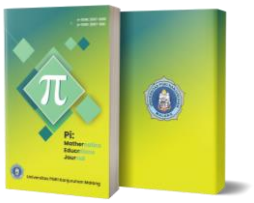
Pi: Mathematics Education Journal
e-ISSN: 2597-6915 p-ISSN: 2597-5161
Vol. 9, No. 2, 2026 | Pp. 22 – 41



Developing PISA-Like Mathematics Problems Using the Context of Islamic Architecture for Junior High School Students

Mar'atus Syakdia[a,1], Agustiany Dumeva Putri[b,2], Harisman Nizar[c,3,*], Liana Septy[d,4]
[a,b,c,d]Department of Mathematics Education, Fakultas Ilmu Tarbiyah dan Keguruan, Universitas Islam Negeri Raden Fatah Palembang, South Sumatra, Indonesia
[1]syakdiam@gmail.com
[2]agustianydumevaputri_uin@radenfatah.ac.id
[3]harismannizar_uin@radenfatah.ac.id
[4]lianasepti_uin@radenfatah.ac.id
*Corresponding Author

ARTICLE HISTORY
Received:
July 26, 2026

Revised:
October 21, 2026

Published:
October 31, 2026

Keywords:
PISA-like mathematics problems
Islamic architecture
Validity
Practicality

ABSTRACT
This study aims to develop PISA-like mathematics problems set in an Islamic-architecture context that are valid and practical for junior high school students. A development-research design with Tessmer's formative evaluation was used, consisting of self-evaluation, expert review, one-to-one, small-group, and field-test stages. In total, 32 ninth-grade students of SMPN 01 Payaraman took part: 3 in the one-to-one stage, 6 in the small-group stage, and 23 in the field test. Five items were developed and mapped to PISA content categories, contexts, mathematical processes, and proficiency levels 4–6. Data were collected through validation sheets, walkthroughs, interviews, observations, and students' written work. Five expert validators examined content, construct, and language, and every item was judged acceptable after revision. Practicality was examined through students' interviews and observations in the three trial stages, which showed that the items were understandable and workable for students, with the Kufic-calligraphy item being the most demanding. The findings indicate that the items are valid and practical within this limited setting. The effects of the items on students' mathematical reasoning or problem-solving ability were not tested.

Copyright © 2026 (Mar'atus Syakdia, Agustiany Dumeva Putri, Harisman Nizar, & Liana Septy). All Right Reserved

Syakdia, M., Putri, A. D., Nizar, H., Septy, L. (2026). Developing PISA-Like Mathematics Problems Using the Context of Islamic Architecture for Junior High School Students. *Pi: Mathematics Education Journal, 9*(2), 22-41. https://doi.org/10.21067/pmej.v9i2.14680

This work is licensed under a *Creative Commons Attribution 4.0 International License*. Allows readers to read, download, copy, distribute, print, search, or link to the full texts of its articles and allow readers to use them for any other lawful purpose. The journal hold the copyright.

1. Introduction

Mathematics, as one of the subjects taught in schools, certainly has a role in achieving national education goals. One of the functions of mathematics is as a tool to form a mindset, logical ability, critical thinking, and the ability to work together in facing the challenges of future technological developments [1].

PISA is an assessment designed to evaluate the skills and knowledge of 15-year-old students. It was launched by the Organisation for Economic Co-operation and Development (OECD) as a program launched in 2000 and carried out on a three-year cycle [2]. The assessment is conducted internationally to measure three areas of literacy, the assessment

covers three literacy domains-reading, mathematics, and science [3]. PISA aims to encourage countries to learn from each other about education systems so that they can build better and more inclusive school systems efficiently. Every 3 years a 15 years old student from the school is randomly selected around the world. Indonesia participates in measuring literacy in reading, mathematics, science, finance, and global competence ([4],[5]).

The limited ability shown by Indonesian students in answering PISA 2018 questions reflects their weak literacy skills, in both scientific literacy, mathematics literacy, and reading literacy. In fact, PISA 2018 results for Indonesia were generally lower than the average reading, mathematics, and science scores [6]. Indonesia's low PISA scores stem from students' inability to correctly answer the test questions, especially their struggle in translating everyday problems into formal mathematical form, grasping mathematical structures, and assessing the relevance of mathematical outcomes within real-world situations ([7], [8]).

This low performance is often attributed to the limited number of released PISA items available for classroom practice in Indonesia, so that simply drilling students on the small set of existing PISA questions is unlikely, by itself, to close the gap. Repeated exposure to the same released items builds familiarity with question formats rather than the underlying mathematization and contextual-reasoning skills that PISA is designed to measure [8]. Developing new, locally contextualized PISA-like problems therefore serves a complementary purpose: it gives students and teachers a wider and continually renewable pool of practice material that reflects PISA's characteristics (content, context, and level of cognitive demand) while using situations that are familiar to Indonesian students, which in turn supports the transfer of literacy skills to the original PISA items rather than substituting for them.

To justify the label "PISA-like", it is necessary to state which features of PISA items the developed problems are expected to share. PISA defines mathematical literacy as an individual's capacity to formulate, employ, and interpret mathematics in a variety of real-world contexts [3], [4], [20]. Accordingly, PISA mathematics items are described along four dimensions. First, *content categories*: quantity, uncertainty and data, change and relationships, and space and shape. Second, *contexts*: personal, occupational, societal, and scientific. Third, *mathematical processes*: formulating situations mathematically, employing mathematical concepts, facts, and procedures, and interpreting, applying, and evaluating mathematical outcomes. Fourth, *proficiency levels*, which describe increasing cognitive demand from Level 1 to Level 6 [6]. At Level 4, students work effectively with explicit models of complex concrete situations that may involve constraints or assumptions; at Level 5, they develop and work with models of complex situations, identifying constraints and specifying assumptions; and at Level 6, they conceptualize, generalize, and use information obtained from investigating and modeling complex problem situations [6].

In this study, a problem is therefore called PISA-like when (a) it is set in a real-world context, (b) it belongs to one of the four content categories, (c) it requires students to go through the formulate–employ–interpret cycle instead of applying a single routine procedure, and (d) its intended cognitive demand corresponds to proficiency Levels 4–6. These four criteria are used in this article to map each developed item (Table 4) and to judge its content validity.

