs, a descriptive rather than right/wrong construct, and judged the instruments valid and usable after minor revisions. A quantitative CVI/Aiken's V was not computed for these two questionnaires (see Limitation).

*Data Analysis*

Data analysis was conducted through three interconnected stages: qualitative analysis, quantitative analysis, and integration. These three stages are described in more detail as follows:

1. Qualitative Stage

This stage draws on teacher interview transcripts and coding of teacher-created test-item documents (Table 1), analyzed via content analysis (Miles et al., 2014). Each teacher-created item will be coded against SOLO levels (L1-L5) and the literacy-process dimension: Formulate (FM), Employ (EM), Interpret (IM), and Routinize (RT), by two independent coders, with Cohen's Kappa establishing inter-rater reliability. As noted in Limitation, this coding pahse had not yet been completed at this preliminary stage and will occur in the main research phase; it is distinct from the diagnostic test's PISA/SOLO classification, which follows the official PISA framework directly.

2. Quantitative Phase

Questionnaire responses (1-4 Likert) were analyzed via descriptive statistics (mean, SD, percentage per item). Dianostic-test items were scored dichotomously (0-1) and calibrated with the Dichotomous Rasch Model via Joint Maximum Likelihood Estimation (JMLE). Following standard practice Bond et al., (2021), model quality was evaluated

e-ISSN 2746-5594
p-ISSN 2746-5608

using infit/outfit mean-square (acceptable range 0.5-1.5), Person/Item Reliability and Separation, Point-Measure Correlation (PTMEA), unidimensionality via PCA of standardixed residuals, and Differential Item Functioning (DIF) by gender and domicile, estimated by comparing item difficulties calibrated separately per group (|t|>2.0 flagged as potential DIF). Results appear in the Diagnostic Test Results subsection. Unfavorably worded items were reverse-scored before analysis.

3. Integration Phase

This stage combines quantitative findings (teacher/student questionnaires, diagnostic test) with qualitative findings (teacher interviews, test-document analysis) via a Joint isplay procedure (Creswell & Clark, 2018), producing a convergent need profile as the empirical foundation for the next instrument-development phase.

This study was conducted under a research permission letter issued by the authors' institution, obtained prior to data collection at the participating schools.

## Results and Discussion

This section presents the results of the data analysis from the three quantitative instruments used in the preliminary study: the teacher questionnaire, the student questionnaire, and the PISA-based diagnostic test of mathematical literacy. The presentation of results follows the Exploratory Sequential Mixed-Methods approach. Creswell & Clark (2018) describe how the Exploratory Sequential Mixed-Methods approach works, wherein: each sub-result is presen ted alongside its discussion, and the section concludes with the integration of all findings into a profile of instrument development needs.

*Teacher Questionnaire Results: Perceptions and Needs Regarding Mathematics Literacy Assessment The teacher questionnaire consists of 29 items*

The teacher questionnaire comprises 29 items (1-4) across four dimensions (D1-D4), completed by three teachers. Six items (B3, B9, B12, B18, B24, B27) were reverse-scored. Descriptive statistics are presented in Table 2 below:

Table 2. Descriptive Statistics of the Teacher Questionnaire by Dimension ($\boldsymbol{N = 3}$)

| Kode | Dimension | Items | Number | Mean (Raw Score) | SD | % ≥ 3 | Interpretation |
|---|---|---|---|---|---|---|---|
| **D1** | Understanding of Mathematical Literacy | B1-B7 | 7 | 3.048 | 1.024 | 81.0% | High |
| **D2** | Item-Construction Practice | B8-B15 | 8 | 2.917 | 0.830 | 70.8% | Moderate |
| **D3** | Need for New Instruments | B16-B22 | 7 | 3.238 | 0.768 | 90.5% | High |
| **D4** | Real-World Context Integration | B23-B29 | 7 | 2.905 | 0.944 | 71.4% | Moderate |
| | **Overall** | **B1-B29** | **29** | **3.023** | **0.889** | **78.4%** | **Moderate-High** |

Based on Table 2 above, Need for New Instruments (D3) highest (M=3.238, 90.5% ≥ 3): teachers consistently perceive a strong need for new, better-graded

e-ISSN 2746-5594
p-ISSN 2746-5608

instruments (highest item B16, M=3.67), while the lowest item (B18, reversed, M=2.33) suggests uncertainty about how such instruments should be structured, precisely the gap this needs analysis addresses.

Understanding of Mathematics Literacy (D1) was second-highest (M=3.048, 81.0% ≥ 3), indicating reasonably strong conceptual understanding, consistent with Fitriati et al., (2024), who found teachers' main HOTS challenge lies in practical assessment implementation rather than conceptual understanding, anticipating the gap below.

