Prima: Jurnal Pendidikan Matematika
Vol. 10, No. 3, September 2026, pp. 784 - 803
P-ISSN: 2579-9827, E-ISSN: 2580-2216
Web: http://jurnal.umt.ac.id/index.php/prima

DEVELOPMENT AND EMPIRICAL ANALYSIS OF A MATHEMATICAL LITERACY INSTRUMENT USING THE PISA FRAMEWORK

Muhamad Zulfikar Mansyur[1], Al Jupri [2], Suhendra [3], Bambang Avip Priatna Martadiputra [4]
[1,2,3,4] Department of Mathematics Education, Universitas Pendidikan Indonesia, West Java, Indonesia
e-mail: aljupri@upi.edu

Abstract

This study aimed to describe the process of developing a mathematical literacy instrument based on the PISA (Programme for International Student Assessment) framework and to describe the validity of the instrument developed. This research was a development study comprising the following stages: preliminary analysis, development of the initial instrument design, expert validation, field trial, empirical validation analysis, and analysis of invalid items with proposed revisions. The instrument was tested on 47 students in grades IX and X in Tasikmalaya City. Data were analyzed using both quantitative and qualitative analysis. The results of this study were as follows: 1) expert review indicated that the PISA-model mathematical literacy test instrument was usable; 2) the instrument consisted of 13 items, comprising 7 essay items, 2 multiple-choice items, 1 complex multiple-choice item, 1 true/false item, and 2 fill-in items; 3) the instrument's reliability was 0.770, indicating high reliability, and its validity results showed 1 item with very low validity, 2 items with low validity, 8 items with moderate validity, and 2 items with high validity. The resulting instrument was usable but had shortcomings in several items, so further field testing is needed to determine the impact of the proposed revisions on the instrument's reliability and validity. Theoretically, this instrument can be used to measure students' mathematical literacy, and researchers or practitioners may use it to measure students' mathematical literacy performance, taking into account the proposed improvements.

Keywords: Test Instrument, Mathematical Literacy, PISA, Development, Theoretical Validity, Empirical Validity

Abstrak

Penelitian ini bertujuan untuk mendeskripsikan proses pengembangan instrumen literasi matematis dengan framework PISA (Programme for International Student Assessment) serta menggambarkan validitas instrumen yang dikembangkan. Jenis penelitian ini adalah penelitian pengembangan dengan tahapan: analisis pendahuluan, pengembangan desain awal instrumen, validasi ahli, uji coba lapangan, analisis validasi empiris, serta analisis butir tidak valid dan usulan revisinya. Instrumen diujicobakan kepada 47 siswa yang terdiri dari siswa kelas IX dan kelas X di Kota Tasikmalaya. Data dianalisis dengan analisis kuantitatif dan kualitatif. Hasil penelitian ini adalah: 1) review dari ahli menyatakan instrumen tes literasi matematika model PISA dapat digunakan; 2) instrumen ini terdiri atas 13 soal dengan rincian 7 soal esai, 2 soal pilihan ganda, 1 soal pilihan ganda kompleks, 1 soal benar/salah, dan 2 soal isian; hasil reliabilitas instrumen sebesar 0,770 yang berarti instrumen memiliki reliabilitas tinggi, dan hasil validitas instrumen menunjukkan 1 soal berkategori sangat rendah, 2 soal rendah, 8 soal sedang, dan 2 soal berkategori tinggi. Instrumen yang dihasilkan dapat digunakan namun memiliki kekurangan pada beberapa butir sehingga perlu dilakukan uji lapangan lanjutan untuk mengetahui dampak usulan revisi terhadap reliabilitas dan validitas instrumen. Secara teoretis instrumen ini sudah dapat digunakan untuk mengukur literasi matematis siswa sehingga peneliti/praktisi dapat menggunakan instrumen ini untuk mengukur performa literasi matematis siswa dengan memperhatikan usulan perbaikan yang ada.

Kata kunci: Instrumen Tes, Literasi Matematis, PISA, Pengembangan, Validitas Teoritis, Validitas Empiris

INTRODUCTION

Mathematics education plays a strategic role in preparing human resources capable of facing the challenges of the 21st century. Mathematics not only functions as a discipline that develops logical, systematic, critical, and creative thinking skills, but also serves as a

Received: July 23, 2026; Revised: September 3, 2026; Accepted: September 19, 2026

foundation for advances in science, technology, engineering, and innovation. In an increasingly complex and digitalized modern society, mathematical ability has become one of the essential competencies supporting data-based decision-making, problem-solving, and adaptation to rapidly evolving technology (OECD, 2023). Therefore, improving the quality of mathematics education has become an important agenda in the development of education across countries.

Social and technological transformation in the 21st century has changed the competencies required of learners. Beyond mastering conceptual knowledge, learners are expected to possess critical thinking, creativity, communication, and collaboration skills, collectively known as the 4C skills. The development of digital technology and artificial intelligence, along with the use of data in various aspects of life, requires individuals to understand, evaluate, and use quantitative information effectively. In this context, mathematics plays an important role as a means of developing the higher-order thinking skills needed in modern life and the world of work (OECD, 2023; Kappassova et al., 2025) .

