AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 15, No. 2, 2026, 909-923

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

A COGNITIVELY ORIENTED DIAGNOSTIC ASSESSMENT OF NUMERACY LITERACY: DEVELOPMENT AND EXPERT VALIDATION

Rizky Fadillah[1], Rachmat Wasqita[2]

[1.2]STKIP Paracendekia NW Sumbawa, Nusa Tenggara Barat, Indonesia
**Corresponding author: rizkyfh67@gmail.com*

Received 07 November 2025; Revised 23 May 2026; Accepted 23 June 2026

Abstract

Numeracy literacy achievement among Indonesian students remains critically low, yet contextually situated diagnostic instruments are largely unavailable. This study develops a cognitively oriented diagnostic assessment instrument in mathematics, grounded in Sumbawa's local numeracy literacy context, for Phase D students across Grades VII, VIII, and IX. Development followed Thiagarajan's 4D framework: Define, Design, and Develop, involving observation, interviews, and curriculum analysis. The Design phase produced five integrated components test blueprints, item sets, answer keys, scoring rubrics, and interpretive guidelines differentiated across Bloom's cognitive levels (C1–C6) and anchored in local culture. Prototype I was validated by nine experts: three lecturers and two practitioner teachers per grade level, covering content, construct, and language aspects. Quantitative scores were analyzed using Aiken's V; qualitative feedback through thematic analysis. Three aspects achieved highly valid ratings across all grades (mean: 95.09%), with Aiken's V exceeding 0.80 for all items. Validators affirmed alignment with diagnostic purposes, numeracy indicators, and local context, guiding revisions producing Prototype II. Findings show that integrating Bloom's Taxonomy, numeracy literacy, and local culture within a diagnostic framework is theoretically coherent and feasible. The absence of student trials and psychometric analyses are acknowledged, with future directions proposed.

Keywords: Bloom's Taxonomy; Cognitively Oriented Diagnostic Assessment; Local Context; Numeracy Literacy; Phase D.

Abstrak

Capaian literasi numerasi siswa Indonesia masih tergolong rendah, sementara instrumen asesmen diagnostik yang kontekstual masih sangat terbatas. Penelitian ini mengembangkan instrumen asesmen diagnostik berorientasi kognitif dalam matematika, berbasis konteks literasi numerasi lokal Sumbawa, untuk siswa Fase D pada kelas VII, VIII, dan IX. Pengembangan mengikuti kerangka 4D Thiagarajan: Define, Design, dan Develop, yang melibatkan observasi, wawancara, dan analisis kurikulum. Tahap Design menghasilkan lima komponen instrumen yang terintegrasi kisi-kisi, butir soal, kunci jawaban, rubrik penilaian, dan pedoman interpretasi yang didiferensiasikan berdasarkan tingkat kognitif Taksonomi Bloom (C1–C6) dan dikontekstualisasikan pada budaya lokal. Prototipe I divalidasi oleh sembilan validator: tiga dosen dan dua guru praktisi per jenjang kelas, mencakup aspek materi, konstruk, dan bahasa. Skor kuantitatif dianalisis menggunakan Aiken's V; umpan balik kualitatif melalui analisis tematik. Ketiga aspek mencapai kategori sangat valid pada semua jenjang (rata-rata: 95,09%), dengan nilai Aiken's V melebihi 0,80 untuk seluruh butir. Validator mengafirmasi keselarasan instrumen dengan tujuan diagnostik, indikator literasi numerasi, dan konteks lokal, memandu revisi yang menghasilkan Prototipe II. Hasil menunjukkan bahwa pengintegrasian Taksonomi Bloom, literasi numerasi, dan budaya lokal dalam satu kerangka diagnostik bersifat koheren secara teoritis dan layak. Ketiadaan uji siswa dan analisis psikometri diakui sebagai keterbatasan, dengan arah penelitian selanjutnya diusulkan.

***Kata kunci**: Asesmen Diagnostik Berorientasi Kognitif; Fase D; Konteks Lokal; Literasi Numerasi; Taksonomi Bloom.*

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

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

INTRODUCTION

Numeracy literacy is a core mathematical competence encompassing critical thinking, quantitative reasoning, and the application of mathematical concepts in real-world contexts (OECD, 2023). According to the 2022 PISA report, only 18% of Indonesian students reached Level 2 or above in mathematical literacy, compared to an OECD average of 69% (OECD, 2023). At the national level, a large proportion of junior high school students encounter significant challenges in higher-order mathematical processes, including reasoning, analytical thinking, and model construction (Simbolon & Nurjanah, 2024), a pattern also evident in TIMSS results (Chang, 2023). Addressing this challenge requires precise diagnostic instruments capable of identifying specific learning difficulties at the cognitive attribute level. Diagnostic assessment provides teachers with actionable information about students' prior knowledge, reasoning strategies, and misconceptions (Black & Wiliam, 2009; Heritage, 2021), and has been shown to improve instructional tar-geting and adaptive learning efficiency (Zhan et al., 2021). Although Cognitive Diagnostic Models such as the G-DINA (Torre, 2011) and LCDM (Templin & Bradshaw, 2013) represent the frontier of this approach, their application requires large-scale empirical data beyond the scope of instrument deve-lopment studies. The present study occupies an earlier, foundational stage: developing and validating a cognitively oriented diagnostic instrument structu-rally aligned with the attribute-based logic of cognitive diagnosis, designed as a precursor to subsequent CDM analysis.

Despite formal recognition of diagnostic assessment within Indonesia's Merdeka Curriculum, school-based implementations remain largely undifferentiated conducted without reference to students' cognitive levels or local cultural contexts (Supriyadi et al., 2022). This reduces diagnostic precision, particularly in regions such as Sumbawa where mathematical reasoning is embedded in culturally specific everyday practices.

