



Preservice Mathematics Teachers' Contextual Numeracy: Difficulty and Changes in Problem-Solving Patterns

Sri Winarni¹, Kamid^{2*}, Asrial³, Jefri Marzal⁴

^{1,2*,3,4}Universitas Jambi, Indonesia

E-mail: kamid.fkip@unja.ac.id

Abstract

Numeracy among pre-service mathematics teachers is frequently evaluated through test scores alone, making changes in problem-solving processes difficult to capture. This study examined changes in patterns of numeracy difficulties and problem-solving processes from pre-test to post-test following the implementation of DIFUNDES-PROBLEMING. A mixed-methods approach involving 38 mathematics education students at Universitas Jambi was employed. Data were collected using parallel contextual numeracy instruments on systems of linear equations and statistical reasoning, a holistic scoring rubric on a 0–100 scale, coding and categorization informed by grounded theory procedures, and Wilcoxon signed-rank tests. Pre-test results indicated incomplete problem-solving processes, with modelling, verification, and reasoning often absent or only partially demonstrated. Post-test results showed a more structured sequence of modelling, calculation, verification, and conclusion drawing, with mean scores increasing from 45.57 to 76.76. Wilcoxon signed-rank tests indicated a statistically significant improvement. Qualitative analysis of changes across the problem-solving indicators showed the greatest improvements in modelling, calculation, and verification, whereas contextual reasoning remained the primary challenge, particularly on items 1.d, 2.c, and 2.d. These findings indicate that DIFUNDES-PROBLEMING contributed to meaningful improvements in both numeracy achievement and the systematic quality of problem-solving processes.

Keywords: contextual numeracy; DIFUNDES-PROBLEMING; pre-service mathematics teachers; pre-test and post-test.



INTRODUCTION

The Merdeka Curriculum emphasizes essential content, while the Minimum Competency Assessment (AKM) measures numeracy beyond routine calculation: learners are expected to apply mathematical knowledge and reasoning in contextual situations (Kemendikbud, 2020). Despite these policy initiatives, national performance continues to present significant challenges, as reflected in the results of the Programme for International Student Assessment (PISA) 2022 (OECD, 2022). Beyond concerns about achievement, recent educational reforms have also emphasized how students learn and transfer knowledge across contexts. Within this orientation, the Deep Learning approach in education further emphasizes that learning experiences should not only lead to conceptual mastery but also enable learners to transfer their understanding to new situations and construct meaningful learning experiences (Anggraena et al., 2025).

Pre-service mathematics teachers face two related numeracy demands. First, they must solve contextual numeracy problems themselves. Second, they must design instructions that enable students to think, reason, and make decisions using quantitative information in similar ways. However, numeracy preparation across teacher education institutions in Indonesia has not yet developed as systematically as university-level numeracy programs in Australia, South Africa, and Ireland (Forgasz & Hall, 2019; Marx & Winter, 2014; O'Sullivan, 2022). International literature similarly conceptualizes teacher numeracy as the ability to mathematize problems and interpret quantitative information in authentic situations rather than merely perform procedures accurately (Bolstad, 2020; Botha & van Putten, 2018; Dole & Geiger, 2018; Geiger et al., 2015). A gap between formal mathematical knowledge and its application in real-world contexts has also been reported among pre-service mathematics teachers (Marhami et al., 2023; Winarni et al., 2024). These findings indicate that numeracy involves not only procedural proficiency but also interpreting context, constructing models, verifying results, and drawing conclusions.

In Indonesia, PISA-like and AKM-based numeracy instruments have been developed to assess numeracy (Efriani et al., 2019; Oktiningrum et al., 2016), while studies continue to report difficulties in mathematical reasoning among pre-service teachers and challenges in designing contextual tasks among teachers (Aprilianti et al., 2022; Fanani et al., 2022; Herizal et al., 2022; Winarni et al., 2024). Although numeracy training and intervention studies have demonstrated positive outcomes (Getenet, 2022; Novita et al., 2022; Toklu & Hursen, 2021; Walsh et al., 2022), improvements are commonly reported through pre-test and post-test scores (Forgasz & Hall, 2019; Vale & Westaway, 2024). Such evidence provides limited information about how pre-service teachers change their ways of understanding contextual problems, constructing mathematical models, verifying results, and interpreting solutions. Because contemporary numeracy perspectives emphasize these processes alongside achievement outcomes (Geiger & Schmid, 2024), research is needed that documents not only changes in numeracy scores but also shifts in the problem-solving processes underlying those changes.

