EduMa: Mathematics Education Learning And Teaching
July 2025, Vol 14 No 1 Page 330 – 343
https://syekhnurjati.ac.id/jurnal/index.php/eduma/eduma/article/view/21581
p-ISSN: 2086-3918, e-ISSN: 2502-5209
Published by Tadris Matematika UIN Siber Syekh Nurjati cirebon

EduMa
MATHEMATICS EDUCATION LEARNING AND TEACHING

Developing Algebra-Based Test Items for Measuring Students' Mathematical Literacy

Handayani Budi Utami 1[1*], Ellis Salsabila[2], Eti Dwi Wiraningsih[3], Flavia A. Hidajat[4]

1,2,3,4 Universitas Negeri Jakarta

*Corresponding author: Jln.Rawamangun Muka Pulogabung Jakarta, Indonesia;
e-mail addresses: handayaniutami1@gmail.com

article info

How to cite this article:

Utami,et. al. (2025). Developing Algebra-Based Test Items for Measuring Students' Mathematical Literacy. *Eduma : Mathematics Education Learning And Teaching*, 14(1), 330 – 343.
doi:http://dx.doi.org/10.24235/eduma.v14i1.21581

Article history:

Received: 07 01, 2025

Accepted: 07 31, 2025

Published: 07, 2025

abstract

This study aims to develop valid and reliable algebra questions to assess students' mathematical literacy skills. Mathematical literacy is a key indicator in international assessments such as PISA, which emphasizes students' ability to apply mathematical concepts in real-world contexts. The research method used was development research (R&D) employing the 4D model: Define, Design, Develop, and Disseminate (implemented until the Develop stage). The instrument was constructed based on mathematical literacy indicators and validated through expert judgment using Aiken's V. Empirical validation was conducted using Pearson Product Moment correlation and the reliability was tested using Cronbach's Alpha. The results showed that all question items had correlation coefficients ranging from 0.513 to 0.836, exceeding the critical value of r (0.221), indicating good construct validity. The reliability test yielded a Cronbach's Alpha of 0.712, suggesting strong internal consistency. These findings confirm that the developed algebra questions can be effectively used as an instrument to assess students' mathematical literacy skills in a valid and reliable manner.

Keywords:
instrument development; algebra; mathematical literacy; validity; reliability

INTRODUCTION

Mathematical literacy has increasingly become a global concern (Goos & O'Sullivan, 2023). It plays a crucial role in enhancing the quality of education and preparing students to face the challenges of the modern world (Colwell & Enderson, 2016). According to the results of the Programme for International Student Assessment (PISA), conducted by the Organisation for Economic Co-operation and Development (OECD), mathematical literacy is regarded as a key indicator that reflects the educational standards of a country (Hwang et al., 2018).

PISA assesses the ability of 15-year-old students to apply their skills in mathematics, reading, and science to solve real-world problems (OECD, 2023). The assessment not only evaluates students' understanding of fundamental mathematical concepts but also their ability to apply those concepts in various life contexts. For instance, it tests students' competencies in interpreting graphs, making financial plans, and analyzing statistical data (Colwell & Enderson, 2016).

PISA results indicate that countries with high levels of mathematical literacy—such as Finland, Singapore, and South Korea—generally have education systems that support problem-based learning and deep conceptual understanding (Cheung, 2017). Conversely, several lower-performing countries, including Indonesia, face ongoing challenges in improving their educational competitiveness (Faridah et al., 2022). In response, Indonesia has integrated mathematical literacy into its national education policy through the implementation of the Minimum Competency Assessment (AKM), which replaced the national examination in 2021. The AKM focuses on literacy and numeracy, where numeracy refers to the application of mathematical literacy in real-life contexts (PG Dikdas, 2020).

However, despite the implementation of various approaches, improving mathematical literacy remains a significant challenge for both developed and developing countries. Research shows that students often struggle to understand mathematical concepts (Wijaya et al., 2014), connect them to real-life situations (Hwang et al., 2018), and use them to solve complex problems (Maslihah et al., 2020). These challenges reflect a misalignment between school curricula and the skills required in everyday life (Goos & O'Sullivan, 2023).

The gap between conceptual understanding and contextual application highlights the need for systematic efforts to assess and develop students' mathematical literacy(Wijaya et al., 2014). One way to achieve this is by developing assessment instruments that align with the characteristics of mathematical literacy, as defined by PISA. However, mathematics questions used in schools are still largely procedural and memorization-based, rather than focused on applying concepts in real-world contexts (Oktiningrum & Hartono, 2016). This condition hinders the development of students' mathematical literacy competencies.