Contextually and socially situated assessment promotes deeper mathematical understanding by connecting abstract concepts to students' lived experiences (Koskinen & Pitkäniemi, 2022; Mahmuti et al., 2025). In Indonesia, integrating local cultural elements has been shown to strengthen relevance and deepen conceptual understanding (Fajri et al., 2022; Nur Qomaria & Fatanur Baity Tsulutsya, 2024; Rezky et al., 2022). However, despite the richness of Sumbawa's cultural resources including traditional practices such as *Barapan Kebo* (buffalo racing), local trade contexts, and regional landmarks these elements have not been systematically incorporated into diagnostic instruments (Frentika & Rizki, 2020; Rizki & Frentika, 2021), leaving students without culturally anchored diagnostic assessment in numeracy literacy (Nursakiah et al., 2022; Susanta et al., 2023).

Prior studies have developed diagnostic and numeracy oriented instruments addressing higher order thinking skills, contextual literacy, and local cultural contexts (Alfageh et al., 2024; Amelia et al., 2023; Endang Susetyawati & Kintoko, 2022; Putri, 2021; Supriyadi et al., 2022; Utariningsih, 2018). However, three persistent gaps remain. First, no prior study has simultaneously integrated cognitive level differentiation based on Bloom's revised taxonomy (Anderson & Krathwohl, 2001), and local cultural

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

contextualization within a single multi-grade diagnostic framework. Second, existing Indonesian instruments address a single grade level or topic rather than a cross-grade diagnostic system. Third, instruments at the expert validation stage rarely articulate theoretical connections to the cognitive diagnostic tradition, leaving the path to CDM application unclear. These three gaps constitute the rationale for the present study.

The present study addresses these gaps through an integrated conceptual framework mapping Bloom's revised taxonomy (C1–C6) to two numeracy literacy types: (1) pure numeracy (symbol manipulation and mathematical operations without real-world context dependency) and (2) contextual numeracy (mathematical reasoning within authentic, culturally situated problems). Lower-order levels (C1–C2) engage pure numeracy demands; middle-order levels (C3–C4) engage contextual numeracy requiring interpretation and application within real-world situations; and higher-order levels (C5–C6) engage advanced contextual numeracy requiring evaluation and model construction. Table 1 presents this mapping explicitly, anchoring each cognitive level to specific numeracy indicators, item categories, and diagnostic functions consistent with the structure operationalized in the instrument's test blueprints.

Table 1. Conceptual Mapping of Bloom's Revisied Taxonomy to Numeracy Literacy Indicators, Numeracy Types, and Diagnostic Functions

Bloom's Level	**Numeracy Type**	**Numeracy Literacy Indicator**	**Item Category**	**Diagnostic Function**
C1 – Remember	Pure & Contextual Numeracy	Identifying variables, coefficients, or data from mathematical models and real-world contexts	Prerequisite	Detects gaps in foundational mathematical knowledge
C2 – Understand	Pure Numeracy	Processing and manipulating algebraic or numerical forms; interpreting mathematical symbols	Prerequisite	Detects difficulties in symbol manipulation and mathematical comprehension
C3 – Apply	Contextual Numeracy	Interpreting quantitative information from verbal statements; solving problems in local contexts	Prerequisite & Core	Detects difficulties in translating real-world situations into mathematical representations
C4 – Analyze	Contextual Numeracy	Analyzing relationships among quantitative variables; differentiating relevant from irrelevant contextual information	Core	Detects difficulties in higher-order quantitative reasoning
C5 – Evaluate	Contextual Numeracy	Assessing validity of mathematical results within contextual constraints	Advanced	Detects difficulties in critical reflection on mathematical outcomes

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

Bloom's Level	Numeracy Type	Numeracy Literacy Indicator	Item Category	Diagnostic Function
C6 – Create	Contextual Numeracy	Formulating mathematical models from open-ended problems in local cultural contexts	Advanced	Detects difficulties in mathematical modeling and creative problem-solving

This conceptual mapping serves two interconnected functions within the present instrument. First, it provides the theoretical basis for item construction: each item in the test blueprint is explicitly anchored to a specific Bloom's cognitive level, numeracy type, and numeracy literacy indicator as operationalized in the instrument's blueprints ensuring that the instrument captures a differentiated cognitive profile rather than an undifferentiated difficulty score. Second, it provides the interpretive basis for diagnostic use: patterns of student performance across C1–C6 levels and across pure and contextual numeracy demands generate a cognitive attribute profile that identifies not only *whether* a student is experiencing difficulty, but *at which level of cognitive demand and in which numeracy domain* the difficulty originates.

The present study therefore aims to develop a cognitively oriented diagnostic assessment instrument in mathematics, grounded in the local numeracy literacy context of Sumbawa and differentiated by Bloom's cognitive levels (C1–C6), for Phase D students across Grades VII, VIII, and IX. The specific objectives are: (1) to produce an integrated instrument comprising test blueprints, item sets, answer keys, scoring rubrics, and interpretive guidelines; (2) to validate the instrument's content, construct, and language aspects through expert review; and (3) to establish the instrument's theoretical and structural readiness to serve as the foundation for subsequent empirical trials and cognitive diagnostic modeling in future research.

METHODS

This study employed a Research and Development (R&D) design to develop and validate a mathematics diagnostic assessment instrument grounded in the local numeracy literacy context of Sumbawa and differentiated by Bloom's cognitive levels (C1–C6) for Phase D students across Grades VII, VIII, and IX. A three-phase development model Define, Design, and Develop was adapted from Thiagarajan's 4D framework (Thiagarajan et al., 1974).