To address this need, instructional designs are required that do more than present contextual tasks: they must accommodate variation in learning readiness, align learning goals with assessment evidence, and provide authentic opportunities for modelling, verification, and contextual reasoning. These requirements are theoretically supported by Differentiated Instruction, which adjusts learning experiences to learners' readiness (Tomlinson, 2017); Understanding by Design, which strengthens coherence among intended understandings, assessment, and instructional activities (Wiggins & McTighe, 2005); and Problem-Based Learning and Project-Based Learning, which situate mathematical concepts in authentic decision-making contexts (Permatasari et al., 2019; Viro & Joutsenlahti, 2020). Within the Deep Learning approach in education, such integration is further expected to support meaningful transfer of understanding across situations (Anggraena et al., 2025).

Accordingly, this study implemented DIFUNDES-PROBLEMING (Differentiated Instruction and Understanding by Design integrated with Problem-Based Learning and Project-Based Learning) in instruction on systems of linear equations in two variables and statistics. DIFUNDES-PROBLEMING was selected because its integrated components provide a theoretically coherent response to incomplete problem-solving processes, particularly weaknesses in modelling, verification, and contextual reasoning, among pre-service mathematics teachers with diverse readiness levels. Changes following its implementation were examined through a problem-solving rubric (Tekin-Dede & Bukova-Güzel, 2018)

and paired qualitative analysis of contextual numeracy assessments administered in a pre-test and post-test cycle. Operationally, DIFUNDES-PROBLEMING aligns these components within the Deep Learning approach in education (Anggraena et al., 2025) to connect learning goals, assessment practices, authentic problem-based learning experiences, and pre-service mathematics teachers' diverse levels of readiness. This approach enabled the researchers to investigate whether shifts in numeracy difficulties and problem-solving processes accompanied changes in numeracy achievement.

The novelty of this study lies in its process-oriented examination of contextual numeracy development among pre-service mathematics teachers. Rather than reporting pre-test and post-test improvements primarily as score changes, this study examines changes in patterns of numeracy difficulties and problem-solving processes through comparative qualitative analysis of pre-test and post-test essay responses, supported by holistic rubric scores. Using parallel contextual numeracy instruments in systems of linear equations in two variables and statistics contexts, the study identifies not only whether numeracy achievement increased but also which aspects of problem solving, including understanding, modelling, verification, and contextual reasoning, improved, persisted, or remained challenging after instruction. Thus, the study provides evidence of changes in the organization of problem-solving processes underlying numeracy achievement, rather than representing improvement solely as uniform score gains.

Against this background, the present study aimed to (1) analyze shifts in patterns of numeracy difficulties and problem-solving processes across the stages of understanding, modelling, verifying, and concluding; (2) describe shifts in rubric achievement as supporting quantitative evidence; and (3) identify aspects of numeracy that improved, persisted, or still required further strengthening. By comparing achievement and difficulty patterns between the pre-test and post-test, this article addressed the following research question: How do patterns of numeracy difficulties and problem-solving processes among pre-service mathematics teachers change following the implementation of DIFUNDES-PROBLEMING?

METHOD

Research Type and Design

This article reports a mixed-methods analysis (Creswell, 2014) embedded within the retrospective phase (*reflection on action*) of a broader Didactical Design Research (DDR) project (Brousseau, 1997; Suryadi, 2019). The broader DDR project comprised prospective, metapedadidactic, and retrospective analyses. However, consistent with the specific focus of this article, only the retrospective phase is reported here. The retrospective analysis reported in this article examines changes in patterns of numeracy difficulties and problem-solving processes among pre-service mathematics teachers through comparative analysis of pre-test and post-test data following the implementation of DIFUNDES-PROBLEMING. The prospective and metapedadidactic phases informed the development, refinement, and classroom implementation of DIFUNDES-PROBLEMING but are not reported in detail in this article. Accordingly, this article does not report the entire DDR process but emphasizes the mixed-methods examination of changes in numeracy difficulties and problem-solving processes, with numeracy achievement used as supporting quantitative evidence. The validity of the research instruments was evaluated using the product-quality criteria proposed by Nieveen et al. (2007).

