Jurnal Elemen, 12(1), 182-199, January 2026
https:/doi.org/10.29408/jel.v12i1.32841

e- ISSN: 2442 - 4226

Development of an android-based numeracy literacy test using iSpring Suite in a Bengkulu coastal context

Elwan Stiadi [1 *], Ari Putra [2], M. Zainudin [3], Meidy Putra Areka [4], Ahbi Mahdianing Rum [1], Nader U-Sin [5]

[1] Department of Mathematics Education, University of Bengkulu, Bengkulu, Indonesia
[2] Department of Non-formal Education, University of Bengkulu, Bengkulu, Indonesia
[3] National Research and Innovation Agency (BRIN), Jakarta, Indonesia
[4] State Junior High School 6 Kota Bengkulu, Bengkulu, Indonesia
[5] Ban Lanchang School, Pattani, Thailand

[*]Correspondence: elwanstiadi@unib.ac.id

Abstract

The Minimum Competency Assessment (MCA) is Indonesia's national instrument for measuring foundational competencies, including numeracy literacy. However, findings from the Programme for International Student Assessment and the MCA indicate that many students still perform at low levels, partly because test items rely on limited, less contextualized problem settings. This study developed an Android-based numeracy literacy test using iSpring Suite by embedding items within the historical heritage context of the Bengkulu Coast to support students' assessment preparation. The instrument was designed using the ADDIE model (analysis, design, development, implementation, and evaluation) and piloted with ninth-grade students in Bengkulu City, Indonesia. Data were collected through interviews, questionnaires, documentation, and item try-outs, and analyzed for content validity, reliability, practicality, item difficulty, and discrimination. The expert review confirmed the feasibility of the content, construct, and language. Empirical results for 10 items showed acceptable to strong validity, high reliability (0.781), difficulty levels ranging from easy to moderate, and discrimination indices from moderate to very good. The practicality score reached 84.7% (very practical), indicating that contextualized mobile-based assessment is a valid and practical alternative for numeracy learning.

Keywords: android; Bengkulu coast; ispring suite; numeracy literacy; test instruments

How to cite: Stiadi, E., Putra, A., Zainudin, M., Areka, M. P., Rum, A. M., & U-Sin, N. (2026). Development of an android-based numeracy literacy test using iSpring Suite in a Bengkulu coastal context. *Jurnal Elemen, 12*(1), 182-199. https://doi.org/10.29408/jel.v12i1.32841

Received: 29 October 2025 | Revised: 8 January 2026
Accepted: 30 January 2026 | Published: 19 February 2026

Introduction

The Indonesian Ministry of Education has implemented the National Assessment as a system-level evaluation to monitor and improve educational quality. It comprises three components: the Minimum Competency Assessment, the Character Survey, and the Student Learning Environment Survey (Novita et al., 2021). The Minimum Competency Assessment focuses on foundational competencies particularly reading literacy and numeracy that are conceptually aligned with the competencies assessed in international large-scale assessments such as PISA, where performance is examined through real-world problem-solving contexts (OECD, 2019; Rohim, 2021). As a basic competency assessment, the Minimum Competency Assessment is expected to help students strengthen essential skills needed to participate productively in society (Han et al., 2017).

Developing strong literacy and numeracy is widely regarded as crucial for learners in the 21st century (Rezky et al., 2022). According to Novita et al. (2021), numeracy literacy is the ability students possess to understand mathematical concepts, procedures, and numbers to solve everyday problems. This ability helps students understand the importance of mathematics in solving problems in everyday life. Therefore, numeracy literacy skills are an essential part of the work process.

However, in reality, Indonesian students' numeracy literacy results are still low. Based on test and survey data from the Programme for International Student Assessment (PISA) in 2018, Indonesia's score was still low. Indonesia ranked 74th out of 79 countries. In terms of numeracy literacy, Indonesia achieved an average score of 379, compared to the Organisation for Economic Co-operation and Development (OECD) average score of 489 (OECD, 2019; Stiadi et al., 2022, 2023). Furthermore, based on the the Minimum Competency Assessment test results, Indonesian students' numeracy literacy skills are still low (Delima et al., 2022; Klarita & Syafiah, 2022; Wijaya & Dewayani, 2021). Reported from the official website of the Ministry of Education and Culture, the achievement of numeracy learning outcomes of Indonesian junior high school students is below the minimum competency, namely only less than 50% of students have reached the minimum competency limit (Peraturan Menteri Pendidikan Dan Kebudayaan Tentang Kurikulum 2013 Sekolah Menengah Atas/Madrasah Aliyah., 2014).

A key factor contributing to low numeracy performance is students' difficulty in comprehending the reading demands embedded in the Minimum Competency Assessment items, particularly when contexts and texts are unfamiliar or limited in variety. The lack of diverse texts and problem contexts has been identified as a contributor to students' difficulties in understanding numeracy tasks (Wardhani & Oktiningrum, 2022). Therefore, designing numeracy assessments that use meaningful and locally recognizable contexts may help students connect mathematical ideas with everyday experiences while simultaneously addressing the literacy demands of word problems.

One promising context for contextualized numeracy assessment is the coastal environment of Bengkulu. Coastal areas represent the interface between land and sea (Khofifah & Purwati, 2020). Furthermore, according to Kristie et al. (2013), a coastline is a land area at

the edge of the sea that is still influenced by the sea, such as tides, sea breezes, and seawater infiltration. Bengkulu is a coastal region whose daily socio-economic activities, such as fisheries, seafood trade, and marine tourism, naturally involve quantitative reasoning. Embedding numeracy tasks in such settings may make abstract mathematical concepts more accessible, for example through interpreting catch quantities, estimating profits, or reasoning with tidal patterns. This approach is also consistent with the Independent Curriculum orientation that encourages contextual, project-based, and environment-based learning, potentially allowing the assessment to be relevant not only for Bengkulu but also for other coastal areas in Indonesia.

In addition to contextual challenges, assessment implementation in many schools still relies heavily on paper-based tests, which can be costly and time-consuming to administer and score (Zakaria et al., 2017). The use of technology is not a barrier for students at SMPN 6 Bengkulu City, as the school allows students to use Android devices during the learning process. Research conducted by Kartini and Putra (2020) also states that the use of Android-based learning media significantly influences student learning outcomes.