Item-Contruction Practice (D2, M=2.917) and Real-World Context Integration (D4, M=2.905) scored lowest, confirming a knowledge-practice gap: teachers understand literacy concepts (D1) and want better instruments (D3), but their item-construction practice (D2) and contextual-integration ability (D4) lag behind. Edwar et al., (2023), found HOTS-untrained teachers' competency at only 45.70% pre-workshop, rising to 85.70% post-training, evidence the gap is trainable. Eka et al., (2026), found a similar pattern in West Kalimantan, and Susanta et al., (2023), dan Falani et al., (2025), confirm Indonesian teachers generally struggle to construct contectual items, consistent with low D2/D4 scores here.

*Student Questionnaire Results: Students' Perceptions of Mathematics Learning and Assessment*

The student questionnaire comprised 25 items (1-4 Likert, seven reverse-scored: B2, B5, B9, B14, B17, B21, B24), completed by 200 students (overall M=2.854, 67.9% ≥ 3). Statistics by dimension are presented in Table 3 below:

Table 3. Descriptive Statistics of the Student Questionnaire by Dimension ($\boldsymbol{N = 200}$)

| Kode | Dimension | Items | Number | Mean | SD | % ≥ 3 | Interpretation |
|---|---|---|---|---|---|---|---|
| **D1** | Perception of Math Problems | B1-B6 | 6 | 2.836 | 0.892 | 66.0% | Moderate |
| **D2** | Experience with Contextual Problems | B7-B12 | 6 | 2.908 | 0.864 | 71.0% | Moderate |
| **D3** | Motivation and Engagement | B13-B18 | 6 | 2.859 | 0.867 | 68.2% | Moderate |
| **D4** | Need for Varied Questions | B19-B25 | 7 | 2.817 | 0.868 | 66.6% | Moderate |
| | **Overall** | **B1-B25** | **25** | **2.854** | **0.873** | **67.9%** | **Moderate** |

Based on Table 3 above, Experience with Contextual Problems (D2, M=2.908) scored highest, with its strongest item (B11, M=3.04) indicating some exposure to contextual problems, while its weakest (B9, reversed, M=2.50) shows persistent difficulty as context complexity rises, consistent with OECD (2023b), which reports only about a third of Indonesian students regularly encounter real-life application problems.

Motivation and Engagement (D3, M=2.859) was second-highest; its strongest item (B15, M=3.08) suggests a meaningful base of engagement not yet leveraged by current instruments, while its weakest (B17, reversed, M=2.48) shows engagement drops

e-ISSN 2746-5594
p-ISSN 2746-5608

when tasks feel too difficult, consistent with Bolstad (2023), on how meaningful, contextual problems shape student' orientation toward mathematics.

Perception of Math Problems (D1, M=2.836) and Need for Varied Questions (D4, M=2.817) scored lowest and close together, indicating only moderate overall perception of current problems and an unmet desire for more varied, appropriately challenging questions. Aisyah & Juandi (2022), similarly found moderate-ability students generally reach only PISA Level 3, and Yulia et al., (2021), dound only high-ability students solve problems requiring contextual reasoning, consistent with the fairly uniform, moderate pattern across all four dimensions here.

Harisman et al., (2023), found learning habits and exposure to varied problems significant predictors of literacy skill, helping explain ehy D1 and D4, most tied to the range and perceived relevance of practiced problems, are weakest here. Combined with the diagnostic results below, this indicates students are not merely unable to solve higher-level problems; they also do not yet perceive current materials as sufficiently varied or engaging, reinforcing the case for a redesigned, SOLO-graded instrument.

*Diagnostic Test Results: Profile of Students' Mathematical Literacy Skills*

The diagnostic test comprised 8 dichotomously scored PISA-referenced items (Levels 1c-6, from the 2017/2022 frameworks), analyzed with the Dichotomous Rasch Model via JMLE (Wright & Stone, 1979; Bond et al., 2021). Rasch analysis provides a true interval scale on which ability and difficulty are directly comparable (Sumintono & Widhiarso, 2015). Item-level results appear in Table 4.

The eight items are level-preserving translations of released PISA 2017/2022 items, one per target difficulty band (1a, 1b, 1c, 2-6), preserving each item's official proficiency level and process category (Formulate, Employ, Interpret) while re-contextualizing surface content into settings familiar to Indonesian students (e.g., local units, school and mobile-technology scenarios) rather than the original PISA context verbatim. Because items retain their official classification, validity is inherited from OECD's international field-trial process (OECD, 2023a). Rather than requiring local expert judgment. Each item's SOLO level (Table 4) follows the established PISA-SOLO correspondence (Biggs & Collis, 1982; Dong & Zhang, 2024). As nly one item was retained per band at this preliminary stage, per-level ability estimates remain indicative; expanding to at least three items per band is recommended for the next phase (see Limitation). Item-level results are presented in Table 4 below:

