ISSN 2355-4185(p), 2548-8546(e)
DOI: https://doi.org/10.24815/jdm.v13i2.2681

Developing an SDG-Based Mathematical Literacy Assessment in Space and Shape Domain for Prospective Teachers

Kurnia Putri Sepdikasari Dirgantoro[1*], Robert Harry Soesanto[1], Margaretha Madha Melissa[2]

[1]Mathematics Education Department, Universitas Pelita Harapan, Indonesia
[2]International Education Development Program, Hiroshima University, Japan
*Email: *kurnia.dirgantoro@uph.edu*

Received: 7 July 2026; Revised: 13 September 2026; Accepted: 20 October 2026

Abstract. *Mathematical literacy (ML) is a central competence in mathematics education, particularly for prospective teachers. However, most ML assessment instruments target school students, leaving a gap in instruments designed specifically for prospective teachers. This study aimed to develop and empirically evaluate an ML assessment instrument integrating Sustainable Development Goals (SDGs)-related contexts within the Space and Shape domain. Using a research and development design, the instrument was developed through a literature review, expert validation, readability testing, and a pilot test with 29 prospective mathematics teachers who had completed a geometry course. Data were analyzed using Pearson's product-moment correlation for item validity, Cronbach's Alpha for internal consistency reliability, and descriptive statistics to examine performance by item and mathematical literacy process. Nine of the ten items demonstrated statistically significant item-total correlations, while one item was excluded from the reliability analysis. Cronbach's Alpha was 0.813. Participants' mean performance was 74% in formulating mathematical situations, 65% in employing mathematics, and 58% in interpreting mathematical results. The findings provide initial evidence of the instrument's measurement properties and describe variations in prospective teachers' performance. Further testing with larger and more diverse samples is needed to establish broader applicability.*

Keywords*: mathematical literacy, prospective mathematics teachers, SDG-based assessment, space and shape domain*

How to cite: Dirgantoro, K. P. S., Soesanto, R. H., & Melissa, M. M. (2026). Developing an SDG-based mathematical literacy assessment in space and shape domain for prospective teachers. *Jurnal Didaktik Matematika*, *13*(2), 202-219. DOI: https://doi.org/10.24815/jdm.v13i2.2681

Introduction

Mathematical literacy (ML) refers to an individual's capacity to reason mathematically and to formulate, employ, and interpret mathematics to solve problems in real-world contexts. It involves using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena (OECD, 2023a). ML helps individuals recognize the usefulness of mathematics in everyday life and make informed, responsible decisions as constructive, caring, and critical citizens (Pratama & Yelken, 2024). Research across countries and educational levels has associated ML with problem solving (Muhaimin et al., 2024), mathematical communication (Hamda et al., 2025), mathematical reasoning (Mills, 2020), and creative thinking (Xu et al., 2025). Furthermore, ML helps learners connect mathematical ideas with everyday experiences

© 2026 The Authors(s). This work is licensed under a Creative Common Attribution-NonCommercial 4.0 (CC-BY-NC) International (https://creativecommons.org/licenses/by-nc/4.0/)

(Bolstad, 2023). These findings highlight that ML extends beyond procedural proficiency to the meaningful application of mathematics in real-world situations.

Developing ML is particularly important for teachers and prospective teachers because teachers' competencies influence students' learning outcomes (Podungge et al., 2020). Teachers with strong ML are better equipped to integrate ML into instruction (Lestari et al., 2018) and use questioning strategies that support students' ML development (Domu et al., 2023). Nevertheless, studies in Turkey, South Africa, and China have reported teachers' limited understanding of ML and difficulties connecting mathematics with real-world contexts, reasoning, and interpretation (Fang, 2012; Genc & Erbas, 2019). Similar challenges have been identified in Indonesia, where mathematics instruction tends to emphasize routine procedures rather than mathematical reasoning and contextual modeling (Umbara & Suryadi, 2019). Jupri and Rosjanuardi (2020) found that only around 25% of Indonesian mathematics teachers were familiar with ML-oriented tasks. Teachers also face difficulties designing engaging, contextually relevant, and higher-order thinking skills (HOTS) tasks, partly because systematic evaluation of task development is lacking (Syafrudin & Vitantri, 2024). Other challenges involve instructional model design, learning resources, students' contextual experiences, and assessment of learning outcomes (Pulungan et al., 2024).

These challenges emphasize the importance of developing ML among prospective teachers before they enter the profession. Higher education institutions need valid and reliable assessment instruments to identify prospective teachers' competencies and difficulties and inform instructional improvement. Nabila et al. (2023) assessed prospective teachers' ML using written tests and interview guidelines. Although this study provides valuable foundations, further development is needed to broaden the mathematical content and contextual applications represented in ML assessment for prospective teachers.

One relevant content domain is Space and Shape, which encompasses geometry, spatial visualization, and representations of two- and three-dimensional forms (OECD, 2023a). This domain assesses learners' ability to interpret visual representations, model spatial situations, and evaluate solutions in contexts such as planning, navigation, mapping, and architecture. However, challenges in this domain remain evident. Ma'ruf et al. (2025) reported that Indonesian students demonstrated below-average ML performance on tasks involving two-dimensional shapes. Similarly, Dirgantoro et al. (2025) found that Indonesian prospective teachers demonstrated low ML, particularly in Space and Shape. These findings highlight the need for assessment instruments that examine prospective teachers' ability to apply mathematics in spatial and geometric situations.

