



Development and evaluation of a steam-based digital activity book and numeracy assessment for elementary mathematics learning

Ayu Made Juniwati
Wiratha

Universitas PGRI Adhi Buana,
INDONESIA

Hartono

Universitas PGRI Adhi
Buana, INDONESIA

Via Yustitia*

Universitas PGRI Adhi
Buana, INDONESIA

Article Info

Article history:

Received: June 07, 2026

Revised: August 30, 2026

Accepted: October 01, 2026

Keywords:

Digital Activity Book;
Elementary Mathematics
Education;
Numeracy Assessment;
Psychometric Evaluation;
STEAM-Based Learning.

Abstract

Background: Developing elementary students' numeracy skills requires innovative instructional resources supported by valid and reliable assessment instruments. Integrating Science, Technology, Engineering, Arts, and Mathematics (STEAM) into digital learning environments offers opportunities to promote interactive and contextualized mathematics learning.

Aims: This study aims to develop and evaluate a STEAM-based Digital Activity Book and examine the psychometric quality of its accompanying numeracy assessment instrument.

Method: A Research and Development (R&D) approach employing the ADDIE model was implemented. The development process involved instructional design, expert validation, and empirical assessment. Media feasibility was evaluated by four experts, while item validity and internal consistency were examined using Pearson Product-Moment correlation and Kuder-Richardson Formula 20 (KR-20), respectively.

Results: Expert evaluation yielded an average media feasibility score of 87.5%, indicating high validity. Of the 30 dichotomously scored multiple-choice items developed, 18 were reported as valid. The numeracy assessment demonstrated acceptable internal consistency, with a KR-20 coefficient of 0.78.

Conclusion: The findings provide preliminary evidence supporting the feasibility of the STEAM-based Digital Activity Book and the reliability of its accompanying numeracy assessment, establishing a foundation for further evaluation of mathematics learning in elementary education.

To cite this article: Wiratha, A. M. J., Hartono, & Yustitia, V. (2026). Development and evaluation of a steam-based digital activity book and numeracy assessment for elementary mathematics learning. *Journal of Advanced Sciences and Mathematics Education*, 6(4), 59-79.

INTRODUCTION

Numeracy has become a fundamental competency in twenty-first-century education, enabling individuals to interpret quantitative information, apply mathematical reasoning, and make informed decisions in various real-life situations. In elementary mathematics education, numeracy extends beyond memorizing mathematical formulas and performing routine calculations to encompass the ability to apply mathematical concepts, procedures, and tools to solve contextual problems (Adi & Winarso, 2026; Danuri, 2025; McNeil et al., 2025; Torres-Peña et al., 2025). Developing these competencies during primary education is essential because early mathematical understanding provides a foundation for higher-order thinking, problem-solving abilities, and subsequent academic achievement. Nevertheless, strengthening students' numeracy skills remains a significant educational challenge, particularly in learning environments where instructional practices emphasize procedural knowledge rather than conceptual understanding and meaningful mathematical applications (Yustitia et al., 2024). International assessments, including the Programme for International Student Assessment (PISA) and Trends in International Mathematics

*Corresponding author:

Via Yustitia, Universitas PGRI Adhi Buana, Indonesia
viayustitia@unipasby.ac.id ✉

and Science Study (TIMSS), have highlighted persistent challenges in mathematical achievement across educational systems (Gilleece et al., 2026; Karakolidis et al., 2021). Although these assessments involve different student populations and mathematical domains, their findings emphasize the importance of strengthening foundational mathematical competencies throughout formal education. In Indonesia, educational policies have increasingly emphasized competency-based mathematics learning and the integration of numeracy into authentic situations through the Minimum Competency Assessment framework. However, conventional instructional materials that rely heavily on textual explanations, repetitive exercises, and teacher-centered activities may provide insufficient opportunities for students to explore mathematical concepts through interactive and contextualized learning experiences (Tegegne et al., 2025). This concern is particularly relevant to elementary school students, whose mathematical understanding often benefits from concrete representations, visual illustrations, and developmentally appropriate learning activities (Hadi et al., 2025; Kaminski & Sloutsky, 2020; Purcar et al., 2024). Consequently, developing innovative instructional resources that connect mathematical knowledge with meaningful experiences is essential for supporting numeracy development and improving the quality of elementary mathematics education (Yustitia, Kusmaharti, et al., 2025).

The integration of Science, Technology, Engineering, Arts, and Mathematics (STEAM) provides an interdisciplinary instructional framework that connects mathematical concepts with scientific inquiry, technological applications, engineering design, and creative expression (Belbase et al., 2022). Rather than presenting mathematics as an isolated discipline, STEAM-oriented learning encourages students to investigate authentic problems, establish connections across subject areas, and construct solutions through meaningful learning activities (Belbase et al., 2022). Previous investigations have explored the implementation of STEAM-based approaches in mathematics education, particularly their potential contributions to mathematical literacy, problem-solving competencies, and interdisciplinary learning experiences (Sudirman et al., 2026; Supianti et al., 2025). Within elementary mathematics education, STEAM integration offers opportunities to contextualize numerical operations, measurement, geometric reasoning, and data interpretation through activities that reflect students' everyday experiences (Yustitia, Murti, et al., 2025). Scientific investigations can provide meaningful contexts for applying quantitative reasoning, while engineering activities encourage students to use mathematical concepts when designing, testing, and evaluating possible solutions. Technological applications further facilitate mathematical exploration by providing interactive representations, immediate feedback, and opportunities for individualized learning experiences (Bang et al., 2023; Cirneanu & Moldoveanu, 2024). Meanwhile, artistic elements can enrich mathematical learning by incorporating visual communication, creativity, and alternative representations of mathematical ideas into instructional activities. Integrating these interdisciplinary components into a Digital Activity Book offers opportunities to transform conventional learning materials into interactive resources that combine contextual mathematical problems with visual and technological features. Nevertheless, the development of STEAM-based digital instructional materials should extend beyond technological attractiveness and visual design to ensure that learning activities remain aligned with mathematical competencies and students' cognitive characteristics. Therefore, systematically designed Digital Activity Books incorporating STEAM principles may provide a relevant instructional alternative for supporting contextual, interactive, and developmentally appropriate numeracy learning in elementary education.

The educational potential of STEAM-based Digital Activity Books depends not only on the quality of instructional content and interactive features but also on the appropriateness of the assessment instruments used to evaluate students' numeracy competencies. Digital learning resources can facilitate mathematics instruction through structured activities, interactive exercises, and flexible learning environments, although their educational value requires systematic evaluation

(Cirneanu & Moldoveanu, 2024; Joshi et al., 2023). The Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model provides an instructional development framework for organizing the systematic design, production, implementation, and evaluation of educational materials. Within this framework, expert judgment plays an important role in examining the alignment of instructional content, media characteristics, and assessment indicators with intended learning objectives. Developing a numeracy assessment also requires a systematic process involving the formulation of a test blueprint, construction of assessment items, expert review, empirical testing, and evaluation of measurement characteristics (Chin et al., 2021; Turner et al., 2021). For elementary students, assessment items should represent relevant numeracy domains while considering linguistic clarity, cognitive development, and the contextual appropriateness of mathematical problems. Previous research has emphasized the importance of psychometric evidence in evaluating mathematics assessments, particularly regarding item characteristics, measurement validity, and reliability (Khanipour et al., 2025; Larrain et al., 2026; Outhwaite et al., 2024). Pearson Product–Moment correlation can be employed to examine relationships between individual item scores and total test scores, although corrected item–total correlations provide a more appropriate basis for evaluating item discrimination when the total score includes the item being examined. Furthermore, the Kuder–Richardson Formula 20 (KR-20) provides an estimate of internal consistency for dichotomously scored assessment instruments, offering evidence concerning the consistency of item responses within a particular measurement context. Accordingly, integrating expert evaluation, empirical item analysis, and reliability estimation into the development of STEAM-based Digital Activity Books may strengthen the methodological quality of instructional resources and their accompanying numeracy assessments.