The Define phase was conducted through curriculum document analysis and semi structured interviews with mathematics teachers at SMPN 1 Lape, Sumbawa. . Interviews addressed three themes: the curriculum currently implemented, existing assessment practices, and the extent of local context integration in learning and assessment. Interview data were transcribed and analyzed through thematic content analysis to identify key contextual needs informing the subsequent design phase. Classroom observation data supplemented interview findings, particularly in identifying recurring patterns of student difficulty in numeracy related tasks.

The Design phase produced Prototype I, comprising five integrated components: test blueprints, item sets, answer keys, scoring rubrics, and result

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

interpretation guidelines. All components were developed with explicit reference to Bloom's revised taxonomy cognitive levels (C1–C6) (Anderson & Krathwohl, 2001), numeracy literacy indicators spanning multiple representations (narrative, tabular, graphical, and pictorial), and the local cultural context of Sumbawa. Each item set included items categorized as prerequisite, core, or advanced, enabling the instrument to function as a tiered cognitive map of student readiness.

The Develop phase involved validation of Prototype I by nine validators: three university lecturers in mathematics education serving as content experts, and two practitioner mathematics teachers per grade level (VII, VIII, IX). This composition ensured evaluation of both theoretical alignment and practical applicability of the instrument across all three grade levels. Validation data were collected using a structured questionnaire covering three aspects content, construct, and language each comprising five indicators. Validators assigned scores using a four-point scale (1= no indicator met; 2= one to two indicators met; 3 = three to four indicators met; 4= all five indicators met). In addition to quantitative ratings, validators provided written qualitative comments and suggestions for each instrument component.

Quantitative data were analyzed using two complementary approaches. First, percentage scores were computed for each validation aspect per grade level as a descriptive summary of validator consensus. Second, and more critically for establishing content validity evidence, Aiken's V coefficient was calculated for each item across validators, following the procedure described by Aiken (Aiken, 1980) and recommended as the standard for content validity in educational instrument development (Kania et al., 2024). Aiken's V is defined as:

$$V = \frac{\sum s}{[n(c-1)]} \qquad (1)$$

where s is the difference between each validator's rating and the minimum rating, n is the number of validators, and c is the number of scale points. An item was considered to have established content validity if its Aiken's V value exceeded the critical value of 0.75 for nine validators at $p < .05$ (Aiken, 1980; Kania et al., 2024). Qualitative data from validators' written comments were analyzed through thematic content analysis, categorizing feedback into areas requiring revision and areas of consistent strength, which directly informed the systematic revision of Prototype I into Prototype II.

RESULTS AND DISCUSSION

Define Phase

The Define phase established the foundational needs guiding instrument development. Semi-structured interviews with three mathematics teachers at SMPN 1 Lape revealed that diagnostic assessment had been practiced only sporadically typically as brief written tests administered without structured follow-up interpretation or linkage to adaptive instruction. Teachers acknowledged the necessity of systematic diagnostic assessment to map students' initial abilities but reported lacking appropriate instruments calibrated to cognitive levels or local context.

Classroom observations identified three recurring patterns of student difficulty: processing narrative-based problems, extracting meaning from graphical and tabular data, and selecting appropriate mathematical operations based on contextual information. These

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

difficulties correspond to core numeracy literacy sub-competencies identified in the PISA framework particularly the interpret and evaluate, and employ dimensions and confirm that existing assessment practices were insufficient to capture the specific cognitive nature of students' learning difficulties in numeracy literacy.

Curriculum document analysis confirmed that Phase D competencies spanning numbers, algebra, geometry, measurement, statistics, and probability require application in real life situations connected to students' everyday experiences and local cultural contexts. These findings collectively defined the scope, cognitive demand structure, and contextual orientation of the instrument developed in the subsequent phase.

Design Phase

The Design phase produced Prototype I, comprising five core components. The test blueprints specified the distribution of items across Bloom's cognitive levels (C1–C6), numeracy literacy indicator types (narrative, tabular, graphical, pictorial), item formats, and material categories (prerequisite, core, advanced) for each grade level. Each item was developed according to four design principles: contextualization (embedding mathematical tasks within authentic Sumbawa cultural contexts), variation (representing diverse numeracy indicator types), gradation (mapping to distinct cognitive levels), and numeracy literacy orientation (requiring active interpretation rather than mere calculation).

Answer keys specified complete solution steps, and scoring rubrics employed a 0–2 scale, enabling partial credit and supporting fine-grained analysis of students' reasoning strategies. Interpretive guidelines categorized diagnostic results into high, medium, and low readiness levels, each linked to concrete adaptive follow up learning recommendations. This five-component architecture particularly the explicit linkage between diagnostic scores and adaptive instructional pathways distinguishes the present instrument from prior diagnostic assessments that focused exclusively on item construction without providing operationalized follow up guidance for teachers (Putri, 2021; Rachmah, 2018; Utariningsih, 2018).

Figures 1 and 2 illustrate two representative components of Prototype I: the test blueprint, which operationalizes the conceptual mapping presented in Table 1, and the interpretation guideline, which translates diagnostic scores into tiered readiness profiles linked to adaptive instructional recommendations. The complete instrument package comprising test blueprints, item sets, answer keys, and scoring rubrics for all three grade levels is available upon request from the corresponding author.