In developing ML assessments, context is an important component because ML emphasizes using mathematics in real-world situations. The PISA framework identifies four contexts: personal, occupational, societal, and scientific (OECD, 2023a). Beyond providing a narrative setting, context should create meaningful situations that require learners to apply mathematics. The Sustainable Development Goals (SDGs) offer a promising basis for such contexts. Established as a global framework for addressing social, economic, and environmental challenges during 2015-2030, the SDGs encompass issues such as poverty, education, health, gender equality, and climate change (United Nations, 2026). For example, Space and Shape tasks may address green-space planning under SDG 11 (Sustainable Cities and Communities), climate-related issues under SDG 13 (Climate Action), or clean-water distribution under SDG 6 (Clean Water and Sanitation). Integrating these contexts can connect mathematical reasoning with sustainability challenges relevant to everyday life. Previous research has demonstrated that sustainability-oriented mathematics education can enhance learning relevance and support competencies associated with Education for Sustainable Development (ESD) (Vásquez et al., 2023). This approach is also consistent with the National Council of Teachers of Mathematics' recommendation to connect mathematics with real-world contexts to strengthen conceptual understanding and mathematical modeling.

Despite these developments, the integration of ML assessment, Space and Shape domain, and SDGs contexts remains insufficiently explored, particularly for prospective teachers. Existing instruments have predominantly assessed students' ability to apply mathematics in contextual situations, while assessment specifically designed for prospective teachers remains relatively limited. In addition, contexts in ML instruments are often used as general or realistic settings rather than as integral components of problems addressing global sustainability challenges. Although previous research has examined ML assessment for prospective teachers (Nabila et al., 2023), ML performance in Space and Shape (Dirgantoro et al., 2025; Ma'ruf et al., 2025), and sustainability-oriented mathematics education (Vásquez et al., 2023), these strands have not been sufficiently integrated into a single assessment instrument.

Accordingly, this study's novelty lies in developing an ML assessment instrument for prospective teachers that systematically integrates SDG contexts into the Space and Shape domain. Rather than treating the SDGs merely as thematic backgrounds, the instrument uses sustainability-related situations to engage prospective teachers in formulating, employing, and interpreting mathematics. This study therefore contributes to the development of PISA-oriented ML assessment by connecting geometric and spatial reasoning with meaningful sustainability challenges. Specifically, it aims to develop an ML assessment instrument for prospective teachers in the Space and Shape domain using SDG-based contexts and to examine its validity and

reliability. It also aims to describe prospective teachers' performance on the developed instrument during the pilot test. The research question is: How can an SDG-based ML assessment instrument be developed and validated to assess prospective teachers' ML in the Space and Shape domain? How do prospective teachers perform on the developed assessment instrument during the empirical trial?

Method

This study employed a modified Research and Development (R&D) framework adapted from Borg and Gall, whose original model comprises ten stages. The framework was adapted into five stages to align with the study's focus on developing and conducting an initial empirical evaluation of an SDG-based mathematical literacy (ML) assessment instrument in the Space and Shape domain. The five stages were: (1) preliminary study, to identify the research needs and relevant assessment requirements; (2) planning, to establish the assessment framework and design specifications; (3) instrument development, to construct the assessment items and associated scoring criteria; (4) instrument finalization, to revise the instrument based on expert validation and readability feedback; and (5) validity and reliability testing, to examine the instrument empirically. Compared with the original ten-stage model, the adapted framework consolidated the development and revision activities and limited empirical testing to an initial trial. The subsequent stages involving larger-scale field testing, further product revisions, dissemination, and implementation were not undertaken because the present study focused on developing the instrument and obtaining initial evidence of its validity, reliability, and performance in a pilot trial.

Preliminary Study

The preliminary study involved a literature review of theoretical and empirical studies on mathematical literacy (ML), the Space and Shape domain, and the Sustainable Development Goals (SDGs). The review aimed to identify relevant concepts and synthesize findings from previous studies. It also helped identify research gaps that informed the development of the assessment instrument.

Planning

Based on the preliminary study, the instrument's competencies and characteristics were specified. A conceptual framework was developed integrating mathematical content, SDG-related themes, and four contextual categories: personal, occupational, social, and scientific. Test

blueprint, expert validation sheet, readability assessment, and scoring rubric were subsequently prepared as supporting components for instrument development and evaluation.

Instrument Development

The initial instrument was developed based on the conceptual framework and relevant theoretical and empirical foundations, with particular attention to the characteristics of prospective mathematics teachers. The prototype was evaluated through expert validation and readability testing. Expert validation involved three mathematics education specialists (two professors and one master's-level lecturer), who were assigned to evaluate the instrument in terms of item construction, content relevance, language clarity, and usability, with particular attention to its capacity to assess the processes of formulating, employing, interpreting, and evaluating mathematics in real-world contexts within the Space and Shape domain.

Readability testing involved ten third-year undergraduate students representing high, medium, and low mathematical ability, categorized by GPA ranking. These students had completed most mathematics content courses, including Geometry. Readability was assessed using seven indicators: (1) language clarity, (2) vocabulary appropriateness, (3) clarity of instructions, (4) clarity of context and information, (5) consistency of notation and symbols, (6) visual presentation and formatting, and (7) alignment between item objectives and competency indicators (Li et al., 2025; Patel et al., 2023). Based on these indicators, the research team formulated a set of statements that were administered to the ten respondents, as presented in Table 1. Feedback from experts and respondents was documented and used to revise the instrument and minimize ambiguity or misinterpretation.