Within the national curriculum, algebra is one of the essential topics in mathematics education, with significant potential to be connected to everyday situations—such as budgeting, data analysis, and mathematical modeling(Natsir et al., 2022). Nevertheless, algebra questions used in classroom instruction and assessments are often limited to conventional formats, such as solving equations or substituting values, without placing students in challenging and meaningful contexts. To effectively assess mathematical literacy, questions should be designed to evaluate reasoning, representation, communication, and the use of mathematics in personal, societal, and scientific contexts (OECD, 2023).

Furthermore, previous studies have noted that the development of contextual, literacy-based mathematics tasks has not yet become a common practice in Indonesian education (Husni & Herman, 2024). This situation is exacerbated by the limited availability of valid and reliable assessment tools that explicitly measure mathematical literacy in the domain of algebra. As a result, teachers face difficulties in obtaining an accurate picture of students' literacy abilities on this topic, which affects the precision of instructional planning and interventions.

Based on the above background, there is a clear need to develop algebra questions that authentically and comprehensively assess students' mathematical literacy. This development is expected not only to provide a more representative evaluation tool for students' competencies but also to serve as a reference for teachers in designing literacy-oriented instruction. Therefore, this study focuses on the development of mathematical literacy-based algebra tasks as a contribution to improving the quality of mathematics assessment and instruction in Indonesia.

LITERATURE REVIEW

Mathematical literacy as a concept began to gain widespread attention toward the end of the 20th century, in line with the growing demand for critical thinking skills to address the challenges of the modern world (Machaba, 2017). Previously, mathematics education had primarily focused on procedural mastery and numerical calculations (Hillman, 2014). However, with ongoing social, economic, and technological transformations, there has been a growing awareness that mathematical competence involves more than solving abstract problems—it requires the ability to apply mathematical knowledge in real-life situations (Stacey & Turner, 2015). This shift has led to the development of broader and more contextual definitions of mathematical literacy.

In response to the emerging needs in mathematics education, the National Council of Teachers of Mathematics (NCTM) introduced one of the earliest milestones in defining mathematical literacy in 1989. The NCTM described mathematical literacy as the ability of individuals to explore, conjecture, and reason logically, as well as to use a variety of mathematical methods effectively to solve problems (Stacey & Turner, 2015). This definition emphasizes the importance of applying mathematical knowledge to enhance thinking skills and support informed decision-making.

Mathematical literacy, as articulated by the OECD through successive PISA cycles, highlights the importance of students' ability to formulate, use, and interpret mathematics in a variety of real-world contexts (OECD, 2023). This concept aligns with the demands of the 21st century, which require citizens who are reflective, critical, and constructive thinkers (Bender, 2012). A central aspect of mathematical literacy is the ability to model contextual problems into mathematical form, apply relevant concepts, procedures, and facts to derive solutions, and interpret the results back into the original context (Hillman, 2014). This process relies heavily on a solid understanding of mathematical content—

particularly algebra, which serves as a foundation for symbolic representation and generalization in problem solving.

In line with these definitions, various scholars have contributed additional perspectives to enrich the concept of mathematical literacy. Machaba (2018) describes mathematical literacy as an individual's preparedness to face mathematical challenges across both mathematical and non-mathematical situations. Wijaya (2016) adds that mathematical literacy involves the ability to understand, analyze, evaluate, and model problems mathematically in order to find effective solutions. Stacey and Turner (2015) further emphasize the role of mathematical literacy in supporting reflective and constructive decision-making.

The definition provided by the OECD underscores an individual's capacity to formulate, employ, and interpret mathematics. These three keywords—formulate, employ, and interpret—offer a meaningful and practical structure for organizing mathematical processes (OECD, 2017). This framework describes what individuals do to bridge the context of a problem with the relevant mathematics needed to resolve it.

Algebra is one of the key elements in the mathematics curriculum that supports the development of mathematical literacy (Natsir et al., 2022). Mastery of algebra enables students to think abstractly and to translate concrete situations into mathematical models. Therefore, the development of mathematical literacy-based algebra tasks must take these capabilities into full consideration. To design items that are not only content-valid but also authentically measure literacy skills, this study adopts the 4D development model proposed by Thiagarajan (1974), this model comprises four stages: Define, Design, Develop, and Disseminate, offering a systematic framework for designing, testing, and refining assessment items. Through this approach, the development of algebra tasks aims not only to assess procedural knowledge but also to evaluate students' abilities to understand, apply, and reflect upon mathematics in real-life contexts.

