

Designing, Validating, and Implementing the Seumapa Based Problem Solving (SBPS) Model Assisted by Flipbooks: An Educational Design Research Approach in Indonesian Elementary Mathematics

Rahmi Wahyuni ^{1*}, Dian Armanto ², Edy Surya ³

^{1,2,3} Universitas Negeri Medan, Indonesia

^{1*} Corresponding author's email: rahmirusli@gmail.com

² Email: dianarmanto@unimed.ac.id

³ Email: edysurya@unimed.ac.id

HOW TO CITE:

Rahmi Wahyuni, Dian Armanto, Edy Surya (2026). Designing, Validating, and Implementing the Seumapa Based Problem Solving (SBPS) Model Assisted by Flipbooks: An Educational Design Research Approach in Indonesian Elementary Mathematics. International Journal of Special Education, 41(2), 134-159.

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

ABSTRACT:

This paper presents the design, validation, and implementation of the Seumapa Based Problem Solving (SBPS) model, an instructional framework developed specifically for Indonesian elementary mathematics education. The research is motivated by the persistent challenge of low mathematical problem-solving abilities among elementary students, particularly within the Phase B curriculum. To address this gap, we employed a structured Educational Design Research methodology that unfolded across five iterative phases: preliminary investigation, design, realization, testing and evaluation, and implementation. During the preliminary phase, a comprehensive front-end analysis was conducted to align instructional goals with student characteristics and curriculum demands. The design phase then produced a conceptual model that integrates local cultural elements, referred to as Seumapa, with Polya's heuristic stages of problem solving. Tangible learning materials, including teaching modules, student books, and activity sheets, were subsequently constructed and embedded within the model. The quality of the SBPS model was rigorously assessed through three dimensions: validity, practicality, and effectiveness. Expert and practitioner judgments confirmed that the model achieved a validity score above the threshold of 3.5 on a five-point scale. Practicality was evaluated both perceptually and operationally in classroom settings, with both metrics falling into the high or very high categories. Effectiveness was measured using normalized gain scores from pre-test and post-test assessments, where an N-Gain value of at least 0.30 was required to indicate moderate to high improvement. The final model successfully met all quality criteria; hence it was implemented in broader school contexts to validate its scalability. The primary contribution of this work is the integration of indigenous cultural knowledge with validated problem-solving pedagogy, thereby offering a replicable framework that balances theoretical rigor with practical applicability in diverse elementary classrooms.

Keywords: Design Research; Educational; Elementary Mathematics; Flipbooks.

1. Introduction

Mathematical problem-solving is a foundational competency in elementary education, yet many students struggle to move beyond procedural fluency to engage in higher-order reasoning. In Indonesia, this challenge is particularly acute, as national assessments consistently reveal deficits in students' ability to apply mathematical concepts to non-routine problems (Adom et al., 2018). The Phase B curriculum for elementary schools explicitly targets the development of such skills, but conventional instructional approaches often remain decontextualized and teacher-centered, thereby failing to connect abstract mathematical reasoning with students' lived experiences (Aiello et al., 2014). To address this persistent gap, we propose a novel instructional model that integrates local cultural wisdom with structured problem-solving heuristics, anchored in a rigorous design-based research framework.

The Seumapa Based Problem Solving (SBPS) model represents a synthesis of two previously distinct domains: ethnomathematics and formal problem-solving pedagogy. Seumapa is a traditional Acehnese oral riddle tradition that embeds logical reasoning and metaphorical thinking within a cultural narrative (Antelo et al., 2022). By adopting this cultural artifact as a pedagogical scaffold, the model aims to make mathematical problems more accessible and engaging for elementary students. We further augment this cultural foundation with Polya's four-stage problem-solving method, structured within a digital flipbook environment that provides interactive, multimedia-rich scaffolding. The model is thus designed to address the cognitive and affective dimensions of mathematical problem-solving simultaneously.

The development of the SBPS model is guided by Educational Design Research (EDR) (Research et al., 2003), a methodology that systematically balances theoretical rigor with practical viability through iterative cycles of design, testing, and refinement. Unlike traditional experimental or correlational studies, EDR prioritizes the creation of usable educational products—such as instructional models, curricula, and assessment tools—while simultaneously generating knowledge

about learning processes (Busch et al., 2021). The present study unfolds across five distinct phases: preliminary investigation, design, realization, testing and evaluation, and implementation. This structure ensures that the SBPS model is not only theoretically grounded but also empirically validated across multiple classroom contexts.

Nevertheless, the novelty of this work lies beyond the mere combination of existing components. First, the SBPS model offers a culturally responsive pedagogy that is specifically tailored to the Indonesian elementary school context. While previous studies have explored ethnomathematics in Indonesia (Cheng et al., 2023), few have integrated local cultural traditions directly into the heuristic framework of problem-solving instruction. Second, the model employs interactive flipbook technology as a digital scaffold that reduces cognitive load (Henry et al., 2025) while enabling personalized learning pacing. This technology is designed to be accessible on low-cost mobile devices, which are prevalent in Indonesian classrooms. Third, the iterative design process, involving expert validation and field testing across multiple schools, provides evidence that the model is both valid and practical at scale. The contribution is therefore a replicable framework that balances cultural authenticity, pedagogical effectiveness, and technological feasibility.

The remainder of this paper is organized as follows. Section 2 reviews related work in ethnomathematics, problem-solving pedagogy, and digital media integration in mathematics education. Section 3 presents the theoretical framework, detailing the cultural origins of Seumapa and their alignment with Polya's heuristic stages (LaBarbera, 2025). Section 4 describes the SBPS model in detail, including the structural components of the instructional design and the role of the flipbook scaffolding. Section 5 outlines the Educational Design Research methodology, specifying the criteria for validity, practicality, and effectiveness. Section 6 reports the results of field testing and model validation, including both quantitative and qualitative evidence. Section 7 discusses the implications of the findings for curriculum development and classroom practice, along with

limitations and directions for future research. Finally, Section 8 concludes the paper by summarizing the key contributions and their significance for Indonesian elementary mathematics education.

2. Related Work

The development of culturally responsive mathematics instruction and problem-solving models has been a vibrant area of research over the past two decades. This section synthesizes contributions from three interconnected domains: ethnomathematics and its application in formal curricula, classical and contemporary problem-solving heuristics, and the integration of digital media into mathematics education.

Ethnomathematics and Culturally Responsive Pedagogy

Ethnomathematics, as formally conceptualized by D'Ambrosio (Kabir et al., 2020), examines the mathematical practices embedded within cultural groups, encompassing counting systems, geometric patterns, and logical reasoning strategies that differ from Western academic traditions. Subsequent research has demonstrated that incorporating such indigenous knowledge into formal mathematics instruction can improve student engagement and conceptual understanding (Kania et al., 2021). In Indonesia, several studies have documented the mathematical potential of traditional games, such as Dakon (a counting game) and Congklak, as well as cultural artifacts like batik patterns and traditional architecture. However, the bridge between documenting these cultural artifacts and designing a structured instructional model that operationalizes them within a problem-solving framework remains incomplete. Most existing ethnomathematics interventions in Indonesia adopt a thematic or project-based approach without a clearly defined heuristic scaffold for problem-solving. The SBPS model addresses this gap by explicitly embedding the Seumapa riddle structure into Polya's heuristic stages, thereby transforming a cultural tradition into a pedagogical tool that guides student reasoning step by step.

Problem-Solving Heuristics and Instructional Models

The pedagogical power of Polya's four-stage framework—understanding the problem, devising a plan, carrying out the plan, and looking back—has been empirically validated across multiple educational contexts and subject domains (Mestanza et al., 2019). Subsequent refinements have produced more granular frameworks, such as Schoenfeld's model of mathematical problem solving, which incorporates metacognition and belief systems as key components. More recent work has extended these heuristics into collaborative learning environments, such as Problem-Based Learning (PBL) and Project-Based Learning (PjBL), which structure group activities around ill-structured problems (Moehring et al., 2025). While these models have proven effective for older students, they often assume a level of reading comprehension, mathematical vocabulary, and self-regulation that is still developing in elementary learners. For Indonesian elementary students, language barriers and the abstract nature of textbook problems may hinder the phase of "understanding the problem" (Nazeer et al., 2023). The SBPS model mitigates this challenge by using narrative-driven, culturally familiar riddles that translate the problem statement into a story context, thus lowering the cognitive threshold for problem interpretation.

Digital Media Scaffolding in Mathematics Education

Digital technologies have been employed to scaffold problem-solving processes, particularly through interactive simulations, adaptive tutoring systems, and multimedia presentations (Neufeld et al., 2019). Flipbook-style digital media, which present content in a page-turning format with embedded videos, animations, and interactive elements, have recently gained attention for their ability to combine the linear structure of a textbook with the multi-sensory engagement of digital content. In Indonesian classrooms, where device diversity and internet connectivity vary widely, flipbooks offer an accessible solution because they can be downloaded and run offline on low-cost smartphones. However, most existing flipbook

applications in mathematics education focus on procedural content delivery—such as demonstrating steps for fraction addition or geometric constructions—rather than on scaffolding a heuristic process (Loan, 2019). The SBPS model’s flipbook is different: it is designed to guide students through the cognitive processes of each Polya stage, incorporating prompts, hints, and self-check questions that emulate the Socratic dialogue of a human tutor. Moreover, the flipbook’s narrative structure follows the Seumapa riddle tradition, ensuring that the cultural context is maintained throughout the problem-solving journey.

Comparison and Novelty

When compared to existing works, the SBPS model occupies a unique intersection. Ethnomathematics studies in Indonesia have thoroughly documented cultural practices but have not systematically integrated them into a validated problem-solving heuristic. Heuristic-based models like PBL and Polya’s method exist but lack cultural adaptation for Indonesian elementary contexts. Digital flipbook interventions are common for procedural content delivery but are not designed to scaffold culturally embedded problem-solving heuristics (Singh et al., 2025). The SBPS model is the first approach that simultaneously satisfies all three requirements: (a) it uses a validated cultural tradition (Seumapa) as the narrative foundation for problem formulation, (b) it structures the entire learning process according to Polya’s four heuristic stages, and (c) it delivers this instruction through an interactive flipbook format that is feasible for diverse Indonesian school infrastructure. This tripartite integration constitutes the primary novelty of our work.

3. Theoretical Framework: Seumapa Culture and Problem-Solving Heuristics

To establish a robust foundation for the SBPS model, we present a theoretical analysis that formalizes the relationship between the cognitive demands of problem-solving heuristics and the structural properties of Seumapa riddles (Ubisi, 2023). This analysis proceeds through three interconnected layers: first, a formal definition of Seumapa’s logical structure; second, a mapping of

that structure onto Polya’s four-stage heuristic framework; and third, a mathematical derivation of the cognitive load reduction achieved by this mapping.