Table 1. Readability test statements description

Readability Indicators	Statement Description
Clarity of language	The sentences used in the test items are simple, grammatically clear, and not unnecessarily complex.
Appropriateness of vocabulary relative to the readers' level	Mathematical terminology is used consistently and is appropriate to the students' level of understanding.
	The items do not contain unfamiliar or foreign terms without adequate explanation.
Clarity of instructions	The instructions are clear and explicitly indicate the actions required from respondents (e.g., prove, explain, determine).
Clarity of context and information provided	The information presented in the items is relevant, sufficiently complete, not excessive, and aligned with real-life contexts relevant to prospective teachers.
Consistency of notation and symbols	Mathematical notations and symbols are used consistently and do not create ambiguity or multiple interpretations.
Clarity of visual presentation and formatting	The presentation of the items (text, figures, or tables) is well-organized, easy to read, and supports comprehension of the problem content.
Alignment between the item objectives and the intended competency indicators	The language and structure of the items reflect the targeted mathematical literacy indicators, including the ability to formulate real-world situations into mathematical problems, apply appropriate mathematical procedures, and interpret results within real-world contexts.

Pilot Testing

The revised instrument was administered in a pilot test involving 29 second-year prospective mathematics teachers who had completed a Geometry course. Participants completed the ten assessment items individually in a 150-minute session. The pilot test had two purposes: (1) to examine the validity and internal consistency reliability of the assessment items and (2) to describe prospective teachers' mathematical literacy performance at both the item and process levels. Item validity was examined using Pearson's product-moment correlation, internal consistency reliability was assessed using Cronbach's Alpha, and descriptive statistics were used to analyze participants' performance across the assessment items and the three mathematical literacy processes: formulating, employing, and interpreting mathematics.

Validity and Reliability Testing

The pilot test responses were analyzed to examine item validity and internal consistency reliability. Item validity was assessed using Pearson's product-moment correlation between item scores and total test scores. Items with statistically significant item-total correlations were retained, while those failing to meet the validity criteria were excluded. Internal consistency was assessed using Cronbach's Alpha and interpreted according to the reliability criteria in Table 2.

Table 2. Cronbach's alpha reliability score (Creswell, 2012)

Coefficient of Cronbach's Alpha	Reliability Level
$x < 0.6$	Poor
$0.6 \leq x < 0.7$	Questionable
$0.7 \leq x < 0.8$	Acceptable
$0.8 \leq x < 0.9$	Good
$x \leq 0.9$	Excellent

Results and Discussion

The development process resulted in an ML assessment consisting of ten open-ended items designed for prospective teachers in the Space and Shape content domain. The assessment integrates ML with contexts related to the SDGs. The items were constructed to represent different real-world situations and to engage prospective teachers in applying mathematics to contextual problems rather than merely performing routine mathematical procedures. The assessment covers a range of mathematical topics within the Space and Shape domain and incorporates four context types: personal, occupational, social, and scientific. Each item was also associated with a relevant SDG and designed to address one or more ML processes, namely formulate, employ, and interpret. In addition, the items were classified according to the proficiency levels described in the PISA Mathematics Framework to indicate the level of mathematical demand involved in solving each task. Table 3 summarizes the characteristics of the ten assessment items by

mathematical topic, relevant SDG, context, mathematical content, ML process, and PISA proficiency level.

Table 3. Characteristics of the developed mathematical literacy assessment items based on the PISA mathematics framework

Item	Mathematical Topic	Relevant SDG	Context	Mathematical Content	Mathematical Literacy Process	PISA-aligned Proficiency Level
1	Quadrilaterals	SDG 11 (*Sustainable Cities*)	Personal	Area and cost	Formulate, Employ, Interpret	Level 2
2	Circle	SDG 11 (*Sustainable Cities*)	Occupational	Area of rectangle, circle, and quarter-circles	Formulate, Employ, Interpret	Level 3
3	Triangle	SDG 7 (*Clean Energy*)	Scientific	Area, spatial arrangement, geometric reasoning	Formulate, Employ, Interpret	Level 5
4	Cubes	SDG 6 (*Clean Water*)	Personal	Volume, unit conversion, capacity	Formulate, Employ, Interpret	Level 2
5	Cylinders	SDG 6 (*Clean Water*)	Occupational	Volume, capacity, proportional reasoning, distribution	Formulate, Employ, Interpret	Level 3
6	Truncated pyramid and cuboid	SDG 12 (*Responsible Consumption*)	Societal	Surface area, cost, comparison, decision making	Formulate, Employ, Interpret	Level 4
7	Triangular prism and cylinder	SDG 12 (*Responsible Consumption*)	Occupational	Volume and contextual comparison	Formulate, Employ, Interpret	Level 4
8	Sphere	SDG 15 (*Life on Land*)	Scientific	Volume, unit conversion, proportional reasoning, cost	Formulate, Employ, Interpret	Level 3
9	Measurement	SDG 11 (*Sustainable Cities*)	Societal	Distance, speed, time, inverse relationship	Formulate, Employ, Interpret	Level 3
10	Measurement	SDG 13 (*Climate Action*)	Scientific	Flow rate, unit conversion, percentage change	Formulate, Employ, Interpret	Level 3

As shown in Table 3, the developed assessment integrates the three dimensions targeted in this study: mathematical literacy, Space and Shape content, and SDG-related contexts. The use of different contexts and mathematical topics was intended to provide varied situations in which prospective teachers were required to formulate mathematical problems, apply appropriate mathematical procedures, and interpret their results in relation to the situations presented. Figure 1 presents Item 7 as a representative example of the assessment's structure and key features. This item addresses the volume of a triangular prism and is situated in an occupational context related to SDG 12 (Responsible Consumption and Production).