Table 4. Item Achievement Levels for the Diagnostic Test ($\boldsymbol{N = 200}$)

| Kode Item | PISA Level | SOLO Level | Question Context | Correct (n) | % Correct | Difficulty (logit) | Category |
|---|---|---|---|---|---|---|---|
| **P1** | Level 1c | L1-L2 | Perimeter | 153 | 76.5% | -3.2 | Very Easy |
| **P2** | Level 1b | L1-L2 | Cubes & Prisms | 135 | 67.5% | -2.5 | Easy |
| **P3** | Level 1a | L2 | Calendar/Time | 117 | 58.5% | -1.9 | Somewhat Easy |
| **P4** | Level 2 | L2-L3 | Smartphone Data | 80 | 40.0% | -0.8 | Moderate |
| **P5** | Level 3 | L3 | Beauty of Powers | 38 | 19.0% | 0.7 | Somewhat Difficult |
| **P6** | Level 4 | L4 | Tiling Patterns | 26 | 13.0% | 1.4 | Difficult |

e-ISSN 2746-5594
p-ISSN 2746-5608

| Kode Item | PISA Level | SOLO Level | Question Context | Correct (n) | % Correct | Difficulty (logit) | Category |
|---|---|---|---|---|---|---|---|
| **P7** | Level 5 | L4-L5 | Power of 7 | 8 | 4.0% | 3.0 | Very Difficult |
| **P8** | Level 6 | L5 | Diagonal Navigation | 6 | 3.0% | 3.3 | Very Difficult |
| **Avarage** | | | | **2.82/ Student** | **35.2%** | **-1.17 (θ, n=188 estimable)** | |

Based on Table 4 above, correct-response rates decline steadily from Level 1c to Level 6, indicating the instrument successfully spans a wide range of item difficulty consistent with its intended PISA/SOLO hierarchy. Rasch model quality indicators for this instrument are presented in Table 5 below:

Table 5. Rasch Model Quality Indicators fot the Diagnostic Test (N=200, 8 items)

| Indicator | Value | Interpretation |
|---|---|---|
| Item Reliability | 0.99 | Excellent item separation, items reliably ordered by difficulty |
| Item Separation Index | 9.13 | Items span far more than 3 distinct difficulty strata |
| Person Reliability | 0.52 | Moderate; typical for an 8-item screening instrument (Person separation improves with more items per level) |
| Person Separation Index | 1.04 | ≈ 2 distinguishable ability strata among students |
| KR-20 (Classical Reliability) | 0.54 | Consistent with the Rasch person reliability estimate |
| Infit MNSQ Range (items) | 0.85-1.17 | Within the 0.5-1.5 acceptable range for all items |
| Outfit MNSQ Range (items) | 0.18-1.21 | All items within range except P8 (0.18, overfit due to its very low endorsement rate, n=6/200); no item showed misfit (>1.5) |
| Point-Measure Correlation (PTMEA range) | 0.21 to 0.59 (all positive) | All 8 items positively discriminate the intended construct; lowest for P7 (0.21, ceiling-adjacent very-hard item), highest for P3 (0.59) |
| Raw variance explained by measures | 42.4% | Above the Rasch-model expectation floor; supports a dominant single construct |
| 1st residual contrast eigenvalue | 1.54 | Below the 2.0 Winsteps guideline for a competing dimension, supports unidimensionality |
| Person flagged with outfit >2.0 | 18/100 (9.0%) | Within the normal range (<10-15%) expected for real response data |
| DIF by gender (Male vs Female) | No item \|t\|>2.0 | No significant gender-based item bias detected |
| DIF by domicile (Muaro Jambi vs Kota Jambi) | No item \|t\|>2.0 | No significant gender-based item bias detected |

Based on Table 5 above, these JMLE-estimated indicators show the ability hierarchy in Table 4 is a stable, model-confirmed measure rather than a raw-score artifact: item calibration is highly reliable and well separated, all items positively discriminate the intended construct (PTMEA 0.21-0.59), the instrument behaves essentially unidimensionally, no item misfits, and no item shows gender- or domicile-based DIF, supporting fairness across subgroups. Person reliability (0.52) is moderate, expected given the instrument's brevity; expanding to at least three items per SOLO level (see Limitation) should raise person separation substantially, per the Spearman-Brown prophecy.

The declining-achievement pattern in Table 4 is visualized in the Wright Map, as shown in Figure 2 below:

**EMTEKA**: Jurnal Pendidikan Matematika
Volume 7, No. 2, 2026. 933-949
DOI: https://doi.org/10.24127/emteka.v7i2.12604
e-ISSN 2746-5594
p-ISSN 2746-5608