In line with these demands, the concept of mathematical literacy has become a major focus of international education policy. OECD (2023) defines mathematical literacy as an individual's capacity to reason mathematically and to formulate, use, and interpret mathematics in various real-life contexts. Mathematical literacy encompasses the use of mathematical concepts, procedures, facts, and tools to describe phenomena, solve problems, and make decisions grounded in mathematical reasoning. Mathematical literacy is therefore not merely the ability to perform calculations, but also the ability to use mathematics meaningfully in everyday life.

Mathematical literacy is a multidimensional construct that involves reasoning, mathematical modeling, communication, representation, and contextual problem-solving skills. Niss and Jablonka (2020) explain that a mathematically literate person is able to recognize the role of mathematics in life, critically understand quantitative information, and use mathematics to make responsible decisions. Kolar & Hodnik (2021) add that mathematical literacy develops through learners' engagement in solving authentic and contextual problems, enabling them to connect school mathematics with the real-life situations they encounter in daily life.

The importance of mathematical literacy is increasing in today's data-driven society. Individuals are confronted with information presented in tables, graphs, diagrams, statistical models, and other mathematical representations. Consequently, the ability to understand, interpret, and evaluate such information has become an unavoidable necessity. Hwang (2018) shows that mathematical literacy contributes to the quality of individuals' decision-making across various social and economic contexts. This finding is reinforced by Kappassova et al. (2025), who state that mathematical literacy is an important predictor of individuals' success in facing technological change and the complexity of modern life.

In international education, mathematical literacy is measured through the Programme for International Student Assessment (PISA), administered by the OECD every three years. Unlike school-based assessments, which generally focus on mastery of curriculum content, PISA measures students' ability to use mathematical knowledge and skills to solve problems arising in various real-life contexts. The PISA 2022 framework emphasizes three main processes in mathematical literacy: formulating situations mathematically (formulate), employing mathematical concepts and procedures (employ), and interpreting mathematical results in the context of the problem (interpret) (OECD, 2023). This approach has made PISA one of the primary references for developing mathematical literacy assessments across countries.

Despite various efforts to improve the quality of education, PISA results show that Indonesian students' mathematical literacy still faces significant challenges. According to the PISA 2022 report, Indonesia's mathematics performance remains below the OECD average, and most students have not been able to reach the competency levels that require high-level mathematical reasoning (OECD, 2023). This finding indicates that many students still struggle when they must use mathematics to solve non-routine problems requiring modeling, reasoning, and contextual interpretation. This result is consistent with the studies of Stacey (2011), Widjaja (2011), and Kusmaryono and Kusumaningsih (2023), which show a gap between the demands of mathematical literacy in the PISA framework and mathematics teaching practices that remain predominantly oriented toward routine procedures.

This gap not only points to the need to improve instructional practices, but also indicates the need for a valid and reliable assessment instrument to detect the mathematical literacy profile of Indonesian students based on the characteristics of the PISA framework. Without

such an instrument, teachers and researchers find it difficult to obtain a diagnostic picture of which domain, process, or context of mathematical literacy poses the greatest obstacle for students, so that the resulting instructional improvements tend to be generic and poorly targeted.

Various studies have been conducted to improve students' mathematical literacy through the implementation of innovative learning models. Jannah et al. (2021) found that Project Based Learning made a positive contribution to improving students' mathematical literacy. Similar findings were reported by Susanta et al. (2023), who showed that STEM-based learning significantly improved mathematical literacy. In addition, other approaches such as Problem Based Learning, Realistic Mathematics Education, and contextual learning have also been reported to positively affect the development of students' mathematical literacy. Nevertheless, most of these studies focus on the effectiveness of instruction, while the development of instruments for measuring mathematical literacy has received relatively little attention. Yet the quality of measurement is a critically important aspect of both research and educational practice. Valid and reliable instruments are needed to obtain accurate information about students' mathematical literacy. Suciati et al. (2020) emphasize that the development of mathematical literacy instruments must be carried out systematically through theoretical review, expert validation, and empirical testing, so that the resulting instrument can accurately represent the construct being measured. In addition, the instrument developed needs to accommodate the characteristics of the PISA framework, which covers content domains, processes, and contexts, so that the measurement results are consistent with international standards.

Most studies on the development of mathematical literacy instruments still focus on a specific domain or emphasize only theoretical validity without adequate empirical analysis. Research that comprehensively develops instruments covering all PISA content domains—Change and Relationship, Space and Shape, Quantity, and Uncertainty and Data—while also examining the empirical characteristics of each item, remains relatively limited. On the other hand, a comprehensive instrument is greatly needed to obtain a more complete and accurate picture of students' mathematical literacy.

Therefore, this study aims to develop and analyze the empirical characteristics of a PISA-based mathematical literacy instrument. The instrument was developed with reference to the

PISA 2022 framework and involved a process of theoretical validation and empirical testing to obtain evidence of the instrument's validity and reliability.

METHODS

This study is a development study that adapts the development stages of Bakker (2018) and Borg and Gall (2003) into several stages: 1) preliminary analysis, 2) development of the initial instrument design and theoretical validation, 3) field trial and empirical validation, and 4) analysis and revision of invalid instrument items.