A beverage company wants to design environmentally friendly packaging in the shape of a right triangular prism for its drink product. The planned dimensions are: triangle base = 8 cm, triangle height (on the base) = 6 cm, and prism height = 15 cm. Assume the package is filled almost to capacity (ignore the cap), and the packaging walls are thin so that the internal volume closely approximates the geometric volume of the prism.

a. Calculate the volume of the packaging (in mL).
b. Discuss the advantages and disadvantages of a triangular prism-shaped package compared to a cylindrical package for beverage products, considering aspects of production, packaging, usage function (pouring/holding), and aesthetics.

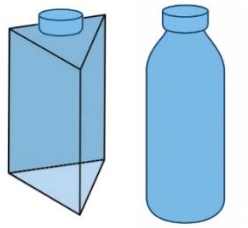


Figure 1. Initial version of item 7

The scoring rubric for Item 7 assesses mathematical literacy through formulating, employing, and interpreting mathematics, as presented in Table 4. Item 7a evaluates mathematical modeling, calculation, and interpretation of a triangular prism's volume, with a maximum score of 5. Item 7b assesses comparing packaging shapes based on relevant aspects and logical reasoning, with a maximum score of 3, for a total maximum score of 8.

Table 4. Scoring rubric

No	Alternative Answers	Components	Indicators	Student's response	Sub Score	Max Score
7a	Base area of the triangular prism: $\frac{1}{2} \times$ base $\times$ height $= \frac{1}{2} \times 8 \times 6 =$ 24 cm^2	Formulate	Formulating a mathematical model for the volume of a triangular prism	No response or irrelevant response	0	2
				States the formula, but it is incomplete or inaccurate	1	
				Correctly formulates the complete model: base area × prism length (height)	2	
	Volume of prism: $= 24 \times 15$ $= 360$ cm^3 $= 360$ mL	Employ	Performing calculations	Incorrect calculation	0	2
				Partially correct calculation	1	
				Correct calculation yielding 360 mL	2	
		Interpret	Expressing the result as the volume of the package	Does not interpret the result	0	1
				Correctly interprets the result	1	
7b	Provides a comprehensive comparison of at least three relevant aspects, discussing the advantages and disadvantages of both packaging shapes with logical, contextually appropriate reasoning.	Interpret	Relating geometric properties to practical aspects of packaging design	No response or irrelevant response	0	3
				Identifies one aspect of comparison in general terms	1	
				Explains at least two aspects of comparison, but the relationships between them are not well established	2	
				Provides a comprehensive comparison covering at least three aspects, supported by logical and relevant reasoning	3	
Total Score for Item 7						8

The ten items constituted the initial prototype and were subsequently subjected to expert validation and readability testing. Expert validation was conducted by three experts in

mathematics education. The validation process covered four major aspects: item construction, content relevance, language clarity, and instrument usability. The experts provided qualitative feedback rather than numerical ratings for individual items; therefore, no item-level expert validity scores were calculated. In general, all experts provided constructive feedback addressing both technical and substantive aspects of the items. Their suggestions included improving the clarity of item instructions to minimize ambiguity, refining the proportionality and accuracy of visual representations, selecting numerical values that better reflect realistic contexts, and making several minor revisions, such as wording improvements, adding unit-conversion guidance, and correcting measurement units.

Item 7 serves as a representative example of the developed assessment and is used again here to illustrate how expert feedback informed the revision process. Specifically, the initial version of Item 7, the three experts' substantive comments, the revisions made in response, and the resulting revised version are presented below. Table 5 summarizes the three experts' substantive feedback on Item 7 and the corresponding revisions.

Table 5. Summary of expert feedback and revisions for Item 7

Expert	Identified Issue	Expert's Comment	Revision Undertaken
Expert 1	Unit conversion in Item 7(a)	Clarify the conversion from cm^3 to mL.	Added the equivalence 1 cm^3 = 1 mL to the question to clarify the required unit conversion.
Expert 1	Scoring rubric for Item 7(b)	Responses to Item 7(b) may vary; therefore, clear scoring guidelines are needed.	Developed an analytic scoring rubric with explicit criteria covering production, packaging, functionality (pouring and handling), and aesthetics to accommodate varied responses and support consistent scoring.
Expert 2	Relevance of the packaging context	Beverage containers commonly have curved surfaces. Packaging with flat sides may be less ergonomic; therefore, the context should be reconsidered.	The triangular-prism packaging context was retained, while the comparison task was clarified to encourage consideration of practical advantages and disadvantages in production, packaging, functionality, and aesthetics.
Expert 2	Clarity of the visual representation	The purpose of the cylindrical packaging shown in the illustration is unclear.	Refined the illustration and specified the cylindrical container's dimensions (d = 8 cm and t = 15 cm) in Item 7(b) to provide a clearer basis for comparison with the triangular-prism packaging.
Expert 3	Overall suitability of Item 7	The item was considered acceptable.	The item was retained, incorporating relevant revisions based on feedback from the other experts.