Setting, Time, and Participants

The study was conducted in the Department of Mathematics Education at Universitas Jambi during the odd semester of the 2024/2025 academic year. Participants consisted of 38 fifth-semester students from the 2023 cohort selected through purposive sampling (Creswell, 2014) based on their willingness to participate in all stages of the study. All participants were pre-service mathematics teachers enrolled in a mathematics teacher education program. Two participants did not complete the post-test and were therefore excluded from post-test categorization, post-test mean-score calculation, and paired pre-test–post-test analyses. Accordingly, pre-test descriptive analyses were based on 38 participants, whereas post-test descriptive analyses and paired analyses were based on the 36 participants who completed both assessments.

Research Instruments

The research instruments consisted of contextual numeracy pre-test and post-test instruments, a holistic problem-solving rubric, and an open-ended essay reflection questionnaire administered after the pre-test. The numeracy assessments were designed as parallel instruments with different contextual settings but equivalent cognitive demands. The parallel forms were developed based on the same AKM numeracy indicators, cognitive processes (*understand*, *apply*, and *reason*), and intended difficulty levels (Kemendikbud, 2020) and were reviewed by three mathematics education experts as part of the instrument validation process. In the pre-test, the contexts involved savings modelling using systems of linear equations in two variables (Raka) and sea-temperature statistics (Banda Sea–Java Sea). In the post-test, the contexts involved sea-temperature statistics (Flores Sea–Sumba Strait) and the modelling of online transportation package costs and voucher usage within a specified budget.

Responses were scored using a holistic problem-solving rubric adapted from Tekin-Dede & Bukova-Güzel (2018). The resulting scores were converted to a 0–100 scale and classified into four achievement categories: *Needs Special Intervention*, *Basic*, *Capable*, and *Advanced*. The pre-test and post-test instruments were validated by three mathematics education experts and achieved an overall validity score of 3.94 out of a maximum score of 4.00, indicating a high level of validity based on the validation criteria used in this study. Parallel items were constructed to enable the identification of changes in patterns of numeracy difficulties and problem-solving processes while minimizing potential practice effects resulting from repeated exposure to the same context.

Research Procedure

The research procedure comprised four stages. First, participants completed a two-hour contextual numeracy pre-test and an open-ended essay reflection questionnaire on a separate day prior to the implementation of DIFUNDES-PROBLEMING. Second, DIFUNDES-PROBLEMING, integrating Differentiated Instruction, Understanding by Design, Problem-Based Learning, and Project-Based Learning within the Deep Learning approach in education (Anggraena et al., 2025), was implemented over four instructional days, comprising two days of instruction on systems of linear equations in two variables and two days on statistics, with approximately 28 effective instructional hours in total. Instruction incorporated differentiation based on learning readiness, authentic problem- and project-based learning activities, reflective learning processes, and assessments aligned with learning goals. Third, participants completed a two-hour contextual numeracy post-test, designed in parallel with the pre-test, on a separate day following the four instructional days of DIFUNDES-PROBLEMING implementation. Fourth, paired pre-test and post-test data from the 36 participants who completed both assessments were comparatively analyzed to identify changes in numeracy difficulties, problem-solving processes, and numeracy achievement.

Data Analysis Techniques

Quantitative analysis included the distribution of achievement categories, mean scores on a 0–100 scale, normality *testing* using the Kolmogorov–Smirnov and Shapiro–Wilk tests, and Wilcoxon signed-rank tests to compare pre-test and post-test scores for the 36 participants who completed both assessments, because the post-test data did not satisfy the normality assumption. Quantitative analysis was used as supporting evidence to describe overall changes in numeracy achievement.