This study developed an instrument to measure students' mathematical literacy based on the 2022 PISA framework (OECD, 2023), namely a person's ability to think mathematically and to formulate, use, and interpret mathematical concepts in order to solve problems in various real-world situations. The focus of this instrument's development was to create an instrument as close as possible to the instrument used by PISA; therefore, this instrument was developed to fully cover the PISA content domains: Change and Relationship, Space and Shape, Quantity, and Uncertainty and Data. The instrument was designed with 13 items, consisting of 7 essay items, 2 multiple-choice items, 1 complex multiple-choice item, 1 true/false item, and 2 fill-in items. Before being administered, the items first underwent theoretical validation by two expert validators: a Mathematics Education lecturer at a state university in Tasikmalaya (holding a doctoral degree) and a mathematics teacher at the junior secondary (MTs) level in Kuningan Regency (holding a bachelor's degree and a professional educator certificate). The instrument was piloted on students at the ninth-grade junior secondary (or equivalent) level and tenth-grade senior secondary (or equivalent) level, with a total of 47 respondents. This study considered the selection of ninth- and tenth-grade respondents because the PISA framework targets individuals aged 15 years. The researchers therefore interpreted this age target, within Indonesia's formal education system, as corresponding to ninth and tenth grade.

This selection of ninth- and tenth-grade respondents is consistent with PISA's age-based (rather than grade-based) sampling approach; PISA officially assesses the population of students aged 15 years 3 months to 16 years 2 months who are enrolled in grade 7 or higher, precisely because grade level cannot serve as an equivalent basis for comparison across education systems (OECD, 2023). Nevertheless, the researchers acknowledge that ninth- and tenth-grade students are exposed to different curricula (junior secondary versus senior

secondary/equivalent curricula), so that differences in achievement across grade levels may reflect differences in learning experience rather than differences in the mathematical literacy being measured. The results of this study should therefore be interpreted as a picture of mathematical literacy among the general 15-year-old age group, rather than as a comparison of achievement between junior and senior secondary levels.

Empirical validity testing was carried out using the Pearson product-moment correlation:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}} \quad (1)$$

Where:

r_{xy} : correlation coefficient

N : number of subjects/respondents

X : item score

Y : total score

(Pearson, 1896)

The magnitude of the correlation coefficient was interpreted using the classification proposed by Evans (1996). The correlation coefficient criteria used are shown in Table 1 below:

Table 1. Correlation Coefficient Criteria

Value	Description
$0 \leq r_{xy} < 0{,}20$	Very Low
$0{,}20 \leq r_{xy} < 0{,}40$	Low
$0{,}40 \leq r_{xy} < 0{,}60$	Moderate
$0{,}60 \leq r_{xy} < 0{,}80$	High
$0{,}80 \leq r_{xy} < 1{,}00$	Very High

Test reliability was calculated using Cronbach's Alpha:

$$r_{11} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum {\sigma_b}^2}{{\sigma_t}^2}\right) \quad (2)$$

Where:

r_{11} : test reliability

k : number of items

$\sum {\sigma_b}^2$: sum of item variances

σ_t^2 : total variance

Cronbach (1951)

The reliability coefficient criteria used are shown in Table 2 below:

Table 2. Reliability Coefficient Criteria

Value	Description
$r_{11} < 0{,}20$	Very Low
$0{,}20 \leq r_{11} < 0{,}40$	Low
$0{,}40 \leq_{11} < 0{,}70$	Moderate
$0{,}70 \leq r_{11} < 0{,}90$	High
$0{,}90 \leq r_{11} < 1{,}00$	Very High

Sugiyono. (2013)

The test results were used to determine whether an item could be used as part of the instrument for measuring students' mathematical literacy. An item was considered usable if its correlation fell into the moderate category or higher. The same criterion was applied to the reliability coefficient, whereby the instrument was considered usable if its reliability fell into the moderate category or higher. Descriptive analysis was used to describe the results of the field trial that had been conducted.

This study used the Classical Test Theory (CTT) approach to analyze the validity and reliability of the instrument. This choice was based on its suitability for the early stage of a development study, namely the limited trial sample size (n = 47) and the need for a simple item analysis that is easy for education practitioners to interpret (Hambleton & Jones, 1993). Nevertheless, CTT has a fundamental limitation compared to Item Response Theory (IRT) or the Rasch model, namely that it is group-dependent: item difficulty and discrimination indices can change depending on the ability of the group tested, making them difficult to compare across samples or to generalize (Hambleton & Jones, 1993). In contrast, the IRT/Rasch approach allows for the estimation of item parameters that are relatively independent of the sample, as well as the estimation of examinee ability on the same scale as the item parameters. Therefore, the item analysis results in this study should be interpreted as an initial picture of the instrument's psychometric characteristics within a specific trial group, and further development is recommended to consider IRT/Rasch-based analysis to obtain more stable item parameter estimates.

RESULTS AND DISCUSSION

Preliminary Analysis

Preliminary analysis was carried out to obtain a conceptual foundation for developing a PISA-based mathematical literacy instrument. This analysis covered a review of the 2022 PISA framework, the characteristics of mathematical literacy, findings from previous studies, and measurement needs within the Indonesian education context. This was done to ensure that the instrument developed truly represented the construct to be measured and reflected mathematical literacy.