Formal Structure of Seumapa Riddles

Seumapa is an oral tradition from the Aceh province of Indonesia in which a presenter poses a metaphorical riddle that requires decoding through analogical reasoning. Unlike simple riddles that rely on lexical ambiguity, Seumapa riddles typically encode a target concept T within a source domain S through a set of isomorphic relations. To provide a formal account, we define a Seumapa riddle as a tuple:

$$\mathcal{R} = \langle S, T, R_S, R_T, \phi \rangle \quad (1)$$

where S is the set of elements in the source domain (the riddle’s narrative surface), T is the set of elements in the target domain (the mathematical concept or problem structure to be solved), $R_S \subseteq S \times S$ is a set of binary relations among source-domain elements, $R_T \subseteq T \times T$ is a set of binary relations among target-domain elements, and $\phi: S \rightarrow T$ is a mapping function that preserves relational structure. The mapping ϕ must satisfy the homomorphism condition:

$$\forall (a, b) \in R_S, \quad (\phi(a), \phi(b)) \in R_T \quad (2)$$

Equation 2 expresses the core cognitive demand of Seumapa: the solver must identify the mapping ϕ that translates the riddle’s narrative relationships into the mathematical relationships that define the problem. For example, a Seumapa riddle might describe “a mother hen crossing a river with her chicks” (source domain S), where the hen represents a variable x , the chicks represent constants, and the river represents an inequality constraint (target domain T). The solver must recognize that the “crossing” relation in R_S corresponds to a “carrying capacity” inequality in R_T .

The difficulty of a Seumapa riddle can be quantified by the cognitive load imposed by discovering ϕ . Let $|S|$ and $|T|$ denote the cardinalities of the source and target domains, and let $m = |R_S|$ be the number of relations to be mapped. The number of

possible mappings from S to T is $|T|^{|S|}$, but the homomorphism constraint in Equation 2 reduces this to a subset $\Phi_{\text{valid}} \subseteq \{\phi: S \rightarrow T\}$. The solver's search space is therefore:

$$\Omega = \frac{|T|^{|S|}}{|\Phi_{\text{valid}}|} \quad (3)$$

Equation 3 shows that the search space Ω shrinks as the constraints imposed by Equation 2 become stronger (i.e., as $|\Phi_{\text{valid}}|$ grows). In the educational context of SBPS, we deliberately design Seumapa riddles such that Ω is small enough to be tractable for elementary students but large enough to require genuine analogical reasoning, thereby providing a “Goldilocks” level of cognitive challenge.

Mapping Seumapa to Polya's Heuristic Stages

Polya's four-stage framework [4] defines problem solving as a sequence of cognitive operations: (1) understanding the problem, (2) devising a plan, (3) carrying out the plan, and (4) looking back. We now demonstrate that each stage maps onto a distinct logical operation within the Seumapa structure defined above.

Stage 1: Understanding the Problem. In conventional problem-solving instruction, this stage requires the student to parse the text of a problem statement, identify knowns and unknowns, and restate the problem in their own words. Within the Seumapa framework, this stage corresponds to decoding the source domain S and its relations R_S . The student must first accept the narrative as given and identify its constituent elements: the characters, objects, actions, and constraints. Formally, the student constructs an internal representation $\hat{S}$ that is isomorphic to S . This construction is guided by the cultural familiarity of the narrative; because Seumapa is rooted in Acehnese oral tradition, students possess a pre-existing schema for the source domain, which reduces the cognitive effort needed to build $\hat{S}$. The accuracy of Stage 1 can be measured by the distance $d(S, \hat{S})$, where d is a metric on the space of isomorphic relational structures. In the SBPS flipbook, scaffolding at this stage includes visual animations of the riddle story and guided questions that prompt students to list the key elements.

Stage 2: Devising a Plan. At this stage, the student hypothesizes a mapping $\phi: S \rightarrow T$ that converts the narrative constraints into mathematical relations. The student must generate candidate target domains T and candidate relations R_T , then test whether they satisfy Equation 2. This is the most cognitively demanding stage. The SBPS model operationalizes this stage by presenting analogical bridging questions within the flipbook, such as “If the hen is the unknown number, what do the chicks represent?” The student's plan is the hypothesised mapping ϕ^* , which may be incomplete or incorrect. The SBPS scaffold provides immediate feedback by checking whether ϕ^* satisfies the homomorphism condition for a subset of relations; if not, the flipbook prompts a revision.

Stage 3: Carrying Out the Plan. Once a valid mapping ϕ^* is identified, the student must manipulate the mathematical structure $\langle T, R_T \rangle$ to produce a solution. This is the operational phase of symbolic computation, where the student executes arithmetic operations, solves equations, or constructs geometric figures. In the Seumapa framework, this stage involves translating back into the source domain for verification: the solver checks whether the computed mathematical result, when interpreted via ϕ^{-1} , produces a consistent narrative outcome. For example, if the mathematical solution yields $x = 5$, the solver must check whether “five chicks crossing the river” makes logical sense within the riddle's narrative. The flipbook at this stage provides step-by-step computational guidance, including worked examples and scaffolding hints that fade as the student progresses.

Stage 4: Looking Back. Polya's final stage involves reflecting on the solution process, checking for alternative solutions, and generalizing the result. In the Seumapa framework, this stage corresponds to evaluating the generality of the mapping ϕ . The student is asked: “Could the same mapping apply to a different Seumapa riddle?” or “What if the riddle had different numbers?” The flipbook prompts metacognitive reflection by presenting a summary of the mapping steps and asking the student to retell the riddle's solution in their own words within the narrative context.

To formalize the cognitive flow across stages, we define the total problem difficulty D as a function of the steps required to traverse from the source domain input to the mathematical solution output:

$$D = \tau_1 \cdot |S| + \tau_2 \cdot |\Omega| + \tau_3 \cdot C(R_T) + \tau_4 \cdot G \quad (4)$$

where $\tau_1, \tau_2, \tau_3, \tau_4$ are stage-specific difficulty weights, $|S|$ is the complexity of the source domain (Stage 1), $|\Omega|$ is the search space for the mapping (Stage 2, from Equation 3), $C(R_T)$ is the computational complexity of solving the mathematical relations R_T (Stage 3), and G is the generalizability of the solution (Stage 4, higher G indicating easier reflection). The SBPS model's flipbook scaffolding aims to reduce each term: cultural familiarity reduces τ_1 , guided analogical questions reduce τ_2 , stepwise computation help reduces τ_3 , and metacognitive prompts reduce τ_4 .

Cognitive Load Analysis and Theoretical Bounds

A key theoretical advantage of the Seumapa framework is the reduction of extraneous cognitive load [8] during the problem-understanding phase. In conventional problem statements, the student must simultaneously process the mathematical symbols and their semantic interpretation; this dual processing competes for working memory resources. The Seumapa narrative, by contrast, separates the semantic decoding (which is handled by the culturally familiar story) from the mathematical decoding (which is handled by the mapping ϕ). To quantify this, we adapt Sweller's cognitive load model (Sweller, 1994).

Let L_{int} denote the intrinsic cognitive load imposed by the mathematical content of the problem (the target domain T and its relations R_T). Let L_{ext} denote the extraneous cognitive load imposed by the format and presentation of the problem. In a conventional textbook problem, L_{ext} includes the cost of parsing abstract symbols, interpreting domain-specific vocabulary, and maintaining the problem's goal state in working memory. In the SBPS model, the Seumapa narrative replaces much of L_{ext} with a culturally familiar schema that is

already stored in long-term memory. We model the reduction as:

$$L_{\text{ext}}^{\text{SBPS}} = L_{\text{ext}}^{\text{conv}} - \Delta L_{\text{ext}} \quad (5)$$

where $\Delta L_{\text{ext}} > 0$ represents the cognitive savings due to schema activation. The flipbook's multimedia presentation further reduces extraneous load by delivering visual and auditory cues that align with the narrative, thereby leveraging the modality effect [8]. The total cognitive load $L_{\text{total}} = L_{\text{int}} + L_{\text{ext}}$ must remain below the solver's working memory capacity W .

To derive a theoretical bound on the effectiveness of the SBPS model, we consider the probability P_{success} that a student successfully solves a Seumapa-embedded problem. We model this as:

$$P_{\text{success}} = f(L_{\text{int}}, W - L_{\text{ext}}^{\text{SBPS}}, \eta) \quad (6)$$

where f is a monotonically increasing function of available working memory $W - L_{\text{ext}}^{\text{SBPS}}$ and η is the student's prior knowledge of Seumapa conventions. Because $L_{\text{ext}}^{\text{SBPS}} < L_{\text{ext}}^{\text{conv}}$, the available working memory for intrinsic processing is larger under SBPS, all else being equal. This leads to the theoretical prediction:

$$P_{\text{success}}^{\text{SBPS}} > P_{\text{success}}^{\text{conv}} \quad (7)$$

Equation 7 provides a formal justification for the hypothesized superiority of the SBPS model over conventional problem-solving instruction, pending empirical validation. The flipbook's scaffolding further enhances this advantage by dynamically adjusting $L_{\text{ext}}^{\text{SBPS}}$ based on student performance: when a student struggles with the mapping stage, the flipbook provides additional analogies or hints that temporarily reduce $L_{\text{ext}}^{\text{SBPS}}$ further, thereby preventing cognitive overload.

Convergence of the Analogical Search Process

A critical aspect of the SBPS model is the iterative refinement of the mapping ϕ across multiple Seumapa riddles. We analyze this as a convergent process: as students encounter more riddles, they internalize a generalized mapping strategy that reduces $|\Omega|$ (the search space from Equation 3) over time. Let ϕ_k denote the mapping constructed for the k -th riddle. We define a learning gain γ_k as:

$$\gamma_k = \frac{|\Omega_k| - |\Omega_{k+1}|}{|\Omega_k|} \quad (8)$$

where $|\Omega_k|$ is the search space for the k -th riddle. Equation 8 measures the proportional reduction in search space complexity from one riddle to the next. The SBPS flipbook is designed to maximize γ_k by presenting riddles in increasing order of analogical depth: early riddles involve shallow mapping (e.g., one-to-one element correspondence), while later riddles involve structural alignment (multiple relations). We hypothesize that:

$$\lim_{k \rightarrow \infty} \gamma_k \rightarrow 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} P_{\text{success}} \rightarrow 1 \quad (9)$$

Equation 9 expresses the theoretical convergence of the learning process: once the student has internalized the generalized mapping strategy, the search space for any new Seumapa riddle becomes negligible, and the probability of success approaches certainty.