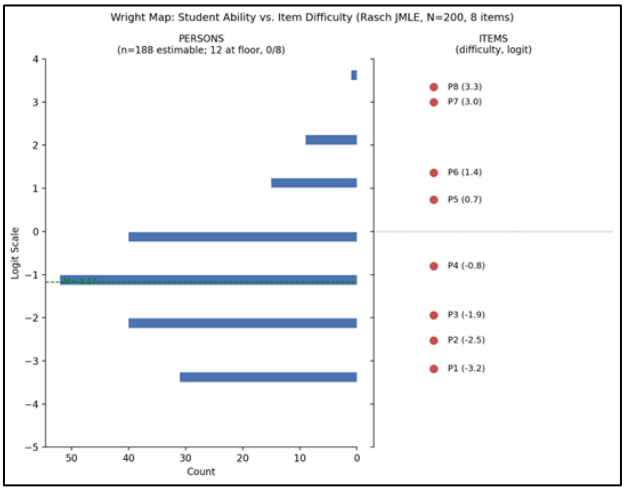



Figure 2. *Wright Map* of Student Ability and Item Difficulty Distributions in the Diagnostic Test ($\boldsymbol{N = 200}$)

Based on Table 4 and Figure 2 above, the results show a sharp decline following the PISA hierarchy, from 76.5% correct at P1 (Level 1c) to 3.0% at P8 (Level 6). Mean ability among 188 student with an estimable score ($\theta$ = -1.17 logit) is well below mean item difficulty (0.00), and the average student answered only 2.82/8 items correctly (35.2%); 12 students (6.0%) scored zero, a floor effect below this short instrument's measurable range. The steepest drop occurs between P4 (Level 2: 40.0%) and P5 (Level 3: 19.0%), a 21-point fall matching the SOLO transition from Unistructural to Multistructural, where students must integrate more than one piece of information (Biggs & Collis, 1982; Dong & Zhang, 2024). This confirms procedurally dominated classroom assessment does not prepare students for this transition, consistent with Ekowati et al., (2023), on unfamiliar-context difficulty and Nurhanurawati et al., (2022) on Bandar Lampung students performing well only at the basic procedural level.

*Distribution of Student Ability by PISA Level and SOLO Taxonomy*

This study categorized students' abilities based on diagnostic test results into PISA levels and the SOLO taxonomy, with the results presented in Table 6 below:

Table 6. Distribution of Student Ability by PISA Level and SOLO Taxonomy SOLO ($\boldsymbol{N = 200}$)

| PISA Level | SOLO Level | $\theta$ Range (logit) | N Siswa | Percentage | Description of Ability |
|---|---|---|---|---|---|
| **Level 1 (1a-1c)** | *Prastruktural* (L1) | $\theta < -1.50$ | 83 | 41.5% | Very simple procedural tasks (incl. 12 students at floor, 0/8 correct) |
| **Level 2** | *Unistruktural* (L2) | $-1.50 < \theta < -0.50$ | 52 | 26.0% | Direct interpretation of context |
| **Level 3** | *Multistruktural* (L3) | $-0.50 < \theta < 0.50$ | 40 | 20.0% | Multi-step tasks, multiple pieces of information |
| **Level 4** | *Relasional* (L4) | $0.50 < \theta < 1.50$ | 15 | 7.5% | Concept integration, coherent reasoning |
| **Level 5** | *Extended Abstract* (L5) | $1.50 < \theta < 2.50$ | 9 | 4.5% | Modeling of complex situations |

e-ISSN 2746-5594
p-ISSN 2746-5608

| PISA Level | SOLO Level | $\theta$ Range (logit) | N Siswa | Percentage | Description of Ability |
|---|---|---|---|---|---|
| **Level 6** | *Extended Abstract* (L5+) | $\theta > 2.50$ | 1 | 0.5% | Highest levels of abstraction and generalization |
| **Overall** | - | - | **200** | **100%** | - |

Based on Table 6 above, 87.5% of student (175/200) concentrated at PISA Levels 1-3; only 7.5% reach Level 4, 4.5% Level 5, and one student (0.5%) Level 6, consistent with the 2022 PISA national report placing Indonesia below the OECD Level-2 threshold (OECD, 2023b) and with Falani et al., (2025), whose instrument likewise concentrated difficulty in the easy-to-moderate range. From a SOLO perspective, most student can only process information in isolation, consistent with the questionnaire's Motivation and Engagement and Needs Variations in Question Dimensions (both moderate, M ≈ 2.86/2.82) and with Harisman et al., (2023), and Samsuddin & Retnawati (2025), who find Indonesian students weaker in Formulate/Interpret than in the procedural Employ process, precisely what characterizes PISA Level 3 and above.

An exploratory check of the relationship between students' questionnaire scores and diagnostic $\theta$ ability (matched individually, n=188) found no significant correlation for any dimension (D1 r = -0.10; D2 r = 0.07; D3 r = 0.01; D4 r = 0.03; overall r = 0.01; all p > 0.19). this null result suggests self-reported perceptions, contextual experience, motivation, and desire for variety do not by themselves track demonstrated ability, reinforcing the Joint Display's core argument that self-report and performance data capture distinct, complementary needs that must be triangulated rather than treated as interchangeable. As Stacey (2011), predicted over a decade ago, without systematic assessment reform Indonesian students would remain focused on low-level procedural skill, confirmed by 2022 PISA data (OECD, 2023b), and this null correlation similarly suggests that raising perceived motivation alone, without a redesigned graded instrument and real-world content, may be insufficient to close the demonstrated ability gap.