Although research on STEAM-based mathematics learning and digital instructional resources has expanded, the systematic integration of instructional media development with empirical evaluation of assessment quality remains an important methodological consideration. Previous studies have examined STEAM-oriented instruction through interdisciplinary learning activities, project-based approaches, and technology-supported mathematics environments (Belbase et al., 2022; Sudirman et al., 2026; Supianti et al., 2025). Other investigations have explored digital mathematics applications, adaptive learning technologies, and the integration of digital resources into instructional and assessment practices (Bang et al., 2023; Cirneanu & Moldoveanu, 2024; Joshi et al., 2023). However, evaluating the feasibility, attractiveness, and technological characteristics of instructional media does not necessarily establish the validity and reliability of the assessment instruments used to measure students' mathematical competencies. Research on mathematics assessment has emphasized the importance of examining psychometric properties, including item characteristics, measurement validity, and internal consistency, to support meaningful interpretations of students' mathematical performance (Khanipour et al., 2025; Larrain et al., 2026; Outhwaite et al., 2024). Nevertheless, the extent to which these psychometric evaluation procedures have been systematically incorporated into the development of STEAM-based Digital Activity Books for elementary numeracy learning remains insufficiently established. This methodological concern is particularly relevant because claims regarding instructional effectiveness may be difficult to interpret when the accompanying assessment instruments lack adequate empirical evidence of measurement quality. Furthermore, combining instructional development procedures with item-level analysis and internal consistency estimation may provide a more transparent approach to evaluating digital mathematics learning resources and their associated assessments. Addressing this concern, the present study integrates the development and expert evaluation of a STEAM-based Digital Activity Book with empirical item analysis using Pearson Product–Moment correlation and reliability estimation through the Kuder–Richardson Formula 20 (KR-20). The principal contribution of this study lies in connecting STEAM-based digital instructional development with initial

psychometric evaluation of a numeracy assessment within a unified research and development framework, thereby providing preliminary evidence concerning media feasibility and assessment reliability as a foundation for subsequent investigations of instructional effectiveness.

This study aims to develop and evaluate a STEAM-based Digital Activity Book and examine the psychometric characteristics of its accompanying numeracy assessment instrument for elementary mathematics learning. The research employs a Research and Development (R&D) approach guided by the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) instructional design framework. First, the study seeks to develop an interactive Digital Activity Book that integrates science, technology, engineering, arts, and mathematics into contextualized learning activities aligned with elementary numeracy competencies. The instructional resource is designed to facilitate mathematical exploration through visual representations, structured exercises, and interdisciplinary problem-solving activities appropriate to students' developmental characteristics. Second, the study aims to evaluate the feasibility of the developed Digital Activity Book through expert judgment focusing on instructional suitability, visual design, navigation, technological functionality, and digital learning features. Third, the research seeks to construct and evaluate a numeracy assessment instrument through systematic test blueprint development, item construction, expert review, and empirical testing. The empirical characteristics of the assessment items are examined using Pearson Product-Moment correlation to identify items demonstrating adequate relationships with overall test performance. Subsequently, the internal consistency of the dichotomously scored numeracy assessment is estimated using the Kuder-Richardson Formula 20 (KR-20). Through these procedures, the study seeks to establish preliminary evidence concerning the feasibility of the STEAM-based instructional resource and the measurement quality of its accompanying numeracy assessment. Ultimately, this research is expected to contribute to the systematic development and evaluation of digital mathematics learning resources while providing a methodological foundation for future studies investigating their practicality, broader psychometric validity, and effectiveness in supporting elementary students' numeracy development.

LITERATURE REVIEW

Numeracy is a fundamental mathematical competency that enables students to interpret quantitative information, apply mathematical reasoning, and solve practical problems across diverse real-life situations. Unlike conventional mathematics instruction that frequently emphasizes procedural calculations and abstract representations, numeracy focuses on the meaningful application of mathematical concepts, numerical relationships, spatial reasoning, measurement, and data interpretation in contextual situations (Rosita et al., 2026; Sakurai & Goos, 2023; Torres-Peña et al., 2025). Developing numeracy competencies during elementary education is particularly important because early mathematical understanding provides a foundation for subsequent mathematical achievement, critical thinking, and problem-solving abilities (Chen, 2025; Liu et al., 2025; Onoshakpokaiye, 2023; Torres-Peña et al., 2025). From a cognitive developmental perspective, elementary school students generally benefit from learning experiences involving concrete objects, visual representations, and meaningful interactions that support the transition toward abstract mathematical reasoning. Consequently, mathematical concepts introduced through authentic situations and manipulable learning materials may facilitate students' understanding of numerical relationships and problem-solving procedures. Nevertheless, elementary mathematics instruction continues to face challenges when learning activities rely predominantly on repetitive exercises, memorization, and teacher-centered explanations without sufficient opportunities for conceptual exploration. International assessments have highlighted persistent challenges in mathematical achievement, reinforcing the importance of strengthening foundational numeracy competencies

through appropriate instructional approaches (Torres-Peña et al., 2025; Wu, 2025). Contextualized mathematics learning provides opportunities for students to connect numerical concepts with familiar experiences, interpret quantitative information, and construct meaningful solutions to practical problems. Furthermore, instructional resources that integrate visual representations, interactive activities, and developmentally appropriate mathematical challenges may support students' engagement and conceptual understanding. Therefore, elementary numeracy education requires systematic instructional innovations that accommodate students' cognitive characteristics while promoting meaningful mathematical reasoning and the application of mathematical knowledge in everyday contexts.

Science, Technology, Engineering, Arts, and Mathematics (STEAM) represents an interdisciplinary educational approach that integrates scientific inquiry, technological applications, engineering design, artistic expression, and mathematical reasoning into coherent learning experiences (Belbase et al., 2022). Rather than treating mathematics as an isolated subject, STEAM-oriented instruction encourages students to investigate authentic problems, establish interdisciplinary connections, and develop solutions through collaborative exploration and creative problem-solving activities (Sudirman et al., 2026; Supianti et al., 2025). Within this framework, science provides opportunities for students to investigate observable phenomena and interpret quantitative evidence, while technology facilitates mathematical exploration through interactive digital tools and visual representations. Engineering introduces design-oriented challenges that require students to apply mathematical concepts when planning, constructing, testing, and evaluating possible solutions. Meanwhile, arts contributes creative expression, visual communication, and aesthetic representation, allowing mathematical ideas to be explored through multiple forms of learning activities. Mathematics serves as the conceptual foundation connecting these interdisciplinary experiences through numerical operations, measurement, geometric reasoning, estimation, and data interpretation. Previous investigations have examined the educational potential of STEAM-based approaches for supporting mathematical literacy, interdisciplinary understanding, and problem-solving competencies (Belbase et al., 2022; Sudirman et al., 2026; Supianti et al., 2025). Integrating these disciplinary components may provide elementary students with opportunities to experience mathematics as a practical and meaningful discipline rather than a collection of isolated procedures. However, the educational contribution of STEAM depends on the alignment between interdisciplinary activities, intended mathematical competencies, and appropriate instructional assessment procedures. Accordingly, STEAM-based mathematics instruction requires carefully structured learning experiences that maintain mathematical coherence while encouraging creativity, inquiry, and contextual problem-solving.

Digital Activity Books represent technology-supported instructional resources that combine structured learning materials, interactive exercises, and multimedia elements within an accessible digital environment. Unlike conventional printed textbooks or static digital documents, interactive activity books can incorporate features such as drag-and-drop exercises, direct numerical input, animated illustrations, and immediate feedback to facilitate active participation in learning activities. These technological features may support elementary students by transforming abstract mathematical concepts into visual and manipulable representations that can be explored through direct interaction. Previous research has investigated digital mathematics applications and technology-enhanced instructional environments, highlighting their potential to support personalized learning, mathematical exploration, and instructional assessment (Bang et al., 2023; Cirneanu & Moldoveanu, 2024; Joshi et al., 2023). Interactive digital activities may also provide opportunities for students to repeat mathematical exercises, receive feedback, and regulate their learning progression according to individual needs. When integrated with STEAM principles, Digital Activity Books can present mathematical challenges through scientific observations, engineering

simulations, creative design activities, and contextual numerical problems (Chai et al., 2020; Juškevičienė et al., 2021; Lavicza et al., 2022; Liston et al., 2022). However, the incorporation of multimedia and interactive features does not automatically guarantee meaningful learning because instructional effectiveness depends on pedagogical alignment, cognitive appropriateness, and the quality of learning activities. The Analysis, Design, Development, Implementation, and Evaluation (ADDIE) framework provides a systematic instructional development process that can guide the identification of learning needs, formulation of instructional objectives, development of digital materials, and evaluation of educational resources. Within this framework, expert evaluation can provide preliminary evidence concerning the suitability of instructional content, visual presentation, navigation, technological functionality, and alignment with elementary numeracy competencies. Therefore, the development of STEAM-based Digital Activity Books should integrate interdisciplinary instructional design, developmentally appropriate interactive features, and systematic evaluation procedures to establish their feasibility as mathematics learning resources.