METHODS

Population and Sample

The population in this study consisted of 11th-grade students from several private high schools in East Bekasi, Indonesia. The sample was selected purposively, comprising students who were willing to participate in the instrument tryout. A total of 79 students were involved in the initial trial of the instrument for the purpose of empirical validation. This number was considered sufficient to examine the validity and reliability during the early stage of instrument development.

Reseach Design

This study is a Research and Development (R&D) study aimed at developing a test instrument to measure students' mathematical literacy, specifically within the domain of algebra. The development model used is the 4D model proposed by Thiagarajan (1974), which consists of four stages: Define, Design, Develop, and Disseminate. However, due to time constraints and the scope of the study, only the first three stages—Define, Design, and Develop—were implemented.

In the Define stage, a needs analysis was conducted through curriculum review, examination of the PISA mathematical literacy framework, and relevant literature to determine the indicators and constructs to be measured. The results of this stage served as the foundation for constructing the test blueprint. The Design stage involved the development of test items based on the predetermined indicators, ensuring contextual relevance and appropriate cognitive levels. The items were constructed in the form of short-answer questions. The Develop stage included content validation by five experts—comprising university lecturers and mathematics teachers—and a limited-scale trial to examine the instrument's empirical validity and reliability. The Disseminate stage was not carried out due to time limitations, and the study was confined to the initial tryout phase.

Frame Work Flow

This study implemented the 4D development model, consisting of the stages Define, Design, and Develop, to construct a valid and reliable test instrument for assessing students' mathematical literacy in the algebra domain. The Disseminate stage was excluded due to time and scope constraints. This section outlines the operational procedures of each development phase along with the data collection and analysis methods employed.

1. Define Stage
   The Define stage aimed to identify the foundational requirements for developing a mathematical literacy assessment instrument. Activities included curriculum analysis, a review of the PISA mathematical literacy framework, and examination of relevant literature to extract measurable indicators. Additionally, interviews with high school mathematics teachers in East Bekasi were conducted to gather practical insights regarding the lack of context-based assessment tools. Findings from this stage served as the basis for constructing the test blueprint and defining content indicators aligned with the PISA framework.
2. Design Stage
   At this stage, test items were constructed based on the indicators formulated during the Define stage. The items were developed in short-answer format to authentically assess students' reasoning and contextual problem-solving skills. Each item was aligned with the defined cognitive processes and real-world contexts, ensuring a balanced coverage of content and skill domains. A scoring rubric was also developed for each item to guide consistent assessment.
3. Develop Stage
   During the Develop stage, the constructed items were validated by five experts, consisting of mathematics education lecturers and secondary school teachers. The validation process assessed the relevance of content, construction, and readability of the items. Expert judgments were analyzed using Aiken's V formula (Aiken, 1980) to determine the level of content validity:

$$V = \frac{\Sigma S}{n(C-1)}$$

   Keterangan:
   S : R – Lo
   V : Aiken's index,
   R : score given by the validator
   Lo : lowest score in the rating scale (1)
   C : highest score in the rating scale (5)

n : number of raters.

Each item was rated on a five-point scale, with the following criteria (Table 1):

Table 1

Aiken's V Score Criteria

Score	Description
1	Not valid
2	Less valid
3	Fairly valid
4	Valid
5	Very valid

The resulting Aiken's V index was then compared with the critical value table to determine whether each item met the criteria for content validity.

In addition to expert validation, empirical validity testing was conducted by calculating the Pearson Product Moment correlation coefficient between each item score and the total score using SPSS (Melati et al., 2023). The formula used was:

$$r_{xy}=\frac{N(\sum XY)-(\sum X)(\sum Y)}{\sqrt{\{N(\sum X^2)-(\sum X)^2\}\{N(\sum Y^2)(\sum Y)^2\}}}$$

Where:

r_{xy} : correlation coefficient between the item score (X) and the total score (Y)

X : individual item score

Y : total score per respondent

An item was considered empirically valid if the calculated r value exceeded the critical value at the 5% significance level or if the p-value was less than 0.05.

Once validity was established, the reliability of the instrument was tested to measure internal consistency among items. Reliability was calculated using Cronbach's Alpha formula.

$$r_{11}=\left(\frac{n}{n-1}\right)(1-\frac{\sum \sigma_i^2}{\sigma_i^2})$$

The reliability coefficient was interpreted based on standard criteria (see Table 2), where higher values indicated stronger internal consistency.