Qualitative analysis of essay responses and reflection questionnaires was conducted through coding and categorization informed by grounded theory procedures, following the stages of open coding, axial coding, and selective coding (Corbin & Strauss, 1990; Creswell, 2014). Paired pre-test and post-test essay responses from the 36 participants who completed both assessments were compared to identify changes in patterns of numeracy difficulties, solution strategies, modelling processes, verification practices, and contextual reasoning, while the post-pre-test reflection questionnaires were used to provide additional evidence of participants' awareness of difficulties and intentions for improvement. The coding results were subsequently synthesized into core categories to characterize shifts in numeracy problem-solving processes among pre-service mathematics teachers.

Data trustworthiness was strengthened through triangulation between test results and reflection questionnaires, systematic documentation of the analytical process, and re-examination of analytical findings using artificial intelligence-assisted tools in comparison with the original research data. In this study, artificial intelligence-assisted tools were used to support data organization, coding management, and verification of analytical consistency. All coding decisions, category development, interpretation, conceptual verification, and conclusion drawing were conducted and verified by the researchers. Therefore, all academic and scientific decisions remained under the authority and responsibility of the researchers.

RESULTS

The results are presented to address the focus of this article, namely changes in patterns of numeracy difficulties and problem-solving processes among pre-service mathematics teachers from pre-test to post-test following the implementation of DIFUNDES-PROBLEMING. This section sequentially presents supporting quantitative evidence through achievement-category distributions and paired-difference tests, patterns of numeracy difficulties identified in the pre-test, response patterns observed in the post-test, and shifts in problem-solving processes revealed through comparisons of responses to parallel items. Comparative qualitative analysis of paired pre-test and post-test responses served as the primary basis for interpreting changes in how participants understood problems, constructed mathematical models, verified results, and drew conclusions in contextual situations, whereas rubric scores and Wilcoxon signed-rank tests were used as supporting quantitative evidence. Consequently, the reported changes were based on broader patterns identified across participants rather than on individual response excerpts alone.

Supporting Quantitative Evidence

A summary of holistic pre-test and post-test achievement is presented in Table 1.

Table 1. Distribution of Pre-test and Post-test Numeracy Achievement Categories

Achievement Category (Rubric)	Pre-test Frequency	Pre-test (%)	Post-test Frequency	Post-test (%)
Needs Special Intervention	1	2.63	0	0.00
Basic	26	68.42	2	5.56
Capable	11	28.95	10	27.78
Advanced	0	0.00	24	66.67
Mean Score (0–100)	45.57	–	76.76	–

Note. Pre-test N = 38; post-test N = 36. Two participants did not complete the post-test and were excluded from post-test frequencies, percentages, and mean-score calculation.

Interpretation of Table 1 indicates a substantial change in the numeracy achievement profile of the participants. In the pre-test, most pre-service mathematics teachers were categorized as Basic (68.42%), whereas in the post-test, most participants were categorized as Advanced (66.67%). The pre-test mean score was 45.57 based on 38 participants, whereas the post-test mean score was 76.76 based on the 36 participants who completed the post-test. These results suggest that, following the implementation of DIFUNDES-PROBLEMING, a greater proportion of participants were able to complete numeracy tasks at higher levels of competence. However, descriptive statistics alone could not determine whether the observed differences were statistically significant. To determine the statistical procedure for the paired comparison, score distributions were examined for the 36 participants who completed both the pre-test and post-test. The results are presented in Table 2.

Table 2. Tests of Normality

Variable	Test Occasion	Kolmogorov–Smirnov Statistic	df	Sig.	Shapiro–Wilk Statistic	df	Sig.
Total Score	Pre-test	.104	36	.200*	.980	36	.744
	Post-test	.185	36	.003	.879	36	.001

Note. $p = .200$ represents the lower bound of the true significance value.

a. Lilliefors Significance Correction.