The broader development of digital assessment has also increased opportunities for efficient and transparent testing practices. Technology-based assessments can offer practical advantages for teachers and students because they are more economical, objective, and transparent (Farman et al., 2020, 2021). Nevertheless, many teachers still face difficulties in implementing technology-integrated assessments (Kamilati, 2018). Mahfud et al. (2019) reported that nearly 92% of 100 teachers did not know what technologies could be integrated into learning processes, particularly for developing test-based assessment instruments. This suggests a need for practical tools and development models that help teachers create technology-supported assessments without excessive technical burden.

One tool that can support such development is iSpring Suite. iSpring Suite enables teachers to design test instruments with varied item formats and multimedia features, including audio, video, and online/offline YouTube integration (Farman et al., 2021) It also provides automated scoring and test result displays (Cahyanti et al., 2019). It is easy to use and doesn't require a lot of time to create. Furthermore, it can be integrated with Microsoft PowerPoint, making it easier for teachers to implement continuous evaluation tests (Afifah et al., 2021). Research conducted by Suprapti (2016) stated that iSpring Suite-based learning devices meet good quality criteria. Likewise, research conducted by Cahyanti et al. (2019) stated that the ispring suite can be used as a learning evaluation tool by creating test instruments using the features available in the ispring suite.

However, existing studies and practices tend to address (a) numeracy outcomes and the Minimum Competency Assessment performance issues, (b) contextual learning materials, or (c) digital tools for assessment, often as separate strands. Less attention has been paid to developing and validating a mobile-based numeracy literacy assessment instrument that simultaneously (1) responds to the Minimum Competency Assessment's literacy-rich numeracy demands, (2) leverages authentic local contexts, such as coastal heritage and coastal livelihoods, and (3) uses an accessible tool that teachers can realistically adopt. Addressing this need, the present study develops an Android-based numeracy literacy test instrument using iSpring Suite

with the Bengkulu coastal context to support students' preparation for the Minimum Competency Assessment.

Methods

This study employed research and development (R&D) to produce an Android-based numeracy literacy test instrument created with iSpring Suite and embedded in the Bengkulu coastal context. The development process followed the ADDIE model—analysis, design, development, implementation, and evaluation. ADDIE was selected because it provides a structured yet flexible workflow that suits iterative development of technology-enhanced assessment instruments: it begins with needs analysis, proceeds to systematic design and prototyping, and ends with implementation and evaluation cycles that support revision based on both expert judgment and empirical trial results (Cahyadi, 2019). Compared with broader development models such as Borg & Gall, ADDIE is more concise and practical for producing a pilot-ready digital assessment product within a school-based setting and a limited development timeframe, while still allowing systematic validation and revision.

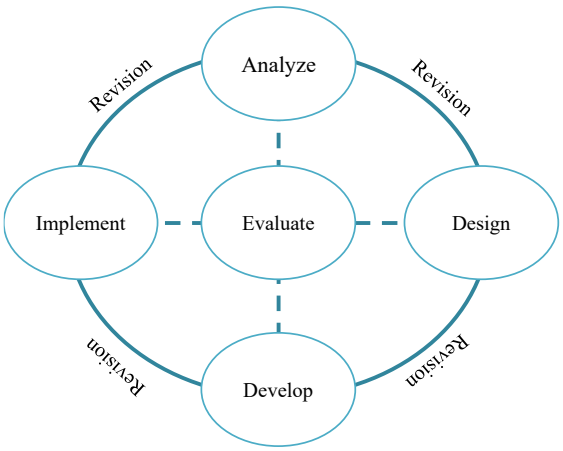

Figure 1. Stages of the ADDIE development model

Participants and sampling

The target users of the instrument were Grade IX students of SMPN 6 Bengkulu City. Participants for the pilot implementation consisted of 24 students selected using purposive sampling. Purposive sampling was used because the trial required students who (1) were enrolled in Grade IX (the intended the Minimum Competency Assessment preparation level in this study), (2) had participated in regular mathematics learning activities at the school, and (3) had access to and were permitted to use Android devices during classroom activities, consistent with the school's learning policy.

The sample size of n=24 was considered adequate for a pilot test whose primary purpose was to evaluate initial empirical item functioning (e.g., difficulty, discrimination, distractor performance) and the practicality of the Android-based delivery, rather than to generalize results to a broader population. This aligns with the role of small-scale trials in early-stage instrument development, where the goal is to identify problematic items and usability issues before larger field testing.

Data collection

Data were collected through interviews conducted during the analysis stage to identify needs, constraints, and user considerations. Expert validation questionnaires were then administered to evaluate the items' content, construct, and language quality, as well as the suitability of the Android/iSpring delivery media. A pilot test (item try-out) was carried out to obtain students' response data for empirical item analysis. Practicality questionnaires were used to assess ease of preparation, ease of use, and ease of processing results within the Android-based testing process. In addition, documentation was compiled to support development records, track revisions, and provide evidence of the improvement process.

Development and validation procedures

Expert validation was conducted in two parts. First, the test items were reviewed for content relevance, construct appropriateness, and language clarity. Second, the Android-based media (iSpring Suite output) was reviewed to ensure the assessment could be delivered and used practically in learning activities. Revisions were made based on validators' feedback before pilot testing.

Empirical item analysis

After the pilot test, item-level empirical analysis was conducted, covering item validity, reliability, difficulty index, discriminating power, and distractor effectiveness. Item validity was examined using the point-biserial correlation between each item score and the total test score (Nitko & Brookhart, 2011). For dichotomous scoring (correct =1, incorrect =0), the point-biserial correlation is:

$$r_{pb} = \frac{M_1 - M_0}{S_t}\sqrt{pq}$$

where M_1 is the mean total score of students who answered the item correctly, M_0 is the mean total score of students who answered incorrectly, S_t is the standard deviation of total test scores, p is the proportion answering correctly, and $q = 1 - p$ (Nitko & Brookhart, 2011).