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 909-923 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

Kisi-kisi Materi Persamaan dan Pertidaksamaan

No. TP	Tujuan Pembelajaran	Level Kognitif	Indikator Soal	Indikator Literasi & Numerasi	Jenis & Konteks Numerasi	Bentuk Soal	Konteks Lokal	No. Soal	Jenis Materi
1	Menyederhanakan bentuk aljabar sederhana melalui operasi penjumlahan, pengurangan, dan perkalian.	C2	Menyelesaikan operasi bentuk aljabar dan menuliskan hasil sederhananya.	Mengolah dan memanipulasi bentuk aljabar.	Numerasi murni – aljabar	Isian	-	1a, 1b, 1c	Prasyarat
2	Menyatakan situasi nyata ke dalam bentuk persamaan atau pertidaksamaan.	C3	Mengubah pernyataan verbal menjadi model matematika menggunakan simbol "=", "<", atau ">".	Menafsirkan informasi kuantitatif dari pernyataan verbal.	Numerasi kontekstual – aljabar	Isian	Istana Dalam Loka, Serunai	2	Prasyarat
3	Mengidentifikasi variabel, koefisien, dan konstanta dalam model matematika.	C1	Menyebutkan variabel, koefisien, dan konstanta dari model matematika yang diberikan.	Mengidentifikasi komponen aljabar dalam model konteks nyata.	Numerasi kontekstual – perdagangan lokal	Uraian	CFD Samota, pelu lenga	3	Prasyarat
4	Menyusun persamaan linear satu variabel dari permasalahan kontekstual dan menyelesaikannya.	C3	Menentukan persamaan dan nilai variabel dari masalah kontekstual.	Menafsirkan dan memecahkan masalah kuantitatif dari informasi kontekstual.	Numerasi kontekstual – aljabar	Uraian	Samota, kaos "Sumbawa Sehat"	4	Inti

Figure 1. Sample Test Blueprint for Grade VII

Interpretasi Materi Pythagoras

Jenis Materi	Kategori	Jumlah Skor Benar	Indikasi Kemampuan	Tindak Lanjut Guru
Prasyarat	Tinggi	4	Siswa mampu mengisi tabel kuadrat dan akar dengan tepat serta mengenal konsep sisi segitiga siku-siku (TP1 & TP2 prasyarat).	Dapat melanjutkan ke Materi Inti TP1 (menggambar segitiga siku-siku dan menentukan sisi miring) dan TP2 (menghitung panjang sisi miring).
	Sedang	2	Siswa menguasai sebagian TP prasyarat, misal mampu mengisi tabel kuadrat tapi belum konsisten menentukan sisi segitiga.	Lanjut ke Materi Inti TP1 saja. Berikan penguatan pada TP2 prasyarat sebelum masuk ke TP2 inti.
	Rendah	0–1	Siswa belum mampu mengisi tabel atau memahami konsep segitiga siku-siku.	Remediasi penuh TP1 & TP2 prasyarat sebelum masuk ke Materi Inti.
Inti	Tinggi	5-6	Siswa mampu menggambar segitiga siku-siku, menentukan sisi miring, dan menghitung panjang sisi dengan tepat (TP1 & TP2 inti).	Dapat melanjutkan ke Materi Lanjut TP1 (menerapkan Pythagoras dalam masalah nyata dan soal garis lurus).
	Sedang	3-4	Siswa menguasai TP1 inti tetapi kesulitan pada TP2 inti (perhitungan sisi miring kompleks atau soal konteks).	Lanjut ke Materi Lanjut TP1 saja. Berikan penguatan pada TP2 inti terlebih dahulu.
	Rendah	0-2	Siswa kesulitan menggambar segitiga dan menghitung sisi miring.	Penguatan penuh TP1 & TP2 inti sebelum ke Materi Lanjut.
Lanjut	Tinggi	2	Siswa mampu menerapkan Pythagoras untuk menentukan hubungan titik-titik nyata dalam soal kontekstual (TP1 lanjut).	Dapat melanjutkan ke topik lanjut berikutnya yang lebih kompleks.
	Sedang	1	Siswa memahami konsep Pythagoras, tetapi belum konsisten menerapkannya dalam konteks nyata.	Fokus penguatan TP1 lanjut dengan contoh nyata sebelum melanjutkan ke topik lanjutan.
	Rendah	0	Siswa belum mampu menerapkan Pythagoras dalam konteks nyata.	Penguatan dari Materi Inti terlebih dahulu, terutama TP1 & TP2 inti, baru perkenalkan TP1 lanjut secara bertahap.

Figure 2. Sample Interpretation Guideline for Grade IX

Develop Phase

Prototype I was submitted for expert validation by nine validators. Table 3 presents the percentage scores and Aiken's V values for each validation aspect across all three grade levels.

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

Table 3. Quantitative Validation Results: Percentage Scores and Aiken's V by Aspect and Grade Level

Validation Aspect	Grade VII (%)	Grade VII Aiken's V	Grade VIII (%)	Grade VIII Aiken's V	Grade IX (%)	Grade IX Aiken's V
Content	95.86	0.86	95.00	0.84	95.71	0.85
Construct	97.86	0.89	95.00	0.84	91.75	0.82
Language	97.29	0.88	95.36	0.85	92.25	0.82
Overall Mean	**97.00**	**0.88**	**95.12**	**0.84**	**93.24**	**0.83**

As shown in Table 3, all validation scores across every aspect and grade level exceeded 81%, and all Aiken's V values exceeded 0.80, placing the instrument firmly within the highest validity category. The content aspect yielded the most consistent scores across grade levels (range: 95.00–95.86%), indicating that item alignment with Phase D competency standards and Bloom's cognitive levels was perceived as uniformly strong regardless of grade. The construct and language aspects showed a declining trend from Grade VII to Grade IX construct scores ranging from 97.86% to 91.75% and language scores from 97.29% to 92.25% suggesting that the greater mathematical complexity at the Grade IX level introduced challenges in both the structural organization of items and the clarity of their linguistic expression. Importantly, however, all scores remained above the 81% validity threshold and all Aiken's V values remained above 0.80, confirming that these differences reflect gradations within the highly valid category rather than validity concerns.

In addition to quantitative ratings, validators provided written qualitative feedback across four dimensions of the instrument. Table 4 presents a thematic summary of this feedback by grade level.

