Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 4, 161-185.
e_ISSN: 2798-1606
DOI: 10.58524/jasme.v6i4.1634

KOSMARA-STEM: Development and evaluation of an integrated RME–PBL–STEM instructional model for supporting numeracy literacy among Indonesian grade V elementary students

Ardi Dwi Susandi*
Universitas Terbuka, INDONESIA

Thesa Kandaga
Universitas Terbuka, INDONESIA

Vivi Rindastri
Universitas Terbuka, INDONESIA

Article Info

Article history:

Received: July 28, 2026
Revised: September 30, 2026
Accepted: October 26, 2026

Keywords:

Elementary Education;
Instructional Model;
KOSMARA-STEM;
Numeracy Literacy;
STEM Education.

Abstract

Background: Numeracy literacy remains a challenge in Indonesian elementary education, highlighting the need for instructional models that integrate contextual reasoning, collaborative problem-solving, and interdisciplinary learning.
Aims: This study aimed to develop and evaluate KOSMARA-STEM, an instructional model integrating Realistic Mathematics Education (RME), Problem-Based Learning (PBL), and STEM to support fifth-grade students' numeracy literacy.
Method: Guided by Plomp's educational design research framework, the study involved 560 students from 20 elementary schools in Cirebon, Indonesia, through small-scale (n = 140) and large-scale (n = 420) trials. Model validity, implementation practicality, and numeracy outcomes were assessed using expert evaluations, classroom observations, questionnaires, and a single-group pretest–posttest design.
Results: Expert evaluations confirmed high model validity (M = 3.74/4.00) and instrument validity (M = 3.77/4.00). Implementation fidelity averaged 85.8%, while 89.5% of student responses were positive. Numeracy scores increased from 62.4 to 82.7 (N-Gain = 0.54; Cohen's dz = 1.58; $p < .001$).
Conclusion: KOSMARA-STEM demonstrated satisfactory validity, classroom feasibility, and promising short-term numeracy improvements. However, its effectiveness requires confirmation through controlled experimental studies.

To cite this article: Susandi, A. D., Kandaga, T. & Rindastri, V. (2026). Development and evaluation of KOSMARA-STEM: An integrated RME–PBL–STEM instructional model for enhancing elementary students' numeracy literacy. *Journal of Advanced Sciences and Mathematics Education, 6*(4), 161-185.

INTRODUCTION

Numeracy literacy has become an essential educational competency because it enables individuals to interpret quantitative information, apply mathematical reasoning, and make informed decisions in diverse real-world situations (Muhaimin et al., 2024; OECD, 2023). Beyond performing routine calculations, numeracy literacy encompasses the ability to formulate mathematical problems, evaluate numerical representations, and communicate evidence-based conclusions across personal, educational, and societal contexts. Developing these competencies during elementary education is particularly important, as early mathematical understanding provides a foundation for subsequent academic achievement and more advanced mathematical reasoning (Liu et al., 2025). Recent systematic evidence further emphasizes the importance of valid numeracy assessments in identifying learning difficulties and supporting instructional decisions during the early stages of mathematical development (Speyer et al., 2024). Nevertheless, improving elementary students' numeracy literacy remains a significant educational challenge in Indonesia, particularly in relation to contextual problem-solving and mathematical interpretation (Wijaya et al., 2024). Results from the Programme for International Student Assessment (PISA) 2022 placed Indonesia 68th among 81

***Corresponding author:**
Ardi Dwi Susandi, Universitas Terbuka, Indonesia
ardi.official@ecampus.ut.ac.id

participating economies, with an average mathematics score of 366 compared with the OECD average of 472 (Huang et al., 2024; OECD, 2023). Similarly, findings from the Indonesian National Assessment, incorporating the Minimum Competency Assessment (*Asesmen Kompetensi Minimum*, AKM), indicated that many elementary students had not achieved the expected proficiency in applying mathematical knowledge to unfamiliar situations (Kemdikbudristek, 2023). These difficulties suggest that mathematical competence cannot be developed adequately through procedural exercises alone, particularly when students encounter problems requiring interpretation, representation, and independent reasoning. Differences between mathematical experiences at home and school may further constrain students' opportunities to connect formal mathematical concepts with meaningful everyday applications (Graven & Jorgensen, 2024; Purbaningrum et al., 2024). Consequently, strengthening elementary numeracy literacy requires instructional approaches that promote conceptual understanding, contextual mathematical reasoning, and the meaningful application of quantitative knowledge.

The persistent challenges associated with elementary numeracy development reflect interactions among socioeconomic conditions, students' learning experiences, teacher preparation, and the instructional practices implemented in classrooms (Graven & Jorgensen, 2024; Wijaya et al., 2024). Among these factors, instructional approaches that prioritize procedural accuracy over conceptual exploration may restrict students' opportunities to develop flexible mathematical reasoning and independent problem-solving abilities (Fauzan et al., 2024; Wijaya et al., 2024). Conventional teacher-centered instruction frequently emphasizes predetermined solution procedures, providing limited opportunities for learners to construct mathematical representations, compare alternative strategies, and justify their reasoning through collaborative discussions (Nusantara et al., 2022). Although contemporary educational reforms increasingly encourage student-centered and inquiry-oriented learning, their implementation remains challenging because of institutional expectations, insufficient professional development, and established pedagogical practices (Fry et al., 2025). Research involving Indonesian elementary classrooms has similarly identified limited mathematical discourse and inadequate connections between instructional content and students' everyday experiences as obstacles to meaningful mathematics learning (Fauzan et al., 2024; Wijaya et al., 2024). From an educational equity perspective, restricted participation in mathematical discussions may further disadvantage students whose informal mathematical experiences differ from the representations commonly emphasized in formal instruction (Abtahi & Planas, 2024). Addressing these limitations requires learning environments in which students actively investigate contextual problems, communicate mathematical arguments, and progressively refine their understanding through interaction with teachers and peers. International research also indicates that students experiencing early mathematical difficulties benefit from structured instructional support that responds to their learning needs and prevents the persistence of conceptual misunderstandings (López-Pedersen et al., 2022; Roos, 2023). Within the Indonesian educational context, these considerations are particularly relevant to the implementation of the *Kurikulum Merdeka*, which emphasizes contextual learning, independent reasoning, and the application of mathematical competencies to authentic situations (Muhaimin et al., 2024). Therefore, developing an instructional framework that systematically connects contextual exploration, collaborative problem-solving, and mathematical formalization represents a relevant strategy for improving the quality of elementary mathematics education.

Several established pedagogical approaches offer complementary theoretical foundations for developing instructional models that support contextual understanding, mathematical reasoning, and interdisciplinary problem-solving. Realistic Mathematics Education (RME), originating from Freudenthal's conceptualization of mathematics as a human activity, emphasizes the progressive construction of mathematical knowledge through meaningful contexts, guided reinvention, and the

transformation of informal strategies into formal representations (Freudenthal, 1968; Van den Heuvel-Panhuizen & Drijvers, 2020). Empirical investigations and meta-analytical studies have demonstrated the potential of RME to strengthen mathematical understanding, representation, and problem-solving abilities, with reported effects varying according to educational level, instructional duration, and implementation characteristics (Fauzan et al., 2024; Juandi et al., 2022; Risdiyanti et al., 2024; Samritin, 2023; Yuanita et al., 2018). Furthermore, mathematical strategy flexibility enables students to select, adapt, and evaluate alternative approaches when transitioning between contextual reasoning and formal mathematical procedures (Verschaffel et al., 2024). Complementing these principles, Problem-Based Learning (PBL) organizes learning around meaningful problems that encourage collaborative inquiry, critical thinking, mathematical communication, and reflective evaluation of solution strategies (Ibad et al., 2025; Yanuarto et al., 2025). Meanwhile, Science, Technology, Engineering, and Mathematics (STEM) education extends mathematical learning beyond disciplinary boundaries by integrating scientific investigation, technological applications, engineering-oriented activities, and quantitative modelling within authentic problem contexts (English, 2023; Kristensen et al., 2024; Ow-Yeong et al., 2023). Previous studies have highlighted the educational potential of STEM-oriented and project-based interdisciplinary activities for supporting students' mathematical reasoning, numeracy development, scientific literacy, and classroom engagement (Maryam et al., 2026; Rehman et al., 2024; Susanta et al., 2025; Suryanti et al., 2024). However, implementing these approaches independently may limit their pedagogical contributions because RME primarily emphasizes progressive mathematization, PBL structures collaborative investigation, and STEM provides interdisciplinary contexts without necessarily ensuring systematic mathematical formalization. Integrating these complementary perspectives offers an opportunity to establish a coherent instructional progression in which students investigate authentic problems, construct and compare mathematical strategies, develop representations, and apply formal concepts to increasingly complex situations. Accordingly, a theoretically grounded instructional model combining RME, PBL, and STEM may provide a systematic framework for supporting elementary students' numeracy literacy while responding to the pedagogical demands of contemporary mathematics education.

Despite growing empirical evidence supporting RME, PBL, and STEM education, existing research has provided comparatively limited explanations of how these approaches can be systematically integrated into a unified instructional model specifically designed for elementary numeracy literacy (Cevikbas et al., 2024; Yang et al., 2025). Previous investigations have predominantly examined individual pedagogical strategies or combinations of two approaches, leaving the instructional relationships among contextual mathematical exploration, collaborative problem-solving, and interdisciplinary modelling insufficiently specified (Muhaimin et al., 2024; Risdiyanti et al., 2024). Recent Indonesian studies incorporating ethnomathematics and culturally contextualized learning have demonstrated promising outcomes, although their instructional designs generally emphasize particular mathematical topics or pedagogical perspectives rather than a comprehensive integration of complementary approaches (Marleny et al., 2026; Wewe et al., 2026). Similarly, research combining STEM-oriented projects or interactive mathematical applications with problem-based instruction has reported improvements in students' numeracy and mathematical reasoning, but has not explicitly established a sequential framework connecting informal mathematical strategies with formal conceptual development (Maryam et al., 2026; Yanuarto et al., 2025). Existing reviews further indicate that implementing contextual and student-centered mathematics instruction requires structured pedagogical guidance that supports teachers in facilitating mathematical discourse, conceptual transitions, and meaningful classroom interactions (Shimizu & Kang, 2025). Moreover, interdisciplinary instructional practices present implementation challenges because teachers' pedagogical knowledge, technological competencies, and professional

experiences influence their capacity to establish meaningful connections across STEM disciplines (Wu et al., 2024). Another limitation concerns the relatively restricted evidence regarding instructional models evaluated simultaneously through theoretical validity, classroom practicality, implementation fidelity, and students' numeracy learning outcomes across multiple elementary schools (Purbaningrum et al., 2024; Wijaya et al., 2024). Consequently, the literature has not sufficiently established how the complementary functions of RME, PBL, and STEM can be operationalized within a consistent instructional sequence that remains feasible for implementation in regular Indonesian elementary classrooms. Addressing this gap requires an educational design research approach that develops a theoretically coherent instructional model, validates its essential components, examines implementation processes, and documents changes in students' mathematical performance without overstating causal effectiveness. Therefore, the present study introduces KOSMARA-STEM as a six-phase instructional model that systematically integrates contextual problem exploration, independent strategy development, collaborative discussion, mathematical modelling, conceptual formalization, and reflective application to support Grade V students' numeracy literacy.