*Integration of Findings: Profile of Needs for the Development of Mathematical Literacy Assessment Instruments*

The integration of findings from the three instruments was conducted using the Joint Display procedure. Creswell & Clark (2018) state that the Joint Display procedure involves comparing findings from all sources to identify patterns of convergence. The integration of the main findings in this study is presented in Table 7 as follows:

Table 7. *Joint Display*: Integration of Findings from Three Quantitative Instruments

| Dimension of Needs | Teacher Questionnaire | Student Questionnaire | Diagnostic Test | Pattern |
|---|---|---|---|---|
| **SOLO-graded, tiered items** | Need for New Instruments (D3) highest, M = 3.24, 90.5% ≥ 3 | Motivation (D3) & Varied Questions (D4) both moderate, M ≈ 2.86/2.82 | 87.5% of students at PISA 1-3; only 12.5% above Level 3 | Convergent |
| **Real-world context items** | Real-World Context | Experience with Contextual | Sharp drop P4 to P5 (40.0% to | Convergent |

e-ISSN 2746-5594
p-ISSN 2746-5608

| Dimension of Needs | Teacher Questionnaire | Student Questionnaire | Diagnostic Test | Pattern |
|---|---|---|---|---|
| | Integration (D4) lowest, M = 2.905, 71.4% ≥ 3 | Problems (D2) highest but still moderate, M = 2.908 | 19.0%) exactly where context complexity rises | |
| **Item-construction practice** | Item-Construction Practice (D2) lowest, M = 2.917, 70.8% ≥ 3 | Perception of Math Problems (D1) lowest, M = 2.836, 66.0% ≥ 3 | P1-P3 (58.5 – 76.5%) far higher than P5-P8 (3.0 – 19.0%) | Convergent |
| **Teacher professional development** | Knowledge-practice gap: D1 (understanding) M = 3.05 vs D2 (practice) M = 2.92 | - | - | Supportive |

Based on Table 7 above, convergence is strongerst for SOLO-graded items: teachers rate perceived need for new instruments as their highest-scoring dimension (D3, M = 3.24, 90.5% ≥ 3), students report only moderate motivation/engagement and desire for variety (D3, D4 ≈ 2.86/2.82), and the diagnostic test confirms 87.5% of students at PISA 1-3 with only 12.5% above Level 3, three independent sources pointing to the same conclusion: a gap between available questions and the graded, SOLO-based instrument teachers themselves say is needed.

Convergence is similarly strong for real-world context items: teachers score lowest on their own contextual item-construction ability (D4, M = 2.905), students' contextual-problem experience is their highest-scoring dimension yet still only moderate (D2, M = 2.908), and the diagnostic test shows its sharpest drop precisely where contextual interpretation begins (P4 to P5). Wang et al., (2023), reviewing 156 PISA-predictor studies, identify "opportunity to learn" as the strongest predictor of PISA performance, implying students rarely exposed to contextual problems cannot be expected to solve PISA items requiring contextual reasoning. Bilad et al., (2024), recommend Indonesia develop locally grounded, context-oriented instruments in response to 2022 PISA results, and Eka et al., (2026), add these must align with Merdeka Curriculum outcomes to be curricularly, not just psychometrically, valid.

A third convergence concerns assessment-practice limits: teachers' item-construction practice (D2) is their second-lowest dimension, students' overall perception of current problems (D1) is their lowest, and the diagnostic test shows a steep decline from near-ceiling procedural performance (P1-P3) to near-floor performance on higher-level items (P5-P8), consistent with classroom assessment still dominated by procedural tasks that do not prepare students for the SOLO L3-L5 transition.

Three convergent needs profiles thus emerge as the empirical foundation for the next development phase: (1) a tiered instrument measuring literacy per the SOLO hierarchy (L1-L5), concentrated at PISA Levels 2-4 to match students' actual ability distribution; (2) items integrating authentic real-life contexts aligned with PISA to train

e-ISSN 2746-5594
p-ISSN 2746-5608

the Formulate-Employ-Interpret process; and (3) systematic support for teachers in developing literacy-based items, given the knowledge-practice gap (D1 M = 3.05 vs. D2 M = 2.92; D3 the highest-rated need overall).

*Implication*

These findings carry three implications, First, a needs profile derived from both teacher and student data provides a more comprehensive empirical foundation than a single-perspective analysis, consistent with the core strength Creswell % Clark (2018), attribute to Exploratory Sequential Mixed-Methods designs.

Second, the critical PISA Level 2 to Level 3 (SOLO L2 to L3) transition has a concrete design implication: item distribution in the next-phase instrument should be strengthened within the PISA Level 2 to Level 4 range, students' zone of proximal development. L'Boy & Khan (2023), show items near taxonomic transition points carry the highest diagnostic power, and Darman et al., (2024), show instreuments Rasch-validated from the development stage achieve more stable separation reliability and require fewer fiels-test revisions, reinforcing the case for building Rasch validation in from the outset. Falani et al., (2025), futher argue Rasch misfit indicates a substantive competency gap, meaning the future instrument should specifically strengthen underrepresented items at PISA Levels 4 to Level 6, as also recommended by Eka et al., (2026).