Test reliability was estimated using the Kuder–Richardson Formula 20 (KR-20), which is appropriate because the items were scored dichotomously (correct/incorrect) (Nitko & Brookhart, 2011). KR-20 estimates internal consistency reliability for a single administration of a test with dichotomous items:

$$KR20 = \frac{k}{k-1}\left(1 - \frac{\sum\ p_i q_i}{\sigma_t^2}\right)$$

where k is the number of items, p_i is the proportion of correct responses on item $i, q_i = 1 - p_i$, and σ_t^2 is the variance of the total test scores.

Item difficulty was calculated using the proportion-correct index:

$$p = \frac{B}{N}$$

where B is the number of students answering correctly and N is the total number of students. Items were then categorized into difficulty levels (easy, moderate, difficult) to ensure an appropriate spread across levels(Nitko & Brookhart, 2011).

Item discrimination was used to evaluate each item's ability to differentiate between higher- and lower-performing students. Discrimination was computed using the upper–lower group approach:

$$D = P_U - P_L$$

where P_U is the proportion of correct responses in the upper group and P_U is the proportion of correct responses in the lower group(Nitko & Brookhart, 2011).

Distractor effectiveness was evaluated by examining whether each incorrect option attracted responses, particularly from students with lower overall performance. Distractors that were rarely chosen were flagged for revision to improve plausibility and diagnostic value(Nitko & Brookhart, 2011).

Practicality analysis

Practicality of the Android-based instrument was assessed using a Likert-scale questionnaire covering ease of preparation, ease of use, and ease of processing results (Mawarinda et al., 2022). The practicality score was converted into a percentage and interpreted using established criteria reported in the relevant literature. In this study, the instrument was categorized as practical when it achieved at least the "practical ($60 < N \leq 80$)" or "very practical ($80 < N \leq 100$)" level according to those criteria. This interpretation ensured that the developed instrument was not only psychometrically acceptable at the pilot level, but also usable for technology-based numeracy learning and preparation for the Minimum Competency Assessment.

Results

Results of the analysis phase

The analysis stage is the first stage in research using the ADDIE development model. This stage aims to identify learning problems and provide solutions to them. The analysis stage involves analyzing needs and curriculum analysis, comparing them to learning outcomes. The results of the analysis stage are explained below.

Needs analysis learning

The analysis phase aims to identify learning problems and appropriate solutions. At SMPN 6 Bengkulu City, teacher interviews and the 2024 the Minimum Competency Assessment report card showed that ninth-grade students' numeracy skills were still low; the majority were in the "red" category (not yet complete). Students' main difficulties were understanding contextual problems, representing them mathematically, and connecting concepts to real-world situations.

Teachers tend to assign procedural questions, leaving students less trained in context-based the Minimum Competency Assessment questions. Therefore, the local context of the Bengkulu coastline is considered relevant to helping students understand the questions because it is close to everyday life and also fosters local wisdom.

Data shows that only 46.67% of students achieved minimum numeracy competency. This situation underscores the need for a locally context-based, Android-based numeracy literacy test instrument. Digital tests are considered more efficient than paper-based ones due to their affordability, quick correction, and practicality. iSpring Suite was chosen because it supports the creation of interactive questions integrated with PowerPoint and various media (text, images, audio, and video).

Thus, the development of Android-based numeracy test instruments and coastal contexts is very necessary to train students to face the Minimum Competency Assessment questions and support 21st-century learning.

Curriculum analysis

The development of numeracy literacy test instruments must align with the Merdeka Curriculum applicable at SMPN 6 Bengkulu City. This curriculum emphasizes competency-based learning, strengthening the Pancasila Student Profile, and literacy and numeracy as fundamental skills across subjects. In Mathematics Phase D (grades VII–IX), learning outcomes include numbers, algebra, geometry, and data/statistics, with an emphasis on contextual problem solving.

Design phase results

In this study, the design stage is the second stage carried out using the ADDIE development model. The design stage consists of several steps, including item design and Android application design. The results of the design stage will be explained below.

Design question items

The item development process begins with determining the test's objective, which is to diagnose numeracy literacy. The stages include establishing cognitive levels (knowing, applying, reasoning), numeracy domains and subdomains (number, algebra, geometry, data/uncertainty), formulating competencies, and linking question indicators to real-world contexts.

The instrument was developed based on the Bengkulu Coastal context to make it more meaningful and relevant to students' lives. The resulting grid consists of 10 multiple-choice questions with four answer options: 3 knowing questions, 5 applying questions, and 2 reasoning questions. Topics include speed, distance, time, averages, social arithmetic, probability, and number operations.

Each correct answer is worth 1 and each wrong answer is worth 0. This instrument will later be implemented in the form of an Android application using iSpring Suite to support local context-based numeracy learning.

Android application design

The Android application design was carried out after conducting the analysis stage and aims to facilitate the development process of the Android application for the numeracy literacy test instrument in the coastal context of Bengkulu. This design stage was carried out by creating an Android application design for the numeracy literacy test instrument in the coastal context of Bengkulu.

The Android application design steps consist of three steps. The first step is determining the flow of the media design being developed. The media design flow is presented in Figure 4. The second step is selecting the media title where the Android application for electronic instruments based on the ARCS design model and the Android ispring suite is called the numeracy literacy test application with the context of the Bengkulu coast. The third step is determining the media layout. Determining the Android application layout is used to design an attractive Android application that is in accordance with the ARCS aspects so that it can increase student learning motivation. There are three steps in determining the layout in the first Android application: preparing the application that will be used in developing the Android application, second: collecting the media components used, and finally determining the font type and size that are appropriate for the Android screen. Thus, the results of the Android application design are obtained which are used to design an attractive Android application for test instruments that are in accordance with the ARCS aspects.

Development phase results

The development stage is the third stage in research using the ADDIE development model. The development stage consists of several steps structured based on the results of the analysis and design stages. The results of the development stage are explained in the section below.