Several studies have developed PISA-like problems in Indonesian contexts. The contexts used so far come mainly from sports, such as Pencak Silat [9], football and table tennis at the 2018 Asian Games [10], and karate [11]; from the COVID-19 pandemic [12]; and from regional tourism [13]. Other studies focused on a particular competency or content area, such as the mathematical communication of elementary school students [14] and problem solving in the quantity content of junior high school [15]. Taken together, these

studies show that PISA-like problems can be developed from various local contexts and improved through formative evaluation. However, among the studies reviewed here, none draws on Islamic architecture, and the studies that specify their focus concentrate on a single competency [14] or a single content category [15]. To the authors' knowledge, it is also still unclear how the geometric and structural features of religious architecture can be linked systematically to PISA content categories and proficiency levels.

The present study addresses these gaps in three ways. Theoretically, it examines whether one cultural context, Islamic architecture, can supply items for all four PISA content categories by making the link between architectural elements and mathematical concepts explicit (Table 1 and Table 4). Methodologically, it reports the validity and practicality indicators, decision criteria, and evidence for each stage of the formative evaluation, so that claims about item quality can be checked. Practically, it offers teachers a small set of locally meaningful items that can be used alongside the released PISA items rather than instead of them.

Islamic architecture blends human culture with devotion to God and is marked by elaborate geometric patterns, layered forms, ornamentation, and symbolic meaning [16]. These features make it a mathematically rich context. Entrances and arches involve counting and symmetry; domes involve solid geometry and patterns of repeated shapes; wall motifs involve tessellation of repeated shapes and number patterns; and Kufic calligraphy panels on facades involve measurement, area, and proportion. Connecting mathematics with Islamic culture is also in line with the view that mathematics helps Muslims to apply one of the branches of knowledge taught in the Qur'an [17]. Table 1 summarizes which mathematical concepts are represented by the architectural elements selected for this study and how they relate to Grade 9 mathematics.

Table 1. Architectural elements, mathematical concepts, and their link to Grade 9 mathematics

Architectural element (item)	**Mathematical concepts represented**	**Link to Grade 9 mathematics (2013 Curriculum)**
Entrance doors of a mosque (item 1)	Counting rules; permutations and combinations under a constraint (entry door ≠ exit door)	Counting rules and probability
Wall motifs of a mosque (item 2)	Repeating star-and-cross patterns (tessellation of plane figures); counting repeating units; linear number pattern and generalization	Number patterns and generalization; properties of plane figures
Triangular lattice of a dome (item 3)	Number patterns; generalization; counting of similar figures of different sizes	Patterns and generalization; quadratic relationships
Kufic calligraphy panels on a facade (item 4)	Area of rectangles and composite regions; unit area; ratio	Area of plane figures; equations; proportional reasoning

Architectural element (item)	Mathematical concepts represented	Link to Grade 9 mathematics (2013 Curriculum)
Marble flooring of a mosque (item 5)	Area of a square and of a floor; unit conversion (cm^2 and m^2); cost computation with unit price and quantity	Area of plane figures, arithmetic with decimals, and rates (prerequisite topics applied in a multi-step costing situation)

Items 2 and 5 in Table 1 follow the final item texts reported in the study. The links to Grade 9 are stated at the level of the topics identified during the design stage (solid figures, counting rules and probability, and number patterns that students had already studied), not as individual basic-competency numbers.

These concepts suit ninth-grade students for three reasons. First, Grade 9 students are about 14–15 years old, which is the age group assessed by PISA. Second, the concepts belong to, or build directly on, the topics taught under the 2013 Curriculum that the research school implements. Third, the situations require knowledge that students already have; the challenge lies in modeling the situation and choosing a strategy, which is the kind of demand expected at PISA Levels 4–6, not in advanced procedures.

To the authors' knowledge, Islamic architecture is rarely used as a source of mathematics tasks in school assessment, so students seldom see the mathematics embedded in the mosques and Islamic buildings around them. For students in a predominantly Muslim country such as Indonesia, it may offer familiar and meaningful situations, although whether this improves learning has to be investigated separately. Developing problems that are both valid and practical is a necessary first step, so that teachers and students can use them without being hindered by unclear wording, figures, or information.

Therefore, this study aims to develop PISA-like mathematics problems in the context of Islamic architecture for ninth-grade students and addresses two questions: (1) Are the developed problems valid in terms of content, construct, and language? (2) Are the developed problems practical for junior high school students? Following Nieveen [21], validity refers to whether the product is based on sound knowledge (content validity) and whether its components are consistently linked (construct validity), and practicality refers to whether users consider the product usable and easy to use in the intended setting. The study is limited to validity and practicality. Whether the problems improve students' mathematical reasoning or problem-solving ability is outside its scope.

2. Methods

Research design

This study used design research with a development-study approach, carried out in two main stages: preliminary and formative evaluation [18]. The formative-evaluation stage follows a flow adapted from [19], consisting of self-evaluation followed by a prototyping cycle comprising expert review, one-to-one evaluation, small-group evaluation, and field testing.

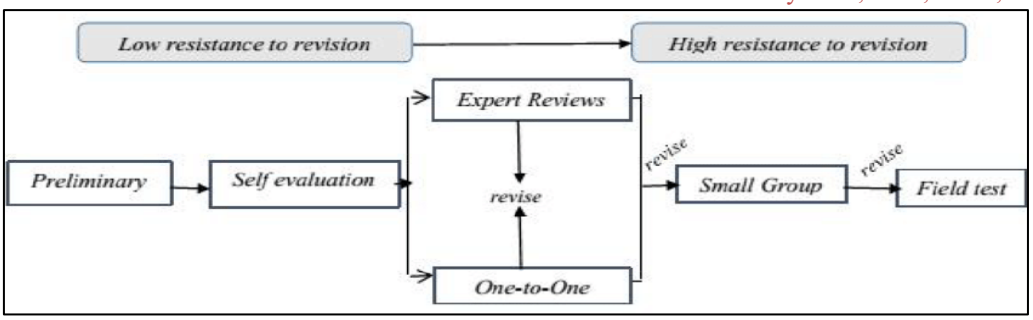


Figure 1. Formative evaluation design flow (Tessmer)

Setting and participants

The study was conducted at SMP Negeri 1 Payaraman during the even semester of the 2021/2022 academic year. The school was selected purposively because it implements the 2013 Curriculum, its ninth-grade students were accessible to the researchers, and no comparable development study had previously been conducted there. Because a development study aims at product quality rather than generalization, a single-site sample is acceptable, but it limits the scope of the findings (see the limitations in the Result and Discussion section). Permission to conduct the study was obtained from the school: the researchers contacted the principal and the mathematics teacher, and the school agreed to serve as the research site. The study documents do not include written informed-consent forms from students' parents or guardians for participation or for the use of photographs; this is acknowledged as a limitation (see Limitations), and students are reported only by code or number.