Table 2 shows that the pre-test scores were approximately normally distributed, whereas the post-test scores showed a significant departure from normality based on both the Kolmogorov–Smirnov and Shapiro–Wilk tests. Given this distributional pattern and the ordinal basis of the rubric scores, the paired pre-test–post-test comparison was conducted using the Wilcoxon signed-ranks test.

Table 3. Wilcoxon Signed-Ranks Test for Pre-test and Post-test Scores

Test Statistic	Post-test – Pre-test
Z	-5.117 ^b
Asymp. Sig. (2-tailed)	.000

Note.

a. Wilcoxon Signed-Rank Test.

b. Based on negative ranks.

The results presented in Table 3 indicate a statistically significant difference in numeracy achievement between the pre-test and post-test among the 36 participants who completed both assessments ($Z = -5.117$, $p < .001$). Post-test numeracy achievement was significantly higher than pre-test achievement following the implementation of DIFUNDES-PROBLEMING. Nevertheless, the primary focus of this article is not the magnitude of score improvement itself but rather the ways in which patterns of numeracy difficulties and problem-solving processes changed following the intervention. Therefore, the subsequent analysis focuses on examining these process changes through coding and categorization of pre-test and post-test responses informed by grounded theory procedures.

Difficulty Patterns in the Pre-test Numeracy Assessment

Pre-test qualitative analysis was conducted using coding and categorization informed by grounded theory procedures, comprising open coding, axial coding, and selective coding. During open coding, essay responses from 38 pre-service mathematics teachers on savings modelling using systems of linear equations in two variables and sea-temperature statistics contexts, together with reflection questionnaires, were segmented into meaning units and assigned labels according to response characteristics, including correct responses (J), errors or misconceptions (K), process limitations (M), solution strategies (A), and improvement-oriented values or intentions (C). During axial coding, these codes were organized into nine axial categories: K1 initial conditions (cognitive/affective), K2 contextual numeracy competence in systems of linear equations in two variables and statistics, K3 task contexts (personal savings-related modelling and scientific sea-temperature statistics), K4 process limitations (incomplete modelling and weak verification), K5 solution strategies, K6 accurate responses, K7 errors or misconceptions, K8 process limitations in performance outcomes, and K9 reflective awareness and improvement intentions. These categories were further organized into five themes: T1 phenomenon (K2), T2 contexts (K3), T3 moderating factors (K1 and K4), T4 strategies (K5), and T5 consequences (K6–K9). The relationships among the categories and themes were characterized as causal, contextual, temporal, and hierarchical. Subsequently, selective coding was used to synthesize these findings into core categories characterizing patterns of numeracy difficulties among pre-service mathematics teachers.

The coding results indicated that the primary difficulty experienced by pre-service mathematics teachers lay less in computational ability than in incomplete problem-solving processes. In many responses, participants were able to perform particular calculations correctly; however, they did not

consistently construct mathematical models, verify results, or connect the procedures used to the contextual demands of the problem being solved. Consequently, contextual reasoning was often fragmented, and some responses ended with answers or claims that were not adequately supported by the solution process.

This pattern was also evident in the reflection questionnaires completed after the pre-test. Several participants acknowledged that they tended to perform calculations immediately without first constructing a mathematical model, verified their answers only superficially, or did not fully understand the relationship between the available data and the problem context. At the same time, most participants demonstrated an emerging awareness of these weaknesses and expressed intentions to improve their approaches to problem solving. These findings suggest that awareness of difficulty had begun to develop, although it had not yet been fully translated into changes in problem-solving processes.

To obtain a more comprehensive understanding of the relationships among initial conditions, solution strategies, numeracy performance, and participant reflection, the results of open coding and axial coding were further synthesized through selective coding into three core categories (KI1–KI3), as visualized in Figure 1. KI1 (contexts, conditions, and strategies) integrates T2–T4 (K3, K1, K4, and K5); KI2 (mastery and performance variation) integrates T1 and T5 (K2 and K6–K8); and KI3 (reflection and improvement directions) integrates T5 (K8 and K9). Figure 1 also represents the temporal problem-solving process from initial conditions and limitations (K1 and K4), through solution strategies (K5) and performance outcomes (K6–K8), to reflection and improvement intentions (K9), together with distinctive difficulty patterns in the personal and scientific numeracy contexts.