Item developmentand android apps

The draft test instrument was arranged based on a grid into 10 multiple-choice questions with 4 options. Next, the Android application was developed through three stages: (1) determining the media and supporting devices, (2) establishing the product identity named the Numeracy Literacy Test with the Bengkulu Coastal Context (LitNum), and (3) developing the product according to the design. The media used included iSpring Suite (main platform), PowerPoint (design), Canva (visual assets), Web 2 Apk Builder (HTML to APK conversion), Google Drive (storage), QR Maker (barcode access), and Gmail (score report). The application consists of three parts: (1) introduction (home page, participant identity, instructions), (2) test (10 questions, question list, results, review), and (3) closing (thanks).

The results of the application development phase include an application product and HTML ready for use on Android. After the product is developed, the next stage is validation of the test items and validation of the Android application media.

Validation of test item and android apps

At this stage, the draft questions were validated by the validators. The validation included validation of the material, construction, and language of the question items. The validation results showed that validator 3 stated that the product was not yet valid in terms of material aspect 4. According to validator 1, the product was not yet declared valid because there were still notes of improvement on the material validity sheet for aspect 6 and the language validity sheet for aspect 1. According to validator 2, the product was not yet declared valid because there were still notes of improvement on the language validity sheet for aspect 1. After improvements were made according to the validator's suggestions, the Android-based

numeracy literacy test instrument with the Bengkulu coastal context was revalidated by the three validators and there were no further improvements. Next, the three validators assessed the questions on the validation sheet. The results of the completed validation sheet were then calculated using the Aiken's V formula. The Android-based numeracy literacy test instrument with the Bengkulu coastal context has been validated by the three validators, and all items were declared very valid with an Aiken's V index ranging from 0.978–1.00. Validation of the application media also showed very valid results with an index of 0.966 based on the ARCS (Attention, Relevance, Confidence, Satisfaction) aspects, so no further revisions were required. Thus, the developed Android test instrument and application are suitable for use in field testing, including empirical validation, feasibility testing, question readability, and application practicality.

Implementation results (feasibility test and readability of question items)

A field test of the Android-based numeracy literacy test instrument with the Bengkulu coastal context was conducted on 24 ninth-grade students of SMPN 6 Bengkulu City. The results of the empirical validity test showed that all items were valid (γ_{pbis} = 0.414–0.811) with very strong, strong, and fairly strong criteria. The instrument's reliability was high (KR-20 = 0.78), so the test was declared consistent. Analysis of the difficulty index showed that the items were in the easy–moderate category, while the discriminating power ranged from moderate to very good. The distractor test showed that all answer options functioned effectively. The readability of the items was very good with an average of 97.7%. Overall, this instrument is valid, reliable, easy to understand, and can be used for research.

Table 1. Conclusion of Field Test Results

Question Items	Validity	r_{11}	P	D	Deception	Readability	Conclusion
1	0.811	0.78	Currently	Very well	Functioning Because it was chosen > 2% test participants	Very good	Used
2	0.414		Easy	Currently		Very good	Used
3	0.647		Currently	Very well		Very good	Used
4	0.683		Easy	Good		Very good	Used
5	0.528		Currently	Currently		Very good	Used
6	0.578		Currently	Good		Very good	Used
7	0.67		Easy	Good		Very good	Used
8	0.428		Currently	Good		Very good	Used
9	0.482		Currently	Currently		Very good	Used
10	0.414		Easy	Currently		Very good	Used

Overall, the field test showed that the 10 questions tested met the criteria for good questions and could therefore be used.

Practicality test of android application for numeracy literacy test

The practicality test of the test items was conducted to measure the ease of use of the test instrument application in the learning process (Rukajat, 2018, p. 138). The practicality assessment was conducted using a use questionnaire that considered aspects of usefulness, ease of use, ease of learning, and user satisfaction (Ningtiyas et al., 2021). The results of the practicality test are presented in Figure 2 below.

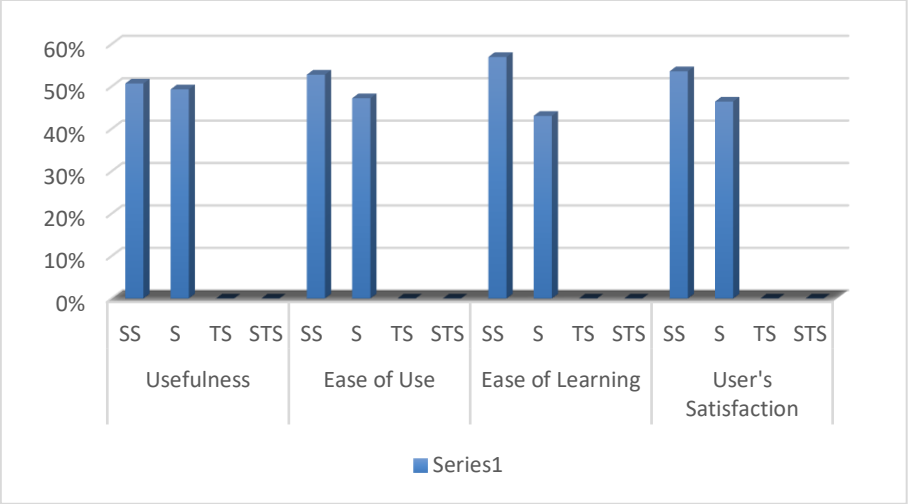

Figure 2. Practicality test results diagram

The results of the practicality test of the Android application in Figure 2 as a whole, show that many students chose the SS (Strongly Agree) and S (Agree) answer options. These SS and S answer options are options that support the practicality of the Android application. So it can be concluded that this Android application is appropriate and practical for students to use in the assessment process. The practicality test was conducted using a use questionnaire. In the usefulness aspect, the overall results obtained were that more students predominantly chose the SS (Agree) option with an overall score of 73.4% for the usefulness aspect with practical criteria. Likewise, in the aspects of ease of use, ease of learning, and user satisfaction, most students chose the SS (Strongly Agree) option with scores of 88.2%, 88.2%, and 88.4% with very practical criteria. The results of the overall practicality test of the Android application obtained a practicality level of the test instrument with the criteria of very practical (84.7%).