Third, the teacher knowledge-practice gap (D3 highest, D2/D4 lowest) has a direct policy implication: instrument development must pair with professional-development programs on constructing and adapting SOLO-graded, contextually rich items, not simply distributing a finished product. Edwar et al., (2023), show structured HOTS workshops raised teacher competency from 45.70% to 85.70%, evidence the gap is trainable, and Juniati & Jamaan (2024), show Problem-Based-Learning-aligned questions consistently improve students' higher-order reasoning. These results point to three concrete audiences: curiculum developers, who can use the SOLO-PISA correspondence (Table 6) to calibrate item-bank difficulty; teacher-training providers, who can target the specific D2/D4 practice gap; and school leaders, who can prioritize PISA Level 2-4 item development first, where are likely highest given the current ability distribution.

*Limitation and Suggestion for Further Research*

This study has acknoeledged limitations. The small teacher sample (N=3) reflects the exploratory design, so teacher-questionnaire findings are indicative rather than generalizable. The qualitative phase (teacher interviews, test-document coding) had not been completed at this preliminary stage and will occur in the main research phase (Creswell & Clark, 2018). The diagnostic test contains only one item per PISA level, so per-level ability estimation remains indicative. The research location, limited to Jambi City and Muaro Jambi, restricts generalizability.

A quantitative CVI/Aiken's V for the teacher and student questionnaires was not computed; validation relied on qualitative expert judgment by two validators (see Instruments), and Cohen's Kappa for SOLO-level coding or teacher-created items remains to be completed in the main research phase. The diagnostic test's validity, by contrast, is inherited from PISA's own official validation; its Rasch reliability, separation, unidimensionality, PTMEA, and DIF indices (Table 5) were calculated directly from response data and are no longer a limitation of this revision.

e-ISSN 2746-5594
p-ISSN 2746-5608

Further research should: (1) develop a SOLO-based instrument with at least three items per level; (2) complete the qualitative phase, including quantitative content-validity review using CVI or Aiken's V (Kania et al., 2024) and inter-rater Cohen's Kappa reporting; (3) field-test with a broader, multi-district sample; and (4) validate using Differential Item Functioning analysis for grade-level and gender bias (Afifah et al., 2024; Fitrah et al., 2024).

## Conclusion and Suggestion

This study analyzed the need for a SOLO Taxonomy-based mathematics literacy instrument through teacher and student perspectives in Jambi City and Muaro Jambi. The teacher questionnaire revealed a knowledge-practice gap: teachers rate their need for new, better-graded instruments highest (D3: M = 3.24/4), while item-construction practice (D2: M = 2.92) and real-world context integration (D4: M =2.91) lag behind conceptual understanding (D1: M = 3.05). the student survey showed an overall moderate, fairly uniform score (M = 2.85/4), with contextual-problem experience the relative strength (D2: M =2.91) and general problem perception (D1) and desire for variety (D4) the relative weaknesses. The Rasch-calibrated (JMLE) diagnostic test confirmed 87.5% of students (175/200) concentrated at PISA Levels 1 to Levels 3 (Pre-structural to Multi-structural), mean $\theta$ = -1.17 logit among estimable students, high item separation reliability (0.99), and no evidence of gender- or domicile- based DIF.

Joint Display integration of the three instruments produced three convergent needs profiles as the empirical foundation for the next development phase: (1) a graded instrument per the SOLO hierarchy (L1 until L5), concentrated at PISA Level 2 until 4; (2) items integrating authentic real-life contexts aligned with PISA to train the Formulate-Employ-Interpret process; and (3) support for teachers' professional development in designing and using literacy-based instruments. Practically, these findings give curriculum developers, teachers-training providers, and school leaders a shared, empirically grounded starting point for prioritizing instrument-development effort. Theoretically, the study integrates SOLO Taxonomy, PISA proficiency levels, and Rasch measurement, not previously triangulated with dual teacher-student needs data in the Indonesian junior high context, underscoring the urgenc and feasibility of a psychometrically validated, SOLO-based matehmatical literacy instrument as a strategis step toward improving mathematics assessment quality in Indonesia.