The development of digital instructional materials should be accompanied by systematic assessment procedures because the quality of educational measurement influences the interpretation of students' mathematical competencies. Numeracy assessment instruments are designed to evaluate students' abilities to apply mathematical concepts, interpret quantitative information, and solve contextual problems through tasks aligned with intended learning outcomes (Demosthenous et al., 2021; Rosyidi et al., 2024). A systematic assessment development process generally includes the construction of a test blueprint, formulation of competency indicators, development of assessment items, expert review, and empirical evaluation of item characteristics. In elementary mathematics education, assessment items should also consider students' cognitive development, linguistic comprehension, and familiarity with contextual mathematical situations. Previous investigations have emphasized the importance of psychometric evaluation in mathematics assessment, particularly regarding item characteristics, measurement validity, and reliability (Khanipour et al., 2025; Larrain et al., 2026; Outhwaite et al., 2024). Expert judgment can contribute content-related validity evidence by examining whether assessment items adequately represent the intended numeracy competencies and are appropriate for the targeted student population. Empirical item analysis using Pearson Product-Moment correlation can provide information concerning relationships between individual item scores and overall test performance, although corrected item-total correlations are preferable when examining item discrimination. For dichotomously scored instruments, in which correct responses receive a score of one and incorrect responses receive a score of zero, the Kuder-Richardson Formula 20 (KR-20) provides an estimate of internal consistency based on item response variability and total test score variance. Nevertheless, evidence of internal consistency alone does not establish comprehensive instrument validity, eliminate potential measurement bias, or demonstrate the effectiveness of an instructional intervention. Consequently, integrating content evaluation, empirical item analysis, and reliability estimation provides a more systematic foundation for assessing the measurement quality of numeracy instruments accompanying STEAM-based digital instructional resources.

The conceptual framework of this study is grounded in the relationship between elementary numeracy development, interdisciplinary STEAM instruction, interactive digital learning resources, and systematic educational assessment. Challenges associated with conventional mathematics instruction indicate the importance of learning environments that provide contextual mathematical experiences, visual representations, and opportunities for active student participation. STEAM-based learning addresses these instructional considerations by connecting mathematical concepts with scientific investigation, technological exploration, engineering design, and creative representation (Belbase et al., 2022; Supianti et al., 2025). Meanwhile, Digital Activity Books provide a technological environment through which interdisciplinary mathematical activities can be organized into

interactive exercises, contextual challenges, and structured learning sequences. The integration of these components is theoretically expected to support meaningful mathematical exploration by allowing students to interact with numerical concepts through visual, contextual, and problem-oriented learning experiences. However, the development of an instructional resource and the evaluation of students' numeracy competencies represent distinct methodological processes that require appropriate procedures and evidence. Within the ADDIE framework, the development and expert evaluation of a STEAM-based Digital Activity Book can provide preliminary information concerning instructional feasibility and the alignment of learning activities with intended numeracy competencies. Complementarily, empirical item analysis and KR-20 reliability estimation can provide initial evidence concerning the measurement characteristics and internal consistency of the accompanying numeracy assessment instrument. This integrated framework distinguishes instructional media feasibility from assessment quality while recognizing that both dimensions contribute to the systematic evaluation of digital mathematics learning resources. Accordingly, the present study develops and evaluates a STEAM-based Digital Activity Book alongside its accompanying numeracy assessment, establishing a methodological foundation for subsequent investigations of practicality and effectiveness in improving elementary students' numeracy competencies.

METHOD

Research Design

This study employed a Research and Development (R&D) approach using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) instructional design model to develop and evaluate a STEAM-based Digital Activity Book and its accompanying numeracy assessment instrument for elementary mathematics learning. The ADDIE framework was selected because it provides a systematic structure for identifying instructional needs, designing learning resources, developing educational products, implementing instructional activities, and evaluating their quality throughout the development process. The research focused on two interconnected components: the development and feasibility evaluation of a STEAM-based Digital Activity Book and the construction and initial psychometric evaluation of a numeracy assessment instrument. The Digital Activity Book was designed to integrate science, technology, engineering, arts, and mathematics into contextualized activities aligned with elementary numeracy competencies. Meanwhile, the assessment development component involved constructing a test blueprint, developing dichotomously scored multiple-choice items, conducting expert reviews, implementing empirical item trials, and estimating instrument reliability. The feasibility of the instructional media was evaluated through expert judgment, whereas the empirical characteristics of the numeracy assessment were examined using Pearson Product-Moment correlation and the Kuder-Richardson Formula 20 (KR-20). These evaluation procedures were intended to provide preliminary evidence concerning the instructional suitability of the developed media and the measurement quality of its accompanying assessment. Accordingly, the research design combined systematic instructional development with initial psychometric evaluation rather than relying exclusively on the visual or technological characteristics of the digital product.

Participant and Research Setting

The study was conducted at SDN Gelam 2 during the first semester of the 2026/2027 academic year. The research participants consisted of second-grade elementary school students who had previously received instruction in foundational numeracy concepts relevant to the assessment content. These students represented the intended users of the STEAM-based Digital Activity Book and the target population for the accompanying numeracy assessment instrument. The empirical

assessment trial involved students at the same educational level as the intended users, allowing the researchers to examine item performance within an age-appropriate learning context. Students participating in the assessment trial were selected from a group outside the intended instructional implementation sample to reduce overlap between preliminary instrument evaluation and subsequent product implementation. In addition, expert evaluators were involved in reviewing the instructional media and assessment materials, particularly their content relevance, instructional suitability, construction, and linguistic appropriateness. The involvement of elementary students and expert evaluators enabled the research to obtain complementary information concerning the characteristics of the instructional resource and the accompanying assessment instrument. The specific numbers of participating students and expert evaluators, together with the sampling procedures, were documented as part of the research implementation and should be reported consistently with the empirical dataset.

Research Instrument and Data Collection

Data were collected using expert evaluation instruments, a STEAM-based Digital Activity Book, and a numeracy assessment instrument developed for second-grade elementary school students. The expert evaluation instruments were designed to examine the feasibility of the Digital Activity Book based on relevant instructional and media characteristics, including content suitability, visual presentation, navigation, technological functionality, and alignment with learning objectives. The numeracy assessment was constructed through a systematic process beginning with the formulation of a test blueprint based on elementary numeracy competencies and their integration with STEAM-oriented learning activities. The blueprint served as a reference for developing assessment indicators and ensuring that the test items represented the intended mathematical content and cognitive demands. Multiple-choice items were subsequently developed using a dichotomous scoring system in which correct responses received a score of one and incorrect responses received a score of zero. The initial assessment items were reviewed by relevant experts in elementary mathematics education and educational assessment to examine their content relevance, construction, linguistic clarity, and appropriateness for the targeted students. Feedback from the expert review was used to identify items requiring revision before empirical testing. The revised assessment was then administered to second-grade students who had previously studied the relevant foundational numeracy material. Students' responses were recorded and converted into item-level scores to establish a dataset for empirical item analysis and reliability estimation. The resulting data were used to evaluate the characteristics of individual assessment items and the internal consistency of the numeracy instrument.

Data Analysis

Data analysis comprised descriptive evaluation of instructional media feasibility and quantitative examination of the psychometric characteristics of the numeracy assessment instrument. Expert evaluation scores were summarized descriptively to determine the feasibility of the developed Digital Activity Book according to the evaluation criteria adopted in the study. The feasibility percentage was calculated by dividing the total observed score by the maximum possible score and multiplying the result by 100%. For the numeracy assessment, empirical item analysis was conducted using Pearson Product-Moment correlation to examine the relationship between individual item scores and total assessment scores. The correlation coefficient was calculated using the following equation:

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

where r_{xy} represents the correlation coefficient between an individual item and the total test score, N denotes the number of respondents, X represents the individual item score, and Y represents the total assessment score. The resulting correlation coefficients were examined to identify items demonstrating adequate relationships with overall test performance according to the statistical decision criteria adopted in the study. Items that did not satisfy the specified criteria were identified for further consideration, including revision or exclusion from the assessment instrument. Because the numeracy assessment employed dichotomous scoring, its internal consistency was estimated using the Kuder–Richardson Formula 20 (KR-20), expressed as follows:

$$r_{11} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{\sigma_t^2} \right) \dots\dots\dots (2)$$

where r_{11} represents the KR-20 reliability coefficient, k denotes the number of assessment items, p_i represents the proportion of students answering item i correctly, $q_i=1-p_i$ represents the proportion answering incorrectly, and σ_t^2 denotes the variance of total assessment scores. The KR-20 coefficient was interpreted as an estimate of the internal consistency of the dichotomously scored numeracy assessment within the observed sample. Higher coefficients generally indicate greater internal consistency, although their interpretation depends on the intended assessment purpose, item characteristics, and sample composition. The results of item analysis and reliability estimation were used to provide preliminary evidence concerning the measurement quality of the developed numeracy assessment. These analyses were not interpreted as sufficient evidence of comprehensive instrument validity or as direct proof of the instructional effectiveness of the Digital Activity Book.