Evaluation stage results

The evaluation stage is a crucial step in the ADDIE development model, which aims to assess product quality from the initial creation process through implementation (Cahyadi, 2019, p. 36). In the research context, evaluation involves assessing various aspects, such as the validity of the test instrument and the practicality of the developed application. The evaluation results from the analysis to implementation stages indicate that the test instrument has been deemed valid and reliable based on the validator's assessment and empirical test results. This is reflected in the good quality of the test items, including the difficulty index, discriminatory power, readability test of the test instrument, and the ability of distractors to function well. In addition, the practicality test of the Android application also produced a high score of 84.7% with the

very practical criterion. This confirms that the product resulting from the ADDIE process has passed a rigorous evaluation and is ready for effective implementation.

Discussion

The discussion section is designed to explain the extent to which the main questions in the research problem formulation have been answered or still require further development. To achieve this objective, an evaluation was conducted based on data from validity tests and instrument practicality tests. The analysis results show that: (1) the electronic instrument based on the ARCS design model and the iSpring Suite Android application in the numeracy literacy test with the coastal context of Bengkulu Beach is proven valid; (2) the aspects of empirical validity, reliability, level of difficulty, discriminatory power, and distractors of the items meet the eligibility criteria, and have met the readability criteria, namely very good; and (3) the test instrument is considered very practical to use based on the results of the use questionnaire.

Thus, it can be concluded that the problem formulation and research objectives have been answered adequately because the instrument developed is proven valid, feasible, and practical according to the direction of the research. Furthermore, a description of the discussion results will be presented in the following section.

Validity of question itemsand android application of numeracy literacy test instrument with coastal context of Bengkulu

The development process for the Android iSpring Suite-based numeracy literacy test instrument for the Bengkulu coastal context was conducted through logical validation stages involving assessments by validators. Validation covered aspects of material, construction, and language. Initial validation results indicated improvements from the validators, such as the lack of clarity in transportation cost information accompanied by a description of the number of people, unclear image references in certain questions, and inconsistent answer options with the indicators. These findings confirm that the developed instrument requires iterative revision to ensure clarity of wording, conformity to indicators, and contextual relevance. This aligns with Azwar's (2012) opinion that instrument validity is a primary requirement that must be met for accurate interpretation of measurement results.

The instrument was revised based on input from the validators. Key changes included clarifying the transportation cost information to Rp 15,000 per person, reinforcing the image reference in question three, and adjusting the answer options to align with the indicators. This expert feedback-based revision process reflects the principle of instrument development as proposed by Retnawati (2016), which states that expert involvement in logical validation is a crucial step in ensuring the instrument has clear content and robust constructs.

After revision, the instrument was revalidated, with results showing that all items were in the highly valid category, with Aiken's V index ranging from 0.978 to 1.000. These results confirm that the instrument has met logical validity standards, both in terms of material, language, and construction. This finding is consistent with Nieveen (1999) research, which

states that a quality learning instrument must meet three main aspects: validity, practicality, and effectiveness. In the context of this study, the validity stage has been achieved very well.

In addition to item validity, the Android-based application developed to support the test was also validated using the ARCS (Attention, Relevance, Confidence, Satisfaction) model. The media validation results obtained an Aiken's V index of 0.966, categorized as very valid. This means that the application is not only technically valid but also effective in supporting student learning motivation through attention, relevance, confidence, and satisfaction. These results support the research.Marhogi et al. (2025), which shows that Android-based learning media can increase student engagement due to ease of access, interactive displays, and suitability to learning needs.

Thus, it can be concluded that the Android-based iSpring Suite numeracy literacy test instrument, applied to the coastal context of Bengkulu, is logically valid. Both the test items and the media application demonstrated a very high level of validity, making it suitable for use in field trials. This achievement confirms that the developed instrument complies with educational measurement standards and is relevant to support local context-based minimum competency assessments.

Eligibilityand readability of test instrument items

The results of the field test show that the questions of the Android-based numeracy literacy test instrument with the Bengkulu coastal context have met the eligibility criteria from various aspects, namely empirical validity, reliability, difficulty index, discriminating power, distractors, and readability.

First, from the empirical validity aspect, all test items have a point biserial correlation coefficient γ_{pbis} greater than r_{tabel} (0.404), with a range of 0.414–0.811. This indicates that all test items are valid, although at different levels, ranging from moderately strong, strong, to very strong. This finding is in line with research. Anshari et al. (2024) which emphasizes that test instruments with greater validity than can be relied upon to measure students' abilities accurately.

Second, from a reliability perspective, the KR-20 coefficient value was 0.781, which is considered highly reliable. This result demonstrates that the test instrument has strong internal consistency. This is consistent with the findings. Erfan et al. (2020) which states that a reliability value above 0.70 indicates that the instrument can be used consistently in different measurement contexts.

Third, the results of the difficulty index analysis show that all questions are in the easy to moderate category, with no items being too difficult. This is ideal because questions that are too difficult or too easy are less able to measure students' abilities proportionally (Naja & Utami, 2025). In addition, the discrimination aspect shows that all questions can differentiate well between high- and low-ability students, with the categories being moderate to very good. This finding supports the results of the study Zulfayani et al. (2024) who found that instruments with balanced variations in difficulty index and good discriminatory power were able to provide an objective picture of students' numeracy literacy achievements.

Fourth, in the distractor function analysis, all answer options functioned well, as they were chosen by at least 2% of students. This indicates that the distractors worked as intended in leading students who lacked conceptual understanding to choose incorrect answers. This result aligns with the statement. Salsabilla et al. (2025) that well-functioning distractors improve the quality of the instrument because they reduce the likelihood of students randomly guessing answers.

Fifth, in terms of readability, the test instrument demonstrated excellent results, with an average score of 97.7%. This indicates that the questions developed were easy for students to understand, both in terms of language, sentence structure, and presentation. These results align with research. Faradisa et al. (2024) which emphasizes that the readability of the instrument is an important factor so that students can understand the context of the questions without language barriers.

Compared with previous research, the results of this study consistently demonstrate that the Android-based instrument offers ease of access, effectiveness in question presentation, and good reliability. The strength of this study lies in the use of the local context of the Bengkulu coastline, which not only strengthens the relevance of the questions to students' experiences but also increases student engagement in solving them. Thus, this test instrument is not only valid and reliable, but also contextual and applicable to support the preparation of the Minimum Competency Assessment.