Table 2

Interpretation of Correlation Coefficients

Correlation Coefficient Interval	Strength of Relationship
0.00 – 0.199	Very Weak
0.20 – 0.399	Weak
0.40 – 0.599	Moderate
0.60 – 0.799	Strong
0.80 – 1.000	Very Strong

All analyses were conducted using Microsoft Excel and SPSS software. The researcher independently managed the instrument development, validation coordination, data collection, and statistical analysis to ensure methodological integrity.

4. Disseminate Stage

The Disseminate stage was not conducted in this study. The developed and revised instrument was used only in a limited context as part of the initial pilot testing.

RESULT AND DISCUSSION

This section presents the results from each phase of the development process of the mathematics literacy test instrument, along with a discussion of the findings. The development process followed the 4D model, specifically the Define, Design, and Develop stages.

Define Stage

The Define stage aimed to identify and articulate the fundamental needs for developing a mathematical literacy assessment instrument. Key activities at this stage included curriculum analysis, literature review, and interviews with mathematics teachers. Document analysis focused on the national curriculum and the mathematical literacy framework used in international assessments such as PISA. The analysis revealed that mathematical literacy is a critical competence that needs to be cultivated in the classroom, as it directly relates to students' ability to solve real-world problems and make data-based decisions.

In addition, interviews with mathematics teachers from senior high schools in East Bekasi were conducted to explore practical needs in the field. The interviews indicated that teachers often struggled to find assessment tools that could validly and reliably measure students' mathematical literacy skills. Teachers also noted that most test items used in classroom evaluations were procedural and rarely embedded in real-life contexts. One teacher remarked, "Contextual problems are usually in the form of stories that don't reflect real issues, so students don't see the value in solving them." This feedback emphasized the importance of developing contextual problems that are not merely narrative but also relevant and meaningful to students' daily experiences. These findings strengthened the rationale for designing a mathematical literacy instrument aligned with both contextual relevance and expected competency indicators.

Based on curriculum analysis, PISA documents, and teacher input, the researcher formulated a set of indicators to guide item development. These indicators are grounded in the mathematical literacy framework developed by the OECD through the Programme for International Student Assessment (PISA), which defines mathematical literacy as the individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. This includes mathematical reasoning and the use of concepts, procedures, facts, and tools to describe, explain, and predict phenomena (OECD, 2023).

The three main processes in the PISA framework used as the foundation for item development are:

1. Formulating situations mathematically
2. Employing mathematical concepts, facts, procedures, and reasoning
3. Interpreting, applying, and evaluating mathematical outcomes

These three processes served as the basis for the test blueprint and guided the construction of test items aimed at capturing students' mathematical literacy holistically.

Design Stage

The purpose of the Design stage was to create an initial draft of the test instrument based on the indicators and needs identified in the Define stage. A test specification table (blueprint) was developed to ensure alignment between indicators, mathematical content, cognitive processes, and real-life contexts. The instrument took the form of short-answer questions, as this format was considered more suitable for revealing students' reasoning and problem-solving processes in authentic contexts.

Each item was developed to align with the three core processes from the PISA framework: formulating, employing, and interpreting mathematics. A scoring rubric was also designed for each item to guide consistent and fair evaluation. The item blueprint (see Table 3) ensured a balanced coverage of content and cognitive skills and served as the foundation for validation and field testing in the next phase.

Table 3
Instrument Blueprint

Learning Objective	Content	Mathematical Literacy Indicators	Item No.	Score
Solve problems related to linear equations in one variable.	"Belanja di Luar Negri " (Context: Personal)	1. Formulating situations mathematically 2. Employing mathematical concepts, facts, procedures, and reasoning 3. Interpreting, applying, and evaluating mathematical outcomes	1 2	15 15
Identify and solve contextual problems related to systems of linear equations in two variables.	"Transportasi Bandara" (Context: Social)	1. Formulating situations mathematically 2. Employing mathematical concepts, facts, procedures, and reasoning 3. Interpreting, applying, and evaluating mathematical outcomes	3 4 5	20 20 30