Table 4. Thematic Summary of Qualitative Validator Feedback by Grade Level

Evaluation Dimension	Grade VII	Grade VIII	Grade IX
Alignment of test items with the objectives of cognitive diagnostic assessment	Appropriate; minor revision of wording is needed	Appropriate; items are varied, prerequisite items need to be added	Appropriate; some items require simplification of wording
Reflection of numeracy literacy indicators	Indicators of numeracy literacy are well represented; able to assess understanding in various contexts	Indicators of numeracy literacy are well represented; require students to read, interpret, and use data	Indicators of numeracy literacy are well repre-sented; assess the ability to identify, interpret, and analyze information
Representation of Sumbawa local context	Sufficient; utilizes culture and daily activities	Sufficient; integrates local culture (e.g., barapan kebo, Pulau Moyo, etc.)	Sufficient; incorporates culture, environment, village names, and tourist attractions

AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 15, No. 2, 2026, 909-923
ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

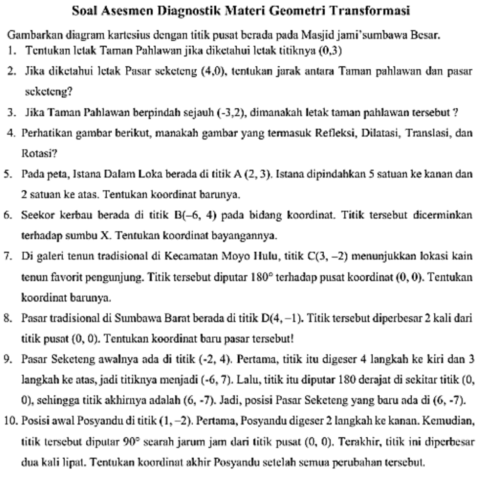
DOI: https://doi.org/10.24127/ajpm.v15i2.14615

Evaluation Dimension	Grade VII	Grade VIII	Grade IX
Mapping students for differentiated learning	Able to map learning levels; items vary from prerequisite to advanced	Able to map learning levels; tiered items support differentiated learning	Most items support mapping; some items require revision

Table 4 reveals that validators consistently affirmed the instrument's alignment with diagnostic purposes, numeracy literacy indicators, and local context across all three grade levels. Qualitative feedback also exposed a systematic pattern: Grade IX items received the most revision-related comments, particularly regarding wording simplification and item completeness. This is consistent with the quantitative finding that Grade IX yielded the lowest scores in both construct and language aspects, and together these findings point to the need for greater attention to linguistic accessibility when constructing items for advanced mathematical content.

Based on the convergent quantitative and qualitative evidence, Prototype I was systematically revised into Prototype II. Revisions were guided by the thematic categories identified in qualitative analysis and focused on three primary areas. First, item wording was simplified in Grades VII and IX, particularly for items at Bloom's C4–C6 levels where language complexity was found to conflate linguistic difficulty with cognitive difficulty, potentially introducing construct-irrelevant variance. Second, prerequisite-level items were added to the Grade VIII battery to ensure adequate coverage of foundational cognitive demands as a diagnostic baseline. Third, the explicitness of local context integration was strengthened in several items where validators indicated that the cultural references were present but insufficiently foregrounded.

Figure 3 illustrates a representative example of a Grade IX test item before and after revision, demonstrating the wording simplification and contextual refinement described above. Similar revisions were applied to items across all three grade levels.

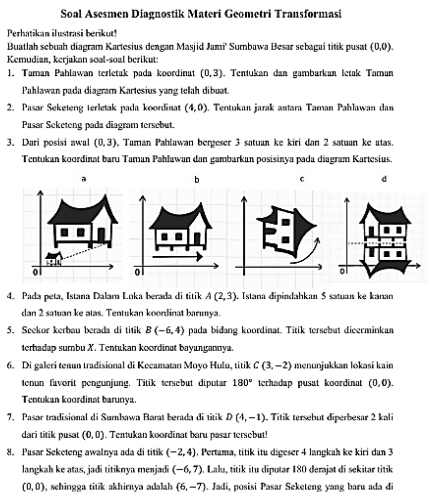


Soal Asesmen Diagnostik Materi Geometri Transformasi

Gambarkan diagram kartesius dengan titik pusat berada pada Masjid jami'sumbawa Besar.

1. Tentukan letak Taman Pahlawan jika diketahui letak titiknya (0,3)
2. Jika diketahui letak Pasar seketeng (4,0), tentukan jarak antara Taman pahlawan dan pasar seketeng?
3. Jika Taman Pahlawan berpindah sejauh (-3,2), dimanakah letak taman pahlawan tersebut ?
4. Perhatikan gambar berikut, manakah gambar yang termasuk Refleksi, Dilatasi, Translasi, dan Rotasi?
5. Pada peta, Istana Dalam Loka berada di titik A (2, 3). Istana dipindahkan 5 satuan ke kanan dan 2 satuan ke atas. Tentukan koordinat barunya.
6. Seekor kerbau berada di titik B(–6, 4) pada bidang koordinat. Titik tersebut dicerminkan terhadap sumbu X. Tentukan koordinat bayangannya.
7. Di galeri tenun tradisional di Kecamatan Moyo Hulu, titik C(3, –2) menunjukkan lokasi kain tenun favorit pengunjung. Titik tersebut diputar 180° terhadap pusat koordinat (0, 0). Tentukan koordinat barunya.
8. Pasar tradisional di Sumbawa Barat berada di titik D(4, –1). Titik tersebut diperbesar 2 kali dari titik pusat (0, 0). Tentukan koordinat baru pasar tersebut!
9. Pasar Seketeng awalnya ada di titik (-2, 4). Pertama, titik itu digeser 4 langkah ke kiri dan 3 langkah ke atas, jadi titiknya menjadi (-6, 7). Lalu, titik itu diputar 180 derajat di sekitar titik (0, 0), sehingga titik akhirnya adalah (6, -7). Jadi, posisi Pasar Seketeng yang baru ada di (6, -7).
10. Posisi awal Posyandu di titik (1, –2). Pertama, Posyandu digeser 2 langkah ke kanan. Kemudian, titik tersebut diputar 90° searah jarum jam dari titik pusat (0, 0). Terakhir, titik ini diperbesar dua kali lipat. Tentukan koordinat akhir Posyandu setelah semua perubahan tersebut.