The research team carefully reviewed and discussed all expert comments. The team then incorporated the revisions to improve the item's contextual coherence, visual accuracy, and clarity. The final version of Item 7 is presented in Figure 2.

A beverage company wants to design environmentally friendly packaging in the shape of a right triangular prism for its beverage product. The proposed dimensions are a triangular base with a base length of 8 cm and a triangle height of 6 cm, and a prism length of 15 cm. Assume that the package is filled almost to capacity (ignore the cap) and that the packaging walls are sufficiently thin so that the internal volume approximates the geometric volume of the prism.

a. Calculate the volume of the package in milliliters. Use the conversion 1 cm^3 = 1 mL.
b. Discuss the advantages and disadvantages of a triangular prism-shaped package compared to a cylindrical package for beverage products, considering aspects of production, packaging, usage function (pouring/holding), and aesthetics.

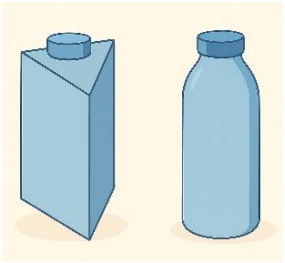


Figure 2. Final version of item 7

Following the expert validation stage, the study proceeded to readability testing to ensure the ML instrument used clear, communicative language appropriate to the characteristics of prospective teachers. This stage also aimed to confirm that the items could be understood without ambiguity and that the contextual situations presented were relevant to students' learning experiences and academic environment. Respondents involved in the readability test examined all items using eight guiding questions derived from the readability indicators, as previously presented in Table 1.

Overall, most respondents commented primarily on the explicitness of item instructions and the adequacy of contextual information, particularly whether the information was relevant, sufficiently detailed, not excessive, and consistent with real-life contexts. Table 6 presents sample feedback obtained from respondents during the readability testing stage.

In general, respondents stated that the ML instrument was contextually meaningful and understandable, although several suggested that additional explanatory information could further enhance clarity and precision. The research team also incorporated all comments from the readability testing stage into its discussions, which informed further refinement of the instrument.

After completing the expert validation and readability testing stages, the developed assessment was subjected to a limited-scale empirical trial involving 29 prospective teachers who were different from those participating in the readability test. All participants completed the assessment. The empirical trial data were analyzed using IBM SPSS Statistics version 25. Item validity was examined using the Pearson product-moment correlation, while internal consistency reliability was assessed using Cronbach’s Alpha.

Table 6. Sample comments on readability test

Guiding Question	Sample Comment
The sentences used in the test items are simple, grammatically clear, and not unnecessarily complex.	The sentences are generally clear; however, they could be simplified further by clarifying that the grass is installed only in the park area, not on the paving blocks. (Respondent 1, Item 1)
Mathematical terminology is used consistently and is appropriate to the students' level of understanding.	The term "hectare" in the item could be accompanied by its equivalent conversion in hm^2, since not all students may know that one hectare equals hm^2. (Respondent 7, Item 8)
The items do not contain unfamiliar or foreign terms without adequate explanation.	The term "panel" is still relatively unfamiliar. (Respondent 8, Item 3)
The instructions are clear and explicitly indicate the actions required from respondents (e.g., prove, explain, determine).	- Note: All respondents indicated no difficulties related to this aspect.
The information in the items is relevant, sufficiently complete, not excessive, and aligned with real-life contexts for prospective teachers.	The phrase "…to avoid …" appears unnecessary. (Respondent 4, Item 3)
Mathematical notations and symbols are used consistently and do not create ambiguity or multiple interpretations.	In item (a), the time unit should be explicitly stated (minutes or seconds). (Respondent 5, Item 9)
The presentation of the items (text, figures, or tables) is well-organized, easy to read, and supports comprehension of the problem content.	Clarify the figure, particularly whether the container has a lid, as this may lead to different interpretations. (Respondent 2, Item 6)
The language and structure of the items reflect the targeted mathematical literacy indicators, including the ability to: formulate real-world situations into mathematical problems apply appropriate mathematical procedures interpret results within real-world contexts.	- Note: All respondents reported no issues related to this aspect.

Item validity was interpreted using the Pearson correlation coefficient based on the critical r-table (N = 29, r = 0.367, $\alpha = 0.05$) and the significance value (Sig.). An item was considered valid when r > r-table or Sig. < α. Table 7 presents the Pearson correlation analysis results. The analysis indicates that one item (Item 3) did not meet the validity criteria and was therefore removed from the instrument.

Table 7. Pearson validity scores

Item	Pearson Correlation (r)	Sig	Interpretation
1	0.405	0.029	Valid (accepted)
2	0.375	0.045	Valid (accepted)
3	0.284	0.136	Not Valid (removed)
4	0.506	0.005	Valid (accepted)
5	0.768	< 0.001	Valid (accepted)
6	0.545	0.002	Valid (accepted)
7	0.767	< 0.001	Valid (accepted)
8	0.745	< 0.001	Valid (accepted)
9	0.651	< 0.001	Valid (accepted)
10	0.742	< 0.001	Valid (accepted)

After Item 3 was excluded, the Cronbach's Alpha coefficient was 0.813. Based on the reliability classification presented in Table 2, this value falls within the "very good" reliability category, indicating strong internal consistency of the developed instrument. Beyond item validity and reliability, the empirical trial responses were analyzed to examine participants' mathematical literacy performance. The analysis focused on participants' performance across the developed assessment items and the mathematical literacy processes required to solve them. The results provide empirical evidence of how prospective teachers responded to the SDG-based mathematical situations and the extent to which they formulated, employed, and interpreted mathematics in the Space and Shape domain. Table 8 presents the mean scores and mean percentages obtained for each of the ten items administered during the empirical trial.