Research Procedure

The research procedure followed the five interconnected stages of the ADDIE model, beginning with an analysis of instructional needs and concluding with an evaluation of the developed instructional resource and its accompanying numeracy assessment. During the Analysis stage, the researchers identified elementary students' numeracy learning needs, examined relevant mathematical competencies, and considered the suitability of STEAM integration for contextualized mathematics instruction. The Design stage involved preparing the instructional structure of the Digital Activity Book, formulating STEAM-oriented learning activities, developing assessment indicators, and constructing a numeracy test blueprint aligned with the intended competencies. During the Development stage, the Digital Activity Book was produced by integrating mathematical content with interactive digital features, visual representations, and interdisciplinary learning activities. At this stage, the initial numeracy assessment items were also constructed and submitted for expert review to examine their content relevance, construction, and linguistic appropriateness. Feedback obtained from expert evaluation was used to refine the instructional media and revise assessment items before empirical testing. The Implementation stage involved administering the revised numeracy assessment to second-grade students and introducing the developed Digital Activity Book within the intended elementary mathematics learning context. Students' responses to the assessment items were collected and scored dichotomously to provide empirical data for evaluating item characteristics and internal consistency. Finally, the Evaluation stage involved analyzing expert feasibility ratings, calculating item–total correlations, estimating KR-20 reliability, and identifying aspects of the instructional resource and assessment instrument requiring further refinement. Through these interconnected procedures, the study sought to establish preliminary evidence concerning the feasibility of the STEAM-based Digital Activity Book and the measurement quality of its accompanying numeracy assessment.

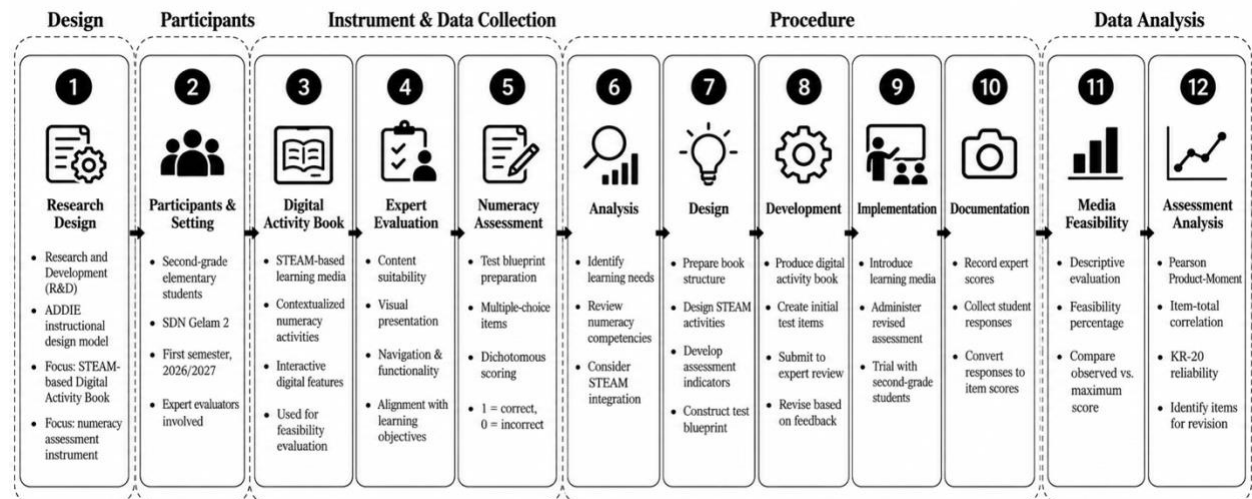


Figure 1. Research Method Flow

RESULTS AND DISCUSSION

Results

Development of the STEAM-Based Digital Activity Book and Expert Evaluation

The development process produced a STEAM-based Digital Activity Book designed to support contextualized numeracy learning among second-grade elementary school students. The instructional resource was developed through the ADDIE framework, integrating science, technology, engineering, arts, and mathematics into structured digital learning activities. The Digital Activity Book incorporated mathematical content and interactive features intended to facilitate students' exploration of numerical concepts through visually supported and contextually relevant exercises. Following the development stage, the prototype underwent expert evaluation to examine its suitability as an instructional resource for elementary mathematics learning. Four media experts independently assessed the Digital Activity Book using the evaluation instrument developed for this study. The evaluation focused on the feasibility of the digital instructional resource, including its visual presentation and technological characteristics. The results indicated an overall mean feasibility score of **87.50%**, suggesting that the developed media demonstrated favorable characteristics according to the expert assessment. The evaluation results are summarized in Table 1.

Table 1. Expert Evaluation Results of the STEAM-Based Digital Activity Book

Evaluation component	Result
Instructional resource	STEAM-Based Digital Activity Book
Number of media experts	4
Evaluation approach	Expert judgment
Mean feasibility percentage	87.50%
Evaluation outcome	Favorable media feasibility

In addition to the quantitative feasibility assessment, the experts provided recommendations for improving the accessibility and visual presentation of the instructional resource. One identified limitation concerned the relatively small instructional text displayed on Android smartphones with smaller screen dimensions. The experts indicated that the text size could reduce readability for elementary school students, particularly when instructional activities were accessed through mobile devices. In response to this feedback, the Digital Activity Book was revised by adjusting its

typography and applying responsive design principles to improve text visibility across different screen sizes. These modifications were intended to enhance the readability of instructional directions without substantially altering the mathematical content or structure of the learning activities. The revision process demonstrated how expert feedback informed the refinement of the digital instructional resource before classroom implementation. Overall, the expert evaluation provided preliminary evidence supporting the feasibility of the developed Digital Activity Book, while the qualitative recommendations contributed to improving its visual accessibility and usability.

Empirical Evaluation of the Numeracy Assessment Instrument

Following the development of the instructional resource, the accompanying numeracy assessment instrument underwent empirical evaluation to examine its item characteristics and internal consistency. The assessment was constructed using multiple-choice items aligned with the intended elementary numeracy competencies and the contextual learning activities incorporated into the STEAM-based Digital Activity Book. Before empirical testing, the assessment items were reviewed by educational evaluation experts to examine their content relevance and suitability for second-grade elementary school students. After the expert review and subsequent revisions, 18 multiple-choice items were administered to 25 second-grade students who had previously studied the relevant foundational numeracy material. Each item was scored dichotomously, with one point assigned to a correct response and zero points assigned to an incorrect response. The resulting student responses were organized into an item-level dataset for empirical analysis. Two complementary procedures were applied to evaluate the instrument: Pearson Product–Moment correlation for examining item–total relationships and the Kuder–Richardson Formula 20 (KR-20) for estimating internal consistency. The findings from these procedures are presented in the following sections.

Item Validity Analysis Using Pearson Product–Moment Correlation

The empirical validity of the numeracy assessment items was examined using Pearson Product–Moment correlation at a significance level of $\alpha = 0.05$. The analysis was conducted to evaluate the relationships between individual item scores and students' overall assessment performance. As presented in Table 2, the available item-validity summary identified 12 items (80.00%) that satisfied the established validity criterion, while three items (20.00%) did not meet the required threshold. The valid items comprised Items 1, 2, 3, 4, 5, 7, 8, 10, 11, 13, 14, and 15, whereas Items 6, 9, and 12 were classified as invalid. Items demonstrating adequate item–total correlations were retained for subsequent assessment procedures, while items failing to satisfy the criterion were excluded from the instrument. These findings indicate that most items included in the reported validity summary demonstrated acceptable relationships with overall test performance according to the adopted statistical criterion. Nevertheless, the identification of three invalid items suggests that certain assessment questions require further consideration regarding their construction and suitability for the intended numeracy competencies. The exclusion of these items resulted in a set of 12 retained assessment items based on the available summary. These retained items formed the basis for the subsequent reliability evaluation, provided that the KR-20 calculation was conducted using the same final item set. Overall, the item-validity analysis provided preliminary empirical evidence for selecting assessment items intended to measure elementary students' numeracy competencies.