This study aims to design, develop, and evaluate KOSMARA-STEM, an integrated instructional model combining Realistic Mathematics Education (RME), Problem-Based Learning (PBL), and Science, Technology, Engineering, and Mathematics (STEM) to support numeracy literacy among fifth-grade elementary school students in Cirebon Region III, West Java, Indonesia. Guided by Plomp's educational design research framework, the study adopts a systematic development process encompassing preliminary investigation, iterative prototyping, expert validation, and classroom-based evaluation (Plomp & Nieveen, 2013). The proposed model comprises six interconnected instructional phases, namely Contextual Problem, Own Production, Strategy Discussion and Comparison, Mathematical Modelling, Formalisation, and Reflection and Reinvention. These phases are designed to facilitate students' progression from interpreting authentic mathematical situations toward constructing independent strategies, evaluating alternative solutions, developing mathematical representations, and applying formal concepts in meaningful contexts. Specifically, the first research objective is to examine the validity of the KOSMARA-STEM instructional model and its supporting assessment instruments through structured evaluations conducted by mathematics education specialists and elementary school practitioners. The second objective is to investigate the practicality of implementing the model in authentic classroom environments by examining instructional fidelity, student learning activities, and learners' responses during small-scale and large-scale trials. The third objective is to evaluate changes in students' numeracy literacy between pretest and posttest assessments following participation in KOSMARA-STEM instruction, while recognizing that the absence of a comparison group limits causal interpretations of the observed outcomes. The study's principal theoretical contribution lies in specifying the complementary instructional functions of RME, PBL, and STEM within a coherent learning sequence, thereby extending existing discussions concerning integrated, contextualized, and interdisciplinary mathematics education (Cevikbas et al., 2024; English, 2023; Van den Heuvel-Panhuizen & Drijvers, 2020). From a practical perspective, the research seeks to provide a structured instructional framework and supporting assessment instruments that may assist elementary teachers in implementing contextual numeracy instruction aligned with the objectives of the *Kurikulum Merdeka* (Fauzan et al., 2024; Muhaimin et al., 2024). Ultimately, the findings are expected to contribute preliminary evidence concerning the validity, feasibility, and associated learning outcomes of KOSMARA-STEM while establishing a foundation for subsequent controlled investigations of its comparative effectiveness and broader applicability in elementary mathematics education.

LITERATURE REVIEW

Numeracy literacy represents a multidimensional mathematical competency that enables students to interpret quantitative information, apply mathematical reasoning, and solve problems situated within authentic personal, social, and educational contexts (Muhaimin et al., 2024; OECD, 2023). Unlike procedural mathematical proficiency, numeracy literacy emphasizes the meaningful application of mathematical knowledge, including numerical interpretation, data representation, proportional reasoning, and informed decision-making. Developing these competencies during elementary education is particularly important because early mathematical understanding significantly influences subsequent academic achievement and the development of advanced mathematical reasoning (Liu et al., 2025). Research on early numeracy assessment highlights the importance of valid measurement instruments for identifying students' mathematical difficulties and supporting evidence-based instructional decisions (Speyer et al., 2024). Nevertheless, students frequently encounter difficulties transferring mathematical knowledge from familiar classroom exercises to unfamiliar situations requiring contextual interpretation and independent reasoning. These challenges indicate that numeracy development requires instructional approaches that connect mathematical concepts with students' everyday experiences rather than emphasizing isolated computational procedures (Graven & Jorgensen, 2024). Within Indonesian elementary education, numeracy literacy has become an essential educational priority through the implementation of the Minimum Competency Assessment (*Asesmen Kompetensi Minimum*, AKM), which emphasizes understanding, application, and mathematical reasoning (Kemdikbudristek, 2023). Effective numeracy instruction should therefore provide opportunities for students to construct mathematical representations, evaluate alternative solutions, and communicate quantitative arguments through meaningful learning experiences. Furthermore, contextualized learning environments can support students in recognizing the relevance of mathematics to everyday decision-making while encouraging greater participation in mathematical discussions (Geiger et al., 2023). Accordingly, strengthening elementary numeracy literacy requires pedagogical frameworks that integrate conceptual understanding, authentic problem-solving, and systematic assessment of mathematical competencies.

Realistic Mathematics Education (RME) provides a theoretical foundation for contextual mathematics instruction by positioning mathematics as a human activity developed through meaningful problem-solving experiences (Freudenthal, 1968). Central to this approach is the principle of guided reinvention, through which students progressively construct mathematical concepts by exploring realistic situations and refining their informal solution strategies (Van den Heuvel-Panhuizen & Drijvers, 2020). RME emphasizes progressive mathematization, involving the transformation of contextual problems into mathematical representations and the subsequent development of increasingly formal mathematical understanding. This progression encourages students to establish conceptual connections between concrete experiences, visual models, symbolic representations, and abstract mathematical principles. Research indicates that students' flexibility in transitioning between informal and formal strategies contributes substantially to mathematical competence and adaptive problem-solving (Verschaffel et al., 2024). Empirical investigations involving elementary mathematics education have demonstrated positive associations between RME-based instruction and improvements in mathematical understanding, reasoning, and numeracy performance (Fauzan et al., 2024; Risdiyanti et al., 2024). Meta-analytic evidence further supports these findings, reporting meaningful improvements in students' mathematical abilities following RME interventions across different educational contexts (Juandi et al., 2022; Samritin, 2023). Additionally, mathematical representation has been identified as an important mechanism connecting students' mathematical beliefs with their problem-solving performance in RME-oriented

learning environments (Yuanita et al., 2018). However, implementing progressive mathematization effectively requires structured instructional guidance that supports teachers in facilitating students' transitions from contextual exploration toward formal mathematical generalization. Consequently, RME provides an essential pedagogical foundation for KOSMARA-STEM, particularly through its emphasis on contextual problem exploration, student-generated strategies, mathematical modelling, and conceptual formalization.

Problem-Based Learning (PBL) is a student-centered instructional approach that promotes knowledge construction through the investigation of meaningful problems requiring critical thinking, collaboration, and reflective decision-making. Rather than presenting mathematical procedures as predetermined solutions, PBL encourages students to identify relevant information, formulate strategies, evaluate alternative approaches, and justify their mathematical conclusions. This instructional orientation aligns with contemporary mathematics education perspectives emphasizing active participation, conceptual understanding, and the development of transferable problem-solving competencies (Ibad et al., 2025). Through collaborative inquiry, students engage in mathematical discourse that enables them to articulate reasoning, compare solution strategies, and negotiate shared interpretations of mathematical relationships. Such interactions can strengthen mathematical communication while creating opportunities for students to recognize the validity and limitations of different problem-solving approaches. Meta-analytic findings indicate that PBL-based instruction positively influences students' problem-solving performance, suggesting its potential contribution to mathematics learning across educational settings (Ibad et al., 2025). Research involving Indonesian elementary students has also demonstrated that integrating PBL with interactive mathematics applications can support improvements in mathematical reasoning and collaborative learning (Yanuarto et al., 2025). Nevertheless, successful PBL implementation requires appropriate instructional scaffolding to ensure that exploratory activities remain connected to intended mathematical concepts and learning objectives. Within KOSMARA-STEM, PBL principles are incorporated through independent strategy development, collaborative discussion, comparison of mathematical solutions, and structured reflection on learning processes. Therefore, integrating PBL into the instructional model is expected to support students' active engagement with contextual mathematical problems while strengthening collaborative reasoning and reflective problem-solving.

Science, Technology, Engineering, and Mathematics (STEM) education promotes interdisciplinary learning by connecting mathematical concepts with scientific inquiry, technological applications, and engineering-oriented problem-solving within authentic contexts. In mathematics education, STEM integration positions mathematical knowledge not only as a computational resource but also as a central learning objective developed through interdisciplinary investigation (English, 2023; Kristensen et al., 2024). This approach encourages students to recognize relationships between mathematical representations and practical situations involving measurement, data interpretation, quantitative modelling, and evidence-based decision-making. Technology-supported learning environments can further facilitate mathematical exploration by providing interactive representations, digital visualization tools, and opportunities to examine relationships among numerical and graphical information (Ow-Yeong et al., 2023). Research has demonstrated that STEM-oriented instructional activities can support mathematical thinking, conceptual understanding, and student engagement when interdisciplinary tasks are aligned with explicit learning objectives (English, 2023; Kristensen et al., 2024). Indonesian educational studies have similarly reported positive learning outcomes associated with digital STEAM inquiry and project-based STEM instruction, particularly in developing mathematical and scientific competencies (Susanta et al., 2025; Suryanti et al., 2024). Furthermore, project-based STEM learning involving authentic design challenges has demonstrated promising outcomes for fifth-grade students' numeracy development compared with conventional instructional approaches (Maryam et al., 2026).

Despite these benefits, interdisciplinary instruction may present implementation challenges because teachers require adequate pedagogical knowledge, technological competence, and the ability to establish meaningful connections across disciplinary boundaries (Wu et al., 2024). Within KOSMARA-STEM, STEM principles are incorporated through contextual problem-solving, technology-supported mathematical modelling, and the application of mathematical concepts to interdisciplinary learning situations. Accordingly, STEM integration complements RME and PBL by extending contextual mathematical exploration toward practical applications that encourage students to interpret, represent, and apply quantitative information across diverse problem contexts.