Table 8. Participants' mathematical literacy performance by assessment item

Item	Maximum Score	Mean Score	Mean Performance (%)
1	10	7.14	71.38
2	6	5.62	93.68
3	11	6.45	58.62
4	9	6.86	76.25
5	12	6.83	56.90
6	8	5.83	72.84
7	8	5.48	68.53
8	6	5.38	89.66
9	12	6.86	57.18
10	10	5.45	54.48

As shown in Table 8, participants' performance varied across the ten assessment items. The mean performance ranged from 54.48% to 93.68%. The highest mean performance was on Item 2 (93.68%), followed by Item 8 (89.66%), while the lowest was on Item 10 (54.48%). Item 3, which was subsequently removed based on the item validity analysis, had a mean performance of 58.62%. The participants' performance was also examined across the three mathematical literacy processes targeted by the assessment: formulate, employ, and interpret. The results are presented in Table 9.

Table 9. Participants' mathematical literacy performance by process

Mathematical Literacy Process	Mean Performance (%)
Formulate	74
Employ	65
Interpret	58

Table 9 shows that participants achieved a mean performance of 74% in the formulate process, 65% in the employ process, and 58% in the interpret process. These results indicate variations in participants' performance across the three mathematical literacy processes, with the highest performance in formulating mathematical situations and the lowest in interpreting mathematical results in context. This pattern provides an initial description of the mathematical

literacy aspects assessed during the pilot test. It also shows the assessment's potential to identify variations in prospective teachers' performance across the targeted mathematical literacy processes. However, further testing with a larger sample is needed to establish the broader applicability of these findings.

The findings indicate that the SDGs-based ML instrument in the space and shape domain developed in this study meets key quality criteria for educational assessment instruments, particularly in terms of validity and reliability. The item-validity analysis showed that nine of the ten items had statistically significant item-total correlations, indicating that these items consistently represent the underlying construct of mathematical literacy. This result aligns with established principles of instrument development, which emphasize the alignment between measurement objectives, competency indicators, and item characteristics (Creswell, 2014). Similarly, Maidiyah et al. (2023) highlighted the importance of empirical evaluation by assessing the validity, practicality, and potential effect of context-based numeracy problems through field testing. These findings underscore the value of systematic instrument development, including literature review, expert validation, and readability testing, in establishing assessment quality.

The failure of one item (Item 3) to meet the validity criterion offers valuable insight into the complexities of integrating real-world contexts into ML assessments. The invalid item incorporated a scientific SDG-related context requiring relatively complex geometric modeling, which may have increased respondents' cognitive load. This finding supports the argument that while authentic contexts can enhance the relevance of assessment tasks, they must be designed carefully so they do not obscure the mathematical constructs being measured (Palm, 2008; Sweller et al., 2011). Therefore, removing this item should not be viewed as a limitation of the instrument, but as an essential refinement step to maintain the construct validity of the assessment as a whole.

From a reliability perspective, the Cronbach's Alpha coefficient of 0.813 indicates strong internal consistency. This level of reliability suggests that the remaining items function coherently in measuring the mathematical literacy construct among prospective teachers. This result is particularly important because ML instruments are used not only for academic measurement but can also serve as diagnostic tools for evaluating prospective teachers' professional readiness to connect mathematical knowledge with real-world situations (Tavakol & Dennick, 2011). With this level of reliability, the instrument is appropriate for assessing prospective teachers' ML.

The pilot results provide an initial description of prospective teachers' mathematical literacy performance across the three processes, with mean scores of 74% for formulating, 65% for employing, and 58% for interpreting mathematics. This pattern corresponds to the mathematical literacy framework described by the OECD (2023a), with the highest performance

in formulating and the lowest in interpreting mathematical results. These findings complement Winarni et al. (2025), who identified conceptual and procedural errors in prospective teachers' contextual numeracy, and Wulan and Astuti (2022), who reported differences in performance across the three ML processes by personality type. Although the present study did not analyze individual errors or personality-related differences, the lower interpreting score may reflect the demands of relating mathematical results to their original context and evaluating their meaning. However, further analysis of individual responses is needed to establish the factors underlying this pattern. Assessment-development research, including Utari and Gustiningsi (2024), also highlights the importance of field testing in evaluating the quality of developed mathematics assessment materials. While these studies differ in their constructs, participants, and purposes, they collectively emphasize the value of empirical assessment. In the present study, the pilot findings demonstrate the instrument's potential to characterize variations in prospective teachers' performance, rather than establish an effect on their mathematical literacy. Further response analysis and testing with larger, more diverse samples are needed to clarify these performance patterns and establish the instrument's broader applicability.

Integrating SDGs contexts within the ML instrument provides a meaningful conceptual contribution to the development of context-based mathematical assessment. Unlike conventional contextual problems that tend to be neutral or merely realistic, the SDG contexts in this study were intentionally designed to require mathematical modeling, interpretation of results, and reflection on social, environmental, and sustainable development challenges. This approach aligns with the conceptualization of ML as the ability to use mathematics functionally and reflectively in real-life situations (OECD, 2023b). Accordingly, the instrument not only assesses procedural competence but also encourages prospective teachers to recognize the role of mathematics in informed and responsible decision-making.