The review shows that mathematical literacy is not only related to the ability to use mathematical concepts and procedures, but also to the ability to formulate, use, and interpret mathematics in various real-life contexts (OECD, 2023). Therefore, the instrument developed needed to accommodate mathematical thinking processes involving reasoning, modeling, representation, and interpretation, rather than merely computational ability. This finding is consistent with the view of Niss and Jablonka (2020), who position mathematical literacy as the ability to use mathematics meaningfully in everyday life.

Analysis of the 2022 PISA framework shows that the measurement of mathematical literacy is built on three main components: mathematical processes, mathematical content, and problem contexts. Within the content dimension, there are four domains that served as a reference for instrument development: Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data (OECD, 2023). These four domains were then used as the basis for developing the instrument blueprint so that the scope of the ability being measured would be more representative.

In addition to the theoretical review, an analysis of previous research on mathematical literacy instruments was also conducted. This review shows that most studies still focus on developing items for a specific domain or context, so that they do not yet comprehensively capture mathematical literacy (Suciati et al., 2020; Kusmaryono & Kusumaningsih, 2023). On the other hand, PISA results show that Indonesian students' mathematical literacy still requires attention, particularly on items that demand reasoning and the application of mathematics in real situations (OECD, 2023). These findings indicate the need for an instrument capable of measuring mathematical literacy more comprehensively, in line with the characteristics of the PISA framework.

Based on this analysis, an instrument specification was developed that integrates content domains, mathematical processes, and problem contexts into each item. This specification formed the basis for developing the blueprint and the instrument prototype, which was subsequently validated by experts and empirically tested to obtain a valid and reliable instrument. The initial design of this instrument is presented as follows:

Table 3. Initial Design Of The Instrument

Item No.	Domain	Process	Context
Item 1	Change and Relationships	Formulate	Societal
Item 2	Change and Relationships	Formulate	Societal
Item 3	Change and Relationships	Employ	Societal
Item 4	Uncertainty and Data	Employ	Personal
Item 5	Uncertainty and Data	Employ	Personal
Item 6	Change and Relationships	Formulate	Personal
Item 7	Space and Shape	Employ	Occupational
Item 8	Space and Shape	Interpret	Scientific
Item 9	Space and Shape	Interpret	Scientific
Item 10	Quantity	Employ	Scientific
Item 11	Quantity	Interpret	Scientific
Item 12	*Uncertainty and Data*	Formulate	Societal
Item 13	*Uncertainty and Data*	Employ	Societal

Theoretical Validity of the Instrument

After the preliminary analysis and initial review—covering the indicators, descriptors, and instrument blueprint—had been carried out, expert validation was conducted by an academic and a practitioner in the field of mathematics education. Theoretical validation was carried out by the two experts using a validation sheet assessing the content, substance, and language of the items on a Likert scale. This theoretical validation was intended to ensure that the instrument was theoretically appropriate in terms of content, substance, and language. The results of the expert validation indicated that the instrument was usable, with the following suggestions:

1. Providing more space within the items to give students room to write their answers
2. The items were considered too difficult for students, although they were consistent with the given indicators and criteria

Based on these results, revisions were made in accordance with the suggestions, after which the instrument was piloted on a limited basis with students as the basis for empirical validation.

Field Trial and Empirical Validation

The field trial involved students from junior and senior secondary schools in Tasikmalaya City, with a total of 47 respondents. After the trial, the data were processed using SPSS 22, yielding the following validity test results:

Table 4. Instrument Validity Results

Item No.	Correlation Coefficient	Correlation	Interpretation
Item 1	0,418	Moderate	Adequate
Item 2	0,753	High	Good
Item 3	0,555	Moderate	Adequate
Item 4	0,542	Moderate	Adequate
Item 5	0,494	Moderate	Adequate
Item 6	0,651	Moderate	Adequate
Item 7	0,366	Low	Poor (not usable)
Item 8	0,151	Very Low	Very Poor (not usable)
Item 9	0,505	Moderate	Adequate
Item 10	0,689	Moderate	Adequate
Item 11	0,732	High	Good
Item 12	0,411	Moderate	Adequate
Item 13	0,373	Low	Poor

Based on the validity test results in Table 3, it was found that 1 item fell into the very low category, 2 items into the low category, 8 items into the moderate category, and 2 items into the high category. The reliability results are presented in Table 4 below:

Table 5. Instrument Reliability Results

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.770	.770	13

Based on Table 4, the reliability result was 0.770, indicating that the instrument has high reliability. Following the validity and reliability testing, an analysis of the invalid items was carried out to examine the probable causes of their invalidity.

It should be emphasized that the reliability coefficient of 0.770 was calculated based on field-trial data for the 13 items in their original form (prior to revision). The revisions to Item 7, Item 8, and Item 13 are proposed based on the analysis of invalid items presented in the

following section; consequently, the reliability and validity of the post-revision instrument have not yet been empirically tested and remain an agenda for further field testing.

Analysis of Invalid Items

This section presents respondents' answers to the invalid items for further analysis. The results of the analysis are then described as the basis for revising the items.