A total of 32 ninth-grade students took part, and each student was involved in only one stage: 3 students in the one-to-one stage, 6 in the small-group stage, and 23 in the field test (3 + 6 + 23 = 32; Table 2). According to the participant lists in the research records, the students in the three stages were different individuals (the small-group and field-test students were not research subjects of the other stages). The students in the one-to-one and small-group stages were selected by the researchers in consultation with the school principal and the mathematics teacher, with the requirement that each stage include students of high, medium, and low mathematical ability. High, medium, and low mathematical ability were determined by a judgment-based classification, because no numerical cut-off scores are recorded in the study documents (see Limitations). The 23 field-test students were one intact ninth-grade class that was not among the subjects of the other stages. Five experts took part as validators in the expert-review stage.

Table 2. Participants in each stage of the formative evaluation

Stage	Prototype tested	Participants	Composition	Purpose and data
Self-evaluation	Initial prototype	Researcher and team	Mr. HM and Mrs. ADP	Check content, construct, and language; result: Prototype 1
Expert review	Prototype 1	5 validators	3 mathematics lecturers, 1 Islamic religious education lecturer, 1 mathematics teacher	Validity (content, construct, language); validation sheets and comments

Stage	Prototype tested	Participants	Composition	Purpose and data
One-to-one	Prototype 1	3 students	1 high, 1 medium, 1 low mathematical ability	Difficulties and impressions; interviews; result (together with expert review): Prototype 2
Small group	Prototype 2	6 students (two groups of 3)	Each group mixed high, medium, and low ability	Practicality; observation, interviews, group written answers
Field test	Prototype 3	23 students	One intact class; 6 students interviewed	Practicality in a classroom setting; observation, interviews, individual written answers

No revision was made after the small-group stage; Prototype 3 is therefore identical in content to Prototype 2, and the label is kept only to follow the naming of the prototype that enters the field test.

Instruments and data collection

Data were gathered using documentation, walkthroughs, validation sheets, interviews, observations, and students' written work. The validation sheet covered three aspects, namely content, construct, and language, and was completed by five validators. The interview guide and observation sheet covered the practicality indicators listed in Table 3. The items were written in Indonesian; the English versions shown in this article are translations. The indicators in Table 3 were derived from the PISA characteristics described in the Introduction and from the quality criteria of Nieveen [21].

Table 3. Indicators, data sources, and decision criteria for validity and practicality

Aspect	Indicators	Data source	Decision criterion
Validity: content	(1) The item matches the intended PISA content category. (2) The context is a real-world situation and the information about the architecture is accurate. (3) The mathematical concept is appropriate for Grade 9 and the 2013 Curriculum. (4) The intended proficiency level (Levels 4–6) is appropriate.	Validation sheet and written comments of five validators	Accepted by all five validators on this aspect, with or without revision, and every suggestion followed up

Aspect	Indicators	Data source	Decision criterion
Validity: construct	(1) The stem gives sufficient information and the question command is clear. (2) Figures and data are consistent with the question. (3) The task requires the intended process (formulate, employ, interpret). (4) The answer key and scoring guide are consistent with the task.	Same as above	Same as above
Validity: language	(1) Sentences are communicative and unambiguous. (2) Spelling and punctuation follow PUEBI. (3) Terms and vocabulary suit Grade 9 students.	Same as above	Same as above
Practicality	(1) Students understand the instruction, context, and question. (2) Figures and information are legible and sufficient. (3) The items can be worked on within the available time. (4) Students are willing and interested to work on the items. (5) Difficulties encountered by students.	Interviews and observation in the one-to-one, small-group, and field-test stages; written work as supporting information only	Students report that the items are understandable and workable, and no difficulty requiring a substantial change remains after revision

Table 3 elaborates the three validation aspects recorded in the validation sheet (content: PISA content; construct: theory and criteria of PISA items; language: PUEBI) and the practicality aspects examined through students' interviews. The indicators were elaborated by the authors from these aspects and the sources cited above. Interviews and students' written work are the main documented practicality data; observation during the sessions was supplementary, and no separate observation sheet or test time limit is documented in the study records.

Data analysis and decision criteria

Validity data were analyzed descriptively and qualitatively. The validators' comments were recorded per item, classified into content-, construct-, and language-related feedback, counted, and followed up with revisions (Table 6). An item was considered valid when all five validators accepted it on the three aspects, with or without revision, and every suggestion was followed up by a revision or a justified decision not to revise. No quantitative rating index was computed in this study.

Practicality data were analyzed descriptively by summarizing the interview and observation records of each stage according to the indicators in Table 3 (Table 7). Students' written answers were scored using a total score per student, with a scoring rubric prepared in the design stage together with the item grid and question cards (item-level maximum scores are not reported in the study documents) and are reported as total scores per student

on a 0–100 scale. These scores and the students' solution strategies are used only as supporting information on whether the items can be solved by students at this level. They are not a measure of practicality, which is judged by the indicators in Table 3, and they are not a measure of students' mathematical ability.

3. Result and Discussion

This study applied the Design Research model of the development type, following a research procedure adapted from Tessmer (1993) that covers the Preliminary stage and the Formative Evaluation stage (Self-Evaluation, Expert Review, One-to-One, Small Group, and Field Test). Findings from each stage of this Tessmer-based procedure [19] are presented below.