Practicality of the instrumentandroid-based numeracy literacy test with a coastal context in Bengkulu

Based on the research results, the practicality test of the Android application for the numeracy literacy test using the use questionnaire instrument showed that the application obtained an overall score of 84.7% with a very practical category. When viewed from each aspect, the usefulness aspect obtained a score of 73.4% with a practical category, while the ease of use, ease of learning, and user satisfaction aspects obtained a score above 88% with a very practical category. This shows that the developed iSpring Suite-based Android application is able to provide a good user experience for students, is easy to understand, and is in accordance with the needs of numeracy literacy-based learning.

This finding is in line with the opinion Rukajat (2018)that practical learning instruments must be easy to use, understand, and not make it difficult for students to implement them. In addition, the results of this study are consistent with the findingsPutra and Tanamal (2020)which states that the use of a use questionnaire is effective in measuring the level of practicality of digital-based learning media, where the aspects of usability, ease, and satisfaction are important indicators of the media's success.

The results of this study are also in line with studiesAbuna et al. (2025) who developed an Android-based assessment instrument and found that ease of use and student satisfaction were key factors supporting the effectiveness of implementing digital applications in learning. Similarly, research Wirawan et al. (2024) regarding numeracy literacy instruments in other local contexts shows that practical instruments not only make it easier for students to work on problems, but also increase their engagement and motivation to learn.

Interviews with teachers indicated that the application is efficient and easy to use. This application is economical because teachers no longer need to prepare a lot of paper for exams. Furthermore, this application is easy for teachers to use for corrections because the question corrections are done automatically by the application. Teachers no longer need to correct students' exam results one by one; they only need to check their email because the learning results are sent directly to the teacher's email, complete with the correct or incorrect results for each question. This makes it very easy for teachers to know students' grades directly. In addition to the immediate learning results, the use of this application also allows teachers to identify students' abilities based on the results of the tests they have taken. Based on the results of questionnaires and interviews with teachers and students, the results obtained were obtained that the developed application has met the criteria of practicality.

Thus, it can be concluded that the iSpring Suite Android-based numeracy literacy test instrument, applied to the coastal context of Bengkulu, has met the criteria for practicality. This instrument is not only relevant in terms of content but also excels in terms of ease of use, ease of learning, and student satisfaction. This practicality is an important indicator that the instrument can be widely used in preparation for the Minimum Competency Assessment and makes a positive contribution to strengthening students' numeracy literacy.

Conclusion

The Android-based numeracy literacy test developed with iSpring Suite and embedded in the Bengkulu coastal heritage context meets theoretical feasibility requirements. Expert validation confirmed that the instrument satisfied validity criteria in terms of content, construct, and language, and was judged suitable to support learning activities and preparation for the Minimum Competency Assessment. The instrument also demonstrated strong empirical quality at the item level. The ten developed items showed acceptable to strong item validity, high test reliability (0.781), difficulty indices ranging from easy to moderate, and discrimination indices from moderate to very good, with distractors functioning according to established criteria. Readability was also very high (≥ 97.7%), indicating that the items can measure students' numeracy literacy consistently and fairly. In addition, the practicality evaluation using a user questionnaire yielded an overall score of 84.7% ("very practical"). Students reported very positive perceptions regarding usefulness, ease of use, ease of learning, and satisfaction, suggesting that the Android-based test is user-friendly, supports learning, and can increase students' motivation when taking numeracy assessments.

Acknowledgment

The authors would like to express their sincere gratitude to all individuals and institutions that contributed to the completion of this study. Special appreciation is extended to the Institute for Research and Community Service (LPPM), Universitas Bengkulu, for providing financial support through the PNBP research grant, which enabled the implementation of this research. The authors also thank the advisors and expert validators for their valuable feedback and

recommendations during the development and validation stages. Appreciation is further given to the participating teachers and students for their cooperation during the pilot testing and data collection process. Finally, the authors acknowledge all parties who provided technical assistance and administrative support that facilitated the completion of this work.

Declarations

Conflicts of Interest	:	The authors declare no conflict of interest.
Generative AI Statement	:	AI Used for Limited, Non-Substantive Support: Generative AI tools, such as Grammarly and Microsoft Copilot, were employed solely for language editing and minor phrasing enhancements. All conceptualization, analysis, and scholarly content were independently developed and verified by the authors.
Funding Statement	:	This work was supported by the Institute for Research and Community Service (LPPM), Universitas Bengkulu, through PNBP research funding [grant number 1806/UN30.15/PT/2025].
Author Contributions	:	**Elwan Stiadi:** Conceptualization, writing - original draft, editing, formal analysis, and methodology; **Ari Putra**: Collecting Data Writing - review & editing; **M. Zainudin**: Validation and supervision; **Meidy Putra Areka**: Validation and supervision; **Ahbi Mahdianing Rum**: Validation and supervision; **Nader U-Sin**: Validation and supervision.

References

Abuna, G., Mulyanto, A., Ayu Ashari, S., Rohandi, M., Dwinanto, A., & Suwandi, ad. (2025). Analisis tingkat kepuasan siswa terhadap penggunaan media pembelajaran interaktif berbasis aplikasi android di SMK Negeri 1 Limboto [Analysis of students' satisfaction with the use of interactive learning media based on an Android application at SMK Negeri 1 Limboto]. *INVERTED: Journal of Information Technology Education*, *5*(2), 171–183.