Analysis of Item 7

Item 7 is an essay item on the topic of geometric solids. Item 7 has a correlation value of 0.366, indicating low validity. This low validity is a result of students' difficulty in answering the item. Some students had difficulty answering because of a spatial misconception, assuming that the solid figure in Item 7 was a rectangular prism (cuboid). Some of the students' errors in answering this item are shown below.

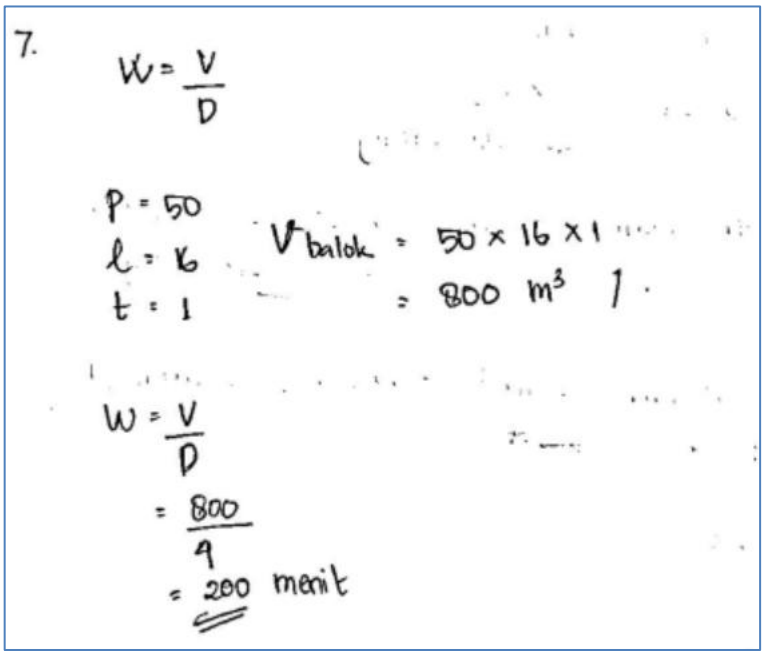


Figure 1. Student's Answer to Item 7

Figure 1 shows an error in calculating the volume using the concept of a cuboid, even though the shape of the swimming pool is not a cuboid. This reflects a conceptual error experienced by the student. To address this error, the revision adds information so that students do not mistakenly assume that the pool is shaped like a cuboid.

Revision of Item 7

Before

The pool is filled with water from a tap that can fill it at a rate of 4 m^3 per minute. How long will it take to fill the swimming pool? (Explain your answer in as much detail as possible!)

After

The pool shown above has varying depths. The shallowest side is 1 m deep and the deepest side is 3 m deep. The pool is filled with water from a tap that can fill it at a rate of 4 m^3 per minute. How long will it take to fill the swimming pool? (Explain your answer in as much detail as possible!)

The proposed revision to Item 7 provides additional explanation intended to give students more information and awareness that the pool is not shaped like a cuboid, so that this conceptual error can be avoided.

Theoretically, students' misconception in assimilating the shape of the swimming pool to a cuboid can be understood through the framework of conceptual change theory, that is, the application of an already-familiar geometric schema to a new situation that is, in fact, different; the addition of explicit information about the varying pool depth is intended to trigger the cognitive dissonance that leads students to abandon the cuboid schema (Posner, Strike, Hewson, & Gertzog, 1982). Misconceptions related to measuring the volume of irregular solids have also been reported at other education levels, indicating that this type of error is common rather than specific to this instrument alone (Tan-Sisman & Aksu, 2016).

Analysis of Item 8

Item 8 is a complex multiple-choice item with more than one correct answer. This item has a validity coefficient of 0.151, categorized as very low. The probable explanation for this very low validity is that students were only accustomed to selecting a maximum of 3 answers without analyzing all the options thoroughly. Overall, the item has 4 correct answers and 1 incorrect answer. Because most students only selected a few correct answers, with the majority selecting only 2, the validity of this item became very low, as students with otherwise high total scores ended up receiving low scores on this item. Some of the students' answers are shown below.

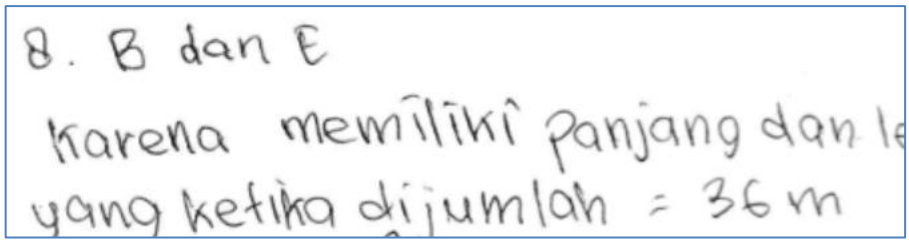


Figure 2. Student Answer 1 to Item 8

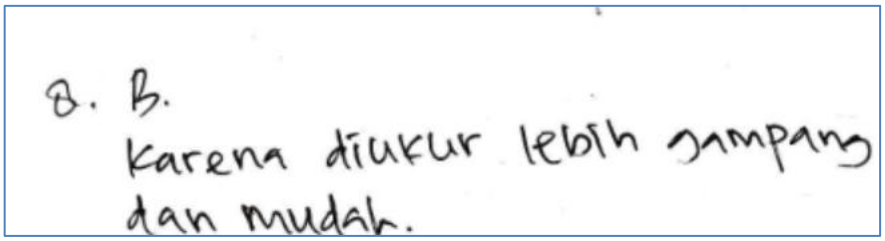


Figure 3. Student Answer 2 to Item 8

As shown in Figures 2 and 3, students were still found to give incomplete answers to Item 8; therefore, the proposed revision reduces the total number of correct answers among the options provided. The number of correct answers was changed to 3 out of 5 options.