The integration of RME, PBL, and STEM provides a conceptual foundation for developing instructional models that combine contextual mathematical understanding, collaborative problem-solving, and interdisciplinary applications within a coherent learning sequence. Although these approaches have independently demonstrated educational benefits, research continues to highlight the need for instructional frameworks that systematically integrate complementary pedagogical principles rather than implementing them as disconnected classroom activities (Cevikbas et al., 2024; Yang et al., 2025). KOSMARA-STEM was developed to address this pedagogical need through six interconnected instructional stages: Contextual Problem, Own Production, Strategy Discussion and Comparison, Mathematical Modelling, Formalisation, and Reflection and Reinvention. The initial stages emphasize authentic problem exploration and independent strategy construction, reflecting RME principles of contextual learning and guided reinvention (Van den Heuvel-Panhuizen & Drijvers, 2020). Subsequent collaborative discussion and strategy comparison incorporate PBL principles by encouraging students to communicate mathematical arguments, evaluate alternative solutions, and construct shared understanding (Ibad et al., 2025). Mathematical Modelling and Formalisation facilitate the transition from informal reasoning to formal mathematical representations while incorporating STEM-oriented applications and technology-supported learning experiences (English, 2023; Kristensen et al., 2024). The final stage emphasizes metacognitive reflection and the application of acquired mathematical knowledge to unfamiliar situations, supporting the development of transferable numeracy competencies. From an instructional design perspective, evaluating such an integrated model requires systematic consideration of theoretical validity, classroom practicality, implementation fidelity, and students' mathematical learning outcomes (Plomp & Nieveen, 2013). Moreover, the development of structured instructional guidance and validated assessment instruments is particularly relevant to Indonesian elementary education, where contextual learning and higher-order mathematical reasoning are emphasized within contemporary curriculum reforms (Purbaningrum et al., 2024; Wijaya et al., 2024). Therefore, KOSMARA-STEM represents a theoretically grounded instructional framework whose development and empirical evaluation may contribute to advancing integrated pedagogical practices and strengthening numeracy literacy in elementary mathematics education.

METHOD

Research Design

This study employed a research and development (R&D) approach, using Plomp's (2013) educational design research model as the overarching methodological framework. Plomp's model was selected for its systematic, iterative structure and its established suitability for developing and evaluating educational interventions in school contexts (Plomp & Nieveen, 2013; Fauzan et al., 2024). The instructional components of KOSMARA-STEM were developed using Joyce and Calhoun's (2024) models-of-teaching framework, which specifies five core elements of an instructional model: syntax, social system, principles of reaction, support system, and instructional and nurturant effects. Product

quality was evaluated against Nieveen's (1999) three criteria of validity, practicality, and effectiveness, which remain the standard benchmark for educational product development in mathematics education R&D (Plomp & Nieveen, 2013). Within the assessment phase, learning outcomes were evaluated using a single-group pretest–posttest design. No comparison group was included, so the outcome data describe within-group change rather than effects attributable to the model. Quantitative data (expert ratings, observation scores, questionnaire responses, and test scores) and qualitative data (validator comments, observation field notes, and teacher interviews) were integrated sequentially: qualitative findings from the preliminary phase informed the design of the prototype, validator comments and small-scale trial notes were used to revise the model, and both data types were considered jointly when interpreting practicality.

Participants and Sampling

The study was conducted in public elementary schools (Sekolah Dasar Negeri) in Cirebon Region III, West Java, Indonesia, during the 2025/2026 academic year. Cirebon Region III comprises five districts/municipalities (kabupaten/kota), all of which were represented in the study. Consistent with Plomp's (2013) prototyping approach, implementation followed a two-stage trial design: a small-scale trial (ujicoba skala kecil) involving one school from each of the five districts/municipalities (n = 5 schools) and a large-scale trial (ujicoba skala besar) involving three schools from each district/municipality (n = 15 schools), for a total of 20 participating schools. Purposive sampling was used at both stages to select schools reflecting the socioeconomic, geographic, and academic diversity of Cirebon Region III (Ahmad & Wilkins, 2024; Creswell & Plano Clark, 2022). Schools were eligible if they were public elementary schools with at least one Grade V class and if their principals and Grade V teachers agreed to take part for the full implementation period. Because selection was purposive and depended on schools' willingness to participate, the sample may over-represent schools that are receptive to instructional innovation, which may introduce selection bias and limits generalisation. Participants were Grade V students (aged 10–11 years) and their classroom teachers. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki for research involving children, and permission to conduct the research was obtained from the principals of all participating schools. Written informed consent was obtained from school principals, participating teachers, and students' parents or guardians after they had received information on the study's aims, procedures, voluntary nature, and confidentiality safeguards; students were informed that they could withdraw at any time without consequences for their grades.

Procedures and Time Frame

Plomp's (2013) model was operationalised in three phases. In the Preliminary Research phase, needs assessment data were collected through classroom observations (n = 18 lessons), semi-structured teacher interviews (n = 6), and analysis of Grade V mathematics assessment records, following Plomp's recommended context and needs analysis procedures (Plomp & Nieveen, 2013). Observation field notes and interview transcripts were analysed thematically (coding, categorisation, and identification of recurring themes), and assessment records were summarised descriptively against AKM-aligned proficiency expectations. Classroom observations showed that mathematics lessons were dominated by teacher explanation and procedural exercises, with few opportunities for students to work on contextual problems or to discuss their strategies. Teacher interviews indicated that teachers were more accustomed to procedural instruction and found it challenging to design contextual, AKM-type problems. Assessment records showed that many Grade V students had difficulty with contextual and non-routine numeracy items, particularly those requiring interpretation of data and proportional reasoning. Taken together, these data indicated a gap

between expected and demonstrated numeracy literacy competencies in the participating schools and pointed to over-reliance on procedural instruction as one prominent contributing factor.

In the Prototyping phase, the theoretical framework integrating RME, PBL, and STEM was elaborated, and the six-syntax KOSMARA-STEM model was conceptualised, with each instructional component aligned with Joyce and Calhoun's (2024) framework: syntax, social system, principles of reaction, support system, and instructional and nurturant effects. Instructional materials, including contextual problem scenarios, student worksheets, teacher guides, and digital learning tools, were designed to operationalise each syntax. Expert validation was conducted against Nieveen's (1999) quality criteria of validity, practicality, and effectiveness (Plomp & Nieveen, 2013; Fauzan et al., 2024). Four validators, namely two mathematics education specialists and two elementary school practitioners, evaluated all components using structured four-point rubrics. Validators were selected on the basis of their expertise and experience: the specialists had academic backgrounds and research experience in mathematics education, including realistic or contextual mathematics approaches, and the practitioners were experienced elementary school teachers familiar with Grade V mathematics and the Kurikulum Merdeka. The materials were then pilot-tested with 30 Grade V students who did not participate in the main study. The pilot focused on the clarity of instructions and problem wording, the time required for each phase, and the functioning of the NLT items (item–total correlations and internal consistency). Based on the pilot, ambiguous problem statements and instructions on the worksheets were reworded, the time allocation for several phases was adjusted, and NLT items that students found unclear were revised before the main trials.

The Assessment phase involved full classroom implementation in all 20 participating schools in two stages: a small-scale trial (5 schools, n = 140 Grade V students) and a large-scale trial (15 schools, n = 420 Grade V students), giving a total of N = 560 students. Each stage comprised eight instructional sessions of 60 minutes each, covering Grade V numeracy topics aligned with the test domains, namely percentages, ratio and proportion, data presentation and interpretation, and measurement. Before each stage, participating teachers attended a preparation workshop of one day (6 hours) on the model's phases and materials and received a teacher guide with lesson plans for every session. To monitor implementation consistency, two trained observers per classroom completed the MIOS and SAOS in every session. The numeracy literacy test was administered before and after implementation to all participants at both trial stages; effectiveness analyses are reported for the large-scale trial (N = 420) because the small-scale trial served a formative purpose. Its results were used to revise the model, so its participants experienced an earlier version of the intervention, and pooling the two stages would mix outcomes from different model versions. In line with Plomp's summative evaluation objectives, this phase synthesised quantitative and qualitative data to assess the model's validity, practicality, and effectiveness, and the small-scale trial findings informed refinements before the large-scale implementation (Plomp & Nieveen, 2013).

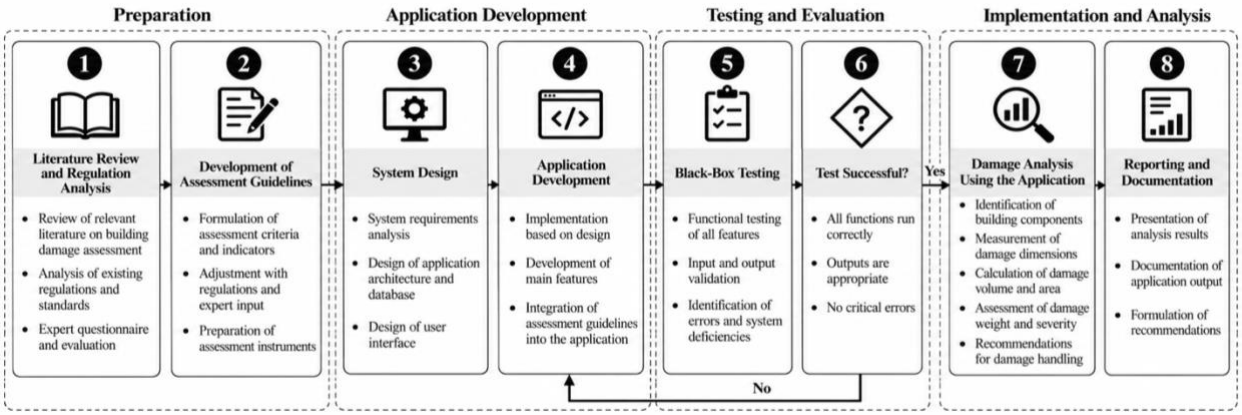


Figure 1. Research method flow

The KOSMARA-STEM Instructional Model

The KOSMARA-STEM model comprises six sequential, theoretically grounded syntaxes, as described below and summarised in Table 1. The name is an acronym of the Indonesian labels of the phases, namely Kontekstual (Contextual Problem), Own Production, Strategi (Strategy Discussion), Modelling, Abstraksi (Formalisation), and Refleksi-Aplikasi (Reflection and Reinvention), combined with STEM. Pedagogically, it denotes a cycle in which an authentic STEM context is progressively transformed, through students' own strategies, into formal mathematics.

Syntax 1: Contextual Problem: The teacher presents a rich, real-world mathematical situation using multimodal representations (images, video, narrative text, or manipulatives). The context is deliberately chosen to be culturally familiar, cognitively challenging, and mathematically rich. Guiding questions prompt students to observe, question, and identify the salient mathematical elements of the context. This syntax operationalises RME's use-of-contexts principle and the problem-activation phase of PBL (Van den Heuvel-Panhuizen & Drijvers, 2020; Ibad et al., 2025).