(a)



Soal Asesmen Diagnostik Materi Geometri Transformasi

Perhatikan ilustrasi berikut!
Buatlah sebuah diagram Kartesius dengan Masjid Jami' Sumbawa Besar sebagai titik pusat (0,0).
Kemudian, kerjakan soal-soal berikut:

1. Taman Pahlawan terletak pada koordinat $(0,3)$. Tentukan dan gambarkan letak Taman Pahlawan pada diagram Kartesius yang telah dibuat.
2. Pasar Seketeng terletak pada koordinat $(4,0)$. Tentukan jarak antara Taman Pahlawan dan Pasar Seketeng pada diagram tersebut.
3. Dari posisi awal $(0,3)$, Taman Pahlawan bergeser 3 satuan ke kiri dan 2 satuan ke atas. Tentukan koordinat baru Taman Pahlawan dan gambarkan posisinya pada diagram Kartesius.

4. Pada peta, Istana Dalam Loka berada di titik $A\ (2,3)$. Istana dipindahkan 5 satuan ke kanan dan 2 satuan ke atas. Tentukan koordinat barunya.
5. Seekor kerbau berada di titik $B\ (-6,4)$ pada bidang koordinat. Titik tersebut dicerminkan terhadap sumbu X. Tentukan koordinat bayangannya.
6. Di galeri tenun tradisional di Kecamatan Moyo Hulu, titik $C\ (3,-2)$ menunjukkan lokasi kain tenun favorit pengunjung. Titik tersebut diputar 180° terhadap pusat koordinat $(0,0)$. Tentukan koordinat barunya.
7. Pasar tradisional di Sumbawa Barat berada di titik $D\ (4,-1)$. Titik tersebut diperbesar 2 kali dari titik pusat $(0,0)$. Tentukan koordinat baru pasar tersebut!
8. Pasar Seketeng awalnya ada di titik $(-2,4)$. Pertama, titik itu digeser 4 langkah ke kiri dan 3 langkah ke atas, jadi titiknya menjadi $(-6,7)$. Lalu, titik itu diputar 180 derajat di sekitar titik $(0,0)$, sehingga titik akhirnya adalah $(6,-7)$. Jadi, posisi Pasar Seketeng yang baru ada di

(b)

Figure 3. Grade IX Test Item (a) Before Revision; (b) After Revision

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 909-923 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

The findings are grounded in four interconnected theoretical traditions. First, diagnostic assessment theory positions assessment not as a summative judgment but as a formative mechanism for identifying the specific nature and location of learning difficulties, enabling teachers to make evidence-based instructional adjustments (Black & Wiliam, 2009; Heritage, 2021). The present instrument operationalizes this function through its interpretive guidelines, which translate diagnostic scores into tiered readiness profiles directly linked to adaptive follow-up recommendations moving beyond instruments that identify difficulties without prescribing instructional responses (Putri, 2021). Second, the numeracy literacy framework underpinning the instrument draws on the PISA mathematical literacy construct, which defines numeracy as the capacity to formulate, employ, and interpret mathematics in real-world contexts (OECD, 2023). The instrument's item design requiring students to process narrative, tabular, graphical, and pictorial representations within culturally situated problems directly operationalizes these three process dimensions across all grade levels (Susanta et al., 2023; Zhang et al., 2025). Third, adaptive learning theory holds that instruction is most effective when calibrated to students' current cognitive readiness level rather than delivered uniformly (Nabila & Widjajanti, 2020; Tamur et al., 2020). The instrument's three-tier item structure prerequisite, core, and advanced provides the diagnostic baseline necessary for teachers to implement differentiated instructional pathways responsive to individual cognitive profiles. Fourth, cognitive diagnosis theory, as formalized in Cognitive Diagnostic Models such as the G-DINA (Torre, 2011) and LCDM (Templin & Bradshaw, 2013), requires that assessment items be explicitly mapped to cognitive attributes through a Q-matrix specification. The present instrument's differentiation across Bloom's C1–C6 levels constitutes a preliminary Q-matrix structure, positioning it as a theoretically coherent precursor to formal CDM analysis in future empirical phases (Zhan et al., 2021).

The high content validity scores (mean 95.34%) confirm that items accurately represented Phase D competency standards and Bloom's cognitive levels across all grade levels, demonstrating that a single framework can simultaneously satisfy cross-grade content alignment and fine-grained cognitive level differentiation a combination rarely achieved within a single instrument in the Indonesian context (Amelia et al., 2023; Utariningsih, 2018). The high construct validity scores (mean 94.87%). reflect the coherence of the five-component architecture as an integrated diagnostic system. Notably, construct and language scores declined from Grade VII to Grade IX (construct: 97.86% to 91.75%; language: 97.29% to 92.25%), reflecting that Grade IX items addressing statistics, probability, and algebraic reasoning at C4–C6 levels required greater structural precision to maintain cognitive level differentiation and avoid conflating language complexity with cognitive complexity (Clark & Watson, 2019). All scores nonetheless remained above the 81% validity threshold, confirming these differences reflect gradations within the highly valid category.