This revision is consistent with findings in the assessment item-development literature indicating that the complex multiple-choice (Type-K) format, which requires students to identify every combination of correct answers, tends to measure test-taking strategy skills (test-wiseness) rather than genuine conceptual mastery. Accordingly, the low validity of Item 8 can be partly attributed to the characteristics of the item format itself, rather than solely to students' mathematical literacy, so that reducing the number of correct-answer options is a recommended revision strategy for this type of item (Towns, 2014).

Analysis of Item 13

Item 13 requires students to fill in true/false statements. This item consists of 4 true/false statements, so the score for this item depends on how many statements the student answers correctly. The correlation coefficient for this item is 0.373, which falls into the low category. Few students were able to correctly identify the statements and their truth values. Two statements contributed to inaccurate answers because they required closer attention: statement (1) and statement (4). Both statements required greater care in analyzing the data. As a result, students obtained fairly low scores on this item, which sometimes did

not correspond to their otherwise high total scores, thereby affecting the item's validity and resulting in a low validity classification. Examples of students' errors on Item 13 are presented in Figures 4 and 5 below:

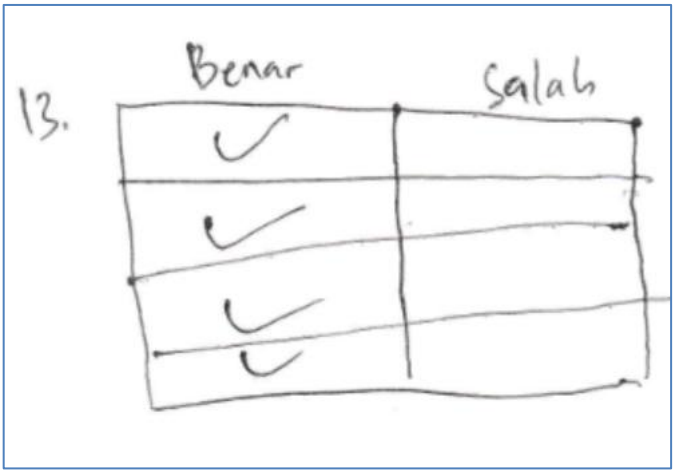


Figure 4. Student Answer 1 to Item 13

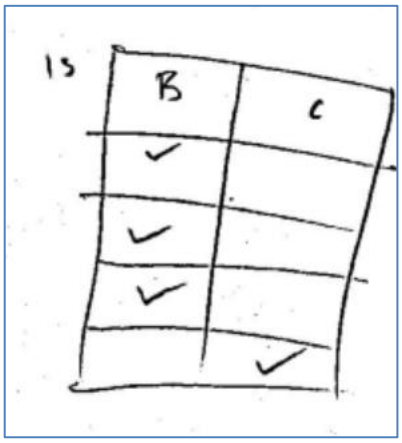


Figure 5. Student Answer 2 to Item 13

Based on this analysis, the following revision is proposed.

Revision of Item 13

Before

Statement	True	False
The open unemployment rate decreased from August 2020 to August 2023 in all categories		
The average male unemployment rate is higher than the average female unemployment rate from August 2020 to August 2023		
The highest open unemployment is found in urban areas		

The largest decrease in the open unemployment rate occurred between August 2020 and August 2021		

After

Statement	True	False
There is a category in which the open unemployment rate increased from August 2020 to August 2023		
The average male unemployment rate is higher than the average female unemployment rate from August 2020 to August 2023		
The highest open unemployment is found in urban areas		
The largest decrease in the open unemployment rate occurred between August 2020 and August 2021		

The proposed revision changes the wording of statement (1), which is expected to increase students' awareness of the need to examine the data for each category more carefully.

In general, the three revisions above are consistent with PISA's item-design principles, which emphasize that an item's complexity should stem from the targeted mathematical process (formulate, employ, or interpret) rather than from contextual ambiguity or item-format issues unrelated to the construct being measured (OECD, 2023).

The analysis of the PISA-based mathematical literacy instrument shows that the instrument is valid, with minor revisions noted. This result is consistent with the studies of Syaifuddin (2022), Sinaga et al. (2024), and Afri and Nurrahmawati (2025), who report that mathematical literacy instruments developed based on PISA or AKM (Minimum Competency Assessment) characteristics show good validity and reliability after undergoing expert validation. In the process of researching and developing this mathematical literacy instrument, it was found that the instrument is theoretically valid based on expert judgment. However, the field trial results indicate invalidity in several items. This condition is common in instrument development research, since theoretical validity is not always followed by good empirical performance across all items (DeVellis & Thorpe, 2021). This is also consistent with the findings of Udil, Jupri, and Nurlaelah (2023); Nubatonis, Jupri, and Mulyaning (2024);

Kirana (2025); and Sari et al. (2024), who found the need to revise several items after empirical analysis, even though the instrument had received a favorable assessment from expert validators. Further analysis was carried out on the invalid items to serve as the basis for revision, so that a better instrument could be obtained for large-scale implementation.