From a pedagogical perspective, the findings suggest that the SDG-based ML instrument has considerable potential to enrich both teaching and assessment practices in prospective teacher education programs. Such instruments enable lecturers to map students' mathematical literacy achievements in a more contextualized and authentic manner, while also providing a reflective basis for designing geometry learning experiences that are more meaningful and socially relevant. This perspective aligns with NCTM recommendations, which emphasize authentic assessments that reflect the application of mathematics in real-world contexts (National Council of Teachers of Mathematics, 2000). Therefore, the instrument developed in this study contributes not only to the research domain of mathematical literacy assessment but also offers practical implications for curriculum development and instructional practices in prospective teacher education.

Conclusion

This study aimed to develop and conduct an initial empirical evaluation of an SDGs-based ML instrument in the Space and Shape domain for prospective mathematics teachers. Nine of the ten items showed statistically significant item–total correlations, and the instrument achieved a Cronbach's alpha of 0.813, providing initial evidence of validity and internal consistency. The pilot results showed mean performance levels of 74% for formulating mathematical situations, 65% for employing mathematics, and 58% for interpreting mathematical results. By integrating SDG-related contexts into problem design, the instrument requires mathematical modeling, reasoning, and interpretation in authentic real-world situations, offering conceptual and pedagogical value for sustainability-oriented mathematics assessment.

Nevertheless, this study has several limitations. Instrument development focused primarily on the space and shape domain, and pilot testing involved a relatively small sample with specific participant characteristics. Future studies are therefore encouraged to expand ML instrument development across other domains, such as quantity, algebra, and statistics, while also conducting external validation with larger and more diverse populations. Such efforts would further strengthen the generalizability and robustness of ML assessment instruments. Overall, the instrument developed in this study is an initial step toward establishing a contextualized, sustainability-oriented ML assessment framework for prospective teachers. By integrating meaningful global contexts with rigorous psychometric validation, this study contributes to the growing body of research aligning mathematics education with the broader goals of sustainable development and future-oriented teacher preparation.

Statements

Author contributions	:	KPSD: conceptualization and writing-original draft; RHS: formal analysis and writing-review & editing; & MMM: writing-review & editing. All authors agreed with the results and conclusions.
Funding (If any)	:	This research was funded by Universitas Pelita Harapan under the performance contract (No. P-044-INT-FIP/VII/2025), dated July 1, 2025.
Ethical Statement	:	This study did not require formal ethical review under applicable institutional guidelines, as it involved minimal risk to participants. Informed consent was obtained from all participants before data collection. Participation was voluntary, and participants' data were used solely for research purposes. The confidentiality of participants' information was maintained throughout the study.

AI Statement	:	The authors stated that they used AI tools solely for English grammar refinement and reference formatting. The authors reviewed the content, take full responsibility for its integrity, and confirm that no AI was used to create or alter research data.
Declaration of interest	:	No conflict of interest is declared by the authors.
Data sharing statement	:	Data supporting the findings and conclusions are available upon request from the corresponding author.

References

Bolstad, O. H. (2023). Lower secondary students' encounters with mathematical literacy. *Mathematics Education Research Journal*, *35*(1), 237–253. https://doi.org/10.1007/s13394-021-00386-7

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods*. Sage.

Dirgantoro, K. P. S., Jupri, A., & Soesanto, R. H. (2025). Tracing the mathematical literacy of prospective teachers: Teaching readiness in the modern era. *RANGE: Jurnal Pendidikan Matematika*, *7*(1), 331–350. https://doi.org/10.32938/jpm.v7i1.9436

Domu, I., Regar, V. E., Kumesan, S., Mangelep, N. O., & Manurung, O. (2023). Did the teacher ask the right questions? An analysis of teacher asking ability in stimulating students' mathematical literacy. *Journal of Higher Education Theory and Practice*, *23*(5), 247–255. https://doi.org/10.33423/jhetp.v23i5.5970

Fang, Q. (2012). A study on high school mathematics teachers' mathematics professional literacy: What does high quality mathematics teaching need?. *Journal of Mathematics Education*, *5*(1), 136–158. https://journalofmathed.scholasticahq.com/article/90299

Genc, M., & Erbas, A. K. (2019). Secondary mathematics teachers' conceptions of mathematical literacy. *International Journal of Education in Mathematics, Science and Technology*, *7*(3), 222–237. https://ijemst.com/index.php/ijemst/article/view/433

Hamda, Faisal, A., Alimuddin, Naufal, M. A., & Hassan, M. N. (2025). Exploring students' mathematical literacy based on mastery of prerequisite material. *Journal of Educational Research and Evaluation*, *9*(2), 234–243. https://doi.org/10.23887/jere.v9i2.89037

Jupri, A., & Rosjanuardi, R. (2020). Analysis of teacher understanding on mathematical literacy problems. *Jurnal Pengajaran Matematika dan Ilmu Pengetahuan Alam*, *25*(1), 1–6. https://doi.org/10.18269/jpmipa.v25i1.21690

Lestari, N. D. S., Juniati, D., & Suwarsono, S. (2018). Gender differences in prospective teachers' mathematical literacy: Problem solving of occupational context on shipping company. In *Journal of Physics: Conference Series*, *1008*(1), 1–11. https://doi.org/10.1088/1742-6596/1008/1/012074