1. Preliminary Study

   In this phase, the researcher prepared and designed valid PISA-like mathematics problems set in an Islamic-architecture context, moving through preparation and design, self-evaluation, and expert review.

   a. Preparation phase

      During preparation, the researcher determined the research subjects and setting, namely SMPN 01 Payaraman, and identified the school's curriculum along with student characteristics relevant to developing the PISA-type items. First, curriculum identification: a short interview with RD, the school's curriculum coordinator, confirmed that SMPN 01 Payaraman implements the 2013 curriculum. Second, identification of student characteristics: this step ensured that the item design matched the traits of the students who would use the PISA-like problems set in the Islamic-architecture context.

   b. Design Stage

      Designing the PISA-like problems in the Islamic-architecture context took roughly one month. It began with gathering references on PISA item formats, content categories, contexts, and proficiency levels, together with material on Islamic architecture. The design then proceeded in three steps: (1) selecting architectural elements that contain mathematical structure suitable for Grade 9 (Table 1); (2) assigning each element to a PISA content category, context, mathematical process, and target proficiency level; and (3) writing the item stem, figure, question command, and scoring guide. This process produced five PISA-type items. Table 4 maps each item to the PISA characteristics defined in the Introduction.

Table 4. Mapping of the developed items to PISA content category, context, processes, and proficiency level

No	Architectural context	Content category	PISA context	Main processes	Level	PISA characteristics represented
1	Nine entrance doors of the Wali Songo Islamic Boarding School mosque; number of ways in which three people can enter and exit through different doors	Uncertainty and data (counting)	Societal	Formulate (model the entry–exit constraint); employ (counting rules)	4	Real-world situation; requires an explicit model with a constraint (entry door ≠ exit door) that is extended from one person to three persons; cannot be solved by a single routine step.
2	Wall motifs of Al-Akbar Mosque, Surabaya (star-and-cross pattern); number of large motifs needed for 50 small motifs, with written reasoning	Space and shape (pattern; properties of shapes in a tessellation)	Societal	Formulate; employ; interpret (justify the answer)	5	Real motif linked to a figure with marked large (green) and small (yellow) motifs; students must identify the repeating unit, build a general relation (the large motifs increase by 3 when the small motifs increase by 2), and justify the result; more than one strategy (table, pattern, or formula).
3	Hemispherical triangular pattern of the dome of the mosque in Ljubljana, Slovenia; number of triangles in a triangular pattern	Change and relationships (pattern, generalization)	Societal	Formulate; employ; reasoning (generalizing a pattern)	6	Non-standard situation without a given formula; students can draw and count, or build smaller cases to find a general pattern; calls for generalization.

No	Architectural context	Content category	PISA context	Main processes	Level	PISA characteristics represented
4	Kufic calligraphy panels on a mosque facade (area 272 m², given length 16 m); area not covered by calligraphy	Space and shape (area, measurement)	Occupational	Formulate; employ; interpret	5	Multi-step problem: a missing dimension must be derived, the number of panels counted from the figure, and the result interpreted in the situation.
5	Marble flooring of the Sheikh Zayed Grand Mosque, UAE (floor area 22,412 m², 60 × 60 cm marble tiles, prices in dirham); number of boxes of marble and installation cost	Quantity	Occupational	Formulate; employ; interpret	4	Realistic costing situation (amounts in dirham); students must convert units (cm² to m²), combine several operations in sequence, and interpret the result as a cost; more than one solution route.

The contexts, processes, and levels in Table 4 were assigned by the authors by comparing the final item texts and the students' solution evidence with the descriptors of PISA content categories, contexts, and proficiency levels (Chapter II of the study documents). They were not rated item by item by the validators, who judged content validity as a whole, and no item-difficulty data are available to confirm the levels empirically. Item 5 was replaced after expert review because validator A1 noted that one content element was missing, which we read as the Quantity category; the marble-flooring item fills it.

As Table 4 shows, the five items cover all four content categories (quantity; uncertainty and data; change and relationships; space and shape) and proficiency Levels 4 to 6. All items are set in real situations related to mosques, require students to interpret a figure or data taken from the architecture, cannot be solved by one routine procedure alone, and allow more than one solution strategy. These characteristics were the basis of the content-validity judgment in the expert-review stage.

2. Prototyping stage with formative evaluation flow

The prototyping stage comprised five phases, self-evaluation, expert review, one-to-one, small group, and field test-each described in turn below:

a. Self evaluation

During self-evaluation, the initial prototype was reviewed jointly by the researcher and the research team (Mr. HM and Mrs. ADP) to check its content,

construct, and language. Feedback from this review served as the basis for revising the items; the revised version is referred to as prototype 1.

b. Expert Reviews

Prototype 1 was then assessed by experts for content, context, and language. Tessmer [19] this validation process can be carried out through three approaches, namely mail review (via email), panel review, and face-to-face review.

Five experts validated Prototype 1: three mathematics lecturers, one Islamic religious education lecturer, and one mathematics teacher at SMPN 01 Payaraman (coded A1–A5 in Table 5). They were chosen to cover the mathematical and PISA aspects, the accuracy of the Islamic-architecture content, and classroom practice. Validation used a combination of the review methods described by Tessmer [19]: face-to-face review (A1, A4, and A5), WhatsApp review (A1 and A3), and mail review (A2); no panel session is recorded. The validation took place during the study period (7 February to March 2022). Each validator assessed the items with the validation sheet (content, construct, and language; Table 3) and wrote comments and suggestions, which are presented in Table 5.

