JURNAL ABSIS
Jurnal Pendidikan Matematika dan Matematika
Vol. 8 No. 3, December 2025

DEVELOPMENT OF A MATHEMATICAL COGNITIVE ASSESSMENT INSTRUMENT ORIENTED TO LITERACY, NUMERACY, AND LOCAL CULTURE

Hariyanto[1], Sutama[2], Yulia Maftuhah Hidayati[3]
[1,2,3] Universitas Muhammadiyah Surakarta, Indonesia
a418230002@student.ums.ac.id

ABSTRACT The paradigm of mathematics assessment has shifted from memorization- and procedure-oriented approaches toward an emphasis on reasoning, problem solving, and mathematical representation. This study aimed to develop a valid and reliable numeracy literacy-oriented cognitive assessment instrument for the statistics topic, specifically measures of central tendency. The study employed a Research and Development (R&D) method using the ADDIE model and was conducted at MTs N 2 Brebes during the 2024/2025 academic year, involving 96 students. Data were collected through interviews, questionnaires, and tests and analyzed using descriptive quantitative analysis. The results show that the developed instrument has very high validity, with material expert validation scoring 0.86 and item expert validation scoring 0.87, while content validity reached 0.57, indicating a moderate category. The reliability coefficients were 0.44 for multiple-choice items (moderate) and 0.57 for descriptive items (high). The items demonstrated good psychometric characteristics, with moderate to high difficulty levels and adequate discrimination power. In terms of practicality, the instrument was rated very practical by five teachers (average score 88.67) and 66 students (average score 83.6), indicating that it is easy to use and well understood. Overall, the findings indicate that the developed assessment instrument is valid, reliable, and practical for wider implementation and contributes to the development of mathematics assessment aligned with contemporary learning paradigms that emphasize numeracy literacy at the secondary school level.

Keywords: assessment instrument, cognitive assessment, numeracy literacy, measures of central tendency

ABSTRAK Paradigma asesmen matematika telah mengalami pergeseran signifikan dari pendekatan yang berorientasi pada hafalan dan prosedural menuju penekanan pada kemampuan penalaran, pemecahan masalah, dan representasi matematis. Penelitian ini bertujuan untuk mengembangkan instrumen asesmen kognitif matematika berorientasi literasi numerasi yang valid dan reliabel pada materi statistika, khususnya ukuran pemusatan

https://doi.org/10.30606/absis.v8i3.3368

data. Penelitian ini menggunakan metode Research and Development (R&D) dengan model ADDIE dan dilaksanakan di MTs N 2 Brebes pada tahun ajaran 2024/2025 dengan melibatkan 96 peserta didik. Data dikumpulkan melalui wawancara, angket, dan tes, kemudian dianalisis menggunakan analisis deskriptif kuantitatif. Hasil penelitian menunjukkan bahwa instrumen yang dikembangkan memiliki validitas sangat tinggi, dengan nilai validasi ahli materi sebesar 0,86 dan ahli soal sebesar 0,87, serta validitas isi sebesar 0,57 yang termasuk kategori sedang. Instrumen juga menunjukkan reliabilitas yang memadai, dengan koefisien 0,44 untuk soal pilihan ganda (kategori sedang) dan 0,57 untuk soal uraian (kategori tinggi). Karakteristik butir soal menunjukkan kualitas yang baik dengan tingkat kesukaran sedang hingga tinggi dan daya pembeda yang memadai. Dari aspek kepraktisan, instrumen dinilai sangat praktis oleh lima guru dengan skor rata-rata 88,67 dan oleh 66 peserta didik dengan skor rata-rata 83,6, yang menunjukkan bahwa instrumen mudah digunakan dan dipahami. Secara keseluruhan, instrumen asesmen yang dikembangkan terbukti valid, reliabel, dan praktis untuk digunakan secara lebih luas serta berkontribusi pada pengembangan sistem asesmen matematika yang selaras dengan paradigma pembelajaran modern yang menekankan literasi numerasi di jenjang sekolah menengah.

Kata-kata kunci: instrumen asesmen, asesmen kognitif, literasi numerasi, ukuran pemusatan data

INTRODUCTION

The transformation of mathematics education over the last decade reflects a fundamental paradigmatic shift from memorization-oriented and procedural learning toward the development of reasoning, problem-solving, and mathematical representation skills. This transformation marks a new era in mathematics education and aligns with the growing emphasis on numeracy literacy, which highlights students' ability to understand, analyze, and apply mathematical information in various real-life contexts. Herbel-Eisenmann et al. (2016) emphasized that the objectives of mathematics learning should encompass five fundamental abilities: mathematical problem solving, mathematical communication, mathematical reasoning, mathematical connections, and mathematical representation. In line with this view, Ekowati et al. (2019) defined numeracy literacy as an individual's ability to use mathematical reasoning, symbols, and language to interpret quantitative information, extending beyond mere mastery of formulas and calculations.

