# Development of Contextualised Numeracy Test Items for the Minimum Competency Assessment for Senior High School Students in Bengkulu

**Pratiwi Disha Stanggo**
Universitas Bengkulu
pratiwi_disha@unib.ac.id
**Edi Susanto**
Universitas Bengkulu
edisusanto@unib.ac.id
**Corresponding Email**: edisusanto@unib.ac.id

## Abstract

The competency outcomes of Indonesian secondary school students, as measured by the Minimum Competency Assessment (AKM) over the past three years, remain suboptimal, particularly in mathematical proficiency. National data from 2023 indicate that less than 50% of students have achieved mastery in mathematical literacy, highlighting the need for improvements in classroom teaching and learning. A key factor in enhancing AKM implementation is the provision of practice test items, which necessitates that teachers develop the ability to design AKM-style questions to familiarise students with problem-solving strategies. This study aims to develop valid and reliable AKM-Numeracy reasoning test items, contextualised to Bengkulu for senior high school students, and to assess their impact on learning outcomes. Employing a development research approach, the study utilised the two-stage Tessmer model, comprising a preliminary stage and a formative evaluation stage that included self-evaluation, prototyping (expert reviews), and a field test. The large-scale trial involved students from three Senior High Schools or *Madrasah Aliyah* in Bengkulu City. Findings indicate that the contextualized AKM-Numeracy test items meet validity criteria based on expert evaluations and have the potential to enhance students' mathematical literacy skills. Student responses were highly favourable, suggesting that the incorporation of local contexts contributes to improved engagement with the test items. contextualized

**Keywords**: *Minimum Competency Assessment Numeracy (AKM-Numeracy), Development Research, Bengkulu Context, Test Items*

## Introduction

International educational surveys consistently reveal that the academic performance of Indonesian students remains suboptimal. For instance, the 2022 results of the Programme for International Student Assessment (PISA) indicate that the average mathematics score of Indonesian students falls below the international benchmark (OECD, 2023). This evidence highlights the urgent need to enhance student competencies through refined pedagogical strategies and improved learning practices. In response, the Indonesian government has introduced the Minimum Competency Assessment (AKM) as a pivotal initiative; however, student performance on this assessment remains below optimal levels. Data from the 2021 AKM at the senior high school level (Level 5) reveal that fewer than 50% of students achieved the minimum competency threshold in numeracy (Ministry of Education and Culture's Centre for Educational Assessment, 2022).

Further empirical evidence demonstrates the limited proficiency of students in addressing AKM-based questions. For example, a study among secondary school students in Bengkulu reported mastery levels for AKM-type questions as follows: number concepts at 54.47%, algebra at 46.44%, geometry at 33.15%, and data and probability at 18.81% (Susanto et al., 2023). These findings suggest that secondary school students, particularly those in Bengkulu, are not yet accustomed to engaging with AKM-style problems. Complementary research further indicates that students' mastery of TIMSS-model questions contextualised to Bengkulu is markedly deficient, with only 8.33% attaining a high proficiency level (Susanta et al., 2021).

The suboptimal performance of students in both international and national surveys highlight the need to prioritise the enhancement of mathematical literacy (numeracy) skills. This competency is widely recognised as a fundamental requirement for the 21st century and a key indicator of national educational standards (Drew, 2012; Holenstein et al., 2021). Such skills enable students to interpret and solve problems in a variety of real-world contexts (OECD, 2017). The limited proficiency of students in addressing AKM-based problems may be partly attributed to their unfamiliarity with the challenges presented. Consequently, educators are responsible for developing assessment instruments that closely align with the AKM-style questions established by the government.

In the development of AKM-based test items an assessment format likely unfamiliar to students it is advantageous to adopt an approach that situates problems within contexts familiar to learners. A robust body of research underscores the value of embedding real-world contexts in educational settings to bolster students' cognitive and literacy skills. For example, (Susanta et al., 2022) establish a strong correlation between the integration of local contexts and the enhancement of students' literacy development. Similarly, (Clarke & Roche, 2018), highlight that real-life contexts serve as a motivational mechanism, encouraging greater student engagement with assigned tasks. This engagement is critical, as (Warisdiono, 2017), notes that students' thinking skills can be further refined through consistent practice in addressing problems drawn from everyday experiences.

Moreover, contextualization acts as a conduit, linking students' prior knowledge to new concepts, thereby facilitating comprehension (Susanta et al., 2023). This approach has demonstrated particular efficacy when applied to test items, as evidenced by (Kadir, 2014). Ultimately, the deliberate emphasis on real-world contexts in educational practices not only enriches the learning experience but also significantly advances literacy outcomes (Kaiser & Willander, 2005). Thus, the strategic incorporation of familiar contexts into AKM-based assessments offers substantial potential to elevate both student engagement and academic performance.

Numerous studies have explored the integration of contextual elements to enhance students' competencies. Research has examined mathematical literacy through cultural perspectives (Monica et al., 2022; Safina & Budiarto, 2022). The development of PISA items framed within Indonesia's natural and cultural heritage (Oktiningrum et al., 2016), the design of PISA items based on the Bangka context (Dasaprawira et al., 2019), and the incorporation of PISA items related to the ASEAN Games (Pratiwi et al., 2019). Collectively, these studies highlight the effectiveness of using contextual frameworks in test item development. However, the present study specifically focuses on designing AKM-numeracy-type items contextualized to the Bengkulu setting. Drawing from this literature review, the research aims to address the following questions.

1. How can valid and reliable AKM-numeracy-type items, contextualized within the Bengkulu setting, be developed for senior high school students?

2. What is the potential impact of AKM-numeracy-type items incorporating Bengkulu-specific contextual problems on students' mathematical literacy skills?

3. What insights can be gained from pilot testing regarding the ability of senior high school students in Bengkulu City to solve AKM-numeracy-type items?

**Research Method**

**Research Design**

This study adopts a research and development (R&D) methodology to design a context-based AKM-numeracy assessment instrument tailored for senior high school students, incorporating the Bengkulu context. The development process follows (Tessmer, 1993) model, which consists of two main phases: the preliminary phase and the formative evaluation phase.

The preliminary phase focuses on preparation and design activities, establishing a foundation for the instrument's development. This is followed by the formative evaluation phase, which involves self-evaluation, prototyping (including expert reviews and one-on-one evaluations), small-group evaluations, and field testing. These iterative stages ensure the continuous refinement and validation of the assessment instrument, enhancing its reliability and applicability. The procedural workflow of this development process is illustrated in Figure 1.
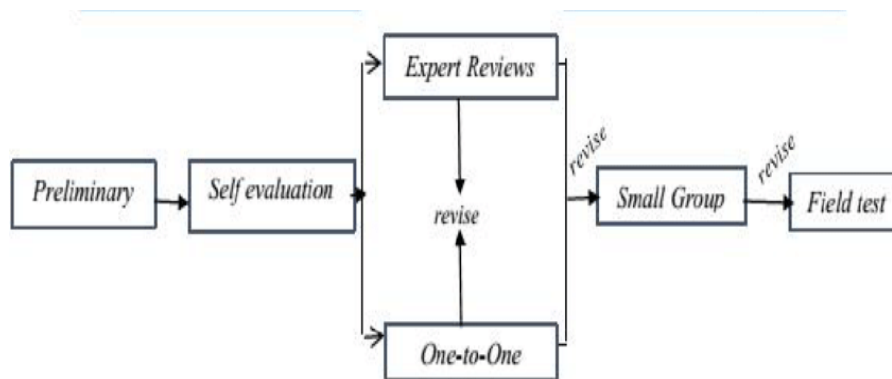


**Figure 1.** Development Stage

**Research Subjects**

The data collection process for developing the context-based AKM-numeracy test items involved a diverse group of participants, including experts (university lecturers), teachers, and students as end-users. The trial participants consisted of second-year senior high

school students from schools in Bengkulu City. Participants were selected according to the research stages, as outlined in Table 1.

**Table 1.** Trial Participants in the Study

| Stage | Purpose | Trial Participants |
|---|---|---|
| Expert and one-on-one review | Evaluation of content validity for the developed items | n=3 (1 lecturer from UIN Fatmawati Bengkulu, 1 lecturer from IAIN Curup, 1 lecturer from FKIP UNIB); n=3 (students) |
| Small-group testing | Assessment of the instrument's readability | n=24 (students) |
| Field test | Analysis of potential impact and students' mathematical literacy skills | n=131 students (School 1: n=36; School 2: n=42; School 3: n=53) |

**Data Collection Techniques and Instruments**

This study employed a mixed-methods approach, combining structured questionnaires and standardized testing to collect data. The questionnaire served two purposes: first, to validate the assessment tool through expert review, and second, to gather student feedback after the trial phase. Concurrently, a criterion-referenced test was administered to senior high school students to evaluate the effectiveness of the newly developed numeracy assessment tool. The research instruments included three key components: a validity evaluation rubric, a post-trial student feedback questionnaire, and a numeracy assessment aligned with Indonesia's National Assessment (AKM) framework.

The validity evaluation rubric utilized a four-point Likert scale (Strongly Disagree to Strongly Agree) to assess critical aspects of the assessment tool, including construct validity, content coherence, clarity of language, and contextual relevance to the Bengkulu region. This rubric enabled detailed analysis of expert judgments across these domains. The post-trial student feedback questionnaire employed a binary response format (Yes/No) to systematically assess students' perceptions. Key areas evaluated included engagement with the test content, relevance of subject matter, language comprehensibility, and cultural appropriateness of the instrument.

The test instrument comprises a suite of assessment items that have undergone rigorous trials across the research subject population. These items are designed in accordance with the Indonesian National Assessment (AKM) framework, emphasizing reasoning skills and

focusing on three core mathematics domains for senior high school students: data and uncertainty, algebra, and geometry. The instrument is structured into two distinct test packages, integrating multiple-choice, complex multiple-choice, and open-response item formats to ensure comprehensive evaluation.

Item development aligns systematically with the AKM proficiency levels for senior high school education, specifically targeting the cognitive domains of knowing, applying, and reasoning. A central performance indicator embedded in the instrument requires students to synthesize diverse representational forms (e.g., symbolic, graphical, contextual) while connecting these to real-world applications, as outlined in guidelines by (Wijaya & Dewayani, 2021).

Scoring criteria are structured into three tiers: responses demonstrating full accuracy and completeness receive 3 points, partially correct or incomplete solutions are awarded 2 points, and incorrect answers or omitted items are assigned 0 points. This rubric ensures consistent evaluation of student performance across all cognitive domains and item formats.

**Data Analysis**

The construction and validation of the AKM-numeracy assessment framework were conducted through a tripartite analytical approach: (1) expert validation analysis, (2) diagnostic evaluation of students' mathematical literacy competencies, and (3) systematic examination of learner feedback. Expert appraisal was conducted using (Aiken, 1980) validity coefficient methodology, with quantitative validity thresholds established through consensus criteria wherein indices exceeding 0.5 were considered psychometrically sound (Retnawati, 2014). This quantitative evaluation was augmented by qualitative analysis of expert commentary, ensuring both statistical rigor and substantive feedback informed subsequent refinements to the instrument.

Student performance was operationalized through a dual analytical lens. Qualitative examination of problem-solving protocols enabled detailed documentation of cognitive processes and error pattern identification. Concurrently, quantitative proficiency stratification was achieved via standardized score conversion (0–100 scale), with performance tiers defined as follows: low proficiency (0–33), moderate proficiency (34–66), and high proficiency (67–100). This bifurcated analytical strategy facilitated both granular diagnostic insights and

standardized achievement categorization. Learner perceptions were systematically examined through quantitative distribution patterns of response frequencies.

## Result and Discussion

### Analysis Phase Outcome

This phase identified the core materials to serve as the foundation for developing test questions. These materials were designed to align with the Minimum Competency Assessment (AKM) framework for Level 5, equivalent to high school standards. The content focuses on three key domains: algebra, geometry, and data analysis. Each domain was reviewed to ensure compatibility with the Merdeka Curriculum's competency goals and the specific requirements of AKM-Numeracy Level 5.

To enhance relevance, the materials were tailored to reflect Bengkulu's regional context. This step was critical, as local contexts often shape the unique characteristics of educational content. Table 2 summarizes how Bengkulu's cultural and environmental elements were integrated into the material development process.

**Table 2.** Content and Context Integration in Material Development

| No | Content Area | Level 5 Materials | Bengkulu Contexts |
|---|---|---|---|
| 1 | Algebra | SPLTV | Batik Basurek (traditional Bengkulu batik), Kue Bay Tat (traditional cake) |
| 2 | Geometry | Solid Figures, Plane Figures | Tabot Bengkulu (traditional ritual artifact), Dol Bengkulu (traditional ceremonial drum) |
| 3 | Data | Data Presentation | Tourism Contexts: Long Beach Bengkulu, Enggano Island Bengkulu |

### Characteristics of the Developed Questions

The AKM-Numeracy test questions formulated for this study were designed through problem-solving tasks anchored in Bengkulu-specific contextual scenarios. The development process resulted in two distinct assessment packages, each consisting of seven items. These questions were scrupulously aligned with the AKM-Numeracy framework to ensure adherence to three key dimensions: content (mathematical principles), context (culturally and geographically relevant Bengkulu-based situations), and question format (structural design and

cognitive complexity). A detailed exposition of the design outcomes for the AKM-Numeracy test questions is provided in Table 3.

**Table 3.** Results of the AKM-Numeracy Test Item Design Based on Content and Context

| Package | Context/Content | Item Number | Question Type |
|---------|-----------------|-------------|---------------|
| A | Bengkulu Batik (Content: Algebra) | 1 | True-False |
| | | 2 | Complex Multiple Choice |
| | | 3 | Essay |
| | Bengkulu Tabot Tradition (Content: Geometry) | 4 | Essay |
| | | 5 | Essay |
| | Bengkulu Tourism (Content: Data) | 6 | Complex Multiple Choice |
| | | 7 | Essay |
| B | Bengkulu Traditional Cuisine (Content: Algebra) | 1 | True-False |
| | | 2 | Complex Multiple Choice |
| | | 3 | Essay |
| | Bengkulu Traditional Musical Instruments (Content: Geometry) | 4 | Essay |
| | | 5 | Essay |
| | Bengkulu Tourism (Content: Data) | 6 | Complex Multiple Choice |
| | | 7 | Essay |

The AKM-Numeracy test items have been meticulously developed in accordance with the AKM (Minimum Competency Assessment) framework, with a particular emphasis on Level 5 reasoning skills that align with the senior high school curriculum. These items encompass key content domains, specifically Algebra, Geometry, and Data Analysis, thereby ensuring a robust and comprehensive evaluation of students' numerical proficiencies.

In addition to their alignment with academic standards, these test items are distinguished by their integration of culturally significant contexts indigenous to Bengkulu Province. This incorporation includes references to prominent tourism attractions, such as Panjang Beach and Enggano Island, alongside cultural heritage elements like the Tabot tradition, Bengkulu batik, and the traditional dol musical instrument. Furthermore, the culinary heritage of the region is represented through the inclusion of the traditional bay tat cake. Each test item is complemented by an informational passage that draws upon these Bengkulu-specific contexts. These passages serve not only to provide a relatable foundation for the mathematical problems but also to foster critical thinking and problem-solving skills by embedding abstract concepts within tangible, real-world scenarios. An illustrative example of such a passage is presented in Figure 2.

**Bengkulu Batik (Content: Algebra)**
Theme 1 (questions 1-3)
Batik Basurek Bengkulu



*source:* detik.com

Basurek cloth is a unique traditional textile originating from Bengkulu, Indonesia. Its history is closely tied to the migration of Prince Sentot Alibasyah, a revered 19th-century figure, who settled in Bengkulu with his family and followers. This cultural heritage is preserved through generations of artisans, many of whom are direct descendants of the prince, ensuring the craft's continuity. The fabric is distinguished by its intricate motifs, which blend natural elements and calligraphy. Common designs include the clove flower, symbolizing prosperity; the tree of life representing growth and connection; the jasmine flower, evoking purity; the kuaw bird, a mythical creature associated with protection; and the rafflesia flower, a rare bloom native to the region. These patterns are artfully interwoven with Arabic calligraphy, reflecting Bengkulu's historical and cultural synthesis of indigenous and Islamic influences (source: detik.com).

**Figure 2.** Design of AKM-Numeracy Test Items in the Bengkulu Context

**Results of the Expert Validation Test**

The test instrument, developed to assess reasoning skills within the AKM-Numeracy framework, underwent a thorough evaluation by a panel of three assessors, consisting of lecturers from UIN Fatmawati Soekarno Bengkulu, IAIN Curup, and Bengkulu University. The evaluation process examined several critical aspects, including content validity to ensure the test accurately measures the intended skills, item construction to assess the quality and structure of the questions, language clarity to verify that the wording is precise and comprehensible, and contextual relevance to confirm alignment with the Bengkulu context. The experts' assessments were converted into numerical scores and analyzed using Aiken's Index, a method applied to determine the validity of each test item by synthesizing consensus among the evaluators. The descriptive statistical results from this analysis are summarized in a table, which presents the Aiken Index scores assigned by the three experts for each question.

**Table 4.** Validity result

| No | Aspect | Average Aiken Value | | Criteria |
|---|---|---|---|---|
| | | **Package A** | **Package B** | |
| 1 | Content | .975 | .914 | Valid |
| 2 | Construction | .890 | .835 | Valid |
| 3 | Language | .890 | .890 | Valid |
| 4 | Bengkulu Context | .963 | .816 | Valid |

As demonstrated in Table 4, the Aiken's validity coefficients for all test items exceeded the critical threshold of 0.05, thereby satisfying the predefined validity criteria. This outcome confirms that each item within the instrument aligns with the stipulated standards for content relevance, structural coherence, linguistic precision, and contextual fidelity to the Bengkulu region. However, while quantitative analysis of expert ratings provided foundational evidence for validity, the validation process extended beyond numerical metrics to incorporate a qualitative-descriptive evaluation of the instrument's feasibility.

Expert recommendations were systematically integrated into iterative revisions, which were further validated through a collaborative focus group discussion (FGD). This multi-stage refinement process ensured the resolution of ambiguities and the incorporation of nuanced insights. Table 5 summarizes the experts' recommendations and corresponding revisions.

**Table 5.** Summary of Expert Responses

| Expert | Recommendations | Revisions |
|---|---|---|
| Expert 1 | ▪ Restructure the test format to differentiate narrative context from problem statements.<br>▪ Enhance question specificity by embedding clear, concise information.<br>▪ Integrate Bengkulu-specific narratives to strengthen contextual relevance. | ▪ Adopt a split format: narrative introduction followed by problem statement.<br>▪ Add supplementary contextual details (e.g., cultural, historical) to each question.<br>▪ Align narratives with authoritative sources on Bengkulu's heritage (e.g., academic journals, cultural archives). |
| Expert 2 | ▪ Include sources for information and images. Strengthen sourcing practices for textual and visual content. | ▪ Cite all information and images using credible references (e.g., peer-reviewed publications, |

| Expert | Recommendations | Revisions |
|--------|-----------------|-----------|
|  | ▪ Include explanatory narratives for each question.<br>▪ Simplify language to ensure accessibility for students. | government cultural databases).<br>▪ Provide a brief narrative preamble for each question to frame its purpose.<br>▪ Revise terminology and syntax to align with student comprehension levels. |
| Expert 3 | ▪ - Provide designated answer spaces to facilitate student responses.<br>▪ - Include the source for each image.<br>▪ - Improve the quality of image presentation.<br>▪ - Address inconsistencies between the narrative and the corresponding question. | ▪ Add designated answer spaces for each test item.<br>▪ - Ensure that sources are properly cited.<br>▪ - Enhance the quality of each image included in the product. |

The data presented in Table 5 reveal that the revisions to the instrument primarily centered on contextual presentation and item construction. This underscores the critical importance of prioritizing three interrelated dimensions when designing locally contextualized test items: (1) the sourcing of visual materials (e.g., images) from credible references, (2) the clarity and sufficiency of contextual information provided, and (3) the alignment of narratives with the assessment objectives. Such considerations are particularly imperative for assessments grounded in regional contexts, where fidelity to local cultural, environmental, or social realities must be meticulously preserved.

An illustrative example of revisions implemented in the research instrument, informed by expert evaluations, is provided below



**Problem**
A Bengkulu souvenir center store sells batik besurek with three distinct motifs, each priced differently per meter, as illustrated below.

Batik 1          Batik 2          Batik 3

(a) Pre-Revision

A Bengkulu souvenir center store sells batik besurek with three distinct motifs, each priced differently per meter, as illustrated below.

Motif 1            Motif 2            Motif 3

(b) Post Revision
**Figure 3.** Expert Review of Question 1: (a) Pre-Revision, (b) Post-Revision

The revisions made to Figure 3 concentrated on refining the image descriptions to clarify Discourse 1 regarding Bengkulu batik. Specifically, the explanation of price differences among the motifs was enhanced with information indicating that each motif utilizes distinct materials, thereby increasing the question's relevance to real-life contexts. Furthermore, the contextual illustrations were updated to reference more reliable and formal sources, such as detik.com, wikipedia.org, and other official publications, to ensure credibility.

In accordance with expert evaluations, the instrument underwent a one-to-one evaluation phase with student participants to verify its alignment with their learning experiences. The results confirmed that the material assessed in the test had been previously studied by the students and was deemed appropriate. Regarding language clarity, students reported that the terminology was accessible and free of ambiguities. Concerning construction quality, the questions were commended for their clear presentation of images and symbols as well as for their unambiguous instructions. Overall, the three student evaluators provided positive feedback on the instrument's material presentation, language accessibility, and instructional clarity.

**Small Group Assessment Result**

Following validation by experts, the research instrument underwent a readability assessment with a small cohort of 24 students. This trial evaluated three core dimensions: (1) the legibility of visual elements (symbols, images, and typography), (2) the clarity of linguistic expression, and (3) the accessibility of contextual framing. The results of the readability

analysis are summarized in Table 6.

**Table 6.** Student Ratings of Readability Criteria

| No | Readability Item | Average Score | Rating |
|----|------------------|---------------|--------|
| 1. | Visual Design (Layout, Symbols, Typeface) | 3.87 | Good |
| 2. | Ease of Navigation | 4.19 | Good |
| 3. | Clarity of Language | 4.21 | Good |
| 4. | Contextual Relevance | 3.74 | Good |

The data indicate that the instrument achieved overall good readability, with all criteria exceeding the threshold for satisfactory usability. Visual design elements including layout, symbols, and typography were rated favorably, suggesting effective integration of graphical and textual components. Similarly, ease of navigation (M = 4.19) and linguistic clarity (M = 4.21) reflect the instrument's user-centric construction. While contextual relevance scored slightly lower (M = 3.74), it remained within the "good" classification, indicating appropriate alignment with students' prior knowledge and cultural familiarity. Collectively, these results demonstrate that the instrument meets the requisite standards for clarity, usability, and contextual accessibility from an end-user perspective.

**Field Test Results**

The implementation of the developed AKM-numeracy test items was conducted across three high schools/Madrasah Aliyah in Bengkulu City, involving 151 students. Data were analyzed using qualitative descriptive methods to characterize students' problem-solving abilities and their engagement with the contextualized test items. To further contextualize performance, student responses were scored on a scale of 0–100, with results summarized in Table 7.

**Table 7.** Students' Numeracy Reasoning Proficiency

| Proficiency Level | Number of Students | Percentage |
|-------------------|--------------------|------------|
| High | 24 | 18,33% |
| Moderate | 74 | 56,48% |
| Low | 33 | 25,19% |

The findings reveal the 56.48% of students (n=74) demonstrated moderate proficiency,

while 18.33% (n=24) achieved high proficiency. Conversely, 25.19% (n=33) scored in the low category. These results indicate that the majority of students exhibited moderate competency in solving the Bengkulu-contextualized AKM-numeracy test items, though significant variability in performance persists. Notably, the moderate average suggests that while students generally grasped the material, challenges remain in achieving higher-order reasoning for a subset of learners.

These data also show that students' abilities at high criteria are relatively small compared to other groups. This indicates that students need to be emphasized when solving non-routine problems. Although the use of the Bengkulu context in the design of the questions has not had a greater impact on students achieving a high level, quantitatively, students have reached a high level of approximately 18%. This is effective when compared to the results of the initial ability survey conducted on students' AKM achievement. Using contextual problems, especially the Bengkulu context in questions, helps students construct ideas in solving questions. Still, the emphasis on the concept of the material must be mastered by each student to solve problems.

To complement the quantitative analysis, qualitative insights into student responses were examined. Figure 4 provides an illustrative example of a student's approach to solving one of the test items, highlighting patterns in reasoning and common errors. This dual-method analysis underscores both the instrument's efficacy and areas for pedagogical improvement.
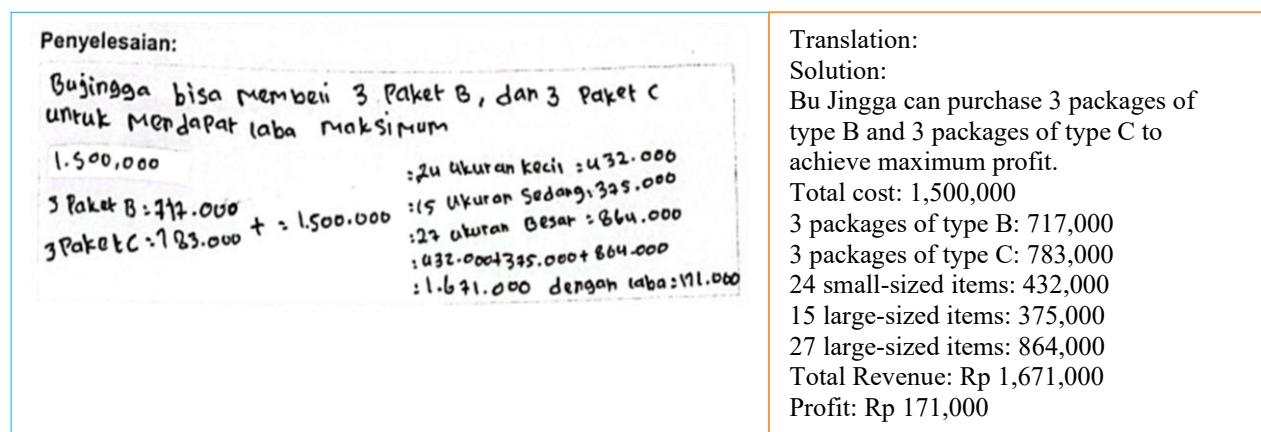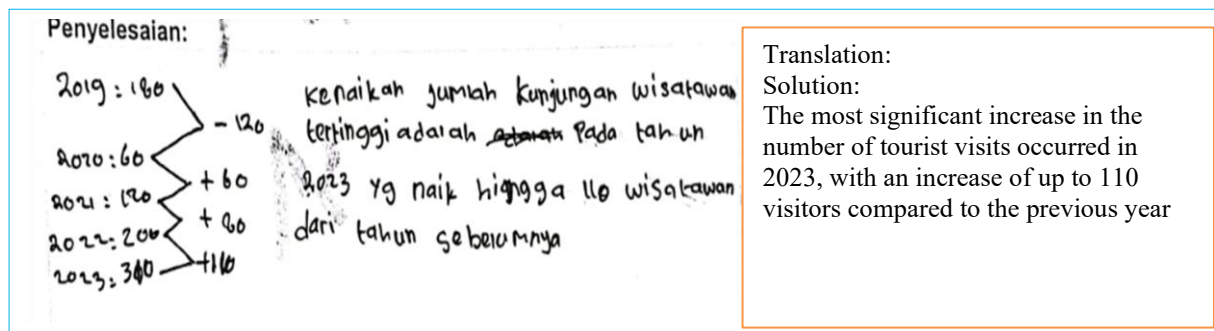


**Figure 4.** Example student's response

The figure above demonstrates the student's successful problem-solving approach in selecting the cookie package that maximizes profit. The student provided a clear rationale to justify both the profit calculation and the final monetary value. However, the response omitted

alternative scenarios that could result in suboptimal profits, limiting a comprehensive evaluation of decision-making nuances.

Figure 5, presented on the following, further elucidates the student's cognitive processes by detailing additional steps and considerations undertaken to resolve the problem.



Translation:
Solution:
The most significant increase in the number of tourist visits occurred in 2023, with an increase of up to 110 visitors compared to the previous year

In addition to outlining students' proficiency levels and elucidating their cognitive processes when addressing the test items, the potential efficacy of the newly developed assessment instruments was evaluated using student response questionnaires. These questionnaires were distributed immediately following the test. A comprehensive summary of the feedback regarding students' experiences with the test items is presented in the table below.

**Table 8.** Analysis of Student Responses to Test Items

| No | Student Response Criterion | Response Rate (%) |
|----|----------------------------|-------------------|
| 1. | Interest in engaging with the presented problems | 63,57 |
| 2. | Seriousness in addressing contextually framed questions | 74,17 |
| 3. | Ease of understanding problem illustrations due to Bengkulu contextualization | 55,62 |
| 4. | Clarity in comprehending task requirements and provided data | 72,84 |

As illustrated in Table 8, more than 60% of the students expressed agreement across all evaluated criteria, suggesting a broadly positive reception of the test items. These findings imply that the AKM-Numeracy test model, contextualized within the Bengkulu framework, effectively fosters student engagement with problem-solving tasks. The results collectively confirm that the test items meet validity standards across several dimensions, including content alignment which reflects relevance to curriculum objective-construct coherence, as evidenced by a logical and consistent task structure, linguistic clarity ensured through accessible and comprehensible language, and contextual applicability achieved by integrating localized Bengkulu elements.

The results of students' responses to solving these questions also show that questions designed using contextual problems, especially the Bengkulu context, have an impact on students' interest or motivation in solving questions. This is to research by Clarke & Roche (2018), which states that using problems with real contexts in assignments can motivate students to complete tasks. In addition, the potential impact of using the questions developed is an increase in students' ability to solve questions, which is indicated by the fact that some students have understood how to solve the questions. This concludes that the use of context in classroom learning is effective in improving students' abilities (Sarwoedi et al., 2018).

This alignment is crucial, as the correspondence between test items and instructional content significantly influences assessment outcomes (Susanta et al., 2022). Furthermore, incorporating familiar contextual scenarios into the AKM-Numeracy questions enhances students' ability to devise solutions, thereby underscoring the pedagogical value of grounding problems in the context of students lived experiences.

**Conclusion**

Drawing upon the findings of this study and the ensuing discussion, it is concluded that the research has successfully developed an AKM-Numeracy test instrument that satisfies the requisite standards of validity and usability, while also demonstrating potential to enhance students' numeracy skills. The validity of the instrument, evaluated through the Aiken index analysis, confirms that all test items surpass the threshold criterion, with values exceeding 0.05, thereby meeting established validity benchmarks. In terms of usability, the developed test has been rated as "good" on average, reflecting its overall ease of application. Furthermore, the test's potential impact is substantiated by the data, with students' average performance classified as adequate and over 60% of respondents offering positive feedback following their engagement with the test items. Additionally, the findings underscore that integrating the Bengkulu context into the design of the AKM test items generally enhances the clarity and relevance of problem-solving scenarios, thereby facilitating students' comprehension and interaction with the material. The implications of the results of this study on learning at the secondary school level, especially in providing variations of questions in preparing for the implementation of AKM. In addition, teachers can use the results of this study as a reference in developing problems in questions or evaluation instruments.

## Suggestion

Based on the findings and conclusions of this study, several recommendations are proposed. First, the development of AKM test items should incorporate real-world contexts that closely reflect students lived experiences, such as local cultural, environmental, and socioeconomic scenarios. Second, it is essential to conduct a comprehensive evaluation of the compatibility between the selected contextual frameworks and the curricular content to ensure both pedagogical relevance and coherence. These measures will enhance the validity of assessments and improve their applicability within regionally specific educational ecosystems.

## Acknowledgment

## References

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4). https://doi.org/10.1177/001316448004000419

Clarke, D., & Roche, A. (2018). Using contextualized tasks to engage students in meaningful and worthwhile mathematics learning. *Journal of Mathematical Behavior*, *51*(September 2017), 95–108. https://doi.org/10.1016/j.jmathb.2017.11.006

Dasaprawira, M. N., Zulkardi, & Susanti, E. (2019). Developing mathematics questions of PISA type using Bangka context. *Journal on Mathematics Education*, *10*(2), 303–314.

Detik.com (2024). Batik basurek budaya Bengkulu. www.detik.com

Drew, S. V. (2012). Open up the ceiling on the common core state standards: preparing students for 21stcentury literacy-now. *Journal of Adolescent & Adult Literacy*, *56*(4), 321–330. https://doi.org/10.1002/JAAL.00145

Holenstein, M., Bruckmaier, G., & Grob, A. (2021). Transfer effects of mathematical literacy: An integrative longitudinal study. *European Journal of Psychology of Education*, *36*, 799–825. https://doi.org/10.1007/s10212-020-00491-4

Kadir, K. (2014). Using context and prior knowledge of mathematics in learning students' creative thinking skills. *Jurnal Pendidikan Matematika*, *5*(1), 52–66. https://doi.org/10.36709/jpm.v5i1.2041

Kaiser, G., & Willander, T. (2005). Development of mathematical literacy: Results of an empirical study. *Teaching Mathematics and Its Applications*, *24*(2–3), 48–60. https://doi.org/10.1093/teamat/hri016

Ministry of Education and Culture's Centre for Educational Assessment. (2022). *Asesmen

*kompetensi minimum [Minimum competency assessment]*. Kemendikbud.

Monica, S. S., Pamugkas, A. S., & Jaenudin, J. (2022). Instrumen Literasi Matematika Model PISA dengan Konteks Budaya Baduy pada Tingkat SMP. *Jurnal Educatio FKIP UNMA*, *8*(4), 1459–1470. https://doi.org/10.31949/educatio.v8i4.3839

OECD. (2017). *Enhancing the Contributions of SMES in a Global and Digitalised Economy*.

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

Oktiningrum, W., Zulkardi, & Hartono, Y. (2016). Developing PISA-like mathematics task with Indonesia natural and cultural heritage as context to assess students' mathematical literacy. *Journal on Mathematics Education*, *7*(1), 1–8. https://doi.org/10.22342/jme.7.1.2812.1-8

Pratiwi, I., Putri, R. I. I., & Zulkardi, Z. (2019). Long jump in Asian Games: Contexts of PISA-Like mathematic problems. Journal on Mathematics Education. *Journal on Mathematics Education*, *10*(1), 81–92. https://doi.org/10.22342/jme.10.1.5250.81-92

Retnawati, H. (2014). *Proving the validity of the instrument in the measurement*.

Safina, D., & Budiarto, M. T. (2022). LITERASI MATEMATIS BERBASIS BUDAYA SIDOARJO DALAM PERSPEKTIF ETNOMATEMATIKA. *MATHEdunesa*, *11*(1), 12–25. https://doi.org/10.26740/mathedunesa.v11n1.p12-25

Shoopie. (2024). Batik Bengkulu. https://shopee.co.id/batikbengkulu

Susanta, A., Sumardi, H., Susanto, E., & Retnawati, H. (2023). Mathematics literacy task on number pattern using Bengkulu context for junior high school students. *Journal on Mathematics Education*, *14*(1), 85–102. https://doi.org/10.22342/jme.v14i1.pp85-102

Susanta, A., Sumardi, H., & Zukardi. (2022). *Development of E-module Using Bengkulu Contexts to Improve Literacy Skills of Junior High School Students*. *16*(2), 171–186. https://doi.org/10.22342/jpm.16.2.17698.171-186

Susanta, A., Sumardi, H., & Zulkardi, Z. (2022). Development of e-module using Bengkulu contexts to improve literacy skills of junior high school students. *Jurnal Pendidikan Matematika*, *16*(2), 171–186. https://doi.org/10.22342/jpm.16.2.17698.171-186

Susanta, A., Susanto, E., & Maizora, S. (2021). The Level of Junior High School Students' Thinking in Solving TIMSS Mathematical Problem in Bengkulu. *Proceedings of the International Conference of Mathematics and Mathematics Education (I-CMME 2021)*, *597*, 9–13. https://doi.org/10.2991/assehr.k.211122.002

Susanto, E., Fransiska, H., & Susanta, A. (2023). *Students ' numerical ability on minimum competency assessment in junior high school*. *6*(1), 47–53.

Tessmer. (1993). *Planning and conducting formative evaluation*. Taylor and Francis.

Warisdiono. (2017). *Module for compiling higher-order thinking skills (HOTS)*. Directorate of High School Development, Directorate General of Primary and Secondary Education, Ministry of Education and Culture.

Wijaya, A., & Dewayani, S. (2021). *Framework Asesmen Kompetensi Minimum (AKM) [Minimum Competency Assessment Framework (AKM)]*. Kementerian Pendidikan dan Kebudayaan.