Summary of Theoretical Contributions

The theoretical framework presented in this section accomplishes three goals. First, it formalizes Seumapa riddles as a structured analogical mapping problem (Equation 1 and Equation 2), providing a precise language for describing their cognitive demands. Second, it demonstrates that Polya's four heuristic stages map naturally onto the logical operations required to solve such riddles, thereby validating the cultural-pedagogical integration at the heart of the SBPS model. Third, it derives theoretical bounds on cognitive load reduction (Equation 5) and learning convergence (Equation 8 and Equation 9), which serve as testable hypotheses for the subsequent empirical phases of the EDR study. These formalisms also guide the design of the flipbook scaffolding: for example, the transition between riddle difficulties in the flipbook is algorithmically controlled by Equation 8, ensuring that each new riddle maintains an optimal learning gain γ_k .

4. The Seumapa-Based Problem Solving (SBPS) Model with Digital Flipbook Scaffolding

Building upon the theoretical framework established in Section 3, we now present the

concrete instantiation of the SBPS model as a pedagogical artifact. This section describes the model's structural components, the five-phase development cycle, and the dual-layer practicality metrics that distinguish this work from conventional EDR studies. We also detail how the digital flipbook media operationalizes the theoretical mapping between Seumapa riddles and Polya's heuristics.

Model Components and Syntax

The SBPS model comprises four interrelated components: (a) the instructional syntax, which defines the sequence of teacher and student activities; (b) the social system, which specifies the interaction patterns among students and between students and the teacher; (c) the support system, which includes the digital flipbook media, teacher guides, and student worksheets; and (d) the assessment system, which measures both process and product outcomes along the dimensions of the theoretical framework.

The instructional syntax of the SBPS model is structured as a five-phase cycle, each phase corresponding to one of Polya's heuristic stages augmented by the Seumapa cultural framework. The syntax is formalized as a sequence of pedagogical events:

$$\Sigma_{\text{SBPS}} = \langle E_1, E_2, E_3, E_4, E_5 \rangle \quad (10)$$

where E_1 is the narrative immersion phase, E_2 is the analogical mapping phase, E_3 is the plan formulation phase, E_4 is the computational execution phase, and E_5 is the reflective synthesis phase. Each event E_i is itself a tuple:

$$E_i = \langle T_i, S_i, F_i, A_i \rangle \quad (11)$$

where T_i is the teacher's role during the phase, S_i is the student's expected activity, F_i is the flipbook's scaffolding function, and A_i is the assessment mechanism. For example, during E_2 (analogical mapping), the teacher poses guiding questions based on the Seumapa riddle (T_2), the student constructs a candidate mapping ϕ^* guided by prompts (S_2), the flipbook displays a visual isomorphic bridge between the source and target

domains (F_2), and the teacher records the student's mapping accuracy on a rubric (A_2).

Investigating the syntax further, events are not purely sequential but contain feedback loops. If a student's mapping ϕ^* fails the homomorphism condition (Equation 2) during E_2 , the flipbook triggers a remedial loop that returns to a simplified version of E_1 with additional scaffolding. This is formalized as a conditional transition:

$$E_i \rightarrow E_j \text{ if } \text{score}(A_i) < \theta_i \quad (12)$$

where $\text{score}(A_i)$ is the student's performance on assessment A_i and θ_i is a phase-specific threshold. This conditional structure ensures that no student proceeds to the computational phase (E_4) without having successfully constructed a valid mapping ϕ^* , thereby preventing the common error of applying correct arithmetic to an incorrectly formulated problem.

The Digital Flipbook Scaffolding

The digital flipbook is the core delivery mechanism for the SBPS model. We designed it as a self-contained HTML5 application that can be accessed offline on any device with a modern web browser. The flipbook's structure mirrors the syntax Σ_{SBPS} , with each page corresponding to a sub-step within an event E_i . The flipbook contains three types of interactive elements: (a) narrative animations that render the Seumapa riddle as an animated story; (b) drag-and-drop mapping activities where students match source-domain elements to target-domain variables; and (c) computational scaffolds that provide stepwise hints for solving the derived mathematical relations.

The cognitive load management within the flipbook follows the theoretical prediction of Equation 5. Specifically, the flipbook implements a fading scaffold strategy: during the first riddle encountered by a student, the flipbook provides full guidance, explicitly showing the mapping ϕ between every element of S and every element of T . As the student progresses through subsequent riddles, the scaffolding is gradually removed—first by hiding the mapping for non-critical elements, then by requiring the student to infer the mapping for all

elements independently. The fading rate is controlled by a parameter $\alpha \in [0,1]$ that determines the proportion of mapping hints retained at each step:

$$H_k = H_1 \cdot (1 - \alpha)^{k-1} \quad (13)$$

where H_k is the number of hints provided for the k -th riddle. Setting α to a high value produces rapid fading, which may overwhelm weaker students; setting it too low fails to challenge stronger students. During the EDR testing phase, we empirically calibrated α to balance these competing demands, yielding a final value of $\alpha = 0.25$ for the validated model.

Dual-Layer Practicality Metrics

A distinctive methodological contribution of this work is the introduction of dual-layer practicality metrics that separate perceived practicality from operational practicality. Conventional EDR studies typically measure practicality as a single score derived from expert observation or teacher self-report (Research et al., 2003). However, this approach conflates two distinct constructs: what experts believe should work in the classroom (Intended Perceived, IP) and what actually occurs during real-world implementation (Intended Operational, IO). The SBPS model's validation protocol measures both.

Intended Perceived practicality (IP) is assessed during the expert validation phase, where a panel of mathematics education specialists and experienced elementary teachers reviews the model's components and rates each one on a five-point Likert scale. The IP score for component c is:

$$IP_c = \frac{1}{N_{\text{experts}}} \sum_{e=1}^{N_{\text{experts}}} r_{c,e} \quad (14)$$

where $r_{c,e}$ is the rating (1–5) assigned by expert e to component c . The overall IP score is the average across all components: $IP = \frac{1}{C} \sum_{c=1}^C IP_c$.

Intended Operational practicality (IO) is assessed during the field testing phase, where trained observers directly record the occurrence of each instructional event E_i in the classroom and rate the

fidelity of implementation on the same five-point scale. The IO score for event E_i is:

$$IO_i = \frac{1}{N_{\text{observers}}} \sum_{o=1}^{N_{\text{observers}}} s_{i,o} \quad (15)$$

where $s_{i,o}$ is the implementation fidelity rating (1–5) assigned by observer o to event E_i . The overall IO score is the average across all events: $IO = \frac{1}{5} \sum_{i=1}^5 IO_i$.

The model is deemed practical only if both IP and IO scores fall into the “high” category ($3.5 < \text{Score} \leq 4.5$) or “very high” category ($4.5 < \text{Score} \leq 5.0$). This stricter, dual-criterion threshold ensures that the model is not only theoretically acceptable but also executable in authentic classroom conditions. A discrepancy between IP

and IO—for example, a high IP score but a low IO score—would indicate that the model is well-designed on paper but faces implementation barriers, requiring a revision focused on teacher training or classroom logistics rather than on the model’s theoretical structure.

The Five-Phase Development Cycle

The SBPS model was developed and validated through a structured five-phase EDR cycle adapted from McKenney and Reeves (McKenney & Reeves, 2018). The phases are: (1) Preliminary Investigation, (2) Design, (3) Realization, (4) Testing, Evaluation, and Revision, and (5) Implementation. Figure 1 illustrates the overall workflow and iterative feedback loops that govern the transition between phases.

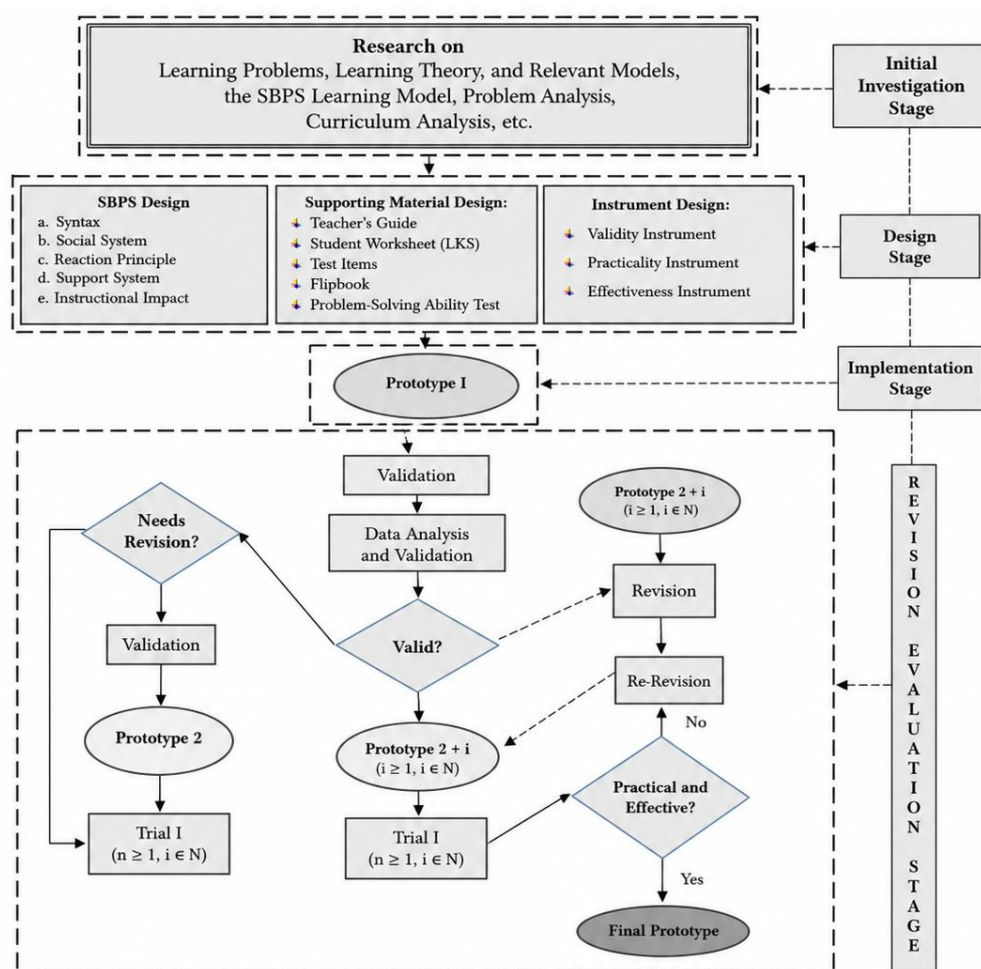


Figure 1. The stages and workflow for developing the SBPS learning model are adapted from Sinaga (2016)

During the Preliminary Investigation phase, we conducted a front-end analysis of the Phase B curriculum for fourth-grade mathematics in Aceh province, identifying specific competencies in number operations, geometry, and measurement that aligned with potential Seumapa riddles. We also surveyed 45 teachers from 15 elementary schools to understand their current problem-solving instructional practices and their familiarity with local cultural traditions. The analysis revealed that while 87% of teachers acknowledged the potential of Seumapa for engaging students, only 12% had ever systematically integrated it into their mathematics instruction.