The implementation of the Merdeka Curriculum in Indonesia further strengthens the orientation toward competency-based rather than content-based learning. Agustyaningrum et al. (2022) explained that meaningful learning achievement should be contextualized and measured through both cognitive and non-cognitive assessments reinforced by numeracy literacy, recognizing that each student experiences unique cognitive development. However, classroom realities indicate a significant gap between these ideals and actual learning practices. The Programme for International Student Assessment (PISA) 2022 results released by the Ministry of Education, Culture, Research, and Technology (2023) revealed a decline in Indonesian students' mathematics literacy scores, from 379 in 2018 to 366 in 2022. This decrease of 13 points signals a deterioration in mathematical literacy and

indicates persistent difficulties among students in understanding problem situations and constructing appropriate mathematical models to solve contextual tasks.

One of the root causes of low numeracy literacy lies in assessment systems that have not fully supported the development of students' cognitive abilities. Rohim (2021) explained that traditional evaluation instruments, such as multiple-choice tests, remain dominant and tend to measure memorization and procedural skills rather than reasoning. Teachers often focus on students' ability to recall formulas instead of fostering progress in higher-order thinking. Similarly, Fiangga et al. (2019) noted that students are generally unfamiliar with literacy-based problems, while Arifin and Retnawati (2017) found that students tend to solve tasks that emphasize memory rather than Higher Order Thinking Skills (HOTS). Rahayu and Friyatmi (2022) further identified that these conditions are exacerbated by limitations in teachers' competence to design and implement numeracy literacy-oriented assessment instruments. Insufficient understanding of numeracy literacy indicators and difficulties in linking them to minimum competency standards pose structural barriers to improving the quality of mathematics learning.

Additional studies have highlighted students' conceptual difficulties in numeracy-based problem solving. Oktavianingtyas (2015) reported that students often make errors in selecting appropriate formulas and produce unstructured solutions when dealing with numeracy literacy tasks. Perdana and Suswandari (2021) identified that students' failure in mathematics learning is largely rooted in their inability to understand underlying concepts. Furthermore, Darmayasa (2020) emphasized that existing assessment systems have not adequately accommodated the integration of local culture into mathematics learning, even though mathematics education at all levels should provide opportunities for cultural preservation through ethnomathematics contexts that are relevant to students' daily lives.

A review of previous studies on numeracy literacy-based assessment instruments also reveals several limitations. Research conducted in various countries—including the Netherlands (de Greef et al., 2015), Germany (Durda et al., 2020), Malaysia (Simamora et al., 2023), the Philippines (Dooma et al., 2024), Iran (Mohsenpur, 2015), Thailand (Chamrat et al., 2019), Latvia (France et al., 2023), and Indonesia (Apipah et al., 2023; Purnomo et al., 2022; Gradini et al., 2021)—generally indicates four major weaknesses. First, the developed instruments are not sufficiently comprehensive in evaluating students' cognitive abilities holistically within numeracy literacy contexts. Second, these instruments are rarely integrated with local culture and ethnomathematical contexts. Third, the material focus remains limited and does not adequately emphasize critical data analysis skills required to address 21st-century challenges. Fourth, assessment functions are still predominantly summative rather than formative, limiting their potential to support learning improvement through meaningful feedback.

https://doi.org/10.30606/absis.v8i3.3368

In the Indonesian context, the lack of integration of local cultural diversity into mathematics assessment presents an additional challenge. The Merdeka Curriculum's emphasis on competency-oriented learning requires assessment instruments capable of accommodating this diversity. Moreover, the need to preserve local culture through mathematics education has become an urgency that cannot be neglected. Black et al. (2003), Black and Wiliam (2009), and Randy E. Bennett and Gitomer (2009) emphasized that assessment for learning—defined as an integrated assessment system embedded within the learning process and used as a basis for instructional improvement—offers a viable solution to these challenges.

Considering the limitations of existing assessment instruments and the urgent need to improve Indonesian students' numeracy literacy, this study proposes the development of an innovative mathematical cognitive assessment instrument. The novelty of the proposed instrument lies in three main dimensions. First, it offers structural innovation by integrating cognitive assessment, numeracy literacy, and local culture into a single instrument that functions both formatively and summatively. Second, it introduces contextual innovation by incorporating ethnomathematics as the foundation for designing tasks aligned with the Merdeka Curriculum, with particular emphasis on data analysis content to prepare students for 21st-century challenges. Third, it presents methodological innovation by developing an assessment-for-learning-oriented instrument that is valid, reliable, practical, and integrable with a Learning Management System.