Li, H., Xing, W., Li, C., Zhu, W., & Oh, H. (2025). Are simpler math stories better? Automatic readability assessment of GAI-generated multimodal mathematical stories validated by engagement. *British Journal of Educational Technology*, *56*(3), 1092–1117. https://doi.org/10.1111/bjet.13554

Ma'ruf, A. H., Triyono, A., Fauziah, N. R., Warohmah, M., Nurimani, & Kusuma, A. P. (2025). Mapping students' mathematical literacy skills in basic geometry: A study based on Stacey and Turner's indicators. *Alpha Math*, *11*(2), 139–158. https://doi.org/10.30595/alphamath.v11i2.28590

Maidiyah, E., Agusta, R. T., & Johar, R. (2023). Development of numeracy problems with the context of bee cultivation activity for junior high school students. *Jurnal Didaktik Matematika*, *10*(2), 300–318.

Mills, J. (2020). Teachers' reflecting on professional knowledge in the numeracy (mathematics) classroom. *Teachers and Curriculum*, *20*(1), 23–30. http://files.eric.ed.gov/fulltext/EJ1266556.pdf

Muhaimin, L. H., Sholikhakh, R. A., Yulianti, S., Ardani, Hendriyanto, A., & Sahara, S. (2024). Unlocking the secrets of students' mathematical literacy to solve mathematical problems: A systematic literature review. *EURASIA Journal of Mathematics, Science and Technology Education*, *20*(4), 1–15. https://doi.org/10.29333/ejmste/14404

Nabila, F., Permadi, H., & Sukoriyanto. (2023). Literasi matematis mahasiswa calon guru dalam menyelesaikan soal numerasi statistik berdasarkan gaya belajar Honey-Mumford. *SJME (Supremum Journal of Mathematics Education)*, *7*(2), 195–209. https://doi.org/10.35706/sjme.v7i2.7757

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.

OECD. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://www.oecd-ilibrary.org/education/pisa-2022-assessment-and-analytical-framework_dfe0bf9c-en

OECD. (2023b). *Program for international student assessment and analytical framework*. OECD Publishing.

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research & Evaluation*, *13*(4), 1–11. https://doi.org/10.7275/0qpc-ws45

Patel, N., Nagpal, P., Shah, T., Sharma, A., Malvi, S., & Lomas, D. (2023). Improving mathematics assessment readability: Do large language models help?. *Journal of Computer Assisted Learning*, *39*(3), 804–822. https://doi.org/10.1111/jcal.12776

Podungge, R., Rahayu, M., Setiawan, M., & Sudiro, A. (2020). Teacher competence and student academic achievement. *In 23rd Asian Forum of Business Education (AFBE)*, *144*(16), 69–74. https://doi.org/10.2991/aebmr.k.200606.011

Pratama, R. A., & Yelken, T. Y. (2024). Effectiveness of ethnomathematics-based learning on students' mathematical literacy: A meta-analysis study. *Discover Education*, *3*, 1–15. https://doi.org/10.1007/s44217-024-00309-1

Pulungan, D. A., Herosian, M. Y., & Harahap, I. H. (2024). Kesulitan guru menerapkan pembelajaran literasi matematika di madrasah ibtidaiyyah kota Medan. *Journal of Education Research*, *5*(1), 651–659. https://doi.org/10.37985/jer.v5i1.849

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4_18

Syafrudin, T., & Vitantri, C. A. (2024). Hambatan guru dalam pengajuan soal literasi matematika bertema word problem. *Jurnal Tadris Matematika*, *7*(2), 207–220. https://doi.org/10.21274/jtm.2024.7.2.207-220

Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. 10.5116/ijme.4dfb.8dfd

Umbara, U., & Suryadi, D. (2019). Re-interpretation of mathematical literacy based on the teacher's perspective. *International Journal of Instruction*, *12*(4), 789–806. https://doi.org/10.29333/iji.2019.12450a

United Nation. (2026). *The 17 goals*. Sustainable Development.

Utari, R. S., & Gustiningsi, T. (2024). Developing of higher order thinking skill in relation and function to support student's creative thinking. *Mathematics Education Journal*, *15*(1), 49–60. https://jpm.ejournal.unsri.ac.id/index.php/jpm/article/view/204

Vásquez, C., Alsina, Á., & Seckel, M. J. (2023). Integrating sustainability in mathematics education and statistics education: A systematic review. *Eurasia: Journal of Mathematics, Science and Technology Education*, *19*(11), 1-16. https://doi.org/10.29333/ejmste/13809

Winarni, S., Kamid, Marzal, J., & Asrial. (2025). Numeracy skill development in prospective mathematics teachers: Challenges and opportunities in real-world contexts. *Journal on Mathematics Education*, *16*(1), 91–110. https://doi.org/10.22342/jme.v16i1.pp91-110

Wulan, E. R., & Astuti, F. (2022). Mathematical literacy with Islamic Nuances based on prospective mathematics teachers ' personality. *Jurnal Didaktik Matematika*, *9*(2), 230–247.

Xu, B., Ma, X., Zhang, Y., & Wu, X. (2025). How does mathematical literacy affect creative thinking? Independent effects and differential impacts across proficiency groups. *Acta Psychologica*, *260*, 1–14. https://doi.org/10.1016/j.actpsy.2025.105509