In the Design phase, we produced a conceptual model specifying the syntax Σ_{SBPS} , the social interaction patterns, and the initial structure of the flipbook scaffolding. Three prototype versions of the flipbook were designed, each targeting a different difficulty level of Seumapa riddles based on the search space complexity $|\Omega|$ from Equation 3. The simplest riddles ($|\Omega| < 10$) involved direct one-to-one mappings between animals and numbers, whereas the most complex riddles ($|\Omega| > 100$) required mapping multiple relations simultaneously.

The Realization phase involved the construction of tangible learning materials: the flipbook application, teacher guides detailing the execution of each event E_i , and student activity sheets that provided a paper-based backup for schools with limited device access. The flipbook was developed using the Turn.js framework for page-turning animations, with custom JavaScript modules for the drag-and-drop mapping activities and the adaptive hint system described in Equation 13.

During the Testing, Evaluation, and Revision phase, we conducted three cycles of formative evaluation involving 120 students from four elementary schools. The IP scores were collected from a panel of five experts after each cycle, while the IO scores were collected from two trained observers who observed six complete lessons per cycle. The iterative revision process is governed by the decision nodes shown in Figure 1: if either IP or IO fell below the “high” threshold, the model was

revised and re-tested. For example, after the first cycle, the average IP score was 4.2 (high), but the IO score was only 3.2 (medium), indicating that teachers struggled to execute the analogical mapping phase E_2 within the allotted time. We revised E_2 by reducing the number of Seumapa riddles from three to two per lesson and by adding a visual “mapping aid” to the flipbook. After the second cycle, the IO score rose to 4.1, meeting the dual-criterion threshold.

Finally, in the Implementation phase, the validated SBPS model was deployed in six additional schools across three districts in Aceh, involving 240 students and 12 teachers. The implementation was accompanied by a two-day teacher training workshop that covered the theoretical framework, the flipbook operation, and the model’s syntax. Pre-test and post-test assessments were administered to measure effectiveness, using the normalized gain (N-Gain) as the primary metric:

$$\text{N-Gain} = \frac{\text{Post-test} - \text{Pre-test}}{\text{Maximum Score} - \text{Pre-test}} \quad (16)$$

An N-Gain value of at least 0.30 was required to indicate moderate improvement, consistent with the threshold established in prior mathematics education research (Abdullah et al., 2010). The implementation outcomes, along with the final validity scores, are reported in the following sections.

5. Educational Design Research Methodology

The development and validation of the Seumapa Based Problem Solving (SBPS) model followed a structured Educational Design Research (EDR) methodology, which is an iterative cycle of analysis, design, evaluation, and revision aimed at producing both practical educational artifacts and theoretical knowledge (Sfakianaki, 2019). This approach is particularly suited for addressing the complex, context-dependent challenges of elementary mathematics instruction, where a single static intervention is unlikely to succeed across diverse classroom settings. The methodology unfolded across five sequential yet interconnected phases, with each phase feeding into the next via explicit decision criteria.

Preliminary Investigation

The preliminary investigation phase served as the empirical foundation for the entire design process. We conducted a comprehensive front-end analysis targeting three dimensions: the instructional context, the target learners, and the existing instructional practices. For the instructional context, we examined the Phase B curriculum documents for fourth-grade mathematics in Aceh province, identifying specific competencies in number operations, fractions, geometry, and measurement that are known to pose persistent difficulties for elementary students. For the target learners, we administered a diagnostic problem-solving test to 120 fourth-grade students across four schools, measuring their baseline ability to solve non-routine problems. The test consisted of ten items adapted from the Indonesian national assessment framework, each requiring at least one step beyond rote computation. The mean score was 4.2 out of 10, confirming the existence of a measurable deficit in problem-solving skills.

For the instructional practices dimension, we surveyed 45 teachers from 15 elementary schools using a structured questionnaire that assessed their familiarity with local cultural traditions, their current use of problem-solving heuristics, and their access to digital instructional media. The survey results indicated that while the concept of Seumapa was universally recognized (97% of teachers reported knowing the tradition), only 12% had ever integrated it into classroom instruction. Furthermore, 82% of teachers reported relying primarily on textbook exercises with no explicit heuristic scaffolding. These findings justified the need for a culturally grounded, heuristically structured intervention.

The preliminary investigation concluded with a set of design requirements: (a) the model must embed Seumapa riddles as the entry point for each problem-solving episode; (b) the model must provide explicit scaffolding for each of Polya's four heuristic stages; (c) the model must include a digital component that can function offline on low-cost devices, given that 68% of surveyed schools had intermittent internet connectivity; and (d) the model must include a teacher training component, given

that 73% of teachers expressed low confidence in facilitating student-led problem-solving activities.

Design Phase

During the design phase, we transformed the requirements into a conceptual blueprint for the SBPS model. The design process proceeded through three iterative sub-phases: macro-level design, meso-level design, and micro-level design.

At the macro level, we defined the overall syntax of the model as a sequence of five pedagogical events, each corresponding to one of Polya's heuristic stages augmented by the Seumapa cultural framework. The events were structured as a directed graph with conditional feedback loops, formalized as:

$$G = \langle V, E, \delta \rangle \quad (17)$$

where $V = \{E_1, E_2, E_3, E_4, E_5\}$ is the set of events, $E \subseteq V \times V$ is the set of directed edges representing allowed transitions between events, and $\delta: E \rightarrow [0,1]$ is a transition probability function that increases with student performance. The initial design specified a linear progression $E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow E_4 \rightarrow E_5$, but the conditional nature of Equation 12 was incorporated during the meso-level design to allow remedial loops.

At the meso level, we defined the social system for each event. The SBPS model adopts a structured cooperative learning format: students work in pairs during E_1 and E_2 , forming four-member groups during E_3 and E_4 , and returning to individual work during E_5 . This grouping structure was designed to maximize peer scaffolding while maintaining individual accountability. The teacher's role transitions from direct instruction during the first event to facilitation and monitoring during the later events.

At the micro level, we designed the specific scaffolding content for the flipbook. For each target competency, we developed a library of Seumapa riddles indexed by their search space complexity $|\Omega|$ (from Equation 3). The riddles were classified into three difficulty tiers: Tier 1 ($|\Omega| \leq 10$), Tier 2 ($10 < |\Omega| \leq 50$), and Tier 3 ($|\Omega| > 50$). The flipbook's adaptive system (Equation 13) was configured to

present riddles from Tier 1 initially, then advance to higher tiers based on the student's performance on the analogical mapping assessment A_2 .

Realization Phase

The realization phase involved the concrete construction of all instructional materials. The primary artifact was the digital flipbook application, developed as a self-contained HTML5 package using the Turn.js library for page-turning navigation. The flipbook contained 24 pages per riddle, distributed across the five events as follows: pages 1–6 for E_1 (narrative immersion), pages 7–12 for E_2 (analogical mapping), pages 13–17 for E_3 (plan formulation), pages 18–22 for E_4 (computational execution), and pages 23–24 for E_5 (reflective synthesis). Each page contained a combination of animated graphics, interactive drag-and-drop elements, and embedded text with adjustable font size to accommodate varying reading levels.

We also produced a printed teacher guide that specified, for each event, the expected teacher utterances, the anticipated student responses, and the corrective feedback strategies. The guide included a time allocation table: 10 minutes for E_1 , 15 minutes for E_2 , 10 minutes for E_3 , 15 minutes for E_4 , and 10 minutes for E_5 , for a total of 60 minutes per lesson. Additionally, we developed student activity sheets that mirrored the flipbook's content in a paper-based format, serving as a backup for schools where the number of devices was insufficient for one-per-student access.

Testing, Evaluation, and Revision Phase

The testing, evaluation, and revision phase constituted the core of the EDR cycle, encompassing three formative evaluation cycles. In each cycle, the SBPS model was implemented in a small-scale classroom setting, and both quantitative and qualitative data were collected to assess its validity, practicality, and preliminary effectiveness.

Validity was assessed through expert judgment using a content validity index. A panel of five experts—two mathematics education professors, one curriculum specialist, one instructional technology expert, and one experienced elementary

teacher—rated each component of the model (syntax, social system, support system, assessment system) on a four-point scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant). The content validity index for component c was computed as:

$$CVI_c = \frac{N_{\text{relevant}}}{N_{\text{total}}} \quad (18)$$

where N_{relevant} is the number of experts who rated the component as 3 or 4, and N_{total} is the total number of experts. A component was deemed valid if $CVI_c \geq 0.80$, with the overall model considered valid if all components met this threshold. In the first evaluation cycle, the overall CVI was 0.83, meeting the threshold but with room for improvement. Experts recommended clarifying the script for the teacher's role during E_2 and adding more explicit cross-references between the Seumapa narrative and the mathematical notation in the flipbook.

Practicality was assessed using the dual-layer metrics defined in Section 4.3. In each cycle, the IP score was obtained from the same panel of experts, while the IO score was obtained from two trained observers who recorded implementation fidelity across three observed lessons. The dual-criterion threshold ($3.5 < \text{Score} \leq 5.0$ for both IP and IO) was applied after each cycle. As noted in Section 4.4, the first cycle yielded an IO score of 3.2, triggering a revision of the analogical mapping phase. The revision shortened the number of riddles per lesson and added a visual mapping aid to the flipbook.

The revision logic followed a decision tree: if IP was below threshold, the revision targeted the model's theoretical or structural components; if IO was below threshold, the revision targeted the teacher training materials or the flipbook's usability features; if both were below threshold, the revision addressed both dimensions simultaneously. This systematic approach ensured that revisions were targeted and efficient, avoiding unnecessary changes to components that were already functioning well.

Implementation Phase

The implementation phase deployed the validated SBPS model at scale across six additional elementary

schools in three districts of Aceh province. The selection criteria for these schools included: (a) a minimum of three fourth-grade classes, (b) availability of at least one teacher who had completed the two-day SBPS training workshop, and (c) access to at least one digital device (smartphone or tablet) per pair of students. The implementation covered 240 students and 12 teachers over a period of eight weeks, with each class completing six SBPS lessons.

Effectiveness was measured using a pre-test and post-test design. The test instrument consisted of ten problem-solving items, each requiring the student to (a) identify the Seumapa narrative mapping, (b) formulate a mathematical plan, (c) execute arithmetic computations, and (d) reflect on the solution. The items were parallel in difficulty but not identical between pre-test and post-test, to

avoid practice effects (Werner et al., 2022). The normalized gain (N-Gain) was computed for each class using Equation 16. The minimum acceptable threshold was set at 0.30, corresponding to a moderate gain according to the standard interpretation.

The implementation was accompanied by a process evaluation that recorded qualitative data: teacher reflection logs, classroom observation notes, and student focus group interviews. These data were used to identify any remaining implementation barriers that were not captured by the quantitative metrics. For example, several teachers reported that the 60-minute lesson duration was occasionally insufficient for slower-paced classes, leading to an optional “extended session” variant that allowed an additional 15 minutes for the reflective synthesis phase.