This study aims to develop a valid, reliable, and practical mathematics cognitive assessment instrument oriented toward numeracy literacy and integrated with local culture for formative assessment purposes. The contribution of this study is not only theoretical, as it addresses gaps in the literature on integrated cognitive assessment instruments, but also practical, as it provides teachers with usable tools to support the implementation of the Merdeka Curriculum. Furthermore, this research is expected to serve as a foundation for developing more comprehensive and contextual mathematics learning systems, ultimately contributing to the improvement of Indonesian students' numeracy literacy and the preservation of local culture through mathematics education.

METHODS

This study employed a Research and Development (R&D) method using the ADDIE model proposed by Dick and Carey, which consists of the analysis, design, development, implementation, and evaluation stages. The research was conducted at MTs N 2 Brebes during the 2024/2025 academic year and involved 96 students as research participants. The research sample was selected purposively from classes IX A, IX F, and IX G, as these classes demonstrated learning motivation and achievement levels that still required improvement.

Data were collected through observations, interviews, questionnaires, and tests. The developed cognitive assessment instrument consisted of various item formats, including multiple-choice, complex multiple-choice, matching, short-answer, and essay questions. The instrument was designed to measure students' numeracy literacy-oriented cognitive abilities, particularly their capacity to interpret and apply mathematical information across diverse contexts.

The data obtained from expert validation questionnaires—completed by subject-matter experts and assessment experts—were analyzed descriptively using percentage analysis and Aiken's V formula to determine the theoretical feasibility of the developed instrument. The validity criteria based on Aiken's V coefficient were categorized into five levels: very valid ($V > 0.84$), valid ($0.68 < V \leq 0.84$), fairly valid ($0.52 < V \leq 0.68$), less valid ($0.36 < V \leq 0.52$), and invalid ($V \leq 0.36$).

Data from the readability and practicality questionnaires were analyzed using descriptive statistics by calculating mean scores and percentages of achievement. The results were then interpreted based on practicality criteria classified into five categories: very practical ($80\% \leq P \leq 100\%$), practical ($60\% \leq P < 80\%$), sufficiently practical ($40\% \leq P < 60\%$), less practical ($20\% \leq P < 40\%$), and not practical ($0\% \leq P < 20\%$).

Furthermore, data obtained from small-scale and large-scale field trials of the cognitive assessment instrument were analyzed quantitatively using SPSS version 25. This analysis aimed to determine the validity and reliability of test items, item difficulty levels, and item discrimination indices. Instrument reliability was classified based on correlation coefficients into very high ($0.80 < r \leq 1.00$), high ($0.60 < r \leq 0.80$), moderate ($0.40 < r \leq 0.60$), low ($0.20 < r \leq 0.40$), and very low ($r \leq 0.20$).

Item difficulty levels were categorized as very difficult ($P = 0.00$), difficult ($0.00 < P \leq 0.30$), moderate ($0.30 < P \leq 0.70$), easy ($0.70 < P \leq 1.00$), and very easy ($P = 1.00$). Meanwhile, item discrimination indices were classified into very good ($0.70 < D \leq 1.00$), good ($0.40 < D \leq 0.70$), sufficient ($0.20 < D \leq 0.40$), poor ($0.00 < D \leq 0.20$), and very poor ($D \leq 0.00$). All analyses were conducted to ensure that the developed assessment instrument possessed adequate quality and could validly and reliably measure students' numeracy literacy skills.

FINDING AND DISCUSSION

The developed numeracy literacy-oriented cognitive assessment instrument was designed to evaluate the numeracy literacy skills of MTs N 2 Brebes students on the topic of measures of central tendency (data centralization). The development followed the ADDIE procedure consisting of five stages: Analyze, Design, Develop, Implement, and Evaluate (Hidayat & Nizar, 2021).

1. Analyze Stage

The analysis stage comprised curriculum analysis, analysis of student characteristics, and material analysis. Based on classroom observations and interviews, MTs N 2

https://doi.org/10.30606/absis.v8i3.3368

Brebes has implemented the Merdeka Curriculum since the 2021/2022 academic year. Several learning and assessment issues were identified: (1) the existing assessment practices were not optimal in measuring students' cognitive abilities because they mainly emphasized memorization and formula application, limited to understanding (C2) and applying (C3); (2) the assessments had not sufficiently supported students' numeracy literacy development; and (3) students' mathematics learning outcomes were not yet optimal. These findings indicated the need for an assessment instrument that could evaluate higher cognitive levels (C4–C6) and support numeracy literacy through integration with the local culture of Brebes.

Interviews with Grade IX mathematics teachers also revealed that students exhibited diverse learning styles (visual, auditory, and kinesthetic) and relatively high enthusiasm for learning mathematics; however, their learning outcomes remained below expectations. Regarding content, the material was systematically mapped to Phase D learning outcomes within the data analysis element, specifically the measures of central tendency topic. Numeracy literacy indicators were then formulated and aligned with minimum competencies as benchmarks for learning improvement.