Table 5. Validators' comments and suggestions on Prototype 1 (A1–A5 = validators)

Test Items		Comments and Suggestions
1	A1	a. Changing questions from 1 person entering to 3 people entering. b. Change the picture by removing the number from the door
	A2	a. Changing the Islamic boarding school into the Wali Songgo Islamic boarding school mosque. b. Changing image source
	A3	a. Fix info regarding 9 doors before getting into the question
	A4	a. Problem is good
	A5	a. Question is good enough
2	A1	a. Add pictures of mosque wall motifs directly b. Fix writing
	A2	a. Focus on the pattern whether the design is on the wall or on the floor. b. Explain that the pattern is from a sketch of a picture
	A3	a. Question is good enough
	A4	a. Fix writing b. Give information about the motifs on the walls
	A5	a. Fix writing
3	A1	a. Show the pattern taken from the dome with an arrow
	A2	a. Add the word half ball b. Eliminate mosque information because it has nothing to do with the question
	A3	a. Writing the preposition in must be separated b. Name of country/city in capital letters
	A4	a. Give information about the mosque from where b. Put a fanah mark on the pattern in question
	A5	a. Correct the word inside
4	A1	a. Fix what is meant by long b. Show kufi calligraphy name

	A2	a. What is meant by calligraphy love focus or love arrows? b. If you focus on the front of the mosque, add information in the question
	A3	a. Writing the preposition in must be separated b. Name of country/city in capital letters
	A4	a. Tell/give the arrow which is kufi calligraphy
	A5	a. Show me which calligraphy it is
5	A1	a. Change the question because there is one content that doesn't exist
	A2	a. The context of the question can be replaced with printing costs etc. b. Delete the word if the data is correct
	A3	a. Question is good enough
	A4	a. Fix writing
	A5	a. Pretty good question

To show how the expert-review evidence supports validity, the 32 comments in Table 5 were classified by the authors into content-, construct-, and language-related feedback (Table 6). No validator recommended rejecting an item; every comment asked for an improvement, or stated that the item was good enough. Eight comments were content-related (for example, changing item 1 from one person to three persons, replacing the Islamic boarding school with its mosque, and stating the origin of the mosque in item 3). Fourteen were construct-related, the largest group, and mostly asked for clearer figures and information (for example, removing the door numbers from the figure, adding arrows to show the dome pattern and the Kufic calligraphy, and clarifying what is meant by "length" in item 4). Ten were language-related (for example, writing prepositions separately and capitalizing names of countries and cities).

Table 6. Summary of expert-review feedback and follow-up per item

Item	Validators suggesting revision (of 5)	Content	Construct	Language	Main follow-up revisions	Decision
1	3 (A1, A2, A3)	2	3	0	Number of persons changed from one to three; setting changed to the mosque of the Wali Songo Islamic Boarding School; door numbers removed from the figure; image source changed; information on the nine doors clarified (Figure 2).	Valid after revision
2	4 (A1, A2, A4, A5)	1	3	3	Pictures of the mosque wall motifs added; focus on the pattern (wall or floor) and its source (a sketch from a picture) stated; information on the motifs added; wording corrected.	Valid after revision

Item	Validators suggesting revision (of 5)	Content	Construct	Language	Main follow-up revisions	Decision
3	5 (A1–A5)	2	3	3	Arrow added to show the pattern taken from the dome; the word "hemispherical" added; information about the mosque reviewed (one validator suggested removing information unrelated to the question, another suggested stating the origin); prepositions and capitalization corrected (Figure 3).	Valid after revision
4	5 (A1–A5)	1	5	2	Meaning of "length" clarified; Kufic calligraphy named and pointed out with arrows; focus on the facade stated; prepositions and capitalization corrected.	Valid after revision
5	3 (A1, A2, A4)	2	0	2	Question changed because one of its content elements did not exist; context may be replaced with painting costs; unnecessary words deleted; wording corrected.	Valid after revision
Total	–	8	14	10	32 comments in total	5 of 5 items accepted

Applying the criterion in Table 3 (all five validators accepted each item, and every suggestion was followed up), the five items were categorized as valid in content, construct, and language, and the suggestions were used to produce Prototype 1 as revised. The validity judgment rests on qualitative expert judgment; no rating scores or agreement index were computed. The validation sheets recorded comments and suggestions rather than numerical ratings: validity was determined descriptively from the validators' notes and the criterion that all experts judged the instrument at least feasible with revision, so no mean ratings are reported.

c. One-To-One

In the one-to-one stage, three students, representing high, medium, and low mathematical ability, worked on the items and were then interviewed about the difficulties they encountered, their understanding of the wording and figures, and their overall impressions. The one-to-one stage used Prototype 1, in parallel with the expert review. Two of the three students were initially confused about the number of doors in item 1, because the text stated nine doors while the figure clearly showed only six (the other three were at the front), and one student found the items challenging; all three described the figures as clear and the information as sufficient, and the green/yellow labels in item 2 as

helpful. For item 1 one student answered correctly and two used sound strategies but obtained inaccurate results; for items 2–4 all used different reasonable strategies but obtained inaccurate results; for item 5 (the earlier version) all answered correctly. Together with the validators' comments, this led to removing the door numbers so that students count the nine doors themselves and to changing item 1 from one person to three persons. Their input, together with the validators' comments, guided further revision of the prototype's practicality, as illustrated in the revised items below.

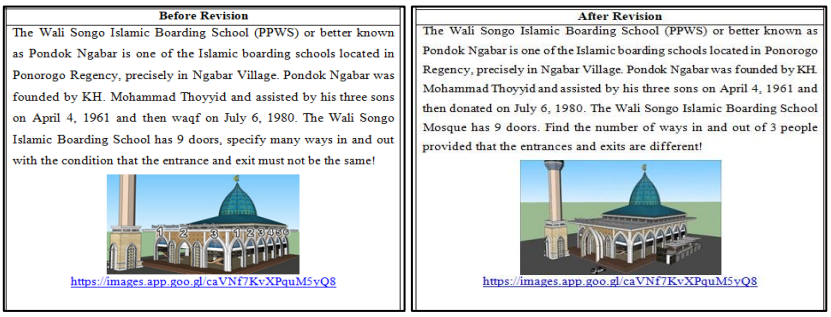


Figure 2. Item 1 before and after revision

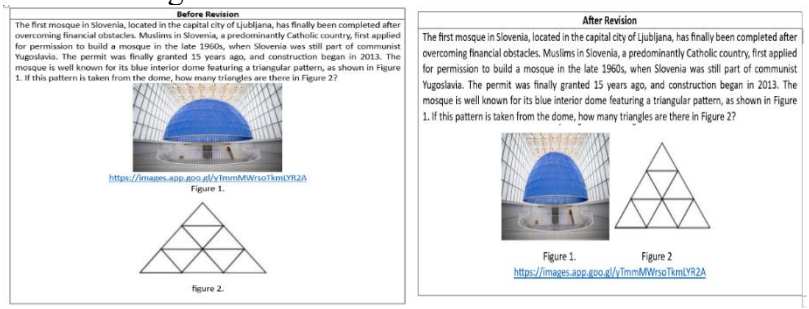


Figure 3. Item 3 before and after revision

Having completed both the Expert Review and One-to-One stages, the researcher used the validation outcomes from these two phases to produce prototype 2, which was carried forward into the Small Group stage

d. Small Group

Prototype 2 was tested with six SMPN 01 Payaraman students split into two groups of three, each group containing a mix of high, medium, and low ability levels; group members worked on the items together and discussed their answers.