Syntax 2: Own Production: Students, individually and in small groups, generate their own strategies and solution pathways for the contextual problem using the available tools (calculators, sketching materials, and digital applications). Teacher facilitation is deliberately non-directive to promote mathematical autonomy and divergent thinking. This syntax enacts RME's student-contributions principle and the guided reinvention philosophy central to progressive mathematisation (Van den Heuvel-Panhuizen & Drijvers, 2020; Risdiyanti et al., 2024).

Syntax 3: Strategy Discussion and Comparison: In a teacher-orchestrated whole-class discussion, the diverse strategies produced in Syntax 2 are presented, compared, and evaluated. Students explain their approaches, respond to peers' questions, and collectively assess the efficiency and correctness of different solution pathways. This syntax enacts RME's interactivity principle and the collaborative sense-making phase of PBL (Ibad et al., 2025; Risdiyanti et al., 2024).

Syntax 4: Mathematical Modelling: Students are guided to formalise their contextual reasoning into mathematical representations equations, graphs, diagrams, or tables that capture the underlying structure of the problem. Digital tools (e.g., spreadsheet applications and graphing utilities) support data analysis and model visualisation. Technology use alone, however, is not treated as STEM integration. The problems are drawn from science and engineering contexts in which students collect or analyse data (for example, collecting, organising, and comparing data from simple observations or measurements of everyday phenomena) and use their mathematical models to make or justify a design decision (for example, planning the size, proportions, or arrangement of a simple product or space using measurements, ratios, and percentages), so that science, technology, engineering, and mathematics each play an explicit role (English, 2023; Kristensen et al., 2024).

Syntax 5: Formalisation: With teacher guidance, students abstract general principles and formal mathematical concepts from the models developed in Syntax 4, connecting informal contextual reasoning to conventional mathematical knowledge. This syntax operationalises RME's vertical mathematisation and supports the development of durable conceptual understanding (Van den Heuvel-Panhuizen & Drijvers, 2020; Samritin, 2023).

Syntax 6: Reflection and Reinvention: Students reflect metacognitively on their problem-solving processes and conceptual gains and then apply the newly formalised knowledge to novel, more complex problem contexts. This syntax is intended to promote learning transfer, metacognitive regulation, and the reinvention of mathematical knowledge in new domains, which are core goals of both RME and PBL; these are design intentions and were not measured as separate outcomes in this study (Susanta et al., 2025; Suryanti et al., 2024).

Table 1. Summary of KOSMARA-STEM Model Syntaxes, Activities, and Theoretical Foundations

No.	Syntax	Key Classroom Activities	Theoretical Basis
1	Contextual Problem	Presentation of real-world problem via multimodal stimuli; guided questioning to activate prior knowledge and contextual understanding.	RME: Use of Contexts; PBL: Problem Activation
2	Own Production	Independent and collaborative strategy generation; open-ended exploration using diverse tools and representations.	RME: Student Contributions; Guided Reinvention
3	Strategy Discussion & Comparison	Whole-class orchestrated discussion; presentation, critique, and comparison of diverse solution strategies.	RME: Interactivity; PBL: Collaborative Sense-Making
4	Mathematical Modelling	Formalisation of contextual reasoning into mathematical representations; digital tool-supported data analysis.	RME: Use of Models; STEM: Science/Engineering Contexts and Technology-Supported Modelling
5	Formalisation	Abstraction of general mathematical principles; connection of informal to formal mathematical knowledge.	RME: Vertical Mathematisation; Conceptual Formalisation
6	Reflection & Reinvention	Structured metacognitive reflection; application of formalised knowledge to novel, complex problem contexts.	RME: Guided Reinvention; PBL: Metacognitive Reflection; STEM: Transfer

Instrument

Four instruments were employed to collect validity, practicality, and effectiveness data.

Numeracy Literacy Test (NLT): The researcher-developed NLT consists of seven items, one of which has two sub-items (7a and 7b), yielding eight scorable responses with a maximum total score of 100. Although the number of items is limited, each item is a multi-step constructed-response task, and together the items cover three of the four AKM content domains (Number, Data, and Geometry & Measurement; Algebra was not assessed) and all three cognitive levels (Table 2). Because some domains are represented by only one or two items, item-level results are interpreted descriptively and with caution. The items are mapped to three cognitive levels of the Indonesian AKM framework understanding (L1), application (L2), and reasoning (L3) and assess numeracy across five contextual domains: proportional reasoning, data interpretation, statistical representation, spatial measurement, and mathematical communication. Content validity was examined by expert validators (n = 3), who rated the relevance of each item on a four-point scale. The item-level CVI (I-CVI) was computed as the proportion of experts giving a rating of 3 or 4, and the scale-level CVI (S-CVI/Ave) as the mean of the I-CVIs (Polit & Beck, 2006). The S-CVI/Ave was 0.91, exceeding the recommended minimum of 0.90. Because Polit and Beck (2006) recommend an I-CVI of 1.00 when three experts are used, items that did not reach this value were revised on the basis of the experts' comments before the pilot test. Internal consistency in the pilot test was good (Cronbach's $\alpha = 0.83$). Construct-related evidence is limited to this content-alignment procedure and internal consistency; no factor-analytic evidence was obtained, which is acknowledged as a limitation. Table 2 presents the item specifications.

Student Response Questionnaire (SRQ): An 11-item, four-point Likert-type instrument (1 = Strongly Disagree; 4 = Strongly Agree) that assesses students' perceptions of the KOSMARA-STEM model in terms of instructional clarity, cognitive engagement, collaborative learning, and motivational impact (CVI = 0.89; Polit & Beck, 2006; Cronbach's $\alpha = 0.81$).

Model Implementation Observation Sheet (MIOS): An 18-item, observer-completed rating scale (1 = Not Implemented; 4 = Fully Implemented) that assesses teachers' fidelity to each of the six syntaxes through three behavioural indicators per syntax. Two trained observers independently rated every item in each observed session. Inter-rater agreement was computed on these paired item-level ratings using Cohen's κ, which is appropriate for two raters assigning categorical ratings

because it corrects for chance agreement, and indicated almost perfect agreement (κ = 0.87; Landis & Koch, 1977), and content validity was established (CVI = 0.92; Polit & Beck, 2006).

Student Activity Observation Sheet (SAOS): An 18-item instrument parallel to the MIOS that assesses student engagement and activity quality across the six syntaxes. Agreement was computed in the same way as for the MIOS and indicated almost perfect agreement (Cohen's κ = 0.84; Landis & Koch, 1977), and content validity was established (CVI = 0.90; Polit & Beck, 2006).

Table 2. Numeracy Literacy Test Item Specifications

Item	Numeracy Indicator	Real-World Context	AKM Domain	Cognitive Level	Max. Score
1	Applying percentage in real-world transactions	Book purchase with 25% discount	Number	L2	10
2	Representing and interpreting ratio	Gender ratio in extracurricular activity	Number	L2	10
3	Constructing a strip/bar chart from data	Student subject preference survey	Data	L2	15
4	Interpreting a pie chart and computing totals	Fruit inventory in a shop	Data	L2	10
5	Distributing quantities using percentage scheme	Prize distribution in competition (50%:30%:20%)	Number	L3	15
6	Solving proportional reasoning problems	Swimming pool depth ratio (3:5)	Geometry & Measurement	L3	15
7a	Applying multiplication in measurement context	Everyday length measurement involving repeated equal lengths	Geometry & Measurement	L1	10
7b	Applying subtraction to measurement problem	Pencil cutting: original 18 cm, remaining 12 cm	Geometry & Measurement	L1	15

Data Collection and Analysis Plan

Data were collected in three phases, following the pre-implementation, implementation, and post-implementation structure recommended for R&D effectiveness studies (Plomp & Nieveen, 2013). Pre-implementation data comprised NLT pre-test scores, baseline classroom observation field notes, and teacher interview transcripts from the Preliminary Research phase. Implementation data comprised MIOS and SAOS ratings collected during all eight instructional sessions in all 20 participating schools; before data collection, observer pairs were trained to criterion (Cohen's κ ≥ 0.80) to ensure rating consistency. Post-implementation data comprised NLT post-test scores and SRQ responses, which were collected within one week of the final session to minimise response decay. No delayed post-test was administered; the data therefore describe immediate post-intervention performance and do not allow conclusions about the retention of learning gains.

Validity was assessed using the mean expert rating (scale 1–4) and categorised according to Nieveen's (1999) criteria: very valid (≥3.50), valid (2.50–3.49), less valid (1.50–2.49), and invalid (<1.50); these criteria are widely used to validate educational products in mathematics education R&D (Plomp & Nieveen, 2013; Fauzan et al., 2024). Practicality was evaluated using the mean MIOS and SAOS scores, categorised on the same scale. Student responses were analysed descriptively, with ratings of 3 or 4 classified as positive. Pretest–posttest change (the "effectiveness" criterion in Nieveen's framework) was examined using paired-samples t-tests and normalised gain (N-Gain) scores, calculated as N-Gain = (post-test − pre-test) / (maximum score − pre-test) and classified according to Hake (1998) as high (≥0.70), medium (0.30–0.69), or low (<0.30); this N-Gain framework is widely used in recent Indonesian mathematics education effectiveness studies (e.g., Maryam et al., 2026). N-Gain was calculated for each student and then averaged (mean individual N-Gain); for transparency, the gain computed from the group means is also reported. Because there was no comparison group, these analyses describe within-group change and do not estimate causal

effects of the model. The effect size for the paired comparison was computed as Cohen's dz, that is, the mean pre–post difference divided by the standard deviation of the difference scores ($dz = t/\sqrt{n}$; Lakens, 2013), and interpreted using Cohen's (1988) benchmarks. Cohen's dz expresses standardised within-group change; it is not comparable to between-group effect sizes and does not indicate comparative effectiveness. The distribution of difference scores was inspected using histograms and Q–Q plots; given the large sample (N = 420), the paired-samples t-test is robust to moderate departures from normality. Ninety-five percent confidence intervals (CIs) are reported for the mean difference, the mean N-Gain, and dz. Effectiveness analyses included only students with complete pretest and posttest data (N = 420). Because the 420 students were nested in 15 schools, the independence assumption of the t-test is not fully met. Therefore, a design-effect sensitivity analysis was conducted in which the t statistic was divided by $\sqrt{DEFF}$, where $DEFF = 1 + (m - 1)\rho$, $m \approx 28$ students per school, and the intraclass correlation ρ was set to conservative values of 0.10, 0.20, and 0.30, with degrees of freedom based on the number of schools (df = 14). This sensitivity analysis was used as an approximation; a full multilevel model with students nested in schools is recommended for subsequent analyses. All analyses were conducted in IBM SPSS Statistics 27.