2. Design Stage

At the design stage, an initial prototype of the assessment instrument was prepared. The prototype included the front cover, foreword, table of contents, concept map, CP and ATP mapping, central tendency material, practice tasks, distribution of numeracy literacy indicators, item specifications (test blueprint), assessment items, scoring rubrics, and an answer key. The instrument layout was developed using Canva and Microsoft PowerPoint.

Figure 1. Draft Cover of the Developed Instrument (Prototype 1).

Figure 1 presents the initial draft of the instrument cover, created prior to expert validation. The cover elements (e.g., assessment sheet icon, diagram/data icon, and student illustration) were intended to represent the focus of the instrument: cognitive assessment in the measures of central tendency topic at the junior high school level.

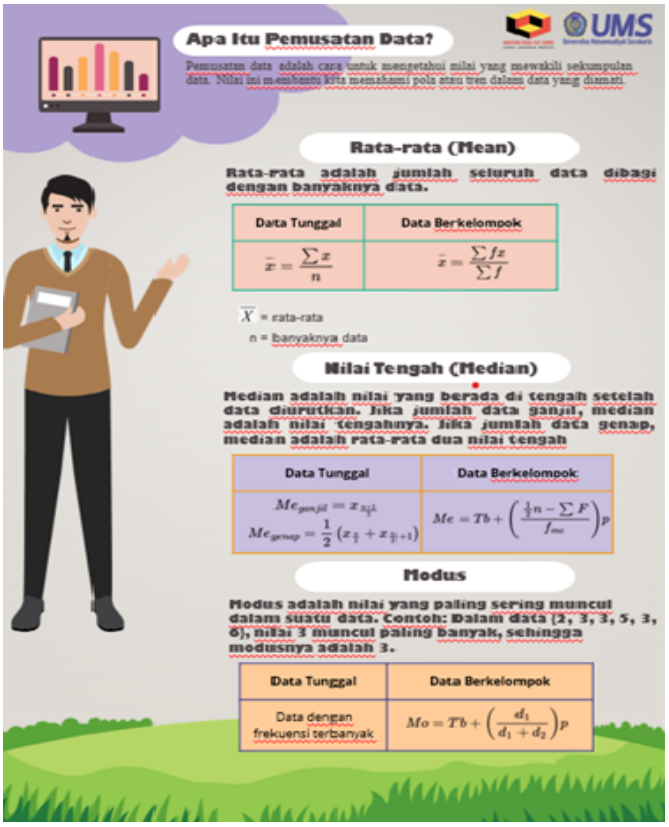

Figure 2. Sample Page of Measures of Central Tendency Material.

Figure 2 shows an example of a content page used as a reference for constructing numeracy literacy-oriented assessment tasks.

To ensure systematic alignment between content and assessment, numeracy literacy indicators were mapped to cognitive levels and item formats.

Table 1. Mapping of Numeracy Literacy Indicators, Cognitive Levels, and Item Formats (Initial Blueprint).

Numeracy literacy indicators	Target cognitive levels	Item formats	Number of items
(1) Understand and interpret numeracy information	C4–C6	Multiple-choice	3
(2) Analyze/interpret numeracy information and use	C4–C6	Complex multiple-choice	3

https://doi.org/10.30606/absis.v8i3.3368

Numeracy literacy indicators	Target cognitive levels	Item formats	Number of items
symbols/numbers in operations to solve problems			
(2) Analyze/interpret numeracy information and use symbols/numbers in operations to solve problems	C4–C6	Matching	3
(3) Evaluate and communicate results of numeracy information	C4–C6	Short-answer	3
(3) Evaluate and communicate results of numeracy information	C4–C6	Essay/constructed response	3
Total			15

Table 1 summarizes the distribution of numeracy literacy indicators for the measures of central tendency topic. Based on this blueprint, 15 items were initially developed to measure C4–C6 cognitive levels using multiple-choice, complex multiple-choice, matching, short-answer, and essay formats and integrating local cultural contexts from Brebes.


INSTRUMEN PENILAIAN KOGNITIF MATEMATIKA

BERORIENTASI LITERASI NUMERASI

I. Berilah tanda silang (x) pada huruf a, b, c, d yang kamu anggap benar pada lembar jawab yang telah disediakan!

1. Seorang guru mata pelajaran matematika sedang menganalisa hasil penilaian harian yang dilaksanakannya dikelas VIII B. Diperoleh nilai matematika dari 32 siswa sebagai berikut:

80	100	80	60	60	50	90	90
70	60	60	80	70	60	60	50
70	80	80	50	90	90	80	60
70	70	70	60	80	70	50	90

Jika data tersebut dibuat dalam bentuk prosentase berupa diagram lingkaran, maka presentasi hasilnya adalah ....

a. Nilai Matematika Siswa VIII B

b. Nilai Matematika Siswa VIII B

c. Nilai Matematika Siswa VIII B

d. Nilai Matematika Siswa VIII B

Figure 3. Example of Numeracy Literacy-Oriented Assessment Items (Prototype 1).