Table 2. Distribution of Item Validity Results for the Numeracy Assessment

Item Validity Status	Item Numbers	Number of Items	Percentage (%)	Follow-up Decision
Valid ($r_{xy} > r_{tabel}$)	1, 2, 3, 4, 5, 7, 8, 10, 11, 13, 14, 15	12	80.00	Retained for the numeracy assessment
Invalid ($r_{xy} < r_{tabel}$)	6, 9, 12	3	20.00	Excluded from the assessment instrument
Total	Items 1–15	15	100.00	—

The preliminary interpretation of the item analysis indicated that some assessment items required further consideration because their response patterns did not adequately correspond with students' overall test performance. The initial evaluation identified potential concerns related to distractor construction and the cognitive demands of selected mathematical problems. In particular, some distractors were considered potentially ambiguous, while certain items appeared to present levels of difficulty that might not have been appropriate for the targeted second-grade students. These observations suggested that the linguistic construction and cognitive complexity of selected assessment items required additional review. However, the precise reasons for inadequate item performance could not be established from item–total correlation coefficients alone. Additional analyses of item difficulty, distractor functioning, and students' response patterns would be necessary to determine whether these characteristics contributed to the observed results. Accordingly, items demonstrating inadequate statistical performance were identified as candidates for revision or exclusion to improve the quality of the numeracy assessment.

Internal Consistency Reliability Analysis Using KR-20

Following the empirical item analysis, the internal consistency of the numeracy assessment instrument was examined using the Kuder–Richardson Formula 20 (KR-20). This procedure was selected because the assessment consisted of dichotomously scored multiple-choice items, with correct responses assigned a score of one and incorrect responses assigned a score of zero. The KR-20 analysis considered the number of assessment items, the proportions of correct and incorrect responses, and the variance of students' total assessment scores. Based on the reported reliability analysis, the instrument obtained a **KR-20 coefficient of 0.78**, indicating a relatively satisfactory level of internal consistency for an initial assessment evaluation. The reliability result is summarized in Table 3.

Table 3. Internal Consistency Reliability of the Numeracy Assessment Instrument

Reliability component	Result
Assessment instrument	Elementary numeracy test
Scoring system	Dichotomous (1 = correct; 0 = incorrect)
Reliability estimation method	Kuder–Richardson Formula 20 (KR-20)
Reliability coefficient	0.78
Interpretation	Satisfactory preliminary internal consistency

The obtained KR-20 coefficient suggested that the assessment items demonstrated a reasonably consistent pattern of responses within the observed student sample. This finding provided preliminary reliability evidence for the numeracy assessment developed to accompany the STEAM-based Digital Activity Book. Nevertheless, the coefficient should be interpreted in relation to the number of retained items, the characteristics of the participating students, and the intended purpose of the assessment. The relatively limited sample of 25 students also indicates that the reliability estimate may be sensitive to sampling variability. Furthermore, internal consistency represents only one aspect of measurement quality and does not independently establish comprehensive validity or the absence of measurement bias. Therefore, the reliability findings

should be interpreted as initial psychometric evidence rather than definitive confirmation that the instrument is standardized for broader educational use. Further evaluation involving larger and more diverse student samples would be valuable for examining the stability and generalizability of the reliability estimate.

Summary of Development and Evaluation Results

The research produced two interconnected outcomes: a STEAM-based Digital Activity Book for elementary mathematics learning and a numeracy assessment instrument subjected to initial empirical evaluation. Expert assessment involving four media evaluators resulted in an overall feasibility percentage of 87.50%, indicating favorable characteristics of the developed instructional resource. Qualitative feedback identified the need to improve the readability of instructional text on smaller Android smartphone screens, leading to revisions in typography and responsive visual design. The accompanying numeracy assessment underwent an empirical trial involving 25 second-grade students who completed 18 dichotomously scored multiple-choice items. Pearson Product-Moment correlation was used to examine the relationships between individual item scores and overall test performance, while the KR-20 procedure was applied to estimate internal consistency. The reported KR-20 coefficient of 0.78 provided preliminary evidence of satisfactory internal consistency within the observed sample. However, the final number of retained assessment items and the complete item-level correlation results require confirmation before definitive conclusions can be drawn regarding item performance. Collectively, these findings provide initial evidence concerning the feasibility of the STEAM-based instructional resource and the reliability of its accompanying numeracy assessment. The results do not establish that the Digital Activity Book improved students' numeracy achievement because comparative learning-outcome data were not presented. Nevertheless, the combined development and evaluation procedures provide a methodological foundation for subsequent investigations of instructional practicality, assessment validity, and classroom effectiveness.

Discussion

The findings of this study provide preliminary evidence concerning the feasibility of a STEAM-based Digital Activity Book as an instructional resource for elementary mathematics learning. The expert evaluation conducted by four media specialists resulted in an overall feasibility score of 87.50%, indicating that the developed instructional resource demonstrated favorable characteristics according to the evaluation criteria adopted in the study. This finding suggests that integrating interdisciplinary STEAM activities with interactive digital features can produce instructional materials that are considered suitable for further implementation in elementary numeracy education. From a theoretical perspective, this result is consistent with Belbase et al. (2022), who emphasized the importance of interdisciplinary curriculum integration, pedagogical processes, and assessment considerations in the development of STEAM education. The conceptual framework discussed by Belbase et al. (2022) highlights that meaningful STEAM integration requires coherent relationships among disciplinary knowledge, instructional objectives, and learning activities rather than merely combining different subject areas. In the present study, these principles were reflected in the development of a Digital Activity Book intended to connect mathematical concepts with scientific exploration, technological interaction, engineering challenges, and creative representation. The findings also complement the research of Bang et al. (2023), who investigated the potential of an adaptive game-based mathematics application to support personalized learning and mathematical development among early elementary school students. However, unlike Bang et al. (2023), whose investigation examined mathematics learning outcomes, the present study primarily evaluated instructional media feasibility and initial assessment quality rather than demonstrating improvements in students' numeracy achievement. This distinction is important

because favorable expert evaluations provide preliminary evidence of instructional suitability but cannot independently establish that a digital learning resource produces measurable improvements in mathematical competence. Therefore, the 87.50% feasibility result should be interpreted as an encouraging developmental outcome that supports subsequent classroom evaluation rather than definitive evidence of instructional effectiveness.

The qualitative feedback obtained from expert evaluation provides additional insight into the importance of accessibility and developmentally appropriate interface design in digital mathematics learning environments. Although the Digital Activity Book received favorable feasibility ratings, the experts identified limitations concerning the readability of instructional text on Android smartphones with relatively small screens. This finding indicates that the instructional suitability of digital learning resources depends not only on the relevance of mathematical content but also on the accessibility of visual and textual information presented through different technological devices. In response, the researchers refined the typography and responsive layout to improve the visibility of instructional directions and support more consistent presentation across screen dimensions. These modifications are relevant to the research of Bang et al. (2023), which examined technology-supported mathematics learning through an adaptive application designed for young learners. Their investigation provides a relevant perspective on the importance of considering students' learning characteristics when designing interactive digital mathematics environments. Nevertheless, the present study extends this discussion toward the developmental importance of interface readability, particularly because younger elementary students may require clear instructions and accessible visual representations when interacting with mathematical activities. The observed need for typographic revision also suggests that technological functionality and pedagogical appropriateness should be evaluated as interconnected dimensions of digital instructional quality rather than as independent design considerations. However, because the present evaluation did not include a reported quantitative usability assessment following the revisions, the extent to which these modifications improved students' interaction with the Digital Activity Book remains undetermined. Consequently, future implementation studies should complement expert feasibility evaluations with direct observations of student interaction, task completion, instructional comprehension, and usability to establish stronger evidence concerning the practical suitability of STEAM-based digital mathematics resources.

The empirical item analysis provides further insight into the importance of evaluating assessment quality alongside the development of innovative mathematics learning resources. Based on the available item-validity summary, 12 of the 15 reported numeracy assessment items, representing 80.00%, satisfied the adopted Pearson Product-Moment correlation criterion, whereas three items, representing 20.00%, were classified as invalid. This distribution indicates that most items included in the reported analysis demonstrated acceptable relationships with overall assessment performance, although several items required further consideration before inclusion in the final instrument. The identification of invalid items is methodologically important because assessment questions that do not demonstrate adequate relationships with overall test performance may weaken the interpretability of students' numeracy scores. These findings are consistent with the broader concerns examined by Outhwaite et al. (2024), whose systematic review investigated the psychometric properties of early mathematics assessments and screening instruments. Their research highlights the importance of considering multiple dimensions of measurement quality when evaluating instruments intended to assess children's mathematical competencies (Outhwaite et al., 2024). This comparison emphasizes that constructing assessment items aligned with instructional objectives is insufficient without empirical procedures that examine whether the resulting scores support the intended interpretations. The present study contributes to this methodological concern by incorporating item-level analysis into the development of a numeracy

assessment accompanying a STEAM-based Digital Activity Book. Nevertheless, Pearson item-total correlations provide limited evidence concerning item functioning and should be supplemented with analyses of item difficulty, distractor performance, and additional validity evidence before the instrument is considered suitable for broader application. Furthermore, because the methodological description reports 18 trial items whereas the available validity summary includes only 15 items, the complete item-level dataset must be reconciled before the final number of valid and excluded items can be established with confidence.