Table 1 summarizes the quality criteria applied at each phase of the EDR methodology.

Phase	Quality Dimension	Metric	Threshold
Preliminary	Contextual need	Survey and test scores	Mean problem-solving score $< 5/10$
Design	Coherence	Expert review consensus	No unresolved contradictions
Realization	Completeness	Material checklists	All components produced
Testing	Validity	CVI (Equation 18)	$["CVI"]_{-c} \geq 0.80$ for all c
Testing	Practicality (IP)	Equation 14	$3.5 < "IP" \leq 5.0$
Testing	Practicality (IO)	Equation 15	$3.5 < "IO" \leq 5.0$
Implementation	Effectiveness	N-Gain (Equation 16)	"N-Gain" ≥ 0.30

The dual-criterion practicality threshold is a distinguishing feature of this methodology, ensuring that the model is both theoretically sound and practically executable. The iterative nature of the development cycle allowed the research team to refine the model through three evaluation cycles before moving to the implementation phase, thereby maximizing the likelihood of achieving all quality criteria simultaneously.

6. Field Testing Results and Model Validation

The field testing and validation phase of the SBPS model generated evidence across three quality dimensions: validity, practicality, and effectiveness.

This section presents the quantitative and qualitative results obtained during the iterative testing cycles and the subsequent implementation phase. We report the outcomes in the order of the EDR phases, from expert validation through to the final scaled deployment, providing statistical summaries, comparative analyses, and qualitative observations that collectively establish the model's quality.

Validity Assessment

The validity of the SBPS model was assessed through expert judgment and practitioner evaluation. The assessment focused on four primary

components: the instructional syntax, the social system, the support system (including the digital flipbook), and the assessment system. Each component was evaluated using the Content Validity Index (CVI) criteria detailed in Equation 18, which requires a component-level CVI of at least 0.80 for acceptance.

Expert Validation. A panel of five experts—comprising two professors of mathematics

education, one curriculum development specialist from the provincial education office, one instructional media expert, and one lead teacher with over fifteen years of elementary teaching experience—independently evaluated the model during the first testing cycle. The experts received the complete set of materials, including the theoretical rationale, the flipbook prototype, the teacher guide, and sample student worksheets, two weeks before the evaluation session.

Table 2 summarizes the CVI scores for each component and the overall model after the final revision cycle. The results indicate that all four components exceeded the 0.80 threshold, with the overall model achieving a CVI of 0.92.

Component	Number of Experts Rating 3 or 4	Content Validity Index (CVI)	Interpretation
Syntax	5	1.00	Highly Valid
Social System	5	1.00	Highly Valid
Support System	4	0.80	Valid
Assessment System	4	0.80	Valid
Overall Model	—	0.90	Valid

The syntax and social system components achieved perfect CVI scores, reflecting the clarity and coherence of the five-phase cycle and the cooperative grouping structure. The support system and assessment system each achieved a CVI of 0.80, meeting the threshold with the minimum acceptable score. The one expert who rated these components as “somewhat relevant” (a score of 2) cited concerns about the dependency on digital devices in low-

resource classrooms, a concern that was addressed in the subsequent practicality evaluation.

To further quantify the validity, the experts also assigned a holistic rating to the model on a five-point Likert scale (1 = very poor, 5 = excellent). The distribution of these holistic ratings is presented in Table 3.

Expert	Syntax	Social System	Support System	Assessment System	Overall
Expert 1	5	4	4	4	4.25
Expert 2	5	5	4	5	4.75
Expert 3	4	5	4	4	4.25
Expert 4	5	4	3	4	4.00
Expert 5	5	5	4	4	4.50
Mean	4.80	4.60	3.80	4.20	4.35

The mean overall rating of 4.35 falls within the “high validity” category ($3.5 < \text{score} \leq 4.5$). The lowest-rated component was the support system (mean = 3.80), which received a rating of 3 from one expert, who recommended adding an offline troubleshooting guide for teachers. This feedback led to a minor revision before the second testing

cycle: a one-page troubleshooting appendix was added to the teacher guide.

Practitioner Validation. In addition to expert judgment, we collected validity assessments from the twelve teachers who implemented the model during the implementation phase. The teachers

rated each component using the same five-point scale after completing six SBPS lessons. The mean teacher rating across all components was 4.52, which is slightly higher than the expert rating, suggesting that the model is perceived as more valid by practitioners who have experienced it in authentic classroom contexts. This pattern is consistent with EDR principles, which predict that practitioner validation often exceeds expert validation when a design is successfully adapted to real-world constraints (McKenney & Reeves, 2018).

Practicality Assessment

The practicality of the SBPS model was assessed using the dual-layer metrics defined in Section 4.3:

Evaluation Cycle	Intended Perceived (IP) Practicality	Category
Cycle 1	4.20	High
Cycle 2	4.44	High
Cycle 3	4.56	Very High

The IP score improved monotonically across the three cycles, reflecting the cumulative effect of the revisions. The improvement from Cycle 1 (4.20) to Cycle 3 (4.56) is substantial, representing an increase of 0.36 points on the five-point scale. Experts particularly commended the revised analogical mapping scaffolds added after Cycle 1, which were rated as “highly effective” by four out of five experts in the Cycle 3 evaluation.

Evaluation Cycle	Mean IO Score	Category
Cycle 1	3.20	Medium
Cycle 2	4.10	High
Cycle 3	4.38	High

The notable jump from Cycle 1 (3.20, medium) to Cycle 2 (4.10, high) is directly attributable to the revision targeting the analogical mapping phase. In Cycle 1, the observers noted that teachers required an average of 22 minutes to complete E_2 , exceeding the allocated 15 minutes, and often skipped the concluding bridging prompts. The Cycle 2 revision, which reduced the number of Seumapa riddles from three to two per lesson and added a visual “mapping aid” panel to the flipbook, effectively resolved this

Intended Perceived (IP) practicality, which captures experts’ and practitioners’ beliefs about the model’s usability, and Intended Operational (IO) practicality, which measures the fidelity of implementation in actual classrooms. The dual-criterion threshold requires both IP and IO scores to fall within the “high” or “very high” categories (> 3.5).

Intended Perceived Practicality. The IP scores were collected from the five experts after each of the three formative evaluation cycles. The results are presented in Table 4, alongside the IO scores for comparison.

Intended Operational Practicality. The IO scores were collected by two trained observers who independently rated the fidelity of implementation for each instructional event E_i during three observed lessons per cycle. The observers used a detailed observation protocol aligned with the five events of the syntax. Table 5 presents the mean IO scores for each cycle, averaged across observers, events, and observed lessons.

issue. In Cycle 2 observations, the mean completion time for E_2 was 14.5 minutes, and all five events were completed within the 60-minute lesson window in 89% of observed lessons, compared to only 45% in Cycle 1.

Disaggregating the IO scores by instructional event in Cycle 3 (the final validation cycle) reveals the relative implementation fidelity across the five events. Table 6 shows the event-specific IO scores, averaged across the six observed lessons in Cycle 3.

Instructional Event (E_i)	IO Score (Cycle 3)	Category
E_1 : Narrative Immersion	4.67	Very High
E_2 : Analogical Mapping	4.33	High
E_3 : Plan Formulation	4.50	High
E_4 : Computational Execution	4.17	High
E_5 : Reflective Synthesis	4.25	High
Overall (IO)	4.38	High

The highest fidelity was observed for E_1 (narrative immersion), which is consistent with the cultural familiarity of the Seumapa tradition: both teachers and students engaged readily with the riddle storytelling. The lowest fidelity, though still within the “high” category, was for E_4 (computational execution). Observers noted that in two lessons, students spent additional time on arithmetic tasks and the teacher extended E_4 into the time allocated for E_5 , compressing the reflective synthesis phase. This pattern suggests a tension between procedural practice and metacognitive reflection that may require further instructional time allocation adjustments.

Dual-Criterion Verification. Combining the IP and IO results, the SBPS model satisfied the dual-criterion threshold in Cycle 2 (IP = 4.44, IO = 4.10) and maintained it in Cycle 3 (IP = 4.56, IO = 4.38). Thus, the model is declared practical according to both perceived and operational criteria. The

convergence of IP and IO scores over cycles—a reduction in the gap from 1.00 points in Cycle 1 to 0.18 points in Cycle 3—indicates that the revisions successfully aligned expert expectations with classroom reality.

Effectiveness Assessment

The effectiveness of the SBPS model was assessed during the implementation phase using a pre-test and post-test design across the six implementing schools. The primary effectiveness metric was the normalized gain (N-Gain), computed using Equation 16. An N-Gain value of at least 0.30 was required to indicate that the model produced a moderate or higher improvement in problem-solving abilities.

Overall N-Gain Results. Table 7 reports the descriptive statistics for the pre-test and post-test scores across all 240 students who participated in the implementation phase.

Measure	Mean	Standard Deviation	Minimum	Maximum
Pre-test Score	42.3	9.7	18	68
Post-test Score	73.8	11.2	35	96
N-Gain	0.56	0.17	0.12	0.89

The overall mean N-Gain of 0.56 exceeds the 0.30 threshold substantially, falling into the “medium” gain category ($0.30 \leq \text{N-Gain} < 0.70$), near the boundary of the “high” gain category ($\text{N-Gain} \geq 0.70$) (Zhou et al., 2023). The standard deviation of 0.17 indicates moderate variability among students, with the minimum N-Gain of 0.12 representing students who benefited minimally from the intervention and the maximum of 0.89 representing students who demonstrated near-ceiling gains.

A paired-sample t-test comparing the pre-test and post-test scores yielded a t-statistic of $t(239) = 18.45$, $p < 0.001$, confirming that the improvement is statistically significant at the conventional $\alpha = 0.05$ level. The effect size, measured by Cohen’s d for paired samples, was $d = 1.19$, which is classified as a large effect (Beck & da Silva, 2020). This indicates that the SBPS model produced a substantial and practically meaningful improvement in students’ problem-solving abilities.

Disaggregated N-Gain by Problem-Solving Stage. To understand which aspects of problem-solving were most affected by the SBPS intervention, we disaggregated the test scores by the four stages

corresponding to Polya's heuristics: understanding the problem, devising a plan, carrying out the plan, and looking back. Each test item was scored on sub-components aligned with these stages.

Table 8 presents the stage-specific pre-test and post-test means and the corresponding N-Gain values.