Figure 3 illustrates sample items developed from the blueprint. An answer key and rubric were subsequently prepared to support objective and transparent scoring.

3. Develop Stage

During the development stage, the instrument underwent expert validation by material experts and assessment (item) experts on November 18–21, 2024. The validators consisted of four experts: two lecturers from the Master's Program in Mathematics Education (FKIP UMS) and two FASDA MGMP Mathematics representatives from Brebes Regency. Revisions were made based on expert feedback, including: (a) improving font and font color on the title page; (b) revising Item 1 (initially multiple-choice) to better fit an essay format; (c) correcting notation and terminology (e.g., writing "average" in full and using "percent" appropriately); and (d) improving distractors in Item 4.

Material expert validation involved 28 indicators across three aspects (material/substance, language, and presentation).

Table 2. Material Expert Validation Framework

No.	Aspect	Sub-aspects	Key indicators (summary)
1	Material / substance	Alignment with curriculum (CP, TP)	Conformity of the instrument with learning outcomes, learning objectives, and the Merdeka Curriculum
		Numeracy literacy alignment	Consistency between numeracy literacy indicators, item blueprint, and assessment items
		Conceptual clarity	Appropriateness of cognitive levels, conceptual accuracy, item independence, relevance to daily-life contexts, and completeness of answer keys
2	Language	Linguistic correctness	Use of communicative, clear, and unambiguous language; correct diction, sentence structure, mathematical symbols, and units
3	Presentation	Illustration and layout	Clarity and relevance of illustrations, visual attractiveness, and alignment between illustrations and numeracy literacy indicators

The material expert validation results indicated that most indicators achieved very high validity, with a small number categorized as high validity.

https://doi.org/10.30606/absis.v8i3.3368

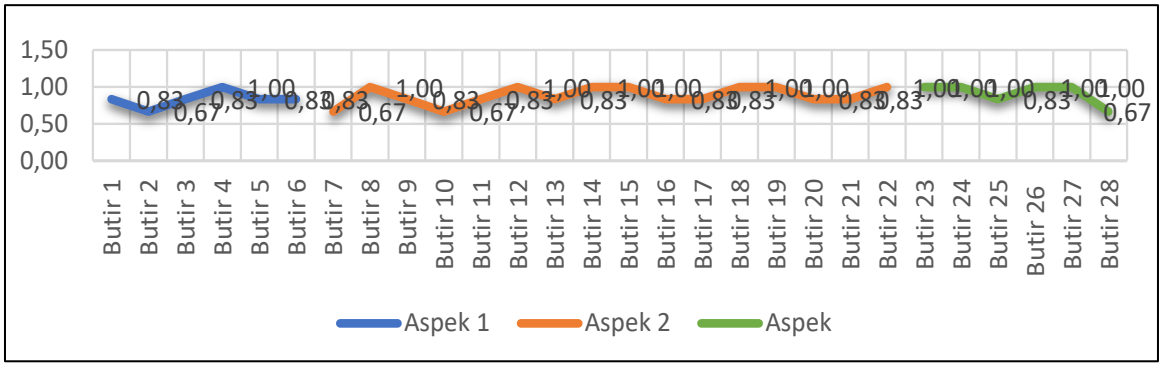

Figure 4. Material Expert Validation Results per Indicator (Aiken's V)

Figure 4 shows that only a few indicators fell into the "high" category (e.g., one indicator in the learning aspect, two in the substance/material aspect, and one in presentation), while the remaining indicators reached "very high" validity.

To provide an overall validity estimate, Aiken's V was computed across all material indicators.

Table 3. Overall Material Expert Validity (Aiken's V Summary).

Indicator range	Assessor 1	Assessor 2	Σ score (S1)	Σ score (S2)	Total ΣS	n(c–1)	Aiken's V	Validity category
Items 01–28	99	102	71	74	145	168	0.863095	Very high

The overall Aiken's V for material expert validation was 0.86, indicating very high validity.

Item expert validation assessed six indicators (e.g., clarity of instructions, alignment with indicators and competencies, cognitive level appropriateness, representation of numeracy literacy indicators, and clarity of illustrations). Item-level validity results are presented below.

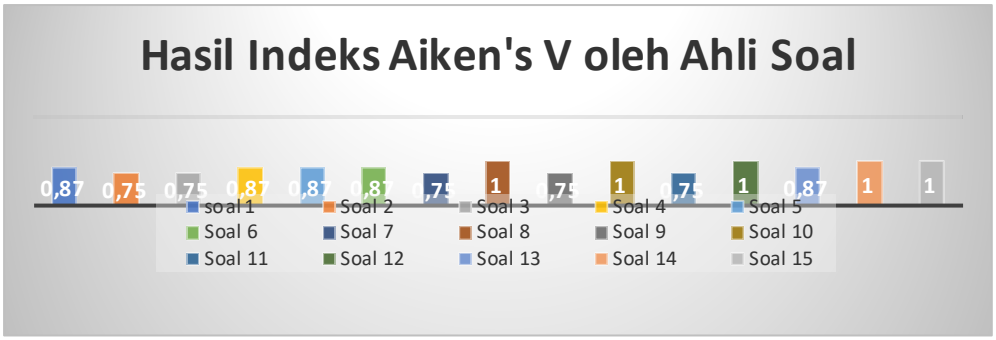

Figure 5. Item Expert Validation Results per Item (Aiken's V).