The internal consistency analysis produced a reported KR-20 reliability coefficient of 0.78, providing preliminary evidence that the numeracy assessment demonstrated a reasonably consistent response pattern within the observed sample. Because the instrument employed dichotomous scoring, KR-20 was an appropriate method for estimating internal consistency based on the variability of item responses and total assessment scores. The obtained coefficient suggests that the assessment may provide sufficiently consistent scores for preliminary research purposes, although this interpretation depends on the number of included items, the characteristics of the respondents, and the intended measurement context. This finding can be considered alongside the research of Larrain et al. (2026), who investigated the development and psychometric validity evidence of a curriculum-based adaptive mathematics assessment instrument. Their investigation provides a relevant methodological comparison by emphasizing the importance of evaluating assessment quality through systematic psychometric procedures rather than relying exclusively on the construction of curriculum-aligned mathematical tasks (Larrain et al., 2026). The comparison demonstrates that reliability estimation constitutes an important component of assessment development but should be integrated with additional evidence when an instrument is intended to support educational decisions across different populations. Similarly, Outhwaite et al. (2024) examined the psychometric characteristics of early mathematics measurement tools, reinforcing the importance of evaluating reliability and validity as complementary aspects of assessment quality. Although the reported KR-20 coefficient of 0.78 represents a promising initial result, it does not independently establish content validity, construct validity, measurement invariance, or the absence of systematic assessment bias (Hosgoren Alici et al., 2026; Navas-Echazarreta et al., 2024; Szablewska et al., 2026; Zhang et al., 2026). Moreover, the relatively small sample and the unresolved discrepancy concerning the number of items included in the final assessment limit the certainty with which the reliability coefficient can be generalized beyond the observed participants. Accordingly, subsequent psychometric investigations should involve larger samples, verify the final item composition, examine additional measurement characteristics, and assess the stability of reliability estimates across different elementary school contexts.

The principal contribution of this study lies in integrating the development of a STEAM-based Digital Activity Book with the initial psychometric evaluation of its accompanying numeracy assessment within a systematic instructional development framework. Rather than evaluating the digital resource exclusively through its technological features or visual presentation, the study incorporated expert judgment, empirical item analysis, and internal consistency estimation to examine complementary dimensions of instructional and measurement quality. This integrated perspective is conceptually aligned with Belbase et al. (2022), who discussed curriculum integration, pedagogical processes, and assessment as important considerations in the development of STEAM education. It also responds to the methodological concerns examined by Outhwaite et al. (2024), whose systematic review investigated the psychometric evidence supporting assessments of children's mathematical competencies. From an instructional perspective, the developed Digital Activity Book provides a preliminary model for organizing contextual numeracy activities through interdisciplinary learning experiences and interactive technological features. From an assessment perspective, the item-analysis and reliability procedures illustrate how empirical evidence can

inform the selection and refinement of numeracy assessment items before their broader application. Nevertheless, the study has several limitations, including its implementation within a single elementary school, the relatively small assessment sample, the absence of reported post-revision usability evidence, and inconsistencies in the documented number of assessment items. Furthermore, the available findings do not establish whether the Digital Activity Book significantly improves students' numeracy performance because comparative achievement data and appropriate effectiveness analyses have not been reported. Future research should therefore employ larger and more diverse samples, incorporate systematic usability evaluations, strengthen psychometric validation, and investigate instructional effectiveness through appropriate experimental or quasi-experimental designs. Overall, the findings provide a preliminary methodological foundation for developing STEAM-based digital mathematics resources while emphasizing that instructional feasibility, assessment reliability, and demonstrated learning effectiveness must be established through distinct but complementary forms of empirical evidence.

Implications

The findings of this study offer several theoretical, pedagogical, methodological, and practical implications for developing and evaluating STEAM-based digital learning resources in elementary mathematics education. From a theoretical perspective, the integration of Science, Technology, Engineering, Arts, and Mathematics (STEAM) into a Digital Activity Book provides a conceptual foundation for connecting numeracy competencies with interdisciplinary learning experiences, contextual problem-solving, and interactive mathematical exploration. The expert feasibility score of 87.50% provides preliminary evidence that the developed instructional resource possesses favorable characteristics for further implementation, although this result should not be interpreted as direct evidence of improved student achievement. From a pedagogical perspective, the Digital Activity Book offers teachers a potential instructional alternative for organizing mathematical activities that incorporate visual representations, contextual numerical problems, and interdisciplinary challenges appropriate to elementary students' developmental characteristics. The expert recommendations concerning instructional text readability further highlight the importance of accessible typography, responsive interfaces, and age-appropriate visual design when developing digital mathematics resources for young learners. These considerations imply that instructional developers should evaluate technological accessibility alongside curriculum alignment and pedagogical suitability to ensure that digital learning materials accommodate students' cognitive and interactional needs. From a methodological perspective, the integration of expert judgment, empirical item analysis, and internal consistency estimation demonstrates the importance of evaluating instructional media and accompanying assessment instruments through complementary but distinct procedures. The available item-validity summary, which identified 12 retained items among 15 reported items, illustrates how empirical item analysis can inform assessment refinement, although the discrepancy with the 18 items described in the methodological section must be resolved before the final instrument composition is established. Furthermore, the reported KR-20 reliability coefficient of 0.78 provides preliminary evidence of internal consistency, emphasizing the relevance of reliability estimation while underscoring the need for additional validity evidence and evaluation across larger student samples. At the institutional level, the findings suggest that elementary schools and instructional development teams may benefit from adopting systematic development frameworks that combine interdisciplinary instructional design, expert evaluation, accessibility improvements, and evidence-based assessment refinement. Nevertheless, the practical application of the developed resource should remain cautious because the available findings establish preliminary media feasibility and assessment reliability rather than demonstrating improvements in students' numeracy performance or confirming the instrument's suitability for widespread standardized assessment. Overall, this study provides a foundation for future research and

educational development initiatives that seek to integrate STEAM-based digital instructional innovation with rigorous assessment practices, thereby supporting the progressive improvement of elementary mathematics learning resources and the quality of numeracy measurement.

Limitations and Suggestions for Future Research

Despite its contributions to the development and preliminary evaluation of a STEAM-based Digital Activity Book and its accompanying numeracy assessment, this study has several limitations that should be considered when interpreting the findings and determining their broader educational implications. First, the research was conducted at a single elementary school, SDN Gelam 2, which limits the generalizability of the findings to educational settings with different student characteristics, technological infrastructures, and instructional practices. Second, the empirical assessment trial involved only 25 second-grade students, representing a relatively small sample that may affect the stability of item-total correlations and internal consistency estimates. Third, although the Digital Activity Book obtained an expert feasibility score of 87.50%, the evaluation primarily relied on judgments from four media specialists without sufficient quantitative evidence concerning students' actual usability experiences, learning engagement, or independent interaction with the digital resource. Fourth, the assessment evaluation was limited to Pearson Product-Moment correlation and KR-20 reliability estimation, without comprehensive analyses of item difficulty, distractor functioning, construct validity, or measurement invariance across different student populations. Fifth, inconsistencies between the reported 18 trial items and the 15 items presented in the validity summary introduce uncertainty regarding the final assessment composition and the number of items included in the reliability calculation. Sixth, the available findings do not establish the instructional effectiveness of the developed Digital Activity Book because comparative learning outcomes, pretest-posttest measurements, and appropriate inferential analyses were not reported. To address these limitations, future research should involve larger and more representative samples from multiple elementary schools to improve the precision, stability, and generalizability of the psychometric findings. Subsequent investigations should also employ comprehensive assessment validation procedures, including expert-based content validity analysis, item difficulty and discrimination indices, distractor analysis, and additional reliability estimation, while ensuring consistency between the number of developed, tested, and retained assessment items. Furthermore, systematic usability testing involving students and teachers is recommended to examine interface accessibility, instructional clarity, navigation efficiency, technological compatibility, and the suitability of interactive features for young learners. Future studies should also adopt experimental or quasi-experimental research designs incorporating appropriate comparison groups, pretest-posttest assessments, and follow-up measurements to investigate the potential effects of STEAM-based Digital Activity Books on numeracy achievement, mathematical reasoning, and learning retention. Finally, longitudinal and mixed-methods investigations are recommended to explore how interdisciplinary digital learning experiences influence students' mathematical engagement, problem-solving processes, and conceptual understanding over time, thereby providing more comprehensive evidence to support the refinement and broader implementation of STEAM-based instructional resources in elementary mathematics education.