Figure 4. Small-Group Trial of the PISA-Like Mathematics Problems That Have Been Developed

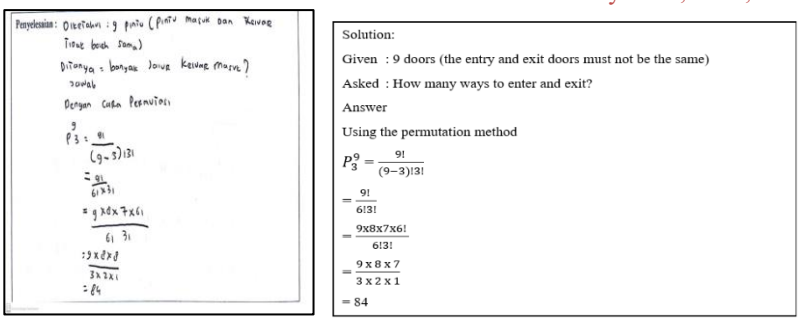


Figure 5. The results of the students' answers at the small group stage

Figure 5 shows the written solution of one group to item 1. The group wrote the permutation notation P(9,3) but applied the formula 9!/((9−3)!·3!), which gives 84. This formula is the number of combinations, C(9,3) = 84, whereas P(9,3) would be 504, and the answer key for the three-person version of item 1 is not reproduced in the study documents, so 84 cannot be confirmed as the intended answer. The mismatch indicates that the wording of item 1 allowed more than one reading (choosing which three doors are used versus counting ordered entry–exit routes for three people). The practicality of item 1 is therefore regarded as only partly established: students understood the context and the figure, but the question command should state whether the order of the doors matters and whether the three people use different doors before the item is used for scoring. On item 3, the two groups used different strategies: Group 1 drew the figure and counted the triangles directly, whereas Group 2 built smaller cases first to derive a general pattern. On items 4 and 5, both groups used the information in the problem to plan their solution steps. These observations come from only two groups and describe solution strategies; they are not evidence that the items improved students' reasoning.

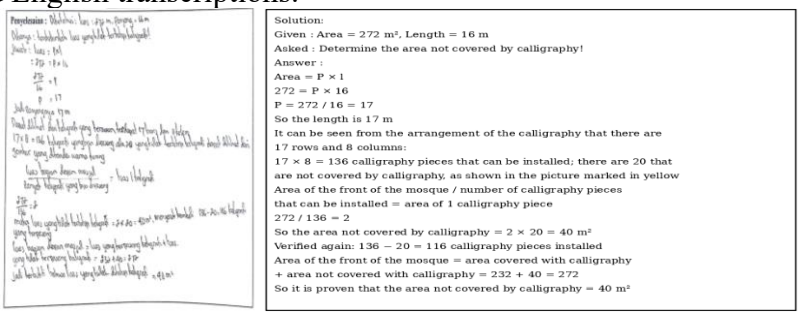
Interview and observation data on practicality showed that students were highly engaged with the items, partly because several questions could be solved through examples or patterns rather than formulas alone. Students also reported that the wording of the items was clear and easy to follow. The exception was item 4 (Kufic-calligraphy area), which students found more demanding because a length had to be determined first before the calligraphy coverage on the facade of the mosque could be estimated. Students' written answers were scored using a total score per student, with a rubric prepared in the design stage (item-level maximum scores are not reported in the study documents), and the two groups obtained an average of 90% of the maximum score (the 90% average is the figure recorded in the study documents for the two groups together; percentages per group were not recorded separately). This score is used only as supporting information that the items can be solved by students at this level; it is not a practicality index, which is judged by the indicators in Table 3. Because the interviews and observations showed no difficulty that required a substantial change, no further revision was made at this stage (practicality was judged from the students' interviews, that is, whether students found the items easy to use; no numerical cut-off was set beforehand, and the decision rested on the 90% average together with the absence of comments that required a change to the items).

e. Field test

Prototype 3 was field-tested with 23 students of SMPN 01 Payaraman, who worked individually on the five items in a classroom setting (Figure 6) (the time allotted and the supervisor of the session are not recorded in the study documents). The researchers observed the session and interviewed six students afterwards.

Figure 6. Field-test session at SMPN 01 Payaraman

Figure 7 shows one student's written solutions to item 4, as handwritten responses and English transcriptions.



Solution:
Given : Area = 272 m², Length = 16 m
Asked : Determine the area not covered by calligraphy!
Answer :
Area = P × l
272 = P × 16
P = 272 / 16 = 17
So the length is 17 m
It can be seen from the arrangement of the calligraphy that there are
17 rows and 8 columns:
17 × 8 = 136 calligraphy pieces that can be installed; there are 20 that
are not covered by calligraphy, as shown in the picture marked in yellow
Area of the front of the mosque / number of calligraphy pieces
that can be installed = area of 1 calligraphy piece
272 / 136 = 2
So the area not covered by calligraphy = 2 × 20 = 40 m²
Verified again: 136 − 20 = 116 calligraphy pieces installed
Area of the front of the mosque = area covered with calligraphy
+ area not covered with calligraphy = 232 + 40 = 272
So it is proven that the area not covered by calligraphy = 40 m²

Figure 7. The results of students' answers at the field test stage (item 4)

In item 4, this student derived the missing side from the area and the given side of 16 m (272 ÷ 16 = 17), determined the number of calligraphy pieces from the 17 rows and 8 columns (136), used the area per piece (272 ÷ 136 = 2 m²) to compute the uncovered area (20 × 2 = 40 m²), and verified the result by recomputing the total area (232 + 40 = 272). This response illustrates one possible solution path for one student; it cannot represent all 23 students. The solution strategies reported for the whole group are summarized in Table 8. In the handwritten solution the student labelled the given 16 m as “length” and the derived 17 m as “the length” as well, and omitted the square in the unit of the area; the item text states 272 m² and gives 16 m as the width of the facade. This is consistent with the validators’ comment that the meaning of “length” in item 4 needed to be clarified (Table 5). The English transcription in Figure 7 follows the handwriting and adds m² to the area. The item-1 response originally shown in Figure 7 (9 × 8 = 72) came from the one-person version used in the one-to-one stage, so it was removed; solutions to the three-person version are summarized in Table 8.