Validators affirmed across all grade levels that items effectively required students to interpret data, apply

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

mathematical symbols within authentic situations, and connect concepts to real-life contexts. This directly addresses a limitation of prior Indonesian numeracy assessments, which rarely integrate procedural fluency, contextual relevance, and explicit cognitive level differentiation simultaneously (Amelia et al., 2023; Endang Susetyawati & Kintoko, 2022). Local contextual integration was similarly affirmed as sufficient across all grade levels, supporting research that culturally grounded instruments enhance construct validity by reducing cultural distance between assessment context and students' experiences (Istihapsari et al., 2023; Nursakiah et al., 2022). However, expert validation of contextual relevance cannot substitute for student-level evidence familiarity with cultural references may vary within a region, and this gap remains an open empirical question for future field trials (Paunno et al., 2025).

Compared to prior studies Amelia et al. (2023) who developed Phase D numeracy items without local cultural context or cross-grade diagnostic structure, Utariningsih (2018) who produced HOTS-oriented diagnostic items without thematic organizational framework or interpretive guidelines, and Nursakiah et al. (2022) and Susanta et al. (2023) who incorporated local context but addressed only single-topic, single-grade instruments the present instrument advances the field by integrating cross-grade, multi-topic, cognitively differentiated diagnostic assessment with adaptive instructional guidance within a single validated framework.

Three limitations must be stated explicitly. First, content validity was established through expert judgment only; construct validity via EFA, CFA, or Rasch modeling remains necessary for broader deployment. Second, the instrument has not been administered to students, so empirical reliability, item difficulty, and usability evidence are absent limitations that reflect the deliberate scope of this study as the development and content validation phase of a multi-phase research program. Third, generalizability is bounded by the Sumbawa context; the framework may be transferable, but specific cultural references require systematic adaptation for other regional settings.

CONCLUSIONS

This study produced a cognitively oriented mathematics diagnostic assessment instrument grounded in the local numeracy literacy context of Sumbawa, developed through a three-phase 3D model (Define, Design, and Develop) for Phase D students across Grades VII, VIII, and IX. Content validation by nine expert validators confirmed that the instrument achieved highly valid ratings across all three aspects content, construct, and language at each grade level, with an overall mean score of 95.09% and Aiken's V values consistently exceeding 0.80 across all items. The five-component integrated architecture test blueprints, item sets, answer keys, scoring rubrics, and interpretive guidelines linked to adaptive learning recommendations represents a structurally coherent and theoretically grounded diagnostic tool designed to support adaptive and differentiated mathematics instruction.

The study makes three contributions to the literature. First, it demonstrates the feasibility of integrating Bloom's Taxonomy cognitive level differentiation, numeracy literacy indicators, and local cultural contextualization within a single multi-grade

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

diagnostic framework. Second, it articulates the instrument's structural alignment with the attribute-based logic of cognitive diagnostic modeling, positioning it as a theoretically coherent precursor to CDM application in future empirical phases. Third, it provides a replicable development methodology for producing contextually valid diagnostic instruments in region-specific educational settings.

Future research should extend this work through: (1) empirical field trials with Phase D students to generate data for reliability analysis, item difficulty calibration, and discriminatory power assessment; (2) construct validity analysis using confirmatory factor analysis or Rasch modeling; (3) application of Cognitive Diagnostic Models (G-DINA or LCDM) using the instrument's cognitive level structure as the Q-matrix specification; and (4) adaptation and validation of the instrument framework for other regional and national contexts to assess the generalizability of the development approach.

REFERENCES

Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Alfageh, D. H., York, C. S., Hodge-Zickerman, A., & Xie, Y. (2024). Elementary teachers' use of adaptive diagnostic assessment to improve mathematics teaching and learning: A case study. *International Electronic Journal of Mathematics Education*, *19*(1), em0768. https://doi.org/10.29333/iejme/14190

Amelia, S., Widiati, I., & Yadrika, G. (2023). Pengembangan Soal Numerasi Untuk Peserta Didik Fase D. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(3), 3048–3063. https://doi.org/10.24127/ajpm.v12i3.7236

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Chang, I. (2023). Early numeracy and literacy skills and their influences on fourth-grade mathematics achievement: a moderated mediation model. *Large-Scale Assessments in Education*, *11*(1). https://doi.org/10.1186/s40536-023-00168-6

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Deda, Y. N., & Maifa, T. (2021). Development of Student Worksheets Using the Context of Local Wisdom on Integers and Fractions. *Mathematics Education Journal*, *15*(1), 71–82. https://doi.org/10.22342/jpm.v.i.12824.71-82

Endang Susetyawati, M., & Kintoko.

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 909-923 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

(2022). Pengembangan Butir Soal Kemampuan Literasi Numerasi Matematika Materi Bangun Ruang Kelas VIII SMP Di Yogyakarta. *Indonesian Journal Of Education And Humanity*, *2*(2), 52–61.

Fajri, H. M., Zulkardi, Z., Putri, R. I. I., & Susanti, E. (2022). Kemampuan Literasi Numerasi Siswa Smp Pada Konteks Pembelajaran Jarak Jauh (PJJ) Pasca Covid-19. *EDU-MAT: Jurnal Pendidikan Matematika*, *10*(2), 162–174. https://doi.org/10.20527/edumat.v10i2.12144

Frentika, D., & Rizki, T. N. H. (2020). Geometri dan Pengukuran dalam Permainan Rakyat Kabupaten Sumbawa Barat. *Jurnal Pendidikan MIPA*, *10*(2), 86–93. https://doi.org/10.37630/jpm.v10i2.346

Heritage, M. (2021). *Formative Assessment: Making It Happen in the Classroom* (2nd ed.). Corwin Press.