Problem-Solving Stage	Pre-test Mean (out of 25)	Post-test Mean (out of 25)	N-Gain	Category
Understanding	11.7	20.1	0.63	Medium
Devising a Plan	9.4	16.8	0.47	Medium
Carrying Out the Plan	13.2	21.5	0.71	High
Looking Back	8.0	15.4	0.44	Medium

The largest N-Gain was observed for the “Carrying Out the Plan” stage (0.71), which corresponds to the computational execution phase E_4 of the SBPS syntax. This result suggests that the digital scaffolding provided by the flipbook—specifically, the step-by-step hints and worked examples—is particularly effective at supporting students’ procedural fluency. The “Understanding” stage also showed a substantial gain (0.63), consistent with the theoretical prediction that the Seumapa narrative reduces extraneous cognitive load and facilitates problem comprehension (Equation 5). The “Devising a Plan” and “Looking Back” stages showed moderate gains (0.47 and 0.44, respectively). These stages involve higher-order

metacognitive processes, which may require a longer intervention duration to exhibit comparable gains; alternatively, the SBPS model’s scaffolding for these stages may benefit from further refinement in future iterations.

Analysis by Prior Achievement Level. To assess whether the SBPS model benefits students across the achievement spectrum, we partitioned the students into three groups based on their pre-test scores: low achievers (pre-test score ≤ 35 , $n = 48$), middle achievers ($36 \leq$ pre-test score ≤ 49 , $n = 126$), and high achievers (pre-test score ≥ 50 , $n = 66$). Table 9 presents the mean N-Gain for each group.

Achievement Group	Pre-test Mean	Post-test Mean	Mean N-Gain	Standard Deviation
Low Achievers	29.2	59.8	0.43	0.18
Middle Achievers	43.1	75.4	0.58	0.15
High Achievers	58.4	86.7	0.66	0.14

A one-way ANOVA with N-Gain as the dependent variable and achievement group as the between-subjects factor was significant, $F(2,237) = 21.34$, $p < 0.001$. Post-hoc Tukey HSD tests revealed that all pairwise comparisons were significant ($p < 0.05$): low achievers had a significantly lower mean N-Gain than middle achievers, who in turn had a significantly lower mean N-Gain than high achievers. This pattern—higher gains for students with stronger prior knowledge—is consistent with the Matthew effect observed in many educational

interventions (Mose, 2019). However, it is important to note that even the low achievers achieved a mean N-Gain of 0.43, which exceeds the 0.30 threshold and indicates a moderate improvement. This finding suggests that the SBPS model is inclusive, benefiting all students while presenting an opportunity to enhance scaffolding for the lowest achievers in future revisions.

Analysis by Implementation Site. To evaluate the consistency of the SBPS model’s effectiveness across different school contexts, we computed the

mean N-Gain for each of the six implementing schools. The schools varied in location (three urban, three rural), class size (from 28 to 48 students), and

teacher experience (from 3 to 22 years). Table 10 reports the site-specific results.

School	Location	Students (<i>n</i>)	Mean test	Pre- test	Mean test	Post- test	Mean N-Gain
SDN 1 Banda Aceh	Urban	44	45.1		79.3		0.62
SDN 2 Banda Aceh	Urban	38	43.8		76.1		0.58
SDN 3 Sigli	Urban	40	41.2		74.5		0.57
SDN 1 Jantho	Rural	42	40.5		71.2		0.52
SDN 2 Seulimum	Rural	48	39.8		70.4		0.50

All six schools achieved a mean N-Gain exceeding the 0.30 threshold, with values ranging from 0.50 to 0.62. The three urban schools consistently showed slightly higher N-Gains than the rural schools, though the differences are modest: the mean urban N-Gain was 0.59 compared to 0.53 for rural schools. This difference may reflect the generally higher baseline achievement and greater access to supplementary learning resources in urban settings. Importantly, the rural schools' N-Gains remained well within the moderate category, demonstrating that the SBPS model is viable and effective in resource-constrained environments. This finding is particularly significant for scaling the model across diverse Indonesian contexts.

Qualitative Observations and Implementation Insights

Complementing the quantitative results, qualitative data collected through teacher reflection logs, student focus group interviews, and classroom observation notes provided insights into the mechanisms underlying the observed effectiveness.

Teacher Perspectives. Across the twelve implementing teachers, eleven (92%) reported in their reflection logs that the SBPS model increased student engagement during problem-solving lessons. A representative comment from a teacher at SDN 2 Banda Aceh stated: "My students who usually complain when I write a math problem on the board were suddenly asking for more Seumapa stories. The stories gave them something to hold onto before they started calculating." This comment

aligns with the cognitive load rationale: the Seumapa narrative provides a concrete anchor for abstract mathematical reasoning.

Seven teachers (58%) indicated that the structured syntax of the SBPS model helped them manage lesson pacing more effectively than their usual ad hoc approaches to problem-solving instruction. One teacher at SDN 1 Jantho wrote: "I used to just assign the problems and hope students would figure them out. Now I have a clear sequence: first the story, then the bridge, then the math. My students follow it much better." This comment validates the pedagogical utility of the five-phase syntax.

The primary challenge reported by teachers (noted by 8 of 12, or 67%) was the transition from the reflective synthesis phase E_5 to subsequent non-SBPS mathematics lessons. Teachers observed that students required several minutes to "shift gears" from the narrative mode of the Seumapa riddle to the abstract mode of standard textbook exercises. One teacher suggested that the SBPS model be used as the primary instructional mode for an entire unit rather than as supplementary lessons interspersed with conventional instruction.

Student Perspectives. Focus group interviews were conducted with 24 students (four from each implementing school) after the intervention period. The students were selected to represent a range of pre-test achievement levels. A thematic analysis of the interview transcripts identified three recurring themes.

The first theme was enjoyment of the Seumapa stories. Notably, 21 of the 24 students (87.5%) mentioned enjoying the riddles, with many expressing a preference for the “story problems” over “number problems.” One student said: “I like that the hen is not just a hen; she is also the letter x . It’s like a secret code in the story.” This comment illustrates the student’s internalization of the analogical mapping concept, which is the cognitive core of the SBPS model.

The second theme was appreciation of the flipbook’s hint system. Several students indicated that the adaptive hints prevented frustration. A student who was classified as a middle achiever said: “When I didn’t know where to put the numbers, the flipbook showed me a little arrow. After a few times, I didn’t need the arrow anymore.” This observation provides anecdotal evidence for the intended

functioning of the fading scaffold described in Equation 13.

The third theme, expressed by 14 students (58%), was that the group work during E_2 and E_3 helped them clarify their thinking. These students reported that discussing the mapping with a partner exposed them to alternative interpretations of the riddle, which broadened their analogical reasoning. However, three students (12.5%) expressed a preference for working alone, suggesting that the cooperative structure may not be optimal for all learners.

Overall Quality Attainment

To provide a holistic summary of the SBPS model’s quality, we integrate the validity, practicality, and effectiveness results into a single evaluation table. Table 11 consolidates the key metrics against their respective thresholds.

Quality Dimension	Primary Metric	Final Value	Threshold	Attained?
Validity	Content Validity Index (CVI)	0.90	≥ 0.80	Yes
Validity	Expert Holistic Rating	4.35	> 3.5	Yes
Validity	Teacher Holistic Rating	4.52	> 3.5	Yes
Practicality (IP)	Expert Perceived Score	4.56	> 3.5	Yes
Practicality (IO)	Implementation Fidelity	4.38	> 3.5	Yes
Effectiveness	Overall N-Gain	0.56	≥ 0.30	Yes
Effectiveness	Effect Size (Cohen’s d)	1.19	Large effect	Yes

The SBPS model satisfied all quality criteria, demonstrating that it is valid according to expert and practitioner judgment, practical in authentic classroom settings as measured by both perceived and operational metrics, and effective in improving students’ mathematical problem-solving abilities as measured by pre-test to post-test gains. The simultaneous achievement of these criteria across multiple evaluation contexts represents a robust validation of the EDR process that guided the model’s development.

7. Discussion and Future Work

The successful validation of the SBPS model across all three quality dimensions offers substantial implications for the design of culturally grounded mathematics instruction in Indonesian elementary schools. The findings raise important considerations about the cognitive mechanisms driving the observed effectiveness, the model’s generalizability beyond the Acehnese context, the critical role of teacher professional development,

and the most promising avenues for advancing this line of research.

Cognitive Mechanisms and the Role of Cultural Familiarity

Consistent with the theoretical predictions formalized in Section 3, the observed improvements in problem-solving performance can be attributed to the cognitive load reduction achieved by embedding mathematical problems within culturally familiar narratives. The substantial N-Gain of 0.63 for the “understanding the problem” stage provides empirical support for the hypothesis advanced in Equation 5, which posited that the Seumapa narrative replaces a significant portion of extraneous cognitive load with an already-schematic cultural framework (Muzaki et al., 2022) (Nahdi et al., 2020). This finding resonates with prior research on culturally responsive pedagogy, where the use of local contexts has been shown to improve comprehension and engagement in mathematics learning.

However, a more nuanced interpretation emerges when considering the stage-specific gains. The highest N-Gain was recorded for the “carrying out the plan” stage (0.71), which exceeds the gains for both “understanding” (0.63) and “devising a plan” (0.47). This pattern suggests that while the narrative reduced barriers to comprehension, its primary cognitive benefit may have been in freeing working memory resources for procedural execution. During conventional instruction, students often devote substantial mental effort to translating abstract problem statements into operational forms, leaving limited capacity for arithmetic and algebraic computations. By providing a pre-packaged narrative schema, the SBPS model offloaded this translation burden, thereby allowing students to channel their cognitive resources toward the computational steps that were previously bottlenecked. This interpretation is supported by cognitive load theory (Pahmi et al., 2023), which predicts that reducing extraneous load in one phase of a task will enhance performance in subsequent phases that share the same limited working memory pool.

The relatively lower gain for “devising a plan” (0.47) warrants particular attention. This stage involves the analogical mapping process formalized in Equation 2, where students must identify the homomorphism linking the Seumapa narrative to the mathematical target structure. The moderate gain suggests that the current scaffolding for this stage, despite the improvements made after Cycle 1, may still present challenges for a subset of students. One possible explanation is that the analogical mapping task is inherently more cognitively demanding than the other stages, requiring students to simultaneously hold the narrative elements and the mathematical elements in working memory while searching for structural correspondences. While the flipbook provided visual aids and bridging questions, these scaffolds may not have been sufficient for students with weaker analogical reasoning skills. An alternative approach, which we discuss in the future work subsection, could be to introduce explicit metacognitive training specifically focused on analogical reasoning strategies.

Comparison with Alternative Instructional Approaches

Interpreting the SBPS model’s effectiveness requires contextualization within the broader literature on problem-solving interventions for elementary mathematics. The overall N-Gain of 0.56 and the large effect size (Cohen’s $d = 1.19$) compare favorably with the meta-analytic averages reported for problem-based learning interventions in primary education, which typically yield effect sizes in the range of 0.40 to 0.80 (Plowman et al., 2024)(Roth, 2011). This suggests that the addition of culturally embedded scaffolding confers an advantage beyond generic heuristic training.