CONCLUSION

Overall, following the analysis and discussion, it can be concluded that this PISA-based mathematical literacy instrument was developed by first reviewing the PISA framework theoretically, which was then translated into items covering all content areas: Change and Relationship, Space and Shape, Quantity, and Data and Uncertainty. The instrument consists of 13 items, comprising 7 essay items, 2 multiple-choice items, 1 complex multiple-choice item, 1 true/false item, and 2 fill-in items. The instrument then underwent a validation process consisting of theoretical validation by two experts, an academic and a mathematics education practitioner, followed by a field trial to analyze its empirical validity, which was in turn followed by analysis and revision of the invalid items. The instrument's reliability was 0.770, indicating high reliability. The validity results showed 1 item with very low validity, 2 items with low validity, 8 items with moderate validity, and 2 items with high validity. The items with very low and low validity were analyzed based on students' answers as the basis for revision, drawing on the probable causes of student errors on the invalid items.

For these three items, the researchers decided to revise the wording/format of the items rather than removing them from the instrument, because their low validity was attributable to the clarity of the item information (Item 7), the complex multiple-choice format being unfamiliar to students (Item 8), and the level of care required by the true/false statements (Item 13)—rather than to the items being irrelevant to the mathematical literacy construct being measured.

This study produced an instrument that is usable, but has limitations up to the stage of empirical validity testing and revision of invalid items. Further trials of the revised items should be carried out in future research. The implementation of this instrument on a wider scale should be postponed until the revised items (Items 7, 8, and 13) are re-piloted empirically to

confirm that the revisions genuinely improve the psychometric characteristics of the items concerned.

Practically, this instrument has the potential to be used by mathematics teachers as a formative and diagnostic assessment tool in the classroom. By mapping items to PISA content domains, processes, and contexts, teachers can identify the specific mathematical domain or process that poses the greatest difficulty for students and use this information to adjust their instructional strategies responsively (Black & Wiliam, 2009). For example, the error patterns observed in Items 7 and 13 can be used as material for classroom discussion to clarify geometric misconceptions and to train students in reading data more carefully, before they are presented with more complex mathematical literacy items.

A further limitation of this study lies in the characteristics of the respondents selected. Nevertheless, the results of this study can serve as a practical reference for developing mathematical literacy instruments using the PISA framework.

REFERENCES

Afri, L. E., & Nurrahmawati, N. (2025). Pengembangan instrumen penilaian literasi matematika mahasiswa calon guru matematika. *Jurnal Pendidik Indonesia*, 6(4), 252–264.

Bakker, A. (2018). Design research in education. In *Design research in education: A practical guide for early career researchers* (pp. 3–22). Routledge.

Berek, M. I., Sumartono, S., & Prastiwi, L. (2023). Profil literasi matematika siswa dalam menyelesaikan masalah matematika ditinjau dari gaya belajar. *Sibatik Journal: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 2(2), 485–490.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. https://doi.org/10.1007/BF02310555

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage Publications.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Publishing Company.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Allyn and Bacon.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Jannah, R. R., Asikin, M., & Zaenuri, Z. (2021). Systematic literature review: Pembelajaran Project Based Learning (PjBL) terhadap kemampuan literasi matematika siswa. *IJoIS: Indonesian Journal of Islamic Studies*, 2(2), 227–234.

Kappassova, S., Abylkassymova, A., Bulut, U., Zykrina, S., Zhumagulova, Z., & Balta, N. (2025). Mathematical literacy and its influencing factors: A decade of research findings (2015-2024). In Eurasia Journal of Mathematics, Science and Technology Education (Vol. 21, Number 7). Modestum LTD. https://doi.org/10.29333/ejmste/16615

Kirana, R. N. (2025). *Pengembangan instrumen tes literasi matematis berbasis kearifan lokal Blitar Raya kelas V sekolah dasar* [Unpublished doctoral dissertation]. Universitas Negeri Malang.

Kolar, V. M., & Hodnik, T. (2021). Mathematical literacy from the perspective of solving contextual problems. In European Journal of Educational Research (Vol. 10, Number 1, pp. 467–483). Eurasian Society of Educational Research. https://doi.org/10.12973/EU-JER.10.1.467

Kusmaryono, I., & Kusumaningsih, W. (2023). Evaluating the results of PISA assessment: Are there gaps between the teaching of mathematical literacy at schools and in PISA assessment? *European Journal of Educational Research*, 12(3), 1479–1493. https://doi.org/10.12973/eu-jer.12.3.1479

Lutfianto, M., & Hartono, Y. (2013). Unfinished student answer in PISA mathematics contextual problem. *Indonesian Mathematical Society Journal on Mathematics Education*, 4(2), 188–193.