Afifah, W., Edy, S., & Huda, S. (2021). Pengembangan alat evaluasi pembelajaran matematika berbasis two tier multiple choice berlevel HOTS menggunakan ispring suite 9 [Development of a mathematics learning evaluation instrument based on two-tier multiple choice with HOTS level using ispring suite 9]. *Postulat : Jurnal Inovasi Pendidikan Matematika*, *2*(2), 175–189. https://doi.org/10.30587/postulat.v2i2.3835

Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI [Analysis of the validity and reliability of odd-semester final summative test items in Islamic Education]. *EDUKATIF : Jurnal Ilmu Pendidikan*, *6*(1), 964–975. https://doi.org/10.31004/edukatif.v6i1.5931

Cahyadi, R. A. H. (2019). Pengembangan bahan ajar berbasis ADDIE model [Development of teaching materials based on the ADDIE model]. *Halaqa: Islamic Education Journal*, *3*(1), 35–42. https://doi.org/10.21070/HALAQA.V3I1.2124

Cahyanti, A. D., Farida, F., & Rakhmawati, R. (2019). Pengembangan alat evaluasi berupa tes online/offline matematika dengan ispring suite 8 [Development of an online/offline

mathematics test evaluation tool using ispring suite 8]. *Indonesian Journal of Science and Mathematics Education*, *2*(3), 363–371. https://doi.org/10.24042/ijsme.v2i3.4362

Delima, N., Kurniasih, I., Tohari, Hutneriana, R., Amalia, F. N., & Arumanegara, E. (2022). *PISA dan AKM literasi matematika dan kompetensi numerasi* [*PISA and AKM: Mathematical literacy and numeracy competence*](Issue June).

Erfan, M., Archi Maulyda, M., Hidayati, V. R., Astria, F. P., Ratu, T., Studi, P., Guru, P., & Dasar, S. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model RASCH [Analysis of the quality of items on distinguishing series and parallel circuits using classical test theory and the RASCH model]. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19.

Faradisa, B. T. Z. V., Kurniasih, S., & Berlian, L. (2024). Pengembangan instrumen tes 4TMC CBT pada materi sistem pernapasan untuk mengukur berpikir kritis siswa SMP kelas VIII [Development of a 4TMC CBT test instrument on the respiratory system to measure eighth-grade junior high school students' critical thinking]. *Jurnal Pendidikan MIPA*, *14*(4), 909–918. https://doi.org/10.37630/jpm.v14i4.2010

Farman, F., Anjelina, S., Putri, Q. T., Mardiah, N. A., & Sari, K. (2021). Pengembangan instrumen penilaian pembelajaran matematika berbasis ispring suite [Development of a mathematics learning assessment instrument based on ispring suite]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *10*(4), 2040–2052. https://doi.org/10.24127/ajpm.v10i4.4288

Farman, Sarimuddin, & Chairuddin. (2020). Meningkatkan kemampuan guru dalam penyusunan soal evaluasi berbasis komputer menggunakan wondershare quiz creator [Improving teachers' ability to develop computer-based evaluation items using Wondershare Quiz Creator]. *SNPKM: Seminar Nasional Pengabdian Kepada Masyarakat*, *2*, 21–27.

Han, W., Susanto, Dicky., Dewayani, Sofie., Pandora, Putri., Hanifah, Nur., Miftahussururi., Nento, M. Noorthertya., & Akbari, Q. Syahriana. (2017). Materi Pendukung Literasi Numerasi [Supporting Materials for Numeracy Literacy]. *Kementrian Pendidikan Dan Kebudayaan, Tim GLN Kemendikbud.*, *8*(9), 1–58.

Kamilati, N. (2018). Analisis komponen penilaian pada rencana pelaksanaan pembelajaran sebagai acuan pengembangan kurikulum diklat teknis sunstantif guru [Analysis of assessment components in lesson plans as a reference for curriculum development in substantive technical teacher training.]. *EDUKASI: Jurnal Penelitian Pendidikan Agama Dan Keagamaan*, *16*(1), 1–17. https://doi.org/10.32729/edukasi.v16i1.440

Kartini, K. S., & Putra, I. N. T. A. (2020). Pengaruh penggunaan media pembelajaran interaktif berbasis android terhadap hasil belajar siswa [The effect of using Android-based interactive learning media on students' learning outcomes]. *Jurnal Pendidikan Kimia Dan Ilmu Kimia*, *3*(2), 8–12. https://doi.org/10.33627/re.v3i2.417

Khofifah, S., & Purwati, I. (2020). The inventory coastal ecosystem in Sipora Mentawai island. Journal of Remote Sensing GIS & Technology, *6*(1), 19–24. https://doi.org/10.5281/zenodo.3742892

Klarita, E. N., & Syafiah, R. (2022). Analisis kemampuan literasi dan numerasi dalam menyelesaikan soal AKM siswa kelas V [Analysis of literacy and numeracy skills in solving AKM items among fifth-grade students]. *Jurnal Pendidikan Guru*, *3*(4), 262–267.

Kristie, S., Tawas, D. H., Tangkudung, H., & Mamoto, J. D. (2013). Analisis karakteristik gelombang pecah terhadap perubahan garis pantai di Atep Oki [Analysis of breaking-wave characteristics on shoreline changes in Atep Oki]. *Jurnal Sipil Statik*, *1*(12), 784–796.

Mahfud, H., Purnama Adi, F., Widianto Atmojo, I. R., & Ardiansyah, R. (2019). Peningkatan kompetensi evaluasi pembelajaran berbasis teknologi pada guru SD di kota Surakarta

[Improving technology-based learning evaluation competence among elementary school teachers in Surakarta City]. *Jurnal Pendidikan Dasar*, *7*(2), 1–5. https://doi.org/https://doi.org/10.20961/jpd.v7i2.35104

Marhogi, A., Haryaka, U., & Kurniawan, K. (2025). Pengembangan media pembelajaran berbasis android dengan bantuan i-spring pada materi matriks [Development of Android-based learning media with the assistance of iSpring on matrix topics]. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *9*(3), 1301–1311. https://doi.org/10.31004/cendekia.v9i3.4191

Mawarinda, I., Andayani, S., & Rahmawati, Y. (2022). Pengembangan media pembelajaran matematika powerpoint interaktif berbasis realistic mathematic education (RME) disertai game [Development of interactive PowerPoint-based mathematics learning media based on Realistic Mathematics Education (RME) with games]. *EMTEKA: Jurnal Pendidikan Matematika*, *3*(2), 219–227.

Naja, A. F., & Utami, M. A. P. (2025). Pengembangan instrumen diagnostik berformat four-tier multiple choice pada materi operasi aljabar [Development of a diagnostic instrument in four-tier multiple-choice format on algebraic operations]. *Suska Journal of Mathematics Education*, *11*(1), 9–22. https://doi.org/10.24014/sjme.v11i1.31306

Nieveen, N. (1999). *Design approaches and tools in education and training.* ICO Cluwer academic publisher.