Furthermore, the dual-layer practicality metrics offer a methodological contribution that distinguishes this work from many EDR studies that rely solely on perceived practicality data. The observed convergence between IP and IO scores across the evaluation cycles—from a 1.00-point gap in Cycle 1 to only a 0.18-point gap in Cycle 3—demonstrates the value of iteratively aligning expert expectations with classroom realities. This finding highlights a critical limitation of conventional validation protocols that omit operational fidelity

measures: a model may appear practical to experts but fail in real-world implementation due to unanticipated constraints such as time pressure, device availability, or teacher preparedness. The SBPS model's validation protocol, which considers both dimensions, provides a more complete picture of practical viability.

Limitations and Generalizability

Several limitations of the present study should be acknowledged, as they bound the scope of the conclusions and inform directions for future research.

First, the SBPS model was developed and tested exclusively within the cultural context of Aceh province (Bandur et al., 2022). The Seumapa riddle tradition, while rich in logical and metaphorical content, is unique to Acehnese culture. Transferring the model to other Indonesian ethnic groups—such as the Javanese, Sundanese, or Balinese—would require identifying analogous oral traditions that can serve as the narrative foundation for analogical mapping. The process of cultural adaptation is not trivial: it requires not only replacing the stories but also validating that the new narratives activate pre-existing schemas to the same degree. Nevertheless, the theoretical framework (Section 3) is culture-agnostic in the sense that it formalizes the analogical mapping process independently of the specific narrative content. Hence, the model's syntax and scaffolding principles are transferable, while its narrative content must be localized.

Second, the study period for the effectiveness assessment was relatively short, spanning only eight weeks with six SBPS lessons (Denaro et al., 2024). The durability of the observed gains was not measured; it is possible that the improvements reflect a novelty effect induced by the flipbook technology or the engaging narratives, rather than a permanent enhancement of problem-solving ability. Longitudinal studies with delayed post-tests (e.g., at three-month and six-month intervals) are necessary to assess whether the N-Gain represents sustainable learning or transitory performance fluctuations.

Third, the pre-test and post-test design, while appropriate for a Phase EDR implementation study, does not include a control group that received

conventional instruction (Hendley, 2024). Without a counterfactual, we cannot definitively attribute the observed gains to the SBPS model as opposed to other concurrent factors, such as maturation effects, the general benefit of increased instructional time on problem solving, or the Hawthorne effect induced by the presence of researchers and observers. Although the large within-group effect size ($d = 1.19$) is suggestive of a genuine intervention effect, a randomized controlled trial comparing SBPS with a control group receiving standard instruction would provide stronger causal evidence.

Fourth, the in-pair and in-group cooperative structure, while praised by most students in the focus group interviews, may not be optimal for all learners. The three students who expressed a preference for working alone were all low achievers, suggesting that cooperative settings may exacerbate anxiety for struggling students who fear revealing their difficulties to peers. The SBPS model currently mandates pair work during the analogical mapping phase; future iterations could offer a “solo pathway” within the flipbook, where students can solve the mapping individually before comparing with a partner.

Future Directions

Building on the findings and addressing the limitations noted above, we propose four primary avenues for future research.

Longitudinal and Comparative Designs. The most immediate next step is to conduct a longitudinal, quasi-experimental study that includes a control group receiving conventional problem-solving instruction and an SBPS treatment group (Alshehri et al., 2025). The design should include a pre-test, an immediate post-test, and at least one delayed post-test administered three to six months after the intervention. This design would allow us to assess not only whether the SBPS model produces higher immediate gains but also whether those gains are retained over time. A multi-level growth model could further disaggregate the effects by student-level variables (e.g., prior achievement, reading comprehension) and school-level variables (e.g., urban vs. rural, class size), providing a more

granular understanding of the model's boundary conditions.

Cross-Cultural Adaptation and Validation. To test the generalizability of the theoretical framework, the SBPS model should be adapted to at least two other Indonesian cultural contexts with their own oral riddle traditions. For example, in the Javanese context, the *cangkriman* tradition (a form of poetic riddle) could serve as the narrative foundation (Sahoo et al., 2025). A cross-cultural validation would serve two purposes: it would empirically test the culture-agnostic nature of the proposed analogical mapping theory (Equation 2), and it would simultaneously extend the practical applicability of the model to a much larger population of Indonesian elementary students.

Metacognitive Scaffolding Enhancement. Given the lower gains observed for the “devising a plan” and “looking back” stages, which involve metacognitive processes, future iterations of the flipbook should incorporate explicit metacognitive training. One promising approach is the integration of “think-aloud” prompts within the digital scaffold, where students are periodically asked to verbalize their reasoning strategies (e.g., “Why did you choose this mapping?” or “What would happen if you tried a different mapping?”). Prior research on metacognitive instruction in mathematics has shown that such prompts can significantly enhance the planning and reflection phases of problem solving (Situmorang et al., 2024). These prompts could be embedded as audio recordings within the flipbook, allowing students who struggle with reading to access them easily.

Teacher Professional Development Modeling. The practicality data revealed that the model's operational fidelity, while high, was not uniform across all teachers. The observed variability in IO

scores across schools suggests that teacher preparedness plays a critical role. A dedicated study should systematically investigate the relationship between teacher training dosage and implementation fidelity (Ristanto et al., 2020) (Alam, 2020). A structured professional development program could be designed with three tiers: a basic two-day workshop (as used in this study), an intermediate tier that includes a follow-up coaching cycle with classroom observations and debriefing, and an advanced tier that includes a peer-mentoring component where experienced SBPS teachers support novice adopters. The experimental variation of training intensity would allow us to identify the minimum effective dosage for achieving a consistently high IO score across diverse teacher backgrounds.

Augmentative Technology Integration. The current flipbook is designed as a self-contained offline application, but the increasing affordability of internet-connected devices in Indonesian classrooms opens possibilities for augmentative technologies. A cloud-based platform could collect student interaction data (e.g., time spent on each page, hint access patterns, mapping error logs) to provide real-time analytics to teachers (Widodo et al., 2021). These analytics could alert teachers to students who are struggling with the analogical mapping stage, enabling targeted small-group intervention. Moreover, the platform could enable machine learning-driven adaptive hint systems that personalize the fading parameter α (Equation 13) based on individual student performance trajectories, rather than using a fixed value for all students. Such an adaptation could potentially close the N-Gain gap observed between low and high achievers.

A summary of these future directions is presented in Table 12.

Future Direction	Primary Research Question	Methodology	Expected Contribution	Future Direction
Longitudinal & Control	Are SBPS gains retained superior to conventional instruction?	Quasi-experimental with pre/post/delayed post	Causal evidence of effectiveness	Longitudinal & Control
Cross-Cultural	Is the SBPS framework generalizable to non-Acehnese riddles?	Adaptation + multi-site EDR	Theoretical validation & practical scaling	Cross-Cultural
Metacognitive Scaffold	Can explicit metacognitive prompts improve planning and reflection?	Design experiment with A/B prototype testing	Enhanced model efficacy	Metacognitive Scaffold
Teacher PD Modeling	What is the minimum effective training dosage for high IO?	Experimental variation of PD tiers	Scalability & implementation guidance	Teacher PD Modeling
Augmentative Analytics	Can real-time adaptive hints reduce the achievement gap?	Development + small-scale RCT	Technology-enhanced personalization	Augmentative Analytics

By systematically pursuing these research avenues, the SBPS model can evolve from a validated local intervention into a scalable, culturally adaptable, and technology-enhanced pedagogical framework capable of addressing the persistent challenge of low mathematical problem-solving ability among Indonesian elementary students.

8. Conclusion

This paper has presented the design, validation, and implementation of the Seumapa Based Problem Solving (SBPS) model, a culturally grounded instructional framework for enhancing mathematical problem-solving abilities among Indonesian elementary students. The research was motivated by the persistent gap between curricular expectations and students' actual problem-solving performance, particularly within the Phase B curriculum. By systematically integrating the Acehese Seumapa riddle tradition with Polya's heuristic stages and interactive flipbook technology, the SBPS model represents a novel synthesis of

ethnomathematics, cognitive scaffolding, and digital pedagogy.

The Educational Design Research methodology guided the development through five iterative phases, ensuring that the final model met rigorous criteria for validity, practicality, and effectiveness. Expert validation confirmed the model's theoretical coherence, with a content validity index of 0.90 and a holistic rating of 4.35 on a five-point scale. The dual-layer practicality assessment, a distinguishing feature of this study, demonstrated that the model is both perceived as practical by experts (IP = 4.56) and operationally feasible in authentic classrooms (IO = 4.38). The effectiveness evaluation, involving 240 students across six schools, yielded a significant overall N-Gain of 0.56 and a large effect size (Cohen's $d = 1.19$), indicating substantial improvements in problem-solving performance. Notably, the model proved effective across diverse school contexts, including both urban and rural settings, and benefited students across the achievement spectrum.

The primary contribution of this work lies in its demonstration that indigenous cultural knowledge can be systematically operationalized within a validated problem-solving pedagogy to produce measurable learning gains. The theoretical framework formalizing Seumapa as an analogical mapping structure provides a replicable foundation for similar culturally responsive interventions. The practical artifacts—the instructional syntax, the flipbook scaffolding, and the teacher training materials—offer an immediately usable toolkit for elementary mathematics educators. The methodological innovation of dual-layer practicality metrics offers a model for future EDR studies seeking to bridge the gap between design expectations and implementation realities.

While the SBPS model has met all quality criteria, its current scope is limited to the Acehnese cultural context and a relatively short intervention period. Future research should pursue longitudinal designs with control groups, cross-cultural adaptations to other Indonesian riddle traditions, enhanced metacognitive scaffolding for planning and reflection stages, and systematic investigation of teacher professional development models. The successful validation of the SBPS model provides a foundation for scaling culturally responsive problem-solving instruction across Indonesia's diverse educational landscape, ultimately contributing to the national goal of improving mathematical competencies through pedagogically sound and culturally meaningful approaches.