CONCLUSION

This study successfully developed and conducted a preliminary evaluation of a STEAM-based Digital Activity Book and its accompanying numeracy assessment instrument for elementary mathematics learning using the Research and Development (R&D) approach guided by the ADDIE instructional design model. The developed Digital Activity Book integrated Science, Technology,

Engineering, Arts, and Mathematics into contextualized and interactive learning activities designed to accommodate the cognitive characteristics of elementary school students while facilitating mathematical exploration and interdisciplinary problem-solving. Expert evaluation involving four media specialists resulted in an overall feasibility score of 87.50%, indicating that the instructional resource demonstrated favorable characteristics in terms of its content presentation, visual design, and technological suitability according to the adopted evaluation criteria. The expert feedback also identified the need to improve instructional text readability on smaller Android smartphone screens, leading to revisions in typography and responsive interface design to enhance the accessibility of the digital learning environment. These findings provide preliminary evidence supporting the feasibility of the developed instructional media, although its practical usability and effectiveness in improving students' numeracy achievement require further empirical investigation. Regarding assessment development, the study constructed a STEAM-oriented numeracy instrument consisting of an initially reported 30 dichotomously scored multiple-choice items designed according to a structured assessment blueprint and relevant elementary numeracy competencies. The research summary reported that 18 items satisfied the Pearson Product–Moment correlation criterion, although this result requires reconciliation with the empirical trial description involving 18 items and the available item-validity table identifying 12 valid items among 15 reported items. Consequently, the final number of retained assessment items cannot be established conclusively until the item-level results and corresponding statistical decisions have been verified against the complete empirical dataset. Furthermore, the internal consistency analysis produced a reported Kuder–Richardson Formula 20 (KR-20) coefficient of 0.78, suggesting satisfactory preliminary reliability within the observed assessment context, provided that the number of items included in the calculation is confirmed. Nevertheless, this reliability estimate does not independently establish comprehensive instrument validity, eliminate potential measurement bias, or demonstrate that the assessment has achieved full psychometric standardization. Collectively, the findings highlight the methodological importance of integrating expert-based instructional media evaluation with empirical item analysis and reliability estimation to strengthen the development of STEAM-based digital mathematics learning resources and their accompanying assessment instruments. Future research should therefore prioritize reconciling the assessment data, conducting comprehensive psychometric validation with larger and more diverse samples, evaluating student-centered usability, and employing appropriate experimental or quasi-experimental designs to determine whether the developed Digital Activity Book significantly improves elementary students' numeracy competencies.

AUTHOR CONTRIBUTIONS STATEMENT

Ayu Made Juniwati Wiratha contributed to the conceptualization of the study, research design, development of the STEAM-based Digital Activity Book, construction of the numeracy assessment instrument, data collection, and preparation of the original manuscript. Hartono contributed to methodological development, supervision of the research process, evaluation of instructional materials, interpretation of empirical findings, and critical revision of the manuscript. Via Yustitia contributed to the development and evaluation of the numeracy assessment instrument, psychometric data analysis, interpretation of item validity and reliability results, and review and editing of the manuscript. All authors contributed to the intellectual development of the study, reviewed the manuscript, and approved the final version for publication.

REFERENCES

- Adi, L., & Winarso, W. (2026). Problems in Mathematics Learning: Fundamental Concepts, Scope, Contributing Factors, Consequences, and Interventions (SSRN Scholarly Paper No. 7547285). Social Science Research Network. <https://doi.org/10.2139/ssrn.7547285>
- Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an Adaptive Game-Based Math Learning App to Support Personalized Learning and Improve Early Elementary School Students' Learning. *Early Childhood Education Journal*, 51(4), 717–732. <https://doi.org/10.1007/s10643-022-01332-3>
- Belbase, S., Mainali, B. R., Kasemsukpipat, W., Tairab, H., Gochoo, M., & Jarrah, A. (2022). At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education: Prospects, priorities, processes, and problems. *International Journal of Mathematical Education in Science and Technology*, 53(11), 2919–2955. <https://doi.org/10.1080/0020739X.2021.1922943>
- Chai, C. S., Rahmawati, Y., & Jong, M. S.-Y. (2020). Indonesian Science, Mathematics, and Engineering Preservice Teachers' Experiences in STEM-TPACK Design-Based Learning. *Sustainability*, 12(21), 9050. <https://doi.org/10.3390/su12219050>
- Chen, W. (2025). Problem-Solving Skills, Memory Power, and Early Childhood Mathematics: Understanding the Significance of the Early Childhood Mathematics in an Individual's Life. *Journal of the Knowledge Economy*, 16(1), 1–25. <https://doi.org/10.1007/s13132-023-01557-6>
- Chin, H., Chew, C. M., & Lim, H. L. (2021). Development and validation of online cognitive diagnostic assessment with ordered multiple-choice items for 'Multiplication of Time.' *Journal of Computers in Education*, 8(2), 289–316. <https://doi.org/10.1007/s40692-020-00180-7>
- Cirneanu, A.-L., & Moldoveanu, C.-E. (2024). Use of Digital Technology in Integrated Mathematics Education. *Applied System Innovation*, 7(4), 66. <https://doi.org/10.3390/asi7040066>
- Danuri, S. B. W. (2025). Numeracy Literacy Reviewed From Mathematical Reasoning In Child-Friendly Mathematics Learning. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, 32(2-June), 504–515.
- Demosthenous, E., Christou, C., & Pitta-Pantazi, D. (2021). Mathematics Classroom Assessment: A Framework for Designing Assessment Tasks and Interpreting Students' Responses. *European Journal of Investigation in Health, Psychology and Education*, 11(3), 1088–1106. <https://doi.org/10.3390/ejihpe11030081>
- Gilleece, L., Karakolidis, A., & Clerkin, A. (2026). The validity of school socio-economic status measures: How do indicators from TIMSS and PISA compare to national administrative data sources? *Large-Scale Assessments in Education*, 14(1), 45. <https://doi.org/10.1186/s40536-026-00305-x>
- Hadi, S., Dolk, M., Kamaliyah, & Hidayanto, T. (2025). Mathematical Reasoning: How Students Learn Mathematics? *Journal on Mathematics Education*, 16(3), 937–954. <https://doi.org/10.22342/jme.v16i3.pp937-954>
- Hosgoren Alici, Y., Özbaran, N. B., Köse, S., Gormez, A., Kandeğer, A., Algedik, P., Erbaş, S., Yenice, H. A., Mançe Çalışır, Ö., Hasanli, J., Asut, G., Mısıır, E., Torenli, Z. K., Bağcaz, A., Ceran, S., Ayraller Taner, H., Emekli, E., Özel, B., Aksoy, U. M., ... Atbaşoğlu, C. (2026). Validity and Reliability of the Turkish Version of the Adult Autism Subthreshold Spectrum Scale. *Intellectual and Developmental Disabilities*, 0(0), 1–16. <https://doi.org/10.1080/19349491.2026.2734441>
- Joshi, D. R., Adhikari, K. P., Khanal, J., Belbase, S., & Khanal, B. (2023). Developing and integrating digital resources in online mathematics instruction and assessment during Covid-19. *Cogent Education*, 10(2), 2230394. <https://doi.org/10.1080/2331186X.2023.2230394>
- Juškevičienė, A., Dagienė, V., & Dolgopolas, V. (2021). Integrated activities in STEM environment: Methodology and implementation practice. *Computer Applications in Engineering Education*, 29(1), 209–228. <https://doi.org/10.1002/cae.22324>
- Kaminski, J. A., & Sloutsky, V. M. (2020). The use and effectiveness of colorful, contextualized, student-made material for elementary mathematics instruction. *International Journal of STEM Education*, 7(1), 6. <https://doi.org/10.1186/s40594-019-0199-7>
- Karakolidis, A., Duggan, A., Shiel, G., & Kiniry, J. (2021). Examining educational inequalities: Insights in the context of improved mathematics performance on national and international assessments at primary level in Ireland. *Large-Scale Assessments in Education*, 9(1), 5. <https://doi.org/10.1186/s40536-021-00098-1>