Figure 5 indicates that out of 15 items, 10 items achieved very high validity, while 5 items achieved high validity.

Table 4. Overall Item Expert Validity

Indicator range	Assessor 1	Assessor 2	Σ score (S1)	Σ score (S2)	Total ΣS	n(c−1)	Aiken's V	Validity category
Items 01-15	68	67	53	52	105	120	0.875	Very high

The overall Aiken's V for item expert validation was 0.87, which also falls in the very high category. Taken together, the results from both material and item experts confirm that the instrument met validity requirements and was appropriate for field testing.

4. Implement Stage

After revisions, the instrument was piloted in Class IX F (30 students) on November 19, 2024. Documentation of the pilot implementation is provided below.

Figure 6. Small-Scale Trial of the Numeracy Literacy-Oriented Cognitive Assessment Instrument.

Following the pilot, item analysis was conducted using SPSS version 25 to examine item validity, reliability, discrimination power, and difficulty level. Based on the small-scale trial, Item 2 showed moderate validity but low difficulty, while Item 5 showed low validity and poor discrimination; therefore, both items were excluded from the larger-scale trial. Items 7, 10, and 13 were also excluded, although they showed moderate validity and good discrimination, because they overlapped with other items in terms of learning objectives and primarily measured only C4.

Reliability estimates from the pilot indicated that objective-format items (multiple-choice, complex multiple-choice, matching) reached 0.601 (high reliability), while constructed-response items (short-answer and essay) reached 0.762 (high reliability). For difficulty levels, the 15 items consisted of 2 easy items, 12 moderate

https://doi.org/10.30606/absis.v8i3.3368

items, and 1 difficult item. For discrimination, 14 items were classified as good and 1 item as poor.

Based on these considerations, 10 items were selected for Prototype 2: Items 1 and 3 (multiple-choice), Items 4 and 6 (complex multiple-choice), Items 8 and 9 (matching), Items 11 and 12 (short-answer), and Items 14 and 15 (descriptive).

Prototype 2 was then tested on a larger scale in two classes (IX A and IX G). These classes were selected based on teacher–researcher discussions indicating that their average learning outcomes were lower than Class IX F. Larger-scale analysis (SPSS 25) indicated that item validity ranged from moderate to high, reliability was acceptable, item difficulty was mostly moderate, and discrimination power was generally good. A decrease in reliability was observed, which may be attributable to increased heterogeneity in student ability levels across classes. As noted by Crocker and Algina (2008), reliability is influenced not only by test consistency but also by participant variability.

5. Evaluate Stage

Formative evaluation was conducted throughout the Analyze–Design–Develop stages to support continuous improvement. A summative evaluation of feasibility and practicality was conducted using teacher and student response questionnaires.

Practicality testing by five mathematics teachers resulted in an average score of 88.67, categorized as very practical. The lowest rating was related to adequacy of time allocation relative to students' ability levels, while the highest rating was related to the clarity of test instructions.

Student readability and practicality were evaluated after the trials and broader implementation involving 99 students. The results are summarized below.

Table 5. Overall Student Readability and Practicality Scores of the Numeracy Literacy-Oriented Cognitive Assessment Instrument

Component	Trial (%)	Usage test (%)	Total (%)
Readability & practicality score	82.60	84.60	83.60

Table 6. Top Five Highest Student Readability and Practicality Items

Rank	Item No.	Statement (short label)	Total (%)
1	14	Images/graphics are clear and help understanding	91.18
2	2	I understood the material before doing the test	90.17
3	1	The math concepts in the test were learned before	90.03

Rank	Item No.	Statement (short label)	Total (%)
4	4	Similar to tasks discussed in class	89.09
5	5	Tasks/answers motivate deeper learning	88.50

Table 7. Bottom Five Lowest Student Readability and Practicality Items

Rank	Item No.	Statement (short label)	Total (%)
1	7	Material in the test is more difficult than examples	74.72
2	19	I repeatedly read and think to solve numeracy items	75.63
3	30	Time provided is sufficient	76.25
4	20	I am not yet used to numeracy literacy problems	76.39
5	21	I feel more confident facing numeracy items after this	78.44

Overall, the student readability and practicality score was 83.6, indicating the instrument was very practical.