RESULTS AND DISCUSSION

Results

Model and Instrument Validity

All model components and research instruments were rated very valid (≥3.50), meeting the quality thresholds established by Nieveen (1999) and widely applied in recent Indonesian mathematics education R&D (Fauzan et al., 2024; Samritin, 2023). Because model components and research instruments serve different purposes, their ratings are reported separately. The six model syntaxes obtained a mean rating of M = 3.74 (SD = 0.08 across syntaxes), ranging from M = 3.60 (Formalisation) to M = 3.85 (Contextual Problem), and the four instruments obtained M = 3.77 (SD = 0.04 across instruments). Across all ten components the mean was M = 3.75; this SD of 0.07 reflects variability among the ten component means, whereas agreement among validators was high (validator means 3.74–3.76). These ratings constitute expert-judgement evidence of content relevance and construct representation. They are complemented by the CVI, reliability, and inter-rater agreement indices reported in the Method section, but they do not substitute for empirical construct validation, which remains a task for future research. Table 3 summarises the validity ratings for all model components and instruments.

The validators' qualitative feedback highlighted three strengths of the KOSMARA-STEM model: (a) the coherent theoretical alignment between RME principles and STEM integration, consistent with the integrated model design advocated by Risdiyanti et al. (2024); (b) the progressive scaffolding across syntaxes from contextual to formal mathematical reasoning, reflecting RME's mathematisation principles (Van den Heuvel-Panhuizen & Drijvers, 2020); and (c) the cultural relevance of the contextual problems for West Java elementary schools, in line with the contextualisation approach recommended for Indonesian numeracy interventions (Nusantara et al., 2022). The validators recommended strengthening the Formalisation syntax by providing explicit teacher guidance on connecting students' informal models to curriculum-aligned formal notation, and this revision was incorporated into the final model. Validator comments were analysed by grouping written remarks into recurring categories (strengths and suggested revisions); the three strengths above were the categories that recurred most consistently across the validators' comments. In their written comments, the validators noted, for example, that the contextual

problems were close to students' daily experiences and that the sequence of phases helped students move gradually from informal strategies to formal mathematical notation.

Table 3. Expert Validity Ratings for KOSMARA-STEM Model Componentsand Research Instruments

Component	V1	V2	V3	V4	Mean
Contextual Problem Syntax	3.90	3.80	3.85	3.83	**3.85**
Own Production Syntax	3.75	3.80	3.70	3.75	**3.75**
Strategy Discussion & Comparison Syntax	3.80	3.70	3.75	3.80	**3.76**
Mathematical Modelling Syntax	3.70	3.75	3.65	3.70	**3.70**
Formalisation Syntax	3.60	3.55	3.65	3.60	**3.60**
Reflection & Reinvention Syntax	3.75	3.70	3.80	3.75	**3.75**
Numeracy Literacy Test	3.85	3.80	3.75	3.85	**3.81**
Student Response Questionnaire	3.70	3.75	3.70	3.75	**3.73**
Model Implementation Observation Sheet (MIOS)	3.80	3.85	3.75	3.80	**3.80**
Student Activity Observation Sheet (SAOS)	3.75	3.70	3.80	3.75	**3.75**
Model syntaxes (mean of 6)	3.75	3.72	3.73	3.74	**3.74**
Research instruments (mean of 4)	3.77	3.77	3.75	3.79	**3.77**
OVERALL MODEL VALIDITY	**3.76**	**3.74**	**3.74**	**3.76**	**3.75**

Note. V1–V4 = Validators 1–4. Validity scale: 1 = Invalid; 2 = Less Valid; 3 = Valid; 4 = Very Valid. Ratings ≥3.50 indicate Very Valid status (Nieveen, 1999). Subtotals are reported separately for the six model syntaxes and the four instruments.

Validation proceeded in two stages: the initial prototype was rated very valid (M = 3.68; Nieveen, 1999), and the ratings in Table 3 (M = 3.75) were obtained after the validators' suggested revisions, mainly to the Formalisation syntax, had been incorporated. The small-scale trial (n = 140 Grade V students in 5 schools) then yielded an initial implementation fidelity of 79.4%. Qualitative feedback from the small-scale trial led to three targeted revisions: (1) structured teacher prompts were added to the Formalisation syntax to support the transition from informal to formal mathematical notation; (2) digital tool guidance in the Modelling syntax was clarified with worked examples; and (3) the contextual problems in Syntaxes 1 and 4 were adjusted to better reflect the socioeconomic contexts of rural districts in Cirebon Region III. These revisions were incorporated before the large-scale implementation, in keeping with the iterative refinement prescribed by Plomp's (2013) prototyping approach.

Practicality of Implementation

Table 4. Session-by-Session MIOS and SAOS Mean Scores Across Large-Scale Trial

Session	MIOS Mean	Implementation Fidelity (%)	SAOS Mean
1	3.16	79.0	2.94
2	3.27	81.8	3.05
3	3.34	83.5	3.26
4	3.42	85.5	3.37
5	3.48	87.0	3.42
6	3.52	88.0	3.51
7	3.60	90.0	3.57
8	3.67	91.7	3.61
Mean (Sessions 1–8)	3.43	85.8	3.34

Note. MIOS = Model Implementation Observation Sheet; SAOS = Student Activity Observation Sheet (scale 1–4). Implementation fidelity (%) = MIOS mean / 4 × 100.

MIOS data from the large-scale trial showed a progressive increase in implementation fidelity across the eight instructional sessions (Table 4). Mean MIOS scores rose from M = 3.16 (Session 1) to M = 3.67 (Session 8). Averaged across all eight sessions, the mean MIOS score was M = 3.43 and mean implementation fidelity was 85.8% (SD = 4.2 across sessions; range 79.0%–91.7%), which falls in the practical category. Fidelity was summarised across all participating schools for each session;

school-level fidelity was not analysed separately. Fidelity reached 91.7% in the final session, a trajectory consistent with the learning curve reported for the adoption of new instructional approaches (Fauzan et al., 2024; Fry et al., 2025). The Contextual Problem and Own Production syntaxes showed the highest fidelity from the early sessions (M ≥3.50 by Session 2), possibly reflecting teachers' greater familiarity with presenting contextualised problems, although this interpretation was not tested with teacher-level data. The Formalisation syntax started at the lowest level and showed the largest descriptive increase, from M = 2.83 (Session 1) to M = 3.58 (Session 8); these session-level trends are descriptive and were not tested statistically. This pattern is in line with prior RME professional development research identifying formal mathematisation as the most challenging phase for teachers to implement (Risdiyanti et al., 2024).

SAOS data followed the same pattern, with student engagement increasing from M = 2.94 in the first session to M = 3.61 in the final session. Student participation in Strategy Discussion and Comparison, assessed through the quality of peer argumentation and the frequency of student-initiated comparisons, showed the largest improvement across sessions (Δ = 0.89 scale points). This indicates increasing observed participation in mathematical discussion with repeated engagement in the collaborative discussion structure. Because discourse competence was not assessed directly, this trend should not be interpreted as evidence of competency development, although it is consistent with the documented social learning benefits of problem-based learning (Ibad et al., 2025).

Student Responses

SRQ responses from the large-scale trial (N = 420) were analysed by classifying each item response of 3 (Agree) or 4 (Strongly Agree) as positive; the overall percentage was computed as the number of positive responses divided by all responses (420 students × 11 items = 4,620 responses). Overall, 89.5% of responses were positive, indicating high perceived acceptability (Plomp & Nieveen, 2013). Accordingly, negative responses (ratings of 1 or 2) accounted for 10.5% of all responses. At the item level, the highest positive response rates were for Item 4 (freedom to express mathematical ideas: 92.3%), Item 7 (teacher responsiveness to student questions: 91.8%), and Item 10 (collaborative learning and peer respect: 90.6%), indicating that students perceived the classroom as open and interactive, which is consistent with RME's interactivity principle; these are self-reported perceptions rather than direct evidence of the learning climate (Van den Heuvel-Panhuizen & Drijvers, 2020). The lowest, though still positive, rates were for Item 8 (increased motivation to study mathematics: 83.7%) and Item 11 (ability to apply learning creatively: 82.5%). One hypothesis, which could not be tested with the present data, is that changes in motivation require a longer implementation period.

Effectiveness: Numeracy Literacy Gains

Table 5 presents descriptive statistics for the pre-test and post-test NLT scores in the large-scale trial (N = 420). The mean score increased from M = 62.4 (SD = 11.3) on the pre-test to M = 82.7 (SD = 8.9) on the post-test. A paired-samples t-test showed that this observed gain was statistically significant, $t(419) = 32.47$, $p < .001$ (two-tailed), mean difference = 20.3 points, 95% CI [19.1, 21.5], with a large standardised within-group change (Cohen's dz = 1.58, 95% CI [1.44, 1.72]; Cohen, 1988). Because the design had no comparison group, this value describes how much students' scores changed, not how much more effective the model is than other instruction. In the design-effect sensitivity analysis accounting for the clustering of students in 15 schools, the gain remained statistically significant even under a conservative intraclass correlation of 0.30 (DEFF = 9.1; effective N ≈ 46; adjusted $t = 10.76$, $df = 14$, $p < .001$). This large within-group change should be interpreted cautiously. Besides the relatively low pre-test baseline (M = 62.4), which left substantial room for improvement, several alternative explanations cannot be excluded in an uncontrolled design: testing

effects from taking the test twice, maturation and regular instruction during the implementation period, and regression to the mean among lower-scoring students. Large gains have also been reported in recent Indonesian elementary numeracy interventions (Maryam et al., 2026; Wewe et al., 2026). The mean of the individual students' N-Gain scores was 0.54 (SD = 0.13; 95% CI [0.53, 0.55]), which falls in the medium category of Hake's (1998) classification; the N-Gain computed from the group means, (82.7 − 62.4)/(100 − 62.4), was also 0.54. At the item level, which should be read descriptively because each content area is represented by a single item or item pair, the largest gains were on the data representation item (Item 3: Δ = +31.2 percentage points) and the proportional reasoning item (Item 6: Δ = +27.8 percentage points). The measurement operation items (Items 7a–7b) showed the smallest absolute gain (Δ = +15.3 percentage points) because of their high pre-test baseline (M = 79.4); relative to the room available for improvement, however, this gain was high (N-Gain = 0.74).