Table 7. Summary of practicality evidence per trial stage

Stage	n	Data	Findings on practicality	Decision
One-to-one	3	Interviews; observation of the students working on the items	Two of three students were initially confused by the number of doors in item 1 (nine in the text, six clearly visible in the figure); one found the items challenging; figures and information were judged clear.	Revised; Prototype 2 produced

Stage	n	Data	Findings on practicality	Decision
Small group	6 (two groups)	Observation; interviews; group written answers	Students were engaged; items could be solved through examples or patterns; wording was clear; item 4 was more demanding. Average score 90% of the maximum score (supporting information only).	No further revision
Field test	23	Observation; interviews; individual written answers (Table 8, Figure 7)	Information, figures, and language were clear and sufficient; students found the items interesting and requiring critical thinking and were willing to work on them; the items added to their knowledge of PISA-type problems, which they had not known before. No comment required a substantial change; written work showed a unit-conversion slip in item 5.	Practical; no further revision

Table 7 summarizes the practicality evidence across the three trial stages. Practicality is judged from the students' interviews and the observations against the indicators in Table 3, whereas the scores and solution strategies in Table 8 serve only as supporting information on whether the items could be solved.

Table 8. Solution strategies reported at the field test per item (n = 23)

Item	Typical solution strategies and difficulties
1	Counted the nine doors (three at the front, six at the sides) and, for three people, listed entry–exit door pairs in a table or used the permutation notation with the formula $\frac{9!}{((9-3)!\cdot 3!)}$; both routes gave 84. Initial difficulty: miscounting the doors. The permutation/combination mismatch shows that the question command needs refinement.
2	Counted the large and small motifs in the figure and tabulated them: the large motifs increase by 3 each time the small motifs increase by 2 (6 large with 2 small, 9 with 4, …), giving the relation 3 + 3n for 2n small motifs, so 50 small motifs need 78 large motifs.
3	Redrew the triangular figure and counted all triangles, obtaining 13; others used a counting pattern based on the number of triangles per row. Main demand: counting triangles of different sizes.
4	Found the missing side (272 ÷ 16 = 17 m), counted 17 × 8 = 136 calligraphy pieces with 20 not covered, computed the area per piece (2 m²) and the uncovered area (40 m², with 232 m² covered). The most demanding item in the small group; the term “length” and the order of the steps caused difficulty.

Item	Typical solution strategies and difficulties
5	Divided the floor area (22,412 m²) by the area of one box of tiles (taken as four 60 × 60 cm tiles) and multiplied by the unit prices. The recorded number of boxes (155.64) suggests a decimal-place slip when converting the box area from cm² to m²; the number of tiles per box is not stated in the item text reported in the study and should be added.

Item-level scores were not recorded in the study documents; only one total score per student (0–100 scale) is available. The 23 total scores ranged from 65 to 90 with a mean of 79.9 (sum 1,838, recomputed from the individual scores): 17 students scored 80 or higher and 6 scored between 65 and 75 (65: one student; 70: four students; 75: one student). These scores are supporting information only (see Data analysis), and the strategies in Table 8 come from the researchers' record of the students' written answers.

This study developed five PISA-like problems in an Islamic-architecture context and examined their validity and practicality through Tessmer's formative evaluation. Validity evidence came from five experts: no item was rejected, and all items were accepted after revision on the content-, construct-, and language-related comments summarized in Table 6. Practicality evidence came from interviews and observations in three trial stages involving 32 students (Table 7). This evidence supports the claim that the items are valid and practical in this setting. It does not allow claims about students' learning outcomes.

The expert comments were mostly about construct (14 of 32), followed by language (10) and content (8). Most of them asked for clearer figures, labels, and information, and not for a different mathematical task, with the changes to the number of persons in item 1 and to item 5 as the main exceptions. This suggests that the main challenge in building items on real architectural contexts is presenting authentic information and figures clearly enough for students to formulate the problem, which deserves attention in future development of context-based items.

The use of expert validation and iterative revision is consistent with earlier PISA-like development studies that used other contexts, such as the Asian Games [10], [11] and COVID-19 [12]. What the present study adds is an explicit mapping showing that one cultural context can supply items for all four PISA content categories (Table 4), and a report of the criteria and evidence used to judge validity and practicality (Tables 3, 6, and 7).

The students' responses suggest that different solution strategies were possible. In item 3, one group drew the figure and counted directly while the other built smaller cases to find a pattern, and in item 4 the sample response in Figure 7 shows a multi-step solution with verification. These observations are consistent with the intention that the items elicit reasoning and not only the recall of a formula, and they suggest that Islamic architecture can provide meaningful contexts for such items. However, they are based on two groups and one selected response, not on a systematic analysis of all students' solution processes, and the study did not measure reasoning, problem solving, or critical thinking. These possibilities should therefore be tested in a separate effectiveness study.

The difficulty students showed with item 4, which involves Kufic-calligraphy motifs, suggests that items with non-standard shapes may need more carefully designed information, such as a clearly marked figure and a clear definition of the given length, as the validators also pointed out.

Limitations. The study has several limitations. (1) It was conducted in one school with 32 students, so the findings cannot be generalized. (2) Validity rests on the qualitative judgment of five validators, without a quantitative index or an agreement coefficient. (3) Practicality was judged from interviews and observations, and the field-test evidence is

limited to the data reported in Tables 7 and 8. (4) No psychometric analysis, such as item difficulty, discrimination, or reliability, was carried out. (5) The items were written in Indonesian and are presented here in English translation. (6) The effect of the items on students' mathematical reasoning, problem solving, or PISA performance was not tested. (7) Participant selection and ability grouping were judgment-based without documented cut-off scores; written consent from parents or guardians, the time allotted, the item-level scores, and the answer key for item 1 are not documented, and the wording of item 1 needs refinement. Future research should test effectiveness with larger and more varied samples, report psychometric properties, and examine whether other cultural contexts can be linked to PISA content categories in the same way.