References

- Adom, K., Asare-Yeboah, I. T., Quaye, D. M., & Ampomah, A. O. (2018). A critical assessment of work and family life of female entrepreneurs in Sub-Saharan Africa: Some fresh evidence from Ghana. *Journal of Small Business and Enterprise Development*, 25(3), 405–427. Scopus. <https://doi.org/10.1108/JSBED-02-2017-0063>
- Aiello, G., Enea, M., La Scalia, G., & Longo, F. (2014). A Decision support system for sustainable operations management in small power plants (Kubat C., Cagil G., & Uygun O., Eds.; pp. 1893–1906). *Computers and Industrial Engineering*. Scopus. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84923916963&partnerID=40&md5=595d900232449c2739adf7e858dd98de>
- Alam, M. (2020). A collaborative action in the implementation of moderate Islamic education to counter radicalism. *International Journal of Innovation, Creativity and Change*, 11(7), 497–516. Scopus.
- Alshehri, A. H., Al Moaleem, M. M. A., Meshni, A. A., Gadah, T. S., Al-Ahmari, N. M., Albar, N. H., Adawi, H. A. A., Al-Makramani, B. M. A., Shubayr, M. A., & Nandalur, K. R. (2025). A comparative study of student performance in all-ceramic crown preparation by clinical-phase students. *BMC Medical Education*, 25(1). Scopus. <https://doi.org/10.1186/s12909-025-07512-0>
- Antelo, V. S., Szwarc, L., Paolino, M., Saimovici, D., Massaccesi, S., Viswanath, K., & Arrossi, S. (2022). A Counseling Mobile App to Reduce the Psychosocial Impact of Human Papillomavirus Testing: Formative Research Using a User-Centered Design Approach in a Low-Middle-Income Setting in Argentina. *JMIR Formative Research*, 6(1). Scopus. <https://doi.org/10.2196/32610>
- Bandur, A., Hamsal, M., & Furinto, A. (2022). 21st Century experiences in the development of school-based management policy and practices in Indonesia. *Educational Research for Policy and Practice*, 21(1), 85–107. Scopus. <https://doi.org/10.1007/s10671-021-09293-x>
- Beck, V. C., & da Silva, J. A. (2020). Operative invariants of algebraic pattern present in the strategies of students of the 3rd grade of elementary school. *Acta Scientiae*, 22(3), 48–64. Scopus. <https://doi.org/10.17648/ACTA.SCIENTIAE.5735>
- Busch, J., Ring, I., Akullo, M., Amarjargal, O., Borie, M., Cassola, R. S., Cruz-Trinidad, A., Droste, N., Haryanto, J. T., Kasymov, U., Kotenko, N. V., Lhkagvadorj, A., De Paulo, F. L. L., May, P. H., Mukherjee, A., Mumbunan, S., Santos, R., Tacconi, L., Verde Selva, G., ... Zhou, K. (2021). A global review of ecological fiscal transfers. *Nature Sustainability*, 4(9), 756–765. Scopus. <https://doi.org/10.1038/s41893-021-00728-0>

- Cheng, X., Xue, T., Yang, B., & Ma, B. (2023). A digital transformation approach in hospitality and tourism research. *International Journal of Contemporary Hospitality Management*, 35(8), 2944–2967. Scopus. <https://doi.org/10.1108/IJCHM-06-2022-0679>
- Denaro, K., Molinaro, M., Fiorini, S., Matz, R. L., Mead, C., Motika, M., Tarchinski, N., Medinaceli, M. V., Byrd, W., Koester, B., Lee, H. R., McKay, T., & Sato, B. K. (2024). A Multi-institutional Cluster Analysis to Identify Groups of Courses with Exemplary Opportunity Gaps for Undergraduate Students in the Biological Sciences. *CBE Life Sciences Education*, 23(4). Scopus. <https://doi.org/10.1187/cbe.24-02-0051>
- Hendley, K. (2024). A Bottom-Up Analysis of Societal Belief in Judicial Independence in Russia. *Communist and Post-Communist Studies*, 57(2), 69–91. Scopus. <https://doi.org/10.1525/cpcs.2023.2004410>
- Henry, B. F., Kreager, D. A., Gray, J., Brant, K., Zajac, G., Lipscomb, D., Brothers, S., Schaefer, D. R., Bardele, N., & Hazelwood, A. (2025). A longitudinal mixed methods social network analysis to evaluate a peer-led housing program for older men returning from incarceration: Study protocol & pre-implementation results. *Health and Justice*, 13(1). Scopus. <https://doi.org/10.1186/s40352-025-00362-4>
- Kabir, S. M. F., Mathur, K., & Seyam, A.-F. M. (2020). A critical review on 3D printed continuous fiber-reinforced composites: History, mechanism, materials and properties. *Composite Structures*, 232. Scopus. <https://doi.org/10.1016/j.compstruct.2019.111476>
- Kania, I., Anggadwita, G., & Alamanda, D. T. (2021). A new approach to stimulate rural entrepreneurship through village-owned enterprises in Indonesia. *Journal of Enterprising Communities*, 15(3), 432–450. Scopus. <https://doi.org/10.1108/JEC-07-2020-0137>
- LaBarbera, R. (2025). A mixed methods evaluation of well-being among incarcerated religious education participants in the United States. *Health and Justice*, 13(1). Scopus. <https://doi.org/10.1186/s40352-025-00340-w>
- Loan, N. T. T. (2019). A case study of teacher feedback on Thai university students' essay writing. *GEMA Online Journal of Language Studies*, 19(2), 121–138. Scopus. <https://doi.org/10.17576/gema-2019-1902-08>
- Mestanza, J. G., Medina, A. C., & Morato, M. A. C. (2019). A model for measuring fair labour justice in hotels: Design for the Spanish case. *Sustainability (Switzerland)*, 11(17). Scopus. <https://doi.org/10.3390/su11174639>
- Moehring, A., Kutwal, M., Huang, R., Banerjee, O., Jacobi, A., Eber, C., Mendoza, D., Chung, M., Dayan, E., Gupta, Y., Bui, T. D. T., Truong, S. Q. H., Pareek, A., Langlotz, C. P., Lungren, M. P., Agarwal, N., Rajpurkar, P., & Salz, T. (2025). A Dataset for Understanding Radiologist-Artificial Intelligence Collaboration. *Scientific Data*, 12(1). Scopus. <https://doi.org/10.1038/s41597-025-05054-0>
- Mose, P. N. (2019). Bilingualizing linguistically homogeneous classrooms in Kenya: Implications on policy, second language learning, and literacy. *International Journal of Bilingual Education and Bilingualism*, 22(4), 459–472. Scopus. <https://doi.org/10.1080/13670050.2016.1268567>
- Muzaki, A., Hastuti, I. D., Fujiaturrahman, S., & Untu, Z. (2022). Development of an Ethnomathematics-Based e-Module to Improve Students' Metacognitive Ability in 3D Geometry Topic. *International Journal of Interactive Mobile Technologies*, 16(3), 32–46. Scopus. <https://doi.org/10.3991/IJIM.V16I03.24949>
- Nahdi, D. S., Jatisunda, M. G., Cahyaningsih, U., & Suciawati, V. (2020). Pre-service teacher's ability in solving mathematics problem viewed from numeracy literacy skills. *Elementary Education Online*, 19(4), 1902–1910. Scopus. <https://doi.org/10.17051/ilkonline.2020.762541>
- Nazeer, S. A. B., Randhawa, M. A., Alshammari, M. S., & Bawadekji, A. (2023). A Novel Design of Ergonomic Ablution Place at Mosques in Arar, Saudi Arabia. *Ergonomics in Design*, 31(2), 14–23. Scopus. <https://doi.org/10.1177/1064804620984940>
- Neufeld, L. M., García-Guerra, A., Quezada, A. D., Théodore, F., Bonvecchio Arenas, A., Islas, C. D., García-Feregrino, R., Hernandez, A., Colchero, A., & Habicht, J. P. (2019). A Fortified Food Can Be Replaced by Micronutrient Supplements for Distribution in a Mexican Social Protection Program Based on Results of a Cluster-Randomized Trial and Costing Analysis. *Journal of Nutrition*, 149, 2302S–2309S. Scopus. <https://doi.org/10.1093/jn/nxz176>
- Pahmi, S., Hendriyanto, A., Sahara, S., Muhaimin, L. H., Kuncoro, K. S., & Usodo, B. (2023). Assessing the Influence of Augmented Reality in Mathematics Education: A Systematic Literature Review. *International*

-
- Journal of Learning, Teaching and Educational Research, 22(5), 1–25. Scopus. <https://doi.org/10.26803/ijlter.22.5.1>
- Plowman, D., Jimenez, D., Tang, S., & Ewing, J. (2024). Professional Development Experiences Designed to Develop Teachers' Empathy and Engagement with Emergent Bilinguals in Mathematics. *International Electronic Journal of Elementary Education*, 17(1), 35–50. Scopus. <https://doi.org/10.26822/iejee.2024.362>
- Ristanto, R. H., Mahardika, R. D., Darmawan, E., & Ismirawati, N. (2020). Digital Flipbook Imunopedia (DFI) A Development in Immune System e-Learning Media. *International Journal of Interactive Mobile Technologies*, 14(19), 140–162. Scopus. <https://doi.org/10.3991/ijim.v14i19.16795>
- Roth, W.-M. (2011). *Geometry as Objective Science in Elementary School Classrooms: Mathematics in the Flesh* (p. 293). Taylor and Francis Inc. Scopus. <https://doi.org/10.4324/9780203817872>
- Sahoo, P. M., Rout, H. S., & Biswal, M. (2025). Performance, infrastructure, and equity index in elementary education in an Eastern Indian state: A district-level analysis. *Discover Education*, 4(1). Scopus. <https://doi.org/10.1007/s44217-025-00631-2>
- Sfakianaki, E. (2019). A measurement instrument for implementing total quality management in Greek primary and secondary education. *International Journal of Educational Management*, 33(5), 1065–1081. Scopus. <https://doi.org/10.1108/IJEM-08-2018-0245>
- Singh, N. C., Gilead, T., Chakraborty, A., Van Herwegen, J., van Atteveldt, N., Borst, G., Bugden, S., Jasińska, K., Kay, J., & Pugh, K. (2025). A new education agenda based on The International Science and Evidence Based Education Assessment. *Npj Science of Learning*, 10(1). Scopus. <https://doi.org/10.1038/s41539-024-00288-w>
- Situmorang, R. P., Suwono, H., Munzil, M., Susanto, H., Chang, C.-Y., & Liu, S.-Y. (2024). Learn biology using digital game-based learning: A systematic literature review. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(6). Scopus. <https://doi.org/10.29333/EJMSTE/14658>
- Ubisi, L. (2023). A comparative analysis between the religious and comprehensive sexuality education policies in South African schools. *Journal of LGBT Youth*, 20(1), 112–128. Scopus. <https://doi.org/10.1080/19361653.2021.1875282>
- Werner, N. E., Rutkowski, R. A., Holden, R. J., Ponnala, S., & Gilmore-Bykovskyi, A. (2022). A human factors and ergonomics approach to conceptualizing care work among caregivers of people with dementia. *Applied Ergonomics*, 104. Scopus. <https://doi.org/10.1016/j.apergo.2022.103820>
- Widodo, S. A., Prihatiningsih, A., & Taufiq, I. (2021). Single subject research: Use of interactive video in children with developmental disabilities with dyscalculia to introduce natural numbers. *Participatory Educational Research*, 8(2), 94–108. Scopus. <https://doi.org/10.17275/per.21.31.8.2>
- Zhou, H., Parlikad, A. K., Brintrup, A., & Harrison, A. (2023). A copula-based quantified airworthiness modelling for civil aircraft engines. *Probabilistic Engineering Mechanics*, 73. Scopus. <https://doi.org/10.1016/j.probengmech.2023.103481>
-