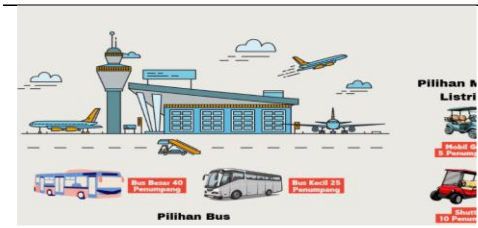
Develop Stage

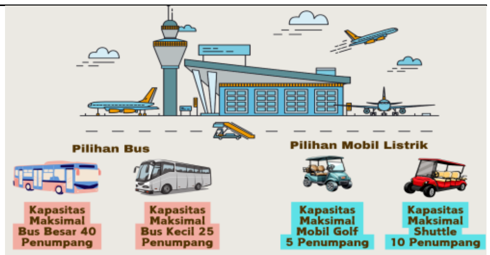
In the Development stage, the researcher began drafting the mathematical literacy test instrument based on the blueprint developed during the Design stage. The draft instrument was then submitted to several experts, including mathematics lecturers and secondary school teachers, for review regarding content validity, language clarity, and the appropriateness of each test item. Feedback and suggestions from these experts were collected and analyzed to determine which items needed improvement. Based on these inputs, the researcher revised several test items. A consolidated summary of the expert suggestions and the corresponding item revisions is presented in Table 4.

Table 4
Expert Feedback and Item Revisions

Before Revision	Suggestion	After Revision
"Before traveling abroad, many people exchange money into foreign currencies. Ani went on vacation with her friend to Japan. Before leaving, she exchanged her money." *"Sebelum bepergian ke luar negeri, banyak orang menukarkan uang ke mata uang asing. Ani berlibur*	Provide clarification that the exchanged currency is only from rupiah to yen.	"Before traveling abroad, many people exchange their rupiah into the currency of their destination country. Ani, who was going on a vacation to Japan with her friend, exchanged some money into yen before departure." *"Sebelum bepergian ke luar negeri, banyak orang menukarkan rupiah ke mata uang negara tujuan. Ani, yang akan berlibur ke Jepang bersama*

<table>
<tr><th>Before Revision</th><th>Suggestion</th><th>After Revision</th></tr>
<tr><td>dengan temannya ke Jepang. Sebelum berangkat ia menukarkan uangnya"</td><td></td><td>temannya, menukarkan sejumlah uang ke yen sebelum keberangkatan."</td></tr>
<tr><td>"What was the exchange rate at the time Ani bought the shoes? Would the shoes have been cheaper if she had paid with the cash she already exchanged earlier?" "Berapakah nilai tukar yang berlaku saat Ani melakukan pembelian sepatu tersebut? Apakah sepatu tersebut akan lebih murah, seandainya ia bisa membayar dengan uang tunai yang sudah dimilikinya?"</td><td>Ensure that the term "cash" refers specifically to cash in yen</td><td>"What was the exchange rate from yen to rupiah at the time of Ani's shoe purchase if the shoes were priced at ¥27,500? In your opinion, would the shoes have been cheaper if Ani had paid using the yen cash she had previously exchanged in Indonesia? Support your answer with a calculation." "Berapakah nilai tukar yen ke rupiah yang berlaku saat transaksi pembelian sepatu tersebut jika harga sepatu itu adalah ¥27.500? Menurut Anda, jika Ani membayar sepatu tersebut dengan uang tunai yen yang telah ditukarkannya di Indonesia, apakah harganya akan lebih murah? Jelaskan dengan perhitungan."</td></tr>
<tr><td>"Passengers will take a bus to the terminal after disembarking from the plane. The airport operates two types of buses: large buses and small buses. Given that some gates are quite far from the check-in area, electric vehicles serve as a faster and more convenient transportation alternative. There are two types of electric vehicles: electric golf carts and electric shuttles." "penumpang akan menggunakan bus untuk menuju terminal setelah turun dari pesawat. Bandara mengoperasikan dua jenis bus: bus besar dan bus kecil.

"Mengingat beberapa gate berada cukup jauh dari area check-in, mobil listrik ini menjadi alternatif transportasi yang lebih cepat dan nyaman. Ada dua jenis mobil listrik: mobil golf listrik dan shuttle listrik."</td><td>Clarify that there are two different types of transportation, each used in different areas of the airport.</td><td>"At Nusantara Airport, various transportation services are provided to facilitate passenger mobility, both inside the terminal building (indoor) and around the runway area (outdoor). The airport operates two types of buses, large and small, with different passenger capacities. In addition, for mobility within the terminal, electric vehicles are available to transport passengers from the check-in area to the departure gate." "Di Bandara Nusantara, tersedia berbagai layanan transportasi untuk memudahkan mobilitas penumpang, baik di dalam area bandara (indoor) maupun di area dekat landasan (outdoor). Bandara mengoperasikan dua jenis bus, yaitu bus besar dan bus kecil, yang memiliki kapasitas angkut berbeda. Selain itu, untuk mobilitas dalam gedung bandara, tersedia mobil listrik yang digunakan untuk mengantar penumpang dari area check-in ke gate keberangkatan."</td></tr>
<tr><td>