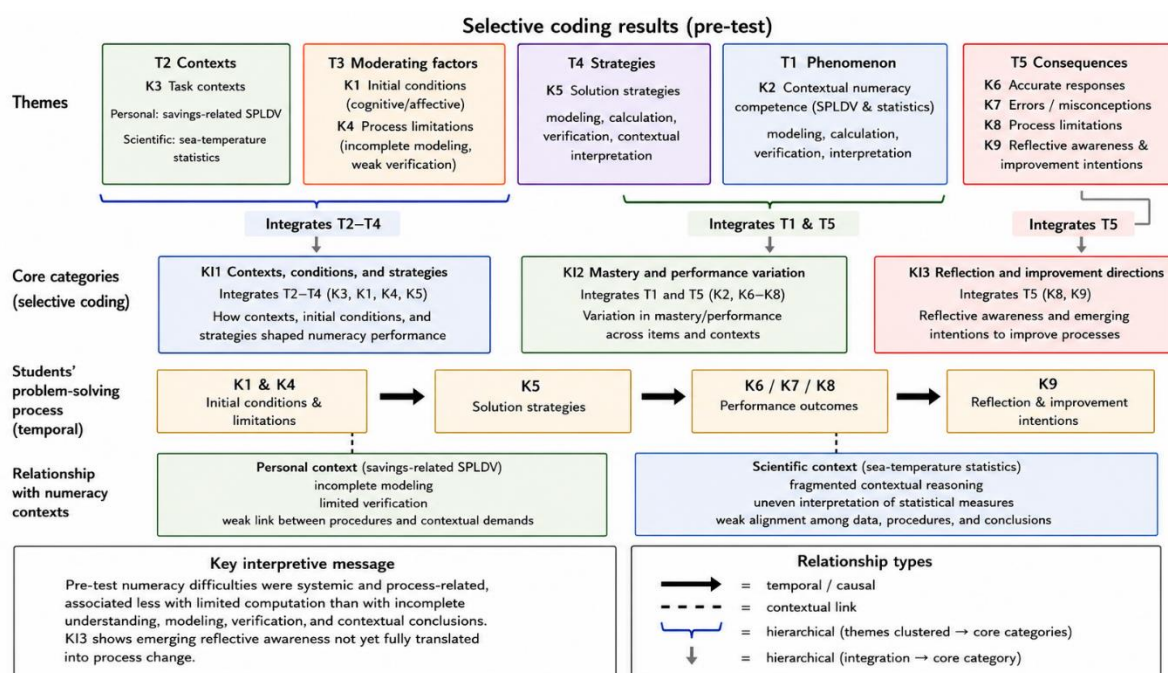


Figure 1. Integrative Representation of Selective Coding Results from the Pre-test Contextual Numeracy Assessment

Interpretation of the integrative representation in Figure 1 suggests that the numeracy difficulties observed in the pre-test were systemic and process-related. KI1 describes the contexts, conditions, and strategies that influenced numeracy performance; KI2 characterizes variation in mastery and numeracy performance across items and contexts; and KI3 represents reflective processes and emerging directions for improvement identified by participants. The relationships among these three core categories indicate that the predominance of the Basic category in Table 1 was not solely attributable to limited computational ability. Rather, it was more closely associated with incomplete processes of understanding problems, modelling situations, verifying results, and drawing contextual conclusions.

Response Patterns in the Post-test Numeracy Assessment

Post-test qualitative analysis was conducted using coding and categorization informed by grounded theory procedures, comprising open coding, axial coding, and selective coding. During open coding, 63 codes (OC-PT) derived from 36 analyzable response sheets were mapped onto a continuum of correct, incomplete, incorrect, and unfinished responses across items. During axial coding, these codes were synthesized into nine axial categories using the Strauss and Corbin paradigm to examine conditions, phenomenon, context, intervening conditions, strategies, consequences, and the relationships between strategies and consequences. In selective coding, the overall findings were condensed into six themes: T1 procedural readiness / initial conditions, T2 achievement phenomenon (contextual numeracy competence in systems of linear equations in two variables and statistics), T3 task contexts (scientific sea-temperature statistics and personal transportation–voucher decisions), T4 remaining metacognitive and procedural limitations, T5 solution strategies (modeling, calculation, and verification), and T6 alignment of model, context, and conclusion. These themes were synthesized into two core categories, KI1 (numeracy problem-solving processes) and KI2 (variation in numeracy performance), which were used to characterize post-intervention numeracy response patterns among pre-service mathematics teachers.