Overall Product Quality

Overall, the results demonstrate that the developed numeracy literacy-oriented cognitive assessment instrument is valid, reliable, and practical. Expert validation yielded Aiken's V coefficients of 0.86 (material) and 0.87 (items), both in the very high category. High validity indicates that the items align with learning indicators and the intended numeracy literacy constructs (Retnawati, 2016). Reliability testing across trial phases showed slight differences but remained within acceptable ranges; this variation is plausibly explained by differences in student ability distribution across trial groups (Crocker & Algina, 2008).

The instrument also includes five item formats (multiple-choice, complex multiple-choice, matching, short-answer, and essay), enabling measurement across a range of cognitive levels and numeracy skills. In particular, essay and descriptive items are appropriate for assessing higher-order thinking such as evaluating and creating (C5–C6), consistent with Anderson and Krathwohl's (2001) revision of Bloom's Taxonomy and the use of open-ended formats to assess complex cognition (Wilson, 2016).

A key feature of this instrument is the integration of Brebes local culture into item contexts (e.g., traditional market practices and local commodities). This contextualization supports meaningful learning by linking abstract mathematics to

https://doi.org/10.30606/absis.v8i3.3368

students' lived experiences and contributes to local cultural preservation through ethnomathematics-based contexts (D'Ambrosio, 2020). Teacher (88.67) and student (83.6) practicality results further indicate that the instrument is understandable, feasible for classroom use, and supported by clear illustrations and communicative language. However, some students reported that numeracy literacy items were challenging and unfamiliar, suggesting the need for continued habituation and scaffolding.

Finally, one limitation of this development is that the study has not yet tested the instrument's effectiveness in improving students' learning outcomes, nor has it compared performance across groups experiencing different instructional approaches.

CONCLUSIONS AND RECOMMENDATIONS

This study successfully developed a numeracy literacy-oriented mathematical cognitive assessment instrument for the topic of measures of central tendency. Using a Research and Development (R&D) approach, the instrument demonstrated high validity, with Aiken's V values ranging from 0.863 to 0.875, acceptable reliability coefficients between 0.440 and 0.572, and very high practicality based on teacher (88.67%) and student (84.59%) evaluations. The ten selected items exhibited appropriate psychometric qualities, including moderate difficulty levels and adequate discrimination power, indicating that the instrument is suitable for measuring students' numeracy literacy competencies in accordance with the specified learning objectives.

The main contribution of this study lies in the integration of local cultural contexts into mathematics assessment, which helps students relate mathematical concepts to their everyday experiences and supports more meaningful, contextual learning. However, this study was limited to a single mathematical topic and a specific research setting, which may restrict the generalizability of the findings. Therefore, future research is recommended to expand the development of similar instruments to other mathematical topics, apply them in broader educational contexts, and investigate their long-term effects on students' numeracy literacy development.

REFERENCES

Agustyaningrum, N., Pradanti, P., & Yuliana. (2022). Piaget and Vygotsky's development theory: Implications for elementary school mathematics learning. *Absis Journal: Journal of Mathematics and Mathematics Education*, *5*(1), 568–582. https://doi.org/10.30606/absis.v5i1.1440

Apipah, I., Nindiasari, H., & Sukirwan. (2023). Development of numeracy literacy question instruments on number material to improve higher-order thinking skills of Grade VIII MTs students. *Jurnal Cendekia: Journal of Mathematics Education*, *7*(3), 3083–3092. https://doi.org/10.31004/cendekia.v7i3.2606

Arifin, Z., & Retnawati, H. (2017). Development of a measuring instrument for higher-order thinking skills in mathematics for Grade X high school students. *Pythagoras: Journal of Mathematics Education*, *12*(1), 98–109. https://doi.org/10.21831/pg.v12i1.14058

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press.

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

Chamrat, S., Manokarn, M., & Thammaprateep, J. (2019). STEM literacy questionnaire as an instrument for STEM education research: Development, implementation, and utility. *AIP Conference Proceedings*, *2081*, 030004. https://doi.org/10.1063/1.5094011

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

D'Ambrosio, U. (2020). What is ethnomathematics, and how can it help children in schools? *Teaching Children Mathematics*, *7*(6), 308–310. https://doi.org/10.5951/tcm.7.6.0308

Darmayasa, J. B. (2020). Ethnomathematics: Mathematical concepts in making and using klakat. *PRISMA: Prosiding Seminar Nasional Matematika*, *3*, 252–257. https://journal.unnes.ac.id/sju/index.php/prisma/

de Greef, M., Segers, M., Nijhuis, J., Lam, J. F., van Groenestijn, M., van Hoek, F., van Deursen, A. J. A. M., Bohnenn, E., & Tubbing, M. (2015). The development and validation of testing materials for literacy, numeracy, and digital skills in a Dutch context. *International Review of Education*, *61*(5), 655–671. https://doi.org/10.1007/s11159-015-9519-4