## Reference

Afifa, M., Khoirunnisa, R., Pratiwi, S. M. V., & Meitaza, D. (2024). Utilizing Rasch Model To Analyze A Gender Gap In Students' Scientific Literacy On Energy. *JPFI: Jurnal Pendidikan Fisika Indonesia, 20(1),* 85-95. https://doi.org/10.15294/jpfi.v20i1.44472

Aisyah, & Juandi, D. (2022). The Description Of Indonesian Student Mathematics Literacy In The Last Decade. *International Journal of Trends in Mathematics Education Research, 5(1),* 105-110. https://doi.org/10.33122/ijtmer.v5i1.114

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A Taxonomy For Learning, Teaching, And Assessing: A Revision Of Bloom's Taxonomy Of Educational Objectives. Longman. https://id.scribd.com/document/369062951

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the Quality of Learning: the SOLO Taxonomy (Structur of the Observed Learning Outcome).* Academic Press.

https://resolutionresources.com.au/wp-content/uploads/2022/02/Biggs-Collis-1982.pdf

Bilad, M. R., Zubaidah, S., & Prayogi, S. (2024). Addressing The Pisa 2022 Results: A Call For Reinvigorating Indonesia's Education System. *International Journal of Essential Competencies in Education, 3(1),* 1-12. https://doi.org/10.36312/ijece.v3i1.1935

Bolstad, O. H. (2023). Lower Secondary Students' Encounters With Mathematical Literacy. *Mathematics Education Research Journal, 35,* 237-253. https://doi.org/10.1007/s13394-021-00386-7

Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences (3rd ed.)*. Routledge. https://doi.org/10.4324/9781410614575

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the rasch model: Fundamental measurement in the human sciences (4th ed.)*. Routledge. https://www.routledge.com/9 780367141424

Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research (3rd ed.)*. SAGE Publications.

Darman D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development And Validation Of Scientific Inquiry Literacy Instrument (Sili) Using Rasch Measurement Model. *Education Science, 14(3),* 1-28. https://doi.org/10.3390/educsci14030322

Dong, J., & Zhang, Y. (2024). Research Review: The Solo Taxonomy (Structure Of The Observed Learning Outcome) In Mathematics Education. *Academic Journal of Mathematical Sciences, 5(1),* 30-34. https://doi.org/10.25236/AJMS.2024.050105

Edwar, Putri, R. I. I., Zulkardi, & Darmawijoyo. (2023). Developing A Workshop Model For High School Mathematics Teachers Constructing HOTS Questions Through The Pendidikan Matematika Realistic Indonesia Approach. *Journal on Mathematics Education, 14(4),* 603-626. https://doi.org/10.22342/jme.v14i4.pp603-626

Eka, F., Suratman, D., Rifat, M., Siregar, & N., Hamdani. (2026). Design And Validation Of A Pisa Oriented Instructional Module For Linear Functions Aligned With The Merdeka Curriculum. *Al-Jabar: Jurnal Pendidikan Matematika, 17(1),* 233-248. https://doi.org/10.24042/ajpm.v17i1.29709

Ekowati, V. I., Nurhayati, E., Suwarna, & Saputri, I. I. (2023). Program For International Student Assessment-Based Analysis For Javanese Test. *IJERE: International Journal of Evaluation and Research in Education, 12(2),* 1122-1135. https://doi.org/10.11591/ijere.v12i2.24426

Falani, I., Ramalisa, Y., Sainuddin, S., & Kriswantoro. (2025). Development of a PISA-Based Mathematical Literacy Instrument For Indonesian Students Using Item Response Theory. *Journal of Physics: Conference Series, 3148, 012003.* https://doi.org/10.1088/1742-6596/3148/1/012003

Fitrah, M., Sofroniou, A., Ofisnto, Judijanto, L., & Widihastuti. (2024). Reliability And Separation Index Analysis Of Mathematics Questions Integrated With The Cultural Architecture Framework Using The Rasch Model. *Journal of Education and e-Learning Research, 11(3),* 499-509. https://doi.org/10.20448/jeelr.v11i3.5861

Fitriati, F., Rosli, R., Iksan, Z., & Hidayat, A. (2024). Exploring Challenges In Preparing Prospective Teachers For Teaching 4c Skills In Mathematics Classroom: A School-University Partnership Perspectives. *Cogent Education, 11(1),* 1-21. https://doi.org/10.1080/2331186X.2023.2286812

e-ISSN 2746-5594
p-ISSN 2746-5608

Harisman, Y., Mayani, D. E., Armiati, Syaputra, H., & Amiruddin, M. H. (2023). Analysis Of Student's Ability To Solve Mathematical Literacy Problems In Junior High Schools In The City Area. *Infinity: Journal of Mathematics Education, 12(1),* 55-68. https://doi.org/10.22460/infinity.v12i1.p55-68

Juniati, H., & Jamaan, E. Z. (2024). Problem-Based Learning In Improving Critical Mathematical Thinking Skills: Independent Curriculum Teaching Module. *Al-Jabar: Jurnal Pendidikan Matematika, 15(2),* 493-503. https://doi.org/10.24042/ajpm.v15i2.23788

Kania, N., Kusumah, Y.S., Dahlan, J.A., Nurlaeelah E., Gurbuz, F., & Bonyah, E. (2024). Constructing And Providing Content Validity Evidence Throught The Aiken's V Index Based On The Experts' Judgments Of The Instrument To Measure Mathematical Problem-Solving Skills. REID: Research and Evaluation in Education, 10(1), 45-58. https://doi.org/10.21831/reid.v10i1.71032

L'Boy, D., & Khan, R. N. (2023). A Rasch-Mode-Based Hierarchical Framework For Statistical Literacy And Learning. *International Journal of Mathematical Education in Science and Technology, 54(9),* 1874-1887. https://doi.org/10.1080/0020739X.2023.2261453

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7(2), 328*. https://www.researchgate.net/publication/235361463

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook (3rd ed.)*. SAGE Publications.