The coding results revealed noticeable changes in response patterns compared with the pre-test. Solution procedures involving mean calculation, range determination, and modelling using systems of linear equations in two variables were more frequently presented sequentially and explicitly. Nevertheless, some difficulties remained. Challenges were particularly evident in items requiring integrative reasoning and contextual argumentation, especially items 1.d, 2.c, and 2.d. These remaining difficulties indicated that more systematic modelling and calculation did not necessarily ensure complete contextual arguments across items.

These changes were also evident in the characteristics of participant responses. Some participants were able to reject claims based on mean or range calculations, construct transportation–voucher models systematically, and compare alternative decisions using the available data. However, some participants still demonstrated inconsistencies between calculation results and written explanations, errors in identifying appropriate comparison objects, or conclusions that were not fully supported by argumentation. These findings indicate that improvements were observed primarily in the sequencing and organization of solution processes, whereas the ability to construct complete contextual arguments continued to vary among participants.

To obtain a more comprehensive understanding of the relationships among initial conditions, solution strategies, performance variation, and numeracy response quality following the intervention, the results of open coding and axial coding were further synthesized through selective coding into two core categories (KI1–KI2), as visualized in Figure 2. KI1 integrates T1, T4, and T5 and captures increasingly structured modelling–calculation–verification sequences following the intervention. KI2 integrates T2, T3, and T6 and characterizes continuing variation in numeracy performance across items and contexts, including incomplete contextual arguments. Figure 2 also represents the temporal problem-solving process from understanding the context through modelling, calculation, verification, and drawing a contextual conclusion or argument to the resulting performance outcome. In the scientific context, stronger sequencing of mean and range procedures was observed, while integrative contextual argumentation remained difficult, especially on item 1.d. In the personal transportation–voucher context, modelling using systems of linear equations in two variables became more systematic, while full scenario comparison and argumentation remained uneven, especially on items 2.c and 2.d.