Niss, M. (2015). Mathematical competencies and PISA. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: The PISA experience* (pp. 35–55). Springer.

Niss, M., & Jablonka, E. (2020). Mathematical literacy. In *Encyclopedia of mathematics education* (pp. 548–553). Springer.

Nubatonis, O. E., Jupri, A., & Mulyaning, E. A. (2024). Interpretasi validitas dan reliabilitas instrumen kemampuan berpikir aljabar. *Jambura Journal of Mathematics Education*, 5(1), 16–28.

OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Ojose, B. (2011). Mathematics literacy: Are we able to put the mathematics we learn into everyday use? *Journal of Mathematics Education*, 4(1), 89–100.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, 187, 253–318. https://doi.org/10.1098/rsta.1896.0007

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227. https://doi.org/10.1002/sce.3730660207

Pratiwi, D., Trapsilasiwi, D., Oktavianingtyas, E., Sunardi, S., & Murtikusuma, R. (2020). Level literasi matematis siswa dalam menyelesaikan soal PISA konten change and relationship berdasarkan gaya kognitif. *KadikmA*, 10(3), 1–14. https://doi.org/10.19184/kdma.v10i3.17401

Rizki, L. M., & Priatna, N. (2019). Mathematical literacy as the 21st century skill. *Journal of Physics: Conference Series*, 1157(4), 042088.

Saputra, F. A., & Khotimah, R. P. (2023). Profil kemampuan literasi matematika dalam menyelesaikan soal berorientasi PISA ditinjau dari gaya kognitif siswa. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 7(2), 1675–1688.

Sari, R. H. N. (2015). Literasi matematis: Apa, mengapa, dan bagaimana. In *Seminar Nasional Matematika dan Pendidikan Matematika UNY* (Vol. 8).

Sari, S. A., Maharani, R., & Briones, S. V., II. (2024). Development of the PISA-like mathematics literacy test based on Yogyakarta socio-culture. *Absis: Mathematics Education Journal*, 6(2), 52–64.

Sinaga, N. A., Zahara, Y., Aufa, Z. Y., & Saragih, N. J. (2024). Validity analysis of the mathematical literacy test: Implementation of the Pancasila student profiles in instrument development. *Proceedings of the Malikussaleh International Conference on Multidisciplinary Studies (MICoMS)*, 4, 00004.

Stacey, K. (2011). The PISA view of mathematical literacy in Indonesia. *Indonesian Mathematical Society Journal on Mathematics Education*, 2(2), 95–126.

Styawati, R. D., & Nursyahida, F. (2017). Profil kemampuan literasi matematis siswa berkemampuan matematis rendah dalam menyelesaikan soal berbentuk PISA. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika*, 8(2), 33–42.

Suciati, S., Munadi, S., Sugiman, S., & Febriyanti, W. D. R. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, 9(2), 865–875. https://doi.org/10.12973/eu-jer.9.2.865

Sugiyono, D. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*. Alfabeta.

Susanta, A., Susanto, E., Stiadi, E., & Rusnilawati. (2023). Mathematical literacy skills for elementary school students: A comparative study between interactive STEM learning and paper-and-pencil STEM learning. *European Journal of Educational Research*, 12(4), 1569–1582. https://doi.org/10.12973/eu-jer.12.4.1569

Syaifuddin, M. (2022). Minimum competency assessment to measure mathematical literacy of junior high school students. *Journal of Education Research and Evaluation*, 6(2), 316–326. https://doi.org/10.23887/jere.v6i2.46263

Tabun, H. M., Taneo, P. N., & Daniel, F. (2020). Kemampuan literasi matematis siswa pada pembelajaran model problem based learning (PBL). *Edumatica: Jurnal Pendidikan Matematika*, 10(1), 1–8.

Tan-Sisman, G., & Aksu, M. (2016). A study on sixth grade students' misconceptions and errors in spatial measurement: Length, area, and volume. *International Journal of Science and Mathematics Education*, 14(7), 1293–1319. https://doi.org/10.1007/s10763-015-9642-5

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426–1431. https://doi.org/10.1021/ed500076x

Truong, H. H., Nguyen, T. T., & Nguyen, H. T. (2025). Mathematical literacy: Parallels between the PISA mathematics framework and Vietnam's mathematics curriculum. *Curriculum Journal*. Advance online publication.

Udil, P. A., Jupri, A., & Nurlaelah, E. (2023). Validitas teoritis dan empirik instrumen tes numerasi. *Euclid*, 10(4), 738–750. https://doi.org/10.33603/1q7nme39

Umami, C., Mustangin, M., & El Walida, S. (2021). Profil kemampuan literasi matematis siswa dalam menyelesaikan soal PISA ditinjau dari kecerdasan logis-matematis. *Penelitian, Pendidikan, Dan*, 16(12), 113–122.

Widjaja, W. (2011). Towards mathematical literacy in the 21st century: Perspectives from Indonesia. *Southeast Asian Mathematics Education Journal*, 1(1), 75–84.

Wu, M. (2011). Using PISA and TIMSS mathematics assessments to identify the relative strengths of students in Western and Asian countries. *Journal of Research in Education Sciences*, 56(1), 67.