Table 5. Descriptive Statistics and Effectiveness Indices for Numeracy Literacy Test Scores: Large-Scale Trial (N = 420)

Item/Domain	Pre-Test M	Pre-Test SD	Post-Test M	Post-Test SD	N-Gain	p
Percentage Application (Item 1)	64.2	14.3	83.1	10.2	0.53	< .001
Ratio & Fractions (Item 2)	60.8	12.7	79.4	9.8	0.47	< .001
Data Representation (Item 3)	51.6	16.1	82.8	11.4	0.64	< .001
Pie Chart Interpretation (Item 4)	58.3	13.9	80.6	10.7	0.53	< .001
Percentage Distribution (Item 5)	55.0	17.2	78.3	12.1	0.52	< .001
Proportional Reasoning (Item 6)	53.1	15.8	80.9	9.9	0.59	< .001
Measurement Operations (Items 7a–7b)	79.4	9.4	94.7	5.8	0.74	< .001
OVERALL	**62.4**	11.3	**82.7**	8.9	**0.54**	**< .001**

Note. N-Gain classification (Hake, 1998): ≥0.70 = High; 0.30–0.69 = Medium; <0.30 = Low. Overall paired-samples t-test: t(419) = 32.47, p < .001, Cohen's dz = 1.58. Item-level p values are from paired-samples t-tests; with a Bonferroni-adjusted α of .05/7 = .007, all item-level differences remained significant. 95% CIs: mean difference [19.1, 21.5]; mean N-Gain [0.53, 0.55]; dz [1.44, 1.72].

Discussion

The findings indicate that KOSMARA-STEM met the expert-validity and implementation-feasibility criteria and that Grade V students in Cirebon Region III showed substantial observed gains in numeracy literacy while learning with the model. The overall N-Gain of 0.54 and the large within-group standardised change (Cohen's dz = 1.58) indicate substantial pre–post improvement, but, as noted above, part of this change may reflect testing effects, maturation, and regression to the mean rather than the model itself. The direction, though not the magnitude, of this finding is consistent with meta-analytic evidence of a positive effect of RME on elementary students' mathematical abilities (g = 0.63 across 10 studies; Samritin, 2023) and with documented improvements in literacy and numeracy following RME instruction in Grade III–V classrooms (Fauzan et al., 2024). A two-decade meta-analysis of 38 RME studies likewise reported a large overall effect (ES = 0.97), moderated by intervention length and education level (Juandi et al., 2022). However, because within-group pre–post effect sizes are typically larger than between-group effects, the present estimate should not be compared directly with such meta-analytic benchmarks, which are largely derived from controlled between-group comparisons. Similarly, a meta-analysis of STEM education reported a moderate overall cognitive effect (d = 0.58) across elementary and secondary contexts (Cao et al.,

2025); because that estimate is a between-group effect and the present estimate is a within-group change, the two are not comparable, and the absence of a comparison group means that the present gain cannot be attributed specifically to the integration of RME, STEM, and PBL, and controlled comparisons with STEM-only instruction are needed. Large effects have also been reported for project-based STEM learning with Grade V Indonesian students (between-group d = 0.914; experimental N-Gain = 0.68; Maryam et al., 2026), although differences in design (between-group versus pre–post comparison) and sample size (N = 46 versus N = 420) call for caution in direct comparison. Reviews of STEM integration in teacher education also emphasise embedding STEM within a structured, pedagogically coherent framework rather than implementing it as discrete activities (Polatuly et al., 2026), and recent work on rural STEMM education stresses approaches that respond to the specific characteristics of rural and resource-constrained schools a consideration directly relevant to Cirebon Region III (Starrett et al., 2026). A possible but untested explanation concerns the quality of teacher–student mathematical discourse: structured professional development that promotes high-level teacher talk in primary mathematics improves student outcomes (Brandel et al., 2026), and Syntax 3 was designed to elicit such talk. However, discourse quality was not measured as a mediator in this study, so this remains a theoretical explanation rather than a demonstrated relationship. In addition, the digital modelling tools in Syntax 4 and the structured metacognitive reflection in Syntax 6 were designed to provide representational and reflective resources that are often absent from conventional single-approach instruction (English, 2023; Kristensen et al., 2024; Ibad et al., 2025).

Descriptively, the largest gains were observed on the data representation item (N-Gain = 0.64) and the proportional reasoning item (N-Gain = 0.59). Because each of these areas was measured by a single item, these differences cannot be treated as reliable domain-level effects; they are, however, in the direction predicted by RME's progressive mathematisation. Both competencies are central targets of the Contextual Problem, Mathematical Modelling, and Formalisation syntaxes, in which students progressively move from informal contextual reasoning to formal mathematical representation (Van den Heuvel-Panhuizen & Drijvers, 2020). Students' ability to move flexibly and adaptively between informal and formal strategies is among the strongest predictors of mathematical competence, which is consistent with the mechanism hypothesised for the Formalisation syntax (Verschaffel et al., 2024). This interpretation is supported by quasi-experimental evidence that, under RME instruction, mathematical representation significantly mediates the relationship between students' mathematical beliefs and their arithmetic problem solving (Yuanita et al., 2018), which offers a theoretical reason to expect representation-intensive phases such as Modelling and Formalisation to matter; because no component-specific analysis was conducted, however, the present data cannot link individual syntaxes to particular gains. Similarly, the STEM-based digital tools in Syntax 4 were intended to support data representation by providing technology-mediated visualisation resources that extend students' engagement with contextual mathematical reasoning beyond procedural computation (Ow-Yeong et al., 2023; Risdiyanti et al., 2024); however, no independent technology effect was measured. This pattern is consistent with the view that data representation is a primary site in which mathematics functions both as a computational tool and as a learning objective within STEM activities (Kristensen et al., 2024). With respect to specific content, a PMRI-based learning trajectory for Indonesian elementary students showed that progressing from a concrete cultural context to bar models and 100-grids helped students connect fractions, decimals, and percentages (Marleny et al., 2026). This progression mirrors the Modelling and Formalisation syntaxes of KOSMARA-STEM and is consistent with the observed gains on the percentage-based items (Items 1 and 5).

Within the large-scale trial, implementation fidelity increased from 79.0% in Session 1 to 91.7% in Session 8 (mean 85.8%); the small-scale trial value (79.4%) is not directly comparable

because it involved a different sample and an earlier version of the model. Because fidelity was summarised across schools rather than by school, variation in implementation between schools could not be examined and should be analysed in future studies. One plausible interpretation is that this trajectory reflects a learning curve associated with adopting a substantially different pedagogical approach. Alternative explanations include observers' growing familiarity with the instrument, teachers' awareness of being observed, and differences in the content demands of later sessions, which cannot be disentangled with the present data. Teachers transitioning to inquiry-based, student-centred approaches require structured professional learning support, and their fidelity typically improves progressively across implementation cycles rather than immediately (Fry et al., 2025). The Formalisation syntax, which requires teachers to make explicit the mathematical generalisations implicit in students' informal models, was consistently the most challenging to implement. This is consistent with evidence that orchestrating the transition from informal to formal knowledge demands advanced pedagogical content knowledge that develops incrementally through guided practice (Risdiyanti et al., 2024; Fauzan et al., 2024). Because teacher characteristics (e.g., teaching experience and prior exposure to RME or STEM) were not analysed in relation to fidelity in this study, variation in implementation cannot be attributed to teachers' professional backgrounds. Meta-analytic evidence that STEM teachers' interdisciplinary teaching abilities differ by teaching seniority, subject specialisation, and prior exposure to interdisciplinary learning (Wu et al., 2024) suggests that such characteristics should be recorded and examined in future implementations. In practical terms, schools adopting KOSMARA-STEM should provide targeted pre-implementation professional development focused on the Formalisation syntax, supported by peer coaching and structured post-lesson reflection to accelerate fidelity development (Fry et al., 2025; Brandel et al., 2026).

The highly positive student responses (89.5% positive; N = 420) merit contextualisation. Primary students' engagement in mathematics is closely linked to their motivation and mathematics self-efficacy, and procedural-only instruction disconnected from daily experience has been identified as a key barrier in Indonesian classrooms (Hidayatullah et al., 2024; Wijaya et al., 2024). The model's deliberate use of culturally familiar, real-world problem contexts including local market transactions, regional recreational settings, and school-based surveys may have contributed to students' perception that mathematics was relevant to their lives, as reflected in their positive questionnaire responses. Numeracy engagement improves markedly when home and school mathematical contexts are meaningfully connected, a condition that the contextualised design of KOSMARA-STEM directly addresses (Graven & Jorgensen, 2024). Similarly, a culturally grounded ethnomathematics approach in Indonesian Grade V classrooms increased active participation, teacher–student interaction, and engagement in mathematics alongside numeracy gains (Wewe et al., 2026). That study, however, used a different cultural approach, a smaller sample, and a comparison-group design, so its findings support the plausibility of the present results rather than confirming them. The relatively lower positive response rates for the motivation-related items (Items 8 and 11; 83.7% and 82.5%, respectively) indicate that students were somewhat less positive about motivational aspects than about classroom interaction. Because intrinsic motivation was not measured with a dedicated motivation scale or assessed over time, no motivational effect of the model can be inferred. Integrated instructional approaches have been shown to enhance student engagement alongside conceptual understanding and problem solving (Gebremeskel et al., 2025); whether such engagement translates into durable motivation should be examined in longitudinal research.

The three integrated theoretical traditions offer a proposed explanatory framework for the observed gains, which should be tested rather than assumed. From an RME perspective, the six-syntax sequence from contextual problem engagement through guided reinvention to formal generalisation is designed to provide the scaffolding that students need to construct durable

mathematical understanding (Van den Heuvel-Panhuizen & Drijvers, 2020; Risdiyanti et al., 2024). Strategy flexibility, the capacity to move adaptively between informal and formal approaches, is the core mechanism through which progressive mathematisation produces durable competence gains (Verschaffel et al., 2024), which provides a theoretical rationale for the explicit scaffolding of this transition in Syntaxes 4 and 5; whether these syntaxes contributed more than others could not be determined, because no syntax-specific outcome data were collected. From a STEM perspective, positioning mathematics as both a computational tool and a primary learning objective within authentic problem contexts is intended to increase student engagement with formal mathematical representations, consistent with the framework describing the dual roles of mathematics in STEM activities (Kristensen et al., 2024). Maass et al. (2023) argue that modelling tasks connected to real-world citizenship contexts can increase students' investment in formal mathematical reasoning; however, citizenship-related outcomes were not measured in the present study, so this remains a design rationale rather than a finding. From a PBL perspective, the collaborative strategy exploration in Syntaxes 2 and 3 was accompanied by increasing observed student participation in discussion (SAOS), although discourse competence was not assessed as a learning outcome; this design is consistent with meta-analytic evidence of a large effect of PBL on students' problem-solving skills (Ibad et al., 2025). Experimental evidence from Indonesian Grade V classrooms further shows that combining PBL with an interactive mathematics application significantly improves both mathematical reasoning and collaboration compared with conventional instruction (Yanuarto et al., 2025), which is compatible with, but does not by itself validate, the pairing of collaborative problem exploration with digital tools across Syntaxes 2–4.