Is it possible to transport all passengers from both planes in a single trip, without mixing passengers from Plane X and Plane Y in the same bus? Explain your answer with a calculation. Apakah memungkinkan untuk mengangkut semua penumpang dari kedua pesawat dalam satu kali perjalanan, tanpa mencampur penumpang pesawat X dan Y di bus</td><td>Ensure that the number of bus passengers mentioned is consistent with the initial narrative, to make it more contextual, since the number of passengers on airport buses often does not match the available seating capacity.</td><td>

Is it possible to transport all passengers from both airplanes in a single trip without mixing passengers from Plane X and Plane Y in the same bus? Explain your answer with calculations. (Assume that each bus carries passengers up to its maximum capacity.) Apakah memungkinkan untuk mengangkut semua penumpang dari</td></tr>
</table>

Before Revision	Suggestion	After Revision
yang sama? Jelaskan dengan perhitungan.		*kedua pesawat dalam satu kali perjalanan, tanpa mencampur penumpang pesawat X dan Y di bus yang sama? Jelaskan dengan perhitungan. (Dengan asumsi bahwa bus membawa penumpang dengan kapasitas maksimum yang sesuai)*

After the revisions were made based on expert feedback, a content validity assessment was conducted using the Aiken's V method to quantify the level of agreement among the experts regarding each test item. This method evaluates the extent to which each item is deemed relevant, clear, and appropriate according to expert judgment. Following the revisions, content validity was measured using Aiken's V. The results showed that all five items met the criteria for valid content, with Aiken's V values ranging from 0.925 to 1.000 (see Table 5). These results suggest a strong consensus among the experts that the test items were appropriate and aligned with the intended indicators of mathematical literacy (Wilsa et al., 2023). The detailed results of the Aiken's V analysis are presented in Table 5.

Table 5
Aiken's V Results

Item Number	**Aiken's V Value**	**Conclusion**
1	0.975	Valid
2	1.000	Valid
3	1.000	Valid
4	1.000	Valid
5	0.925	Valid

Following the expert validation and item revisions, the finalized instrument was pilot-tested with 79 eleventh-grade students from several senior high schools in East Bekasi. This limited-scale field trial aimed to collect empirical data for further statistical analysis, focusing on evaluating the construct validity and internal consistency of the instrument. The pilot testing stage was essential to ensure that the developed test items were not only theoretically sound but also functioned effectively in real classroom settings.

The construct validity of each item was assessed using the Pearson Product Moment correlation to determine the strength of the relationship between each item score and the total test score. This analysis aimed to verify whether each item accurately reflected the underlying construct of mathematical literacy in the context of algebra. The results, as visualized in **Figure 1**, show that all correlation coefficients exceeded the critical r-value (r_t = 0.221; n = 79; α = 0.05), indicating statistically significant relationships across all items (Anggraini et al., 2022). This suggests that each item contributed meaningfully to the overall measurement and could be considered empirically valid.

		Q1	Q2	Q3	Q4	Q5	TOTAL
Q1	Pearson Correlation	1	.136	.165	.225*	.345**	.513**
	Sig. (2-tailed)		.232	.146	.046	.002	.000
	N	79	79	79	79	79	79
Q2	Pearson Correlation	.136	1	.328**	.426**	.271*	.594**
	Sig. (2-tailed)	.232		.003	.000	.016	.000
	N	79	79	79	79	79	79
Q3	Pearson Correlation	.165	.328**	1	.248*	.759**	.782**
	Sig. (2-tailed)	.146	.003		.027	.000	.000
	N	79	79	79	79	79	79
Q4	Pearson Correlation	.225*	.426**	.248*	1	.300**	.637**
	Sig. (2-tailed)	.046	.000	.027		.007	.000
	N	79	79	79	79	79	79
Q5	Pearson Correlation	.345**	.271*	.759**	.300**	1	.836**
	Sig. (2-tailed)	.002	.016	.000	.007		.000
	N	79	79	79	79	79	79
TOTAL	Pearson Correlation	.513**	.594**	.782**	.637**	.836**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	79	79	79	79	79	79

Figure 1 Validity Result

The correlation values obtained ranged from 0.513 to 0.836, which fall within the moderate to very strong categories based on the standard interpretation guidelines (Melati et al., 2023). These findings demonstrate that the test items were well-aligned with the intended construct and functioned cohesively within the overall assessment instrument.