Istihapsari, V., Istiandaru, A., Wahyudi, R. S., Sukestiyarno, S., Suyitno, H., & Rochmad, R. (2023). Using Context of Local Environment Maps to Develop Lesson Plan which Foster Mathematical Literacy. *International Journal on Emerging Mathematics Education*, *7*(1), 35. https://doi.org/10.12928/ijeme.v7i1.25764

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., & Gürbüz, F. (2024). Constructing and providing content validity evidence through 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), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Koskinen, R., & Pitkäniemi, H. (2022). Meaningful Learning in Mathematics: A Research Synthesis of Teaching Approaches. *International Electronic Journal of Mathematics Education*, *17*(2), em0679. https://doi.org/10.29333/iejme/11715

Mahmuti, A., Hamzić, D. K., & Thaqi, X. (2025). The impact of contextual teaching and learning on improving student achievement in economic mathematics. *International Electronic Journal of Mathematics Education*, *20*(3), em0833. https://doi.org/10.29333/iejme/16233

Nabila, L. A., & Widjajanti, D. B. (2020). Self-esteem in mathematics learning: How to develop it through contextual teaching and learning approach? *Journal of Physics: Conference Series*, *1581*(1), 012049. https://doi.org/10.1088/1742-6596/1581/1/012049

Nur Qomaria, & Fatanur Baity Tsulutsya. (2024). Pengembangan E-assessment Numerasi Konteks Kearifan Lokal Madura. *Jurnal Pendidikan MIPA*, *14*(4), 1153–1160. https://doi.org/10.37630/jpm.v14i4.2087

Nursakiah, N., Arriah, F., & Dharma, S. (2022). Developing mathematical literacy test with context of Bugis-Makassar local wisdom for junior high school students. *Jurnal Elemen*, *8*(1), 16–28.

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

https://doi.org/10.29408/jel.v8i1.4049

OECD. (2023). *PISA 2022 Results Factsheets Indonesia PUBE*. https://oecdch.art/a40de1dbaf/C108.

Paunno, D., Laamena, C. M., & Moma, L. (2025). Developing Mathematics Problems Of Literacy Numeracy Based On Local Wisdom Context Of Maluku. *Jurnal Magister Pendidikan Matematika (JUMADIKA)*, *6*(2), 59–67. https://doi.org/10.30598/jumadikavol6iss2year2024page59-67

Putri, A. O. (2021). Pengembangan Tes Diagnostik Berbentuk Uraian Berdasarkan Ranah Kognitif Untuk Mengetahui Letak Kesulitan Belajar Siswa Smp Pada Materi Bangun Ruang. *Jurnal Pusaka*, *10*(1), 40–54. https://doi.org/10.35897/ps.v10i1.597

Rachmah, D. Y. (2018). *Pengembangan Instrumen Asesmen Diagnostik Untuk Melihat Pemahaman Konsep Aljabar* (Skripsi). Universitas Islam Negeri Sunan Ampel Surabaya.

Rezky, M., Hidayanto, E., & Parta, I. N. (2022). Kemampuan Literasi Numerasi Siswa Dalam Menyelesaikan Soal Konteks Sosial Budaya Pada Topik Geometri Jenjang Smp. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(2), 1548–1562. https://doi.org/10.24127/ajpm.v11i2.4879

Rizki, T. N. H., & Frentika, D. (2021). Etnomatematika dalam budaya barapan kebo sebagai inovasi pembelajaran matematika. *JRIP: Jurnal Riset Dan Inovasi Pembelajaran*, *1*(2), 252–264. https://doi.org/10.51574/jrip.v1i2.98

Simbolon, A., & Nurjanah, N. (2024). Measuring Mathematical Literacy Skills of Indonesian Junior High School Students in Solving PISA-Like Problems. *Jurnal Pendidikan MIPA*, *25*(3), 1259–1272. https://doi.org/10.23960/jpmipa/v25i3.pp1259-1272

Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan Instrumen Asesmen Diagnostik untuk Persiapan. *Journal of Community Empowerment*, *2*(2), 63–69. https://journal.unnes.ac.id/sju/index.php/jce

Susanta, A., Sumardi, H., Susanto, E., & Retnawati, H. (2023). Mathematics literacy task on number pattern using Bengkulu context for junior high school students. *Journal on Mathematics Education*, *14*(1), 85–102. https://doi.org/10.22342/JME.V14I1.PP85-102

Tamur, M., Jehadus, E., Nendi, F., Mandur, K., & Murni, V. (2020). Assessing the effectiveness of the contextual teaching and learning model on students' mathematical understanding ability: A meta-analysis study. *Journal of Physics: Conference Series*, *1657*(1). https://doi.org/10.1088/1742-6596/1657/1/012067

Templin, J., & Bradshaw, L. (2013). Measuring the Reliability of Diagnostic Classification Model Examinee Estimates. *Journal of Classification*, *30*(2), 251–275. https://doi.org/10.1007/s00357-013-9129-4

DOI: https://doi.org/10.24127/ajpm.v15i2.14615

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Indiana University.

Torre, J. D. La. (2011). The Generalized DINA Model Framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

Utariningsih, U. (2018). TPengembangan tes diagnostik kognitif berkarakter HOTS matematika di sekolah menengah pertama. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *6*(2), 171–179. http://jurnal.ustjogja.ac.id/index.php/wd

Zhan, P., Li, F., & Jiao, H. (2021). Editorial : Cognitive Diagnostic Assessment for Learning. *Frontiers in Psychology*, *12*, 806636. https://doi.org/10.3389/fpsyg.2021.806636

Zhang, Y., Zhao, B., Jian, M., & Wu, X. (2025). Cognitive diagnostic analysis of mathematics key competencies based on PISA data. *PLoS ONE*, *20*(2). https://doi.org/10.1371/journal.pone.0315539