Taken together, the theoretical contribution of KOSMARA-STEM lies in specifying a division of labour among the three traditions within one lesson cycle (PBL for organising problem engagement and collaboration, STEM for supplying authentic interdisciplinary contexts and modelling tools, and RME for the progressive formalisation of students' strategies) rather than in demonstrating that the combination is superior to its components. The present data support the feasibility of this sequence and show that students' numeracy scores rose while it was used, but several alternative explanations remain plausible: the additional structure and novelty of the lessons, the extra attention associated with classroom observation, increased time on numeracy tasks, practice with the test format, and regression to the mean. Distinguishing the contribution of the integrated sequence from these factors requires comparison-group and component-level designs.

Implications

Given the uncontrolled design, the implications of these findings concern mainly classroom practice and the design of further research rather than educational policy. The AKM numeracy framework requires students to demonstrate contextualised mathematical reasoning across personal, societal, and scientific domains (Kemdikbudristek, 2023; Muhaimin et al., 2024); the NLT items on which students improved most (data representation and proportional reasoning) address part of this framework, but the seven-item test does not cover the full range of AKM competencies. Numeracy education is also increasingly expected to develop citizenship-oriented mathematical reasoning (Geiger et al., 2023); the contextual and reflective syntaxes of KOSMARA-STEM were designed with this goal in mind, but citizenship-related outcomes were not assessed and remain to be examined. The instrument suite accompanying the model (NLT, SRQ, MIOS, and SAOS) can be used by teachers for preliminary classroom assessment and for monitoring implementation; using it to compare schools or to monitor numeracy across districts would first require evidence of measurement equivalence across school contexts (Purbaningrum et al., 2024; Wijaya et al., 2024). For teacher preparation and professional development, the results suggest three actionable points: (a) teachers need explicit preparation in orchestrating the Formalisation phase, which had the lowest initial fidelity; (b) lesson materials should include worked examples showing how students' informal

strategies can be connected to formal notation; and (c) peer observation using the MIOS can provide structured feedback during the first sessions of implementation, when fidelity was lowest. These recommendations also respond to the need for context-sensitive mathematics education research in Southeast Asian settings (Yang et al., 2025).

Limitations and Suggestions for Future Research

Several limitations should be considered when interpreting these findings. First, the study used a single-group pretest–posttest design without any comparison group. This leaves the main threats to internal validity uncontrolled, namely history (other instructional or school events during implementation), maturation, testing, instrumentation, regression to the mean, and novelty or observer effects, so the observed gains cannot be causally attributed to the model. In particular, repeated administration of the NLT may have raised scores through familiarity with the items, and regression to the mean may have contributed to larger gains among students with low pretest scores; both would inflate the apparent within-group effect. Second, the eight-session implementation was relatively brief and the post-test was administered within one week of the final session; the findings therefore describe short-term within-group improvement, and no conclusions about retention or transfer can be drawn (Shimizu & Kang, 2025). Third, schools were selected purposively from a single geographical area (Cirebon Region III), and participation depended on schools' willingness; this selection procedure may favour schools that are receptive to innovation and further limits external validity beyond the issue of geographic scope (Cevikbas et al., 2024). Fourth, and most importantly for the statistical conclusions, students were nested within classes and 15 schools, but the primary analyses were conducted at the student level. This is a serious analytical limitation because ignoring clustering underestimates standard errors; although the design-effect sensitivity analysis indicated that the pre–post difference remained significant under conservative intraclass correlations, a full multilevel reanalysis is required to obtain accurate estimates. Fifth, the NLT contains seven items, so item-level and domain-level results rest on very few indicators, and construct validity evidence is limited to expert judgement and internal consistency.

Rather than repeating these limitations, future research should address the most important unresolved methodological weaknesses. First, a cluster-randomised or matched quasi-experimental design, with schools as the unit of assignment and KOSMARA-STEM compared with regular Kurikulum Merdeka instruction (and, ideally, an RME-only condition), should use an AKM-aligned numeracy test with broader item coverage as the primary outcome and multilevel models for analysis. Second, delayed post-tests administered approximately three and six months after implementation should measure retention (the difference between immediate and delayed scores) and transfer (performance on untaught contexts). Third, generalisability should be examined through stratified sampling of schools across provinces and urban and rural settings and across Grades III–VI, accompanied by tests of the measurement invariance of the NLT across contexts. Finally, professional development studies should specify the training dose (e.g., number of workshop hours, with or without in-class coaching), track MIOS fidelity by syntax across sessions, and relate the time needed to reach a fidelity criterion (e.g., ≥ 85%) to student outcomes.

CONCLUSION

This study designed, developed, and empirically evaluated KOSMARA-STEM, a six-syntax instructional model that integrates Realistic Mathematics Education (RME), problem-based learning, and STEM , and examined preliminary within-group changes in the numeracy literacy of Grade V elementary school students. The model was trialled in 20 schools in Cirebon Region III with 560 Grade V students in total (small-scale trial: 140 students in 5 schools; large-scale trial: 420 students

in 15 schools), and outcome analyses were based on the 420 students in the large-scale trial. KOSMARA-STEM was rated very valid by experts (model syntaxes: M = 3.74/4.00; instruments: M = 3.77/4.00), was implemented with a mean fidelity of 85.8% (91.7% in the final session), and received positive student perceptions (89.5% of responses). Students' numeracy scores increased from 62.4 at pretest to 82.7 at posttest (N-Gain = 0.54, medium; Cohen's dz = 1.58; $p < .001$). Because there was no control group, these observed improvements cannot be attributed causally to the model, and its effectiveness remains to be confirmed through controlled designs.

The study makes three contributions to the mathematics education literature. First, KOSMARA-STEM specifies how RME, PBL, and STEM can be sequenced within one elementary lesson cycle; the evidence for this contribution consists of expert validation and successful classroom implementation, not a demonstration that the integrated model is superior to its components. Second, the study provides preliminary evidence that an integrated RME–STEM–PBL sequence is feasible in regular classrooms and is associated with substantial within-group numeracy gains; the somewhat larger gains on the data representation and proportional reasoning items should be interpreted cautiously because each was measured by a single item; the specific contribution of each component remains to be tested through component analyses or controlled designs. Third, the fully documented model and its instrument suite (NLT, SRQ, MIOS, and SAOS) give teachers practical tools for preliminary classroom assessment and implementation monitoring, addressing a documented gap in implementation support for regional Indonesian schools under the Kurikulum Merdeka reform.

In practical terms, KOSMARA-STEM is compatible with the Kurikulum Merdeka and can be trialled by teachers within regular lessons. Schools in Cirebon Region III and comparable regional contexts can pilot the six-syntax structure and accompanying materials without substantial resources beyond teacher professional development. Such professional development should particularly target the Formalisation syntax, which showed the steepest learning curve in the large-scale trial and requires the most intensive pedagogical support. The two-stage trial showed that the model could be implemented across schools in five districts, which offers a starting point for, but not evidence of, scalability; larger controlled implementations are needed before district-wide adoption is recommended.

ACKNOWLEDGMENT

The authors express gratitude to thanks are expressed to 1) LPPM Universitas Terbuka as the funder of this research activity and 2) To all parties who have helped carry out this research activity.

AUTHOR CONTRIBUTIONS STATEMENT

Ardi Dwi Susandi contributed to conceptualization, methodology, instructional model development, investigation, data collection, formal analysis, and writing the original draft. Thesa Kandaga contributed to conceptualization, methodology, validation, data interpretation, supervision, and manuscript review and editing. Vivi Rindastri contributed to methodology, validation, investigation, data analysis, visualization, and manuscript review and editing. All authors contributed to the development and evaluation of the KOSMARA-STEM instructional model, critically reviewed the manuscript, and approved the final version for publication.

REFERENCES

Abtahi, Y., & Planas, N. (2024). Mathematics teaching and teacher education against marginalisation, or towards equity, diversity and inclusion. *ZDM - Mathematics Education, 56*(7), 1411-1425. https://doi.org/10.1007/s11858-024-01602-x

Ahmad, M., & Wilkins, S. (2024). *Purposive sampling in qualitative research: a framework for the entire journey*. Quality & Quantity. https://doi.org/10.1007/s11135-024-02022-5

Brandel, N., Deri, M., & Karsenty, R. (2026). The effectiveness of a principle-based professional development programme in promoting teachers' high-level talk in primary mathematics classrooms. *British Journal of Educational Psychology, 96*, 356-381. https://doi.org/10.1111/bjep.70020

Cao, X., Lu, H., Wu, Q., & Hsu, Y. (2025). Systematic review and meta-analysis of the impact of STEM education on students' learning outcomes. *Frontiers in Psychology, 16*, Article 1579474. https://doi.org/10.3389/fpsyg.2025.1579474

Cevikbas, M., Kaiser, G., & Schukajlow, S. (2024). Trends in mathematics education and insights from a meta-review and bibliometric analysis of review studies. *ZDM - Mathematics Education, 56*(2), 165-188. https://doi.org/10.1007/s11858-024-01587-7

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum Associates.

Creswell, J. W., & Plano Clark, V. L. (2022). *Designing and conducting mixed methods research (3rd ed.)*. SAGE Publications.