Dooma, J. L. E., Mantes, J., Misajon, R. M., De Mesa, J. R., Dandan, C. J. A., Santos, J. M., & Faustino, J. B. (2024). Development of MATH-erials for teaching numeracy in early childhood. *International Journal of Research Publications and Reviews*, *5*(5), 5529–5542. https://doi.org/10.55248/gengpi.5.0524.1261

Durda, T., Artelt, C., Lechner, C. M., Rammstedt, B., & Wicht, A. (2020). Proficiency level descriptors for low reading proficiency: An integrative process model. *International Review of Education*, *66*(2–3), 211–233. https://doi.org/10.1007/s11159-020-09834-1

Ekowati, D. W., Astuti, Y. P., Utami, I. W. P., Mukhlishina, I., & Suwandayani, B. I. (2019). Numeracy literacy in Muhammadiyah elementary schools. *ELSE (Elementary School Education Journal)*, *3*(4), 93–103.

https://doi.org/10.30606/absis.v8i3.3368

Fiangga, S., Amin, S. M., Khabibah, S., Ekawati, R., & Prihartiwi, N. R. (2019). Writing numeracy literacy questions for elementary school teachers in Ponorogo Regency. *Anugerah Journal*, *1*(1), 9–18. https://doi.org/10.31629/anugerah.v1i1.1631

France, I., Mikite, M., Burgmanis, G., & Namsone, D. (2023). The development of a numeracy test using a three-dimensional framework to assess Grade 7 numeracy skills. In *Proceedings of the ATEE Conference* (pp. 629–641). https://doi.org/10.22364/atee.2022.42

Gradini, E., Firmansyah, B., & Saputra, E. (2021). Designing a mathematical literacy test using PISA-like questions in local cultural contexts. *Al Qalasadi Scientific Journal of Mathematics Education*, *5*(1), 29–43. https://doi.org/10.32505/qalasadi.v5i1.2945

Herbel-Eisenmann, B., Sinclair, N., Chval, K. B., Clements, D. H., Civil, M., Pape, S. J., Stephan, M., Wanko, J. J., & Wilkerson, T. L. (2016). Positioning mathematics education researchers to influence storylines. *Journal for Research in Mathematics Education*, *47*(2), 102–117. https://doi.org/10.5951/jresematheduc.47.2.0102

Hidayat, F., & Nizar, M. (2021). ADDIE model in Islamic education learning. *Jurnal UIN*, *1*(1), 28–37.

Ministry of Education, Culture, Research and Technology. (2023). *PISA 2022 results: Indonesia country report*. https://www.kemdikbud.go.id/

Mohsenpur, M., et al. (2015). Designing and developing a test for cognitive competencies of Iranian students' mathematics literacy based on PISA studies. *Journal of Theory and Practice in Curriculum*, *4*(2), 5–34.

Oktavianingtyas, E. (2015). Media to make basic arithmetic operations learning effective for elementary school students. *Jurnal Pancaran*, *4*(4), 207–218.

Perdana, R., & Suswandari, M. (2021). Numeracy literacy in thematic learning of upper elementary school students. *Absis: Mathematics Education Journal*, *3*(1), 9–18. https://doi.org/10.32585/absis.v3i1.1385

Purnomo, H., Sa'dijah, C., Hidayanto, E., Sisworo, Permadi, H., & Anwar, L. (2022). Development of a numeracy skills test instrument for the minimum competency assessment in Indonesia. *International Journal of Instruction*, *15*(3), 635–648. https://doi.org/10.29333/iji.2022.15335a

Retnawati, H. (2016). *Quantitative analysis of research instruments*. Parama Publishing.

Rahayu, S., & Friyatmi, F. (2022). E-test: An alternative online assessment for vocational high school students. *Al-Ishlah: Jurnal Pendidikan*, *14*(3), 3821–3828. https://doi.org/10.35445/alishlah.v14i3.1766

Bennett, R. E., & Gitomer, D. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment, and professional support. In *Educational assessment in the 21st century* (pp. 43–61). Springer. https://doi.org/10.1007/978-1-4020-9964-9

Rohim, D. C. (2021). Minimum competency assessment to improve elementary school students' numeracy literacy skills. *Varidika Journal*, *33*(1), 54–62. https://doi.org/10.23917/varidika.v33i1.14993

Simamora, Y., Saragih, S., & Dewi, I. (2023). An instrument to test students' mathematical literacy skills in plane geometry based on Malay culture. *EAI Endorsed Transactions on e-Learning*, 1–7. https://doi.org/10.4108/eai.19-9-2023.2340514

Wilson, L. O. (2016). Anderson and Krathwohl's Bloom's taxonomy revised. *The Second Principle*, *1*(1), 1–8.

Purba, Y. O. (2021). *Educational research instrument testing techniques*. Widini Bhakti Persada.

https://doi.org/10.30606/absis.v8i3.3368