Nurhanurawati, Caswita, Bharata, H., & Widyastuti. (2022). The Analysis Of Junior High School Students' Mathematical Literacy: Field Study In Bandar Lampung. *Al-Jabar: Jurnal Pendidikan Matematika, 13(1),* 199-209. https://doi.org/10.24042/ajpm.v13i1.11659

Nusantara, D. S., Zulkardi, & Putri, R. I. I. (2021). Designing Pisa-Like Mathematics Task Using A Covid-19 Context (Pisacomat). *JME: Journal on Mathematics Education, 12(2),* 349-364. https://jme.ejournal.unsri.ac.id/index.php/jme/article/view/3727/251

OECD. (2023a). *PISA 2022 Assessment and Analystical Framework.* PISA, OECD Publishing, Paris. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023b). *PISA 2022 Results (Volume I):The State of Learning and Equity in Education.* PISA, OECD Publishing, Paris. https://doi.org/10.1787/53f23881-en

Rizbudiani, A. D., Jaedun, A., Rahim, A., & Arief. (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

Samsuddin, A. F., & Rernawati, H. (2025). Literasi Matematika Siswa Smp Negeri Makassar Ditinjau Dari Domain Proses, Konten, Dan Konteks Pisa. *Primatika: Jurnal Pendidikan Matematika, 14(2),* 413-422. https://jurnal.fkip.unmul.ac.id/index.php/primatika/article/ view/6127

Stacey, K. (2011). The pisa view of mathematical literacy in indonesia. *Journal on Mathematics Education, 2(2),* 95-126. https://doi.org/10.22342/jme.2.2.746.95-126

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata. https://www.researchgate.net/publication/282673464

Suparman, Juandi, D., & Martadiputra, B. A. P. (2023). Students' Heterogeneous Mathematical Critical Thinking Skills In Problem-Based Learning: A Meta-Analysis

e-ISSN 2746-5594
p-ISSN 2746-5608

Investigating The Involvement Of School Geographical Location. *Al-Jabar: Jurnal Pendidikan Matematika, 14(1),* 37-53. https://doi.org/10.24042/ajpm.v14i1.16200

Susanta, A., Sumardi, H., Susanto, E., & Retnawati. (2023). Mathematics Literacy Task On Number Pattern Using Bengkulu Context For Junior High School Students. *JME: Journal on Mathematics Education, 14(1),* 85-102. https://doi.org/10.22342/jme.v14i1.pp85-102

Tanujaya, B., Prahmana, R. C. I., & Mumu, J. (2021). Mathematics Instruction To Promote Mathematics Higher-Order Thinking Skills Of Students In Indonesia: Moving forward. *TEM Journal: Technology, Education, Management, Informatics, 10(4),* 1945-1954. https://doi.org/10.18421/TEM104-60

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors Predicting Mathematics Achievement In Pisa: A Systematic Review. *Large-Scale Assessments in Education, 11(24),* 11-24. https://doi.org/10.1186/s40536-023-00174-8

Wijaya, T. T., Hidayat, W., Hermita, N., Alim, J. A., & Talib, C. A. Exploring Contributing Factors To Pisa 2022 Mathematics Achievement: Insights From Indonesian Teachers. *Infinity: Journal of Mathematics Education, 13(1),* 139-156. https://doi.org/10.22460/infinity.v13i1.p139-156

Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement.* MESA Press. https://research.acer.edu.au/cgi/viewcontent.cgi?params=/context/measurement/article/1000

Yanuarto, W. N., Setyaningsih, E., & Umiatun, K. N. (2023). Kahoot! As A Collaborative Gamified Mathematics Learning Platform For Indonesian Junior School Students. *Al-Jabar: Jurnal Pendidikan Matematika, 14(1),* 99-108. https://doi.org/10.24042/ajpm.v14i1.16623

Yuanita, P., Maimunah, Kartini, & Arif, N. (2023). Analysis Of Solo Taxonomy-Based Test Instruments In Approaching The Merdeka Curriculum. *JRAMathEdu: Journal of Research and Advances in Mathematics Education, 8(1),* 12-29. https://doi.org/10.23917/jramathedu.v8i1.1270

Yulia, Kustati, M., & Afriadi, J. (2021). Mathematical Literacy: A Case Study On Padang Students' Mathematical Ability. *Al-Jabar: Jurnal Pendidikan Matematika, 12(1),* 109-120. https://doi.org/10.24042/ajpm.v12i1.8538