4. Conclusion

This study produced five PISA-like mathematics problems in the context of Islamic architecture for ninth-grade students and examined their quality through Tessmer's formative evaluation. The items are valid, because five expert validators accepted all of them in terms of content, construct, and language after revision, and they are practical, because interviews and observations in the one-to-one, small-group, and field-test stages showed that students could understand and work on them, with the Kufic-calligraphy item being the most demanding. These conclusions are limited to the validity and practicality of the items in one school with 32 students. The study did not test whether the items improve students' mathematical reasoning, problem-solving ability, or PISA performance, and such effects should not be inferred from these findings. Further research is needed to test their effectiveness and psychometric quality with larger and more varied samples.

Thank You Note

The researchers gratefully acknowledge Mr. HM and Mrs. ADP for serving on the research team and reviewing the initial prototype during self-evaluation; RD, the curriculum coordinator at SMPN 01 Payaraman, for assistance with curriculum identification; the five expert validators, three mathematics-education lecturers, one Islamic-education lecturer, and one mathematics teacher at SMPN 01 Payaraman, for their input during the expert-review stage; and the students of SMPN 01 Payaraman who took part in the one-to-one, small-group, and field-test trials. None of these individuals are authors of this article. The researchers welcome constructive feedback for future work and hope this study offers useful knowledge for both researchers and readers

References

[1] E. Suherman, *Strategi Pembelajaran Matematika Kontemporer*. Bandung: JICA - Universitas Pendidikan Indonesia (UPI), 2001.

[2] G. Shiel, R. Perkins, S. Close, and E. Oldham, “PISA Mathematics: A Teacher’s Guide,” Government Publications, Dublin, 2007.

[3] OECD, “PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving (Revised edition),” OECD Publishing, Paris, 2017. doi: 10.1787/9789264281820-en.

[4] OECD, “PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy,” OECD Publishing, Paris, 2013. doi: 10.1787/9789264190511-en.

[5] K. Stacey, “The International Assessment of Mathematical Literacy: PISA 2012 Framework and Items,” 2012.

[6] OECD, “PISA 2018 Results (Volume I): What Students Know and Can Do,” OECD Publishing, Paris, 2019. doi: 10.1787/5f07c754-en.

[7] S. I. Edo, Y. Hartono, and R. I. I. Putri, "Investigating secondary school students' difficulties in modelling problems PISA-model level 5 and 6," *J. Math. Educ.*, vol. 4, no. 1, pp. 41–58, 2013.
[8] A. Jupri and P. Drijvers, "Difficulties in initial algebra learning in Indonesia," *Math. Educ. Res. J.*, vol. 26, no. 4, pp. 683–710, 2016.
[9] H. Nizar, "Konteks Pencak Silat pada Soal Matematika Tipe PISA," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 3, pp. 2826–2835, 2021, [Online]. Available: https://j-cup.org/index.php/cendekia/article/view/980
[10] H. Nizar, R. I. I. Putri, and Zulkardi, "Developing PISA-like mathematics problem using the 2018 Asian Games football and table tennis context," *J. Math. Educ.*, vol. 9, no. 2, pp. 183–194, 2018, doi: 10.22342/jme.9.2.5246.183-194.
[11] H. Nizar, R. I. I. Putri, and Zulkardi, "PISA-like mathematics problem with karate context in Asian Games," *J. Phys. Conf. Ser.*, vol. 1088, p. 12063, 2018, doi: 10.1088/1742-6596/1088/1/012063.
[12] S. Andita, L. Septy, H. Nizar, and R. S. Nery, "Validitas Soal Matematika Tipe PISA Konteks Covid-19," *JEMST (Journal Educ. Math. Sci. Technol.*, vol. 3, no. 2, pp. 89–100, 2020, doi: 10.30631/JEMST.V3I2.38.
[13] M. N. Dasaprawira, "Pengembangan Soal Matematika Tipe PISA Menggunakan Konteks Wisata Banyumas dan Cilacap," *Prisma*, vol. 10, no. 1, pp. 88–96, 2021, doi: 10.35194/jp.v10i1.1128.
[14] D. Mardhiyanti, "Pengembangan Soal Matematika Model PISA Untuk Mengukur Kemampuan Komunikasi Matematis Siswa Sekolah Dasar," *J. Pendidik. Mat.*, vol. 5, no. 1, pp. 1–14, 2011.
[15] B. Febriani, "Pengembangan Soal Matematika Model PISA Pada Konten Quantity Untuk Mengukur Kemampuan Pemecahan Masalah Matematika Siswa Sekolah Menengah Pertama," *J. Gantang*, vol. 2, no. 1, pp. 63–77, 2017.
[16] A. Fikriarini, "Arsitektur Islam: Seni Ruang dalam Peradaban Islam," *el-Harakah*, vol. 12, no. 3, pp. 194–206, 2010, doi: 10.18860/el.v0i0.452.
[17] S. Maarif, "Integrasi Matematika dan Islam Dalam Pembelajaran Matematika," *J. Ilm. Progr. Stud. Mat. STKIP Siliwangi*, vol. 4, no. 1, pp. 223–236, 2015.
[18] Zulkardi, "Developing a Learning Environment on Realistic Mathematics Education for Indonesian Student Teachers," University of Twente, Enschede, 2002.
[19] M. Tessmer, *Planning and Conducting Formative Evaluations: Improving the Quality of Education and Training*. London: Kogan Page, 1998. [Online]. Available: https://www.routledge.com/Planning-and-Conducting-Formative-Evaluations/TessmerMartin/p/book/9780749408015
[20] OECD, *PISA 2022 Assessment and Analytical Framework*. Paris: OECD Publishing, 2023. doi: 10.1787/dfe0bf9c-en.
[21] N. Nieveen, "Prototyping to reach product quality," in *Design Approaches and Tools in Education and Training*, J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, and T. Plomp, Eds. Dordrecht: Kluwer Academic Publishers, 1999, pp. 125–135.