English, L. D. (2023). Ways of thinking in STEM-based problem solving. *ZDM - Mathematics Education, 55*(7), 1219-1230. https://doi.org/10.1007/s11858-023-01474-7

Fauzan, A., Harisman, Y., Yerizon, Suherman, Tasman, F., Nisa, S., Sumarwati, Hafizatunnisa, & Syaputra, H. (2024). Realistic mathematics education (RME) to improve literacy and numeracy skills of elementary school students based on teachers' experience. *Infinity Journal, 13*(2), 301-316. https://doi.org/10.22460/infinity.v13i2.p301-316

Freudenthal, H. (1968). Why to teach mathematics so as to be useful. *Educational Studies in Mathematics, 1*(1-2), 3-8. https://doi.org/10.1007/BF00426224

Fry, K., Nakar, S., & Zorn, K. (2025). Professional learning interventions for inquiry-based pedagogies in primary classrooms: a scoping review (2012-2022). *Mathematics Education Research Journal, 37*(4), 853-887. https://doi.org/10.1007/s13394-024-00516-x

Gebremeskel, A. A., Ayele, M. A., & Wondimuneh, T. E. (2025). Student engagement, conceptual-understanding, and problem-solving ability in learning plane geometry through an integrated instructional approach. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(5), em2634. https://doi.org/10.29333/ejmste/16391

Geiger, V., Gal, I., & Graven, M. (2023). The connections between citizenship education and mathematics education. *ZDM - Mathematics Education, 55*(5), 923-940. https://doi.org/10.1007/s11858-023-01521-3

Graven, M., & Jorgensen, R. (2024). Early numeracy opportunities through number stories with marginalised families. *ZDM - Mathematics Education, 56*(3), 319-333. https://doi.org/10.1007/s11858-023-01537-9

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. https://doi.org/10.1119/1.18809

Hidayatullah, A., Abidin, R., & Muqit, A. (2024). Motivation and behavioral engagement: the mediating role of mathematics self-efficacy in primary education. *Journal on Efficiency and Responsibility in Education and Science, 17*(3), 237-246. https://doi.org/10.7160/eriesj.2024.170306

Huang, Y., Zhou, Y., Chen, J., & Wu, D. (2024). Applying machine learning and SHAP method to identify key influences on middle-school students' mathematics literacy performance. *Journal of Intelligence, 12*(10), Article 93. https://doi.org/10.3390/jintelligence12100093

Ibad, Z., Sukestiyarno, Y. L., & Hidayah, I. (2025). Cross-disciplinary effects of problem-based learning on problem-solving skills in mathematics and beyond: a comprehensive meta-analysis and

bibliometric. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(3), em2578. https://doi.org/10.29333/ejmste/15862

Joyce, B., & Calhoun, E. (2024). *Models of teaching (10th ed.)*. Routledge. https://doi.org/10.4324/9781003455370

Juandi, D., Kusumah, Y. S., & Tamur, M. (2022). A meta-analysis of the last two decades of realistic mathematics education approaches. *International Journal of Instruction, 15*(1), 381-400. https://doi.org/10.29333/iji.2022.15122a

Kemdikbudristek. (2023). *Laporan hasil asesmen nasional 2022: potret pendidikan Indonesia [National assessment results report 2022: portrait of Indonesian education]*. Pusat Asesmen dan Pembelajaran, Kemdikbudristek.

Kristensen, M. A., Larsen, D. M., Seidelin, L., & Svabo, C. (2024). The role of mathematics in STEM activities: syntheses and a framework from a literature review. *International Journal of Education in Mathematics, Science and Technology, 12*(2), 418-431. https://doi.org/10.46328/ijemst.3357

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

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

López-Pedersen, A., Mononen, R., Aunio, P., Scherer, R., & Melby-Lervåg, M. (2022). Improving numeracy skills in first graders with low performance in early numeracy: a randomised controlled trial. *Remedial and Special Education, 44*(2), 126-136. https://doi.org/10.1177/07419325221102537

Maass, K., Zehetmeier, S., Weihberger, A., & Flößer, K. (2023). Analysing mathematical modelling tasks in light of citizenship education using the COVID-19 pandemic as a case study. *ZDM - Mathematics Education, 55*(1), 133-145. https://doi.org/10.1007/s11858-022-01440-9

Marleny M, A. S., Zulkardi, Putri, R. I. I., & Hartono, Y. (2026). Developing a PMRI-guided learning trajectory for elementary students' understanding of percentages within the melemang cultural context. *Journal of Advanced Sciences and Mathematics Education, 6*(1), 59-72. https://doi.org/10.58524/jasme.v6i1.1095

Maryam, C., Turmudi, & Iriawan, S. B. (2026). Project-based STEM learning through vertical garden design: enhancing elementary students' numeracy and mathematical problem-solving. *Journal of Advanced Sciences and Mathematics Education, 6*(3), 114-130. https://doi.org/10.58524/jasme.v6i3.1420

Muhaimin, L. H., Sholikhakh, R. A., Yulianti, S., Ardani, Hendriyanto, A., & Sahara, S. (2024). Unlocking the secrets of students' mathematical literacy to solve mathematical problems: a systematic literature review. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(4), em2428. https://doi.org/10.29333/ejmste/14404

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), Design approaches and tools in education and training (pp. 125-135). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-4255-7_10

Nusantara, D. S., Zulkardi, & Putri, R. I. I. (2022). Designing PISA-like mathematics problem in COVID-19 context for Indonesian students. *Journal on Mathematics Education, 13*(1), 1-16. https://doi.org/10.22342/jme.v13i1.pp1-16

OECD. (2023). *PISA 2022 results (Volume I): the state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Ow-Yeong, Y. K., Yeter, I. H., & Ali, F. (2023). Learning data science in elementary school mathematics: a comparative curriculum analysis. *International Journal of STEM Education, 10*, Article 8. https://doi.org/10.1186/s40594-023-00397-9

Plomp, T. (2013). Educational design research: an introduction. In T. Plomp & N. Nieveen (Eds.), Educational design research - Part A: An introduction (pp. 10-51). SLO - Netherlands Institute for Curriculum Development.

Plomp, T., & Nieveen, N. (Eds.). (2013). *Educational design research - part a: an introduction*. SLO - Netherlands Institute for Curriculum Development.

Polatuly, S., Ramankulov, S., Yedilbayev, Y., Çoruh, A., & Kuanbayeva, B. (2026). STEM-TVET integration for developing pre-service teachers' professional competence: a systematic literature review. *Frontiers in Education, 11*, Article 1812251. https://doi.org/10.3389/feduc.2026.1812251

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497. https://doi.org/10.1002/nur.20147

Purbaningrum, M., Arliani, E., Ramadhan, S., & Azizah, N. L. (2024). Tracing Indonesian students' mathematical literacy through the minimum competency assessment (AKM): a literature review. *AIP Conference Proceedings, 3201*, 040006. https://doi.org/10.1063/5.0241707

Rehman, N., Huang, X., Mahmood, A., AlGerafi, M. A. M., & Javed, S. (2024). Project-based learning as a catalyst for 21st-century skills and student engagement in the math classroom. *Heliyon, 10*(23), e39988. https://doi.org/10.1016/j.heliyon.2024.e39988

Risdiyanti, I., Zulkardi, Z., Putri, R. I. I., Prahmana, R. C. I., & Nusantara, D. S. (2024). Ratio and proportion through realistic mathematics education and pendidikan matematika realistik Indonesia approach: a systematic literature review. *Jurnal Elemen, 10*(1), 158-180. https://doi.org/10.29408/jel.v10i1.24445

Roos, A. (2023). Promoting basic arithmetic competence in early school years: using a response to intervention model. *Journal of Research in Special Educational Needs, 23*(4), 316-329. https://doi.org/10.1111/1471-3802.12602

Samritin, S. (2023). The effectiveness of RME approach to elementary school students' mathematical abilities: a meta-analysis study. *Indonesian Journal of Educational Research and Review, 6*(1), 176-186. https://doi.org/10.23887/ijerr.v6i1.59150

Shimizu, Y., & Kang, H. (2025). Research on classroom practice and students' errors in mathematics education: a scoping review of recent developments for 2018-2023. *ZDM - Mathematics Education, 57*(4), 695-710. https://doi.org/10.1007/s11858-025-01704-0

Speyer, R., Hakkarainen, A., Yoon, S., Kim, J.-H., Windsor, C., Wilkes-Gillan, S., Littlefair, D., & Cordier, R. (2024). Content validity of measures in early numeracy in children up to eight years: a COSMIN systematic review. *PLOS ONE, 19*(9), e0308874. https://doi.org/10.1371/journal.pone.0308874

Starrett, A., Saw, G., Meador, A., & Harmon, H. (2026). Editorial: rural STEMM education research: bridging between uniqueness and universality: rural-focused and rural-reimagined approaches. *Frontiers in Education, 10*, Article 1736198. https://doi.org/10.3389/feduc.2025.1736198

Suryanti, Nursalim, M., Choirunnisa, N. L., & Yuliana, I. (2024). STEAM-project-based learning: a catalyst for elementary school students' scientific literacy skills. *European Journal of Educational Research, 13*(1), 1-14. https://doi.org/10.12973/eu-jer.13.1.1

Susanta, A., Susanto, E., Rusnilawati, Sumardi, H., & Ali, S. R. B. (2025). Literacy skills through the use of digital STEAM-inquiry learning modules: a comparative study of urban and rural elementary schools in Indonesia. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(4), em2615. https://doi.org/10.29333/ejmste/16170

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic mathematics education. In S. Lerman (Ed.), Encyclopedia of mathematics education (2nd ed., pp. 713-717). Springer. https://doi.org/10.1007/978-3-030-15789-0_170

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2024). Strategy flexibility in mathematics. *ZDM - Mathematics Education, 56*(1), 115-126. https://doi.org/10.1007/s11858-023-01491-6

Wewe, M., Bela, M. E., Wangge, M. C. T., & Bhoke, W. (2026). Enhancing elementary students' numeracy skills through a Ngadhu Bhaga ethnomathematics-based learning approach. *Journal of Advanced Sciences and Mathematics Education, 6*(1), 264-276. https://doi.org/10.58524/jasme.v6i1.998

Wijaya, T. T., Hidayat, W., Hermita, N., Alim, J. A., & Talib, C. A. (2024). Exploring contributing factors to PISA 2022 mathematics achievement: insights from Indonesian teachers. *Infinity Journal, 13*(1), 139-156. https://doi.org/10.22460/infinity.v13i1.p139-156

Wu, X., Yang, Y., Zhou, X., Xia, Y., & Liao, H. (2024). A meta-analysis of interdisciplinary teaching abilities among elementary and secondary school STEM teachers. *International Journal of STEM Education, 11*, Article 38. https://doi.org/10.1186/s40594-024-00500-8

Yang, X., Chan, M. C. E., Kaur, B., Deng, J., Fan, L., & Leung, F. K. S. (2025). International comparative studies in mathematics education: a scoping review of the literature from 2014 to 2023. *ZDM - Mathematics Education, 57*(1), 1-22. https://doi.org/10.1007/s11858-025-01700-4

Yanuarto, W. N., Dewi, S. P., Anggoro, S., Jazuli, A., & Fukui, M. (2025). Problem-based learning and Mathtastic app integration to improve reasoning and collaboration in mathematics learning among primary students: an experimental approach. *Journal of Advanced Sciences and Mathematics Education, 5*(2), 415-427. https://doi.org/10.58524/jasme.v5i2.900

Yuanita, P., Zulnaidi, H., & Zakaria, E. (2018). The effectiveness of realistic mathematics education approach: the role of mathematical representation as mediator between mathematical belief and problem solving. *PLOS ONE, 13*(9), e0204847. https://doi.org/10.1371/journal.pone.0204847