To evaluate the instrument's reliability, Cronbach's Alpha was calculated to assess the internal consistency among the items. As shown in **Figure 2**, the resulting alpha coefficient was 0.712—surpassing the generally accepted threshold of 0.70 for educational assessments. This value indicates that the items consistently measure the same underlying construct and that the instrument is capable of producing stable and dependable results upon repeated administrations (Bujang et al., 2018).

Cronbach's Alpha	N of Items
.712	5

Figure 2 Reliability Result

In summary, the results of both construct validity and reliability analyses, as supported by **Figure 1** and **Figure 2**, confirm that the developed mathematical literacy instrument satisfies accepted psychometric standards. Therefore, it can be concluded that the test is both a valid and reliable tool for assessing students' algebra-related mathematical literacy skills in the context of secondary education.

CONCLUSION AND IMPLICATION

Conclusion

This study aimed to develop a valid and reliable test instrument to assess students' mathematical literacy skills in the context of algebra using the 4D development model. Through the Define, Design, and Develop stages, a set of contextual, literacy-oriented test items was constructed based on the PISA framework. The development process incorporated curriculum analysis, expert reviews, item revisions, and empirical testing.

The findings indicated that the developed instrument met content validity criteria, with Aiken's V values ranging from 0.925 to 1.000. The construct validity was confirmed through Pearson Product Moment analysis, with all items showing significant correlations with the total test score. Furthermore, the reliability analysis using Cronbach's Alpha yielded a coefficient of 0.712, indicating acceptable internal consistency. Overall, the instrument is suitable for measuring students' mathematical literacy in algebra and may support efforts to improve assessment practices aligned with real-world competencies.

Implication

This study has both theoretical and practical implications. Theoretically, the development of a literacy-based assessment instrument enriches the literature on mathematics education, particularly in aligning local assessments with international standards such as PISA. It demonstrates how contextual, open-ended test items can be designed to evaluate students' abilities to reason, model, and interpret mathematical information meaningfully.

Practically, the instrument provides educators with a tool that can more accurately capture students' mathematical thinking in real-life contexts. It addresses the current gap in classroom assessments that tend to focus on procedural knowledge by promoting tasks that require deeper cognitive engagement. This can assist teachers in identifying students' strengths and weaknesses in applying mathematics to authentic problems.

For future implementation, the instrument may be refined further through large-scale trials across diverse student populations and different curriculum contexts. Researchers are encouraged to explore its application in other mathematical domains beyond algebra. Additionally, future studies may investigate the impact of such assessments on students' learning motivation, problem-solving skills, and self-efficacy. One limitation of the current study is its relatively small sample size, which restricts the generalizability of the findings. Expanding the research with broader demographic coverage is recommended for stronger external validity.

Disclosure statement

No potential conflict of interest was reported by the authors.

ACKNOWLEDGMENTS

The author would like to express sincere gratitude to the lecturers and academic advisors at Universitas Negeri Jakarta for their invaluable guidance throughout the research process. Special thanks are also extended to the mathematics teachers and students at the participating senior high schools in East Bekasi for their cooperation and participation in the pilot testing of the instrument.

This research would not have been possible without the support and constructive feedback provided by the expert validators, whose insights greatly contributed to the refinement of the test instrument. The author is also grateful to peers and colleagues who offered continuous encouragement during the development of this study.

REFERENCES

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959.

Anggraini, F. D. P., Aprianti, A., Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran statistika menggunakan software SPSS untuk uji validitas dan reliabilitas. *Jurnal Basicedu*, *6*(4), 6491–6504.

Bender, W. N. (2012). *Project-based learning: Differentiating instruction for the 21st century*. Corwin Press.

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers. *The Malaysian Journal of Medical Sciences : MJMS*, *25*(6), 85–99. https://doi.org/10.21315/mjms2018.25.6.9

Cheung, K. (2017). The effects of resilience in learning variables on mathematical literacy performance: A study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea. *Educational Psychology*, *37*(8), 965–982.

Colwell, J., & Enderson, M. C. (2016). "When I hear literacy": Using pre-service teachers' perceptions of mathematical literacy to inform program changes in teacher education. *Teaching and Teacher Education*, *53*, 63–74.