- Khanipour, H., Geramipour, M., Radfar, M., & Mirahadi, S. S. (2025). Psychometric properties of process assessment of the learner (second edition): Diagnostics for math among Iranian elementary school students. *Applied Neuropsychology: Child*, 0(0), 1–9. <https://doi.org/10.1080/21622965.2025.2461716>
- Larrain, M., Navarro, M., & Pezoa, J. (2026). Development and psychometric validity evidence of a curriculum-based and adaptive mathematics assessment tool. *Assessment in Education: Principles, Policy & Practice*, 33(3), 285–305. <https://doi.org/10.1080/0969594X.2026.2689693>
- Lavicza, Z., Weinhandl, R., Prodromou, T., Anđić, B., Lieban, D., Hohenwarter, M., Fenyvesi, K., Brownell, C., & Diego-Mantecón, J. M. (2022). Developing and Evaluating Educational Innovations for STEAM Education in Rapidly Changing Digital Technology Environments. *Sustainability*, 14(12), 7237. <https://doi.org/10.3390/su14127237>
- Liston, M., Morrin, A. M., Furlong, T., & Griffin, L. (2022). Integrating Data Science and the Internet of Things Into Science, Technology, Engineering, Arts, and Mathematics Education Through the Use of New and Emerging Technologies. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.757866>
- Liu, Y., Peng, P., & Yan, X. (2025). Early numeracy and mathematics development: A longitudinal meta-analysis on the predictive nature of early numeracy. *Journal of Educational Psychology*, 117(6), 863–883. <https://doi.org/10.1037/edu0000925>
- McNeil, N. M., Jordan, N. C., Viegut, A. A., & Ansari, D. (2025). What the Science of Learning Teaches Us About Arithmetic Fluency. *Psychological Science in the Public Interest*, 26(1), 10–57. <https://doi.org/10.1177/15291006241287726>
- Navas-Echazarreta, N., Juárez-Vela, R., Martínez-Sabater, A., Echániz-Serrano, E., Fernández-Rodrigo, M. T., Navarro-Martínez, O., Sancho-Sánchez, C., Cobos-Rincón, A., Rodríguez-Calvo, A., González-Fernández, S., Chover-Sierra, E., & Satústegui-Dordá, P. J. (2024). Measuring Instruments for Media Health Literacy: A Systematic Review of Psychometric Properties. *Nursing Reports*, 14(4), 2795–2818. <https://doi.org/10.3390/nursrep14040206>
- Onoshakpokaiye, O. (2023). Early Childhood Mathematics: An Insight into Strategies for Developing Young Children Mathematical Skills. *Mathematics Education Journal*, 7(1), 16–30. <https://doi.org/10.22219/mej.v7i1.24534>
- Outhwaite, L. A., Aunio, P., Leung, J. K. Y., & Van Herwegen, J. (2024). Measuring Mathematical Skills in Early Childhood: A Systematic Review of the Psychometric Properties of Early Maths Assessments and Screeners. *Educational Psychology Review*, 36(4), 110. <https://doi.org/10.1007/s10648-024-09950-6>
- Purcar, A.-M., Bocoș, M., Pop, A.-L., Roman, A., Rad, D., Mara, D., Crișan, C., Răduț-Taciu, R., Mara, E.-L., Todor, I., Muntean-Trif, L., Neacșu, M.-G., Costache Colareza, C., Maier, M., Tăușan-Crișan, L., Triff, Z., Baciuc, C., Marin, D.-C., & Triff, D.-G. (2024). The Effect of Visual Reasoning on Arithmetic Word Problem Solving. *Education Sciences*, 14(3), 278. <https://doi.org/10.3390/educsci14030278>
- Rosita, C. D., Setiyani, Asmara, A. S., Sumarwati, S., & Suprayo, T. (2026). Mathematical Thinking Processes Based on Decision-Making Types in Grade VIII Numeracy: Foundations for Differentiated Instruction in Geometry. *Journal on Mathematics Education*, 17(1), 1–26. <https://doi.org/10.22342/jme.v17i1.pp1-26>
- Rosyidi, A. H., Sari, Y. M., Fardah, D. K., & Masriyah, M. (2024). Designing mathematics problem-solving assessment with GeoGebra Classroom: Proving the instrument validity. *Journal of Education and Learning (EduLearn)*, 18(3), 1038–1046. <https://doi.org/10.11591/edulearn.v18i3.21191>
- Sakurai, J., & Goos, M. (2023). Revisiting tools in numeracy learning: The role of authentic digital tools. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1291407>
- Sudirman, S., Martadiputra, B. A. P., Faizah, S., Rodríguez-Nieto, C., Soeharto, S., Lavicza, Z., Isnawan, M. G., Sembiring, M. G., Runisah, R., & Suprihatin, R. J. (2026). Integrating steam-based MOOC as a supplement to didactical and pedagogical practices in school mathematics: A design-based research. *Journal of Technology and Science Education*, 16(1), 125–146. <https://doi.org/10.3926/jotse.3633>

- Supianti, I. I., Yaniawati, P., Bonyah, E., Hasbiah, A. W., & Rozalini, N. (2025). STEAM approach in project-based learning to develop mathematical literacy and students' character. *Infinity Journal*, 14(2), 283–302. <https://doi.org/10.22460/infinity.v14i2.p283-302>
- Szablewska, A. W., Czerwińska-Osipiak, A., Popowicz, H., Radzikowski, A., Rozmarynowska, W., Wójcicka, L. P., Stefanowicz-Bielska, A., Pietrzak, K., Krawczyk, A., Michalik, A., Gaworska-Krzemińska, A., Jakimczyk, B., García-Valdivieso, I., & José, F. (2026). Knowledge, Attitudes, and Practices in Pediatric Pain Management: Cross-Cultural Adaptation and Initial Psychometric Evaluation of the Polish HUPEDCARE-Q Among Healthcare Students and Professionals. *Journal of Clinical Medicine*, 15(10), 3678. <https://doi.org/10.3390/jcm15103678>
- Tegegne, W. A., Bizuneh, S. M., & Negasi, R. D. (2025). Exploring the promises and classroom realities of constructivist learning at upper primary school in Ethiopia. *Discover Education*, 4(1), 513. <https://doi.org/10.1007/s44217-025-00967-9>
- Torres-Peña, R. C., Peña-González, D., Lara-Orozco, J. L., Ariza, E. A., & Vergara, D. (2025). Enhancing Numerical Thinking Through Problem Solving: A Teaching Experience for Third-Grade Mathematics. *Education Sciences*, 15(6), 667. <https://doi.org/10.3390/educsci15060667>
- Turner, E. E., Chen, M.-K., Roth McDuffie, A., Smith, J. E., Aguirre, J., Foote, M. Q., & Bennett, A. B. (2021). Validating a student assessment of mathematical modeling at elementary school level. *School Science and Mathematics*, 121(7), 408–421. <https://doi.org/10.1111/ssm.12494>
- Wu, Y. (2025). Unlocking Mathematics Success: Global Lessons on Student Achievement, Teacher Satisfaction, and School Environments. *International Electronic Journal of Mathematics Education*, 20(2). <https://eric.ed.gov/?id=EJ1469671>. <https://doi.org/10.29333/iejme/15900>
- Yustitia, V., Kintoko, K., Pamungkas, M. D., & Wijayanti, S. (2025). Place-based ethnomathematics for early numeracy: Geometric representations from the Tugu Pahlawan monument in Surabaya. *Al-Jabar: Jurnal Pendidikan Matematika*, 16(2), 891–907. <https://doi.org/10.24042/ajpm.v16i2.29234>
- Yustitia, V., Kusmaharti, D., & Wardani, I. S. (2025). Students' critical thinking in numeracy problem-solving through moderate self-efficacy: A mixed-methods study. *Multidisciplinary Science Journal*, 7(8), Article 2025410. <https://doi.org/10.31893/multiscience.2025410>
- Yustitia, V., Rakhmah, Y. N. I., Astuti, I. P., & Untari, E. (2024). Ular tangga numerasi: Inovasi media pembelajaran matematika untuk siswa sekolah dasar. *Jurnal Simki Pedagogia*, 7(1), 32–43. <https://doi.org/10.29407/jsp.v7i1.542>
- Zhang, L., Liu, X., Cui, L., Yang, X., Hou, X., & Hou, L. (2026). Development and Initial Psychometric Evaluation of the Adolescent Oral and Maxillofacial Health Knowledge-Attitude-Practice and Self-Efficacy Questionnaire (AOMH-KAPQ). *Healthcare*, 14(16), 2615. <https://doi.org/10.3390/healthcare14162615>