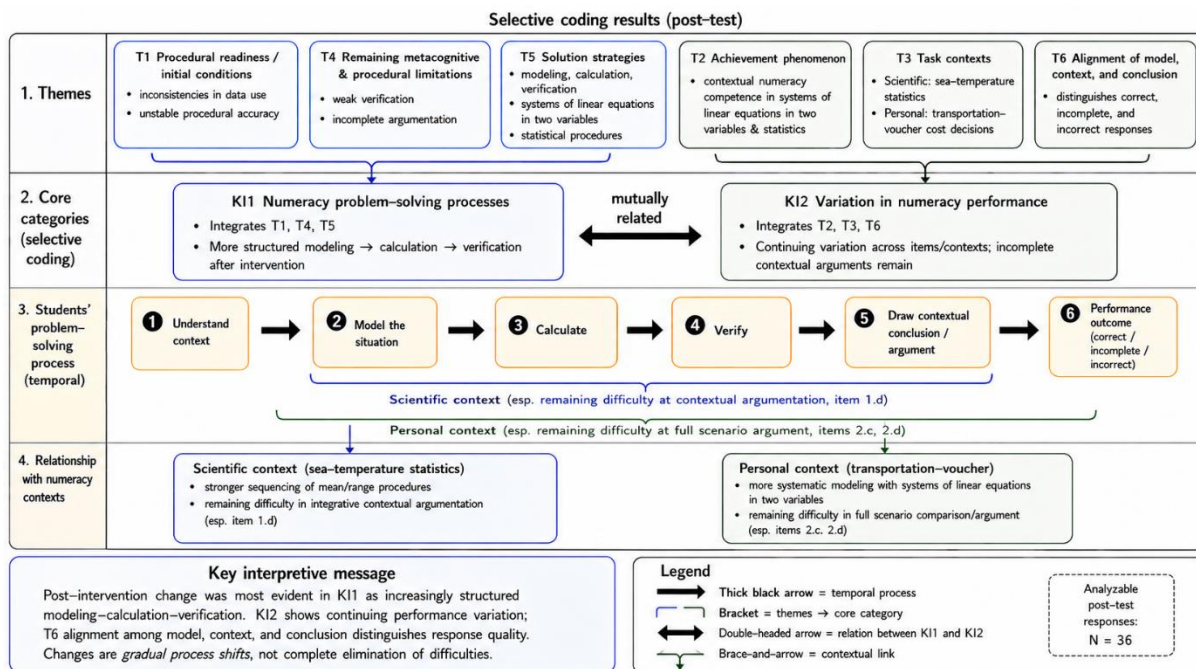


Figure 2. Integrative Representation of Selective Coding Results from the Post-test Contextual Numeracy Assessment

Interpretation of the integrative representation in Figure 2 further suggests that post-intervention change was characterized by an asymmetry between process organization and performance quality. Whereas K11 indicates that modelling, calculation, and verification became more systematically structured, K12 shows that such procedural gains did not automatically produce uniformly complete or correct responses across items and contexts. In this sense, alignment among model, context, and conclusion (T6) emerged as a key interpretive dimension: responses could exhibit improved procedures yet remain incomplete or incorrect when contextual conclusions were not coherently supported. This pattern indicates that post-test numeracy performance should be understood not only in terms of more structured problem-solving processes but also in terms of continuing variation in how participants integrated quantitative results with contextual demands. This variation was subsequently examined through paired pre-test and post-test responses on parallel items.

Pattern Shifts on Parallel Items (Pre-test to Post-test)

The group-level changes illustrated in Figure 2 were further examined through comparisons of individual responses presented in Table 4. The table provides five paired case examples from parallel items with equivalent cognitive demands but different contexts across the pre-test and post-test. In addition to excerpts from responses in both testing phases, the table includes participants' reflections collected after the pre-test. This arrangement makes it possible to examine more clearly the relationship between awareness of difficulty and subsequent changes in performance.

Table 4. Development of Patterns of Numeracy Difficulties and Individual Responses from Pre-test to Post-test (Parallel Items)

Student	Pre-test Response and Reflection Questionnaire	Post-test Response	Description of Change
AMD	Pre-test Response: "I chose 28°C for the Banda Sea and 30°C for the Java Sea because I viewed those numbers as a general picture of the available temperature data. I did	$\frac{29.1+28.9+29.5+29.0+30.1+29.3+29.4}{7} = 29.33^\circ\text{C}$ $\frac{28.7+28.5+29.1+28.9+29.3+28.6+28.8}{7} =$	Shifted from visual estimation without systematic mean calculation to complete and

Student	Pre-test Response and Reflection Questionnaire	Post-test Response	Description of Change
	not calculate the mean of all the data but only selected values that I thought represented each sea based on a brief observation." (M-2a.1-01). Reflection Questionnaire: "My mistake was not only in the final result but mainly in the incomplete process and lack of verification." (C-04); "In the future, I need to follow each problem-solving step more systematically so that my answers are more sequential and accurate." (C-02); "Writing the mathematical model first before calculating." (A-08).	28.84°C. Difference = $29.33 - 28.84 = 0.49^\circ\text{C}$." (OC-PT-1a.1.2).	systematic mean calculation. The improvement intention expressed in the pre-test reflection was reflected on parallel item 1.a.1.
DSW	Pre-test Response: "Wrong." (K-1b-01). Reflection Questionnaire: "I tend to give an answer directly because I feel I already understand the general intent of the item and want to reach the conclusion quickly." (A-01); "I want to try starting by writing the mathematical model first, such as forming a system of linear equations in two variables before calculating." (A-08); "I will also get used to performing final verification so that the conclusion is truly supported by calculation." (A-07).	"Question: Determine whether the following statement is