Faridah, N. R., Afifah, E. N., & Lailiyah, S. (2022). Efektivitas Model Pembelajaran Project Based Learning Terhadap Kemampuan Literasi Numerasi dan Literasi Digital Peserta Didik Madrasah Ibtidaiyah. *Jurnal Basicedu Vol*, *6*(1).

Goos, M., & O'Sullivan, K. (2023). *The Evolution and Uptake of Numeracy and Mathematical Literacy as Drivers for Curriculum Reform BT - Mathematics Curriculum Reforms Around the World: The 24th ICMI Study* (Y. Shimizu & R. Vithal, Eds.; pp. 345–357). Springer International Publishing. https://doi.org/10.1007/978-3-031-13548-4_21

Hillman, A. M. (2014). A literature review on disciplinary literacy: How do secondary teachers apprentice students into mathematical literacy? *Journal of Adolescent and Adult Literacy*, *57*(5), 397–406. https://doi.org/10.1002/jaal.256

Husni, N., & Herman, T. (2024). Analysis of Students' Mathematical Literacy Ability on High-Level Problems of PISA viewed from Gender. *Jurnal Didaktik Matematika*, *11*(2), 219–234. https://doi.org/10.24815/jdm.v11i2.39445

Hwang, J., Choi, K. M., Bae, Y., & Shin, D. H. (2018). Do teachers' instructional practices moderate equity in mathematical and scientific literacy?: An investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*, *16*, 25–45.

Machaba, F. M. (2017). Pedagogical demands in mathematics and mathematical literacy: A case of mathematics and mathematical literacy teachers and facilitators. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 95–108.

Machaba, F. M. (2018). Pedagogical demands in mathematics and mathematical literacy: A case of mathematics and mathematical literacy teachers and facilitators. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 95–108. https://doi.org/10.12973/ejmste/78243

Maslihah, S., Waluya, S. B., & Suyitno, A. (2020). The role of mathematical literacy to improve high order thinking skills. *Journal of Physics: Conference Series*, *1539*(1), 12085.

Melati, R., Destiniar, D., & Rohana, R. (2023). Pengembangan Soal Higher Order Thinking Skills (Hots) Pada Materi Aritmetika Sosial Di Kelas Vii Smp. *Teorema: Teori Dan Riset Matematika*, *8*(1), 116. https://doi.org/10.25157/teorema.v8i1.9140

Natsir, I., Mayasari, D., & Munfarikhatin, A. (2022). Rasch Model: Quality of Test Instruments to Measure Students' Ability in Algebra. *SHS Web of Conferences*, *149*, 01033. https://doi.org/10.1051/shsconf/202214901033

OECD. (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving. *OECD Publishing*. https://doi.org/https://doi.org/10.1787/9789264281820-en

OECD. (2023). Program For International Student (PISA) 2022 Assessment and Analytical Framework. In *OECD (Organisation for Economic Co-operation and Development) Publishing*. https://www.oecd-ilibrary.org/education/pisa-2022-assessment-and-analytical-framework_dfe0bf9c-en

Oktiningrum, W., & Hartono, Y. (2016). Developing PISA-" Like" Mathematics Task with Indonesia Natural and Cultural Heritage as Context to Assess Students Mathematical Literacy. *Journal on Mathematics Education*, *7*(1), 1–10.

PG Dikdas. (2020). *Literasi dan Numerasi Adalah Kompetensi yang Bersifat Mendasar*. https://gurudikdas.kemdikbud.go.id/news/Literasi-dan-Numerasi-Adalah-Kompetensi-yang-Bersifat-Mendasar

Stacey, K., & Turner, R. (2015). Assessing mathematical literacy: The PISA experience. *Assessing Mathematical Literacy: The PISA Experience*, 1–321. https://doi.org/10.1007/978-3-319-10121-7

Thiagarajan, S. (1974). *Instructional development for training teachers of exceptional children: A sourcebook.*

Wijaya, A. (2016). Students' information literacy: A perspective from mathematical literacy. *Journal on Mathematics Education*, *7*(2), 73–82.

Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, *11*(3), 555–584.

Wilsa, A. W., Rusilowati, A., Susilaningsih, E., Jaja, J., & Nurpadillah, V. (2023). Validity, reliability, and item characteristics of cell material science literacy assessment instruments. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(2), 177–188. https://doi.org/10.21831/pep.v27i2.61577