Nitko, A. J. ., & Brookhart, S. M. . (2011). *Educational assessment of students* (6th ed.). Pearson.

Novita, N., Mellyzar, M., & Herizal, H. (2021). Asesmen nasional (AN): Pengetahuan dan persepsi calon guru [National assessment (AN): Preservice teachers' knowledge and perceptions]. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, *5*(1), 172–179. https://doi.org/10.58258/jisip.v5i1.1568

OECD. (2019). PISA 2018 Results : Combined executive summaries volume I, II & III. In *PISA 2009 at a Glance: Vol. I*. https://doi.org/10.1787/g222d18af-en

Peraturan Menteri Pendidikan Dan Kebudayaan Tentang Kurikulum 2013 Sekolah Menengah Atas/Madrasah Aliyah [Regulation of the Minister of Education and Culture on the 2013 Curriculum for Senior High Schools/Islamic Senior High Schools]., Kemdikbud 1 (2014). https://simpuh.kemenag.go.id/regulasi/permendikbud_59_14.pdf

Putra, Y. S. M., & Tanamal, R. (2020). Analisis usability menggunakan metode USE questionnaire pada website ciputra enterprise system [Usability analysis using the USE Questionnaire method on the Ciputra Enterprise System website]. *Teknika*, *9*(1), 58–65. https://doi.org/10.34148/teknika.v9i1.267

Rezky, M., Hidayanto, E., & Parta, I. N. (2022). Kemampuan literasi numerasi siswa dalam menyelesaikan soal konteks sosial budaya pada topik geometri jenjang SMP [Students' numeracy literacy in solving socio-cultural context problems on geometry at the junior high school level]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(2), 1548. https://doi.org/10.24127/ajpm.v11i2.4879

Rohim, D. C. (2021). Konsep asesmen kompetensi minimum untuk meningkatkan kemampuan literasi numerasi siswa sekolah dasar [The concept of the Minimum Competency Assessment to improve elementary students' numeracy literacy]. *Jurnal VARIDIKA*, *33*(1), 54–62. https://doi.org/10.23917/varidika.v33i1.14993

Rukajat, A. (2018). *Teknik evaluasi pembelajaran [Learning evaluation techniques [Learning evaluation techniques]*. Deepublish.

Salsabilla, A. C., Rahmawati, S. C., & Fauza, A. Z. N. (2025). Evaluasi kuantitatif butir soal pilihan ganda sebagai upaya peningkatan kualitas instrumen pembelajaran ipa pada materi usaha, energi, dan pesawat sederhana [Quantitative evaluation of multiple-choice items

as an effort to improve the quality of science learning instruments on the topics of work, energy, and simple machines]. *Sindoro: Cendikia Pendidikan, 15*(10), 1–21.

Stiadi, E., Fitriani, F., & Putra, A. (2023). Analisis kesalahan siswa terkait aspek communication, mathematising , dan representation pada unsur proses literasi matematis PISA [Analysis of students' errors related to the aspects of communication, mathematising, and representation in the PISA mathematical literacy process domain]. *JTMT: Journal Tadris Matematika*, *4*(1), 66–75. https://doi.org/10.47435/jtmt.v4i1.1755

Stiadi, E., Putra, A., Lestary, R., & Rahmasari, H. (2022). Analisis kemampuan siswa dalam menguasai komponen content ketika menyelesaikan soal literasi matematika PISA di SMPN 4. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *6*(3), 440–449. https://doi.org/10.33369/jp2ms.6.3.440-449

Suprapti, E. (2016). Pengembangan perangkat pembelajaran matematika model kooperatif tipe STAD dengan media powerpoint Ispring pada materi jajargenjang, layang-layang dan trapesium di kelas VII SMP [Development of mathematics learning tools using the STAD cooperative model with iSpring PowerPoint media on parallelogram, kite, and trapezoid topics in Grade VII junior high school]. *MUST: Journal of Mathematics Education, Science and Technology*, *1*(1), 57–68. https://doi.org/10.30651/must.v1i1.98

Wardhani, D. A. P., & Oktiningrum, W. (2022). Pengembangan soal AKM bermuatan ethnomatematika dengan media canva untuk mengukur kemampuan literasi numerasi siswa sekolah dasar [Development of AKM items with ethnomathematics content using Canva to measure elementary students' numeracy literacy ability]. *AKSIOMA Jurnal Program Studi Pendidikan Matematika*, *11*(4), 3860–3871. https://doi.org/10.24127/ajpm.v11i4.6241

Wijaya, A., & Dewayani, S. (2021). Framework asesmen kompetensi minimum (AKM) [Minimum Competency Assessment (AKM) framework]. *Badan Penelitian Dan Pengembangan Dan Perbukuan Kementerian Pendidikan Dan Kebudayaan*, 1–108.

Wirawan, N., Nindiasari, H., & Sukirwan, S. (2024). Pengembangan instrumen literasi numerasi konten budaya banten materi perbandingan untuk siswa SMP [Development of a numeracy literacy instrument with Banten cultural content on ratio topics for junior high school students]. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *8*(2), 1148–1161. https://doi.org/10.31004/cendekia.v8i2.2634

Zakaria, Hadiarti, D., & Fadhilah, R. (2017). Pengembangan instrumen evaluasi berbasis CBT dengan software ispring quizmaker pada materi kesetimbangan kimia [Development of a CBT-based evaluation instrument using iSpring QuizMaker software on chemical equilibrium topics]. *Jurnal Pendidikan Matematika Dan Sains*, *1866*(2), 178–183. https://doi.org/10.21831/JPMS.V5I2.16709

Zulfayani, Z., Ariawan, R., Nufus, H., Nurdin, E., & Fitriani, D. (2024). Pengembangan instrumen kemampuan literasi numerasi berbasis HOTS [Development of a HOTS-based numeracy literacy ability instrument]. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *8*(1), 789–800. https://doi.org/10.31004/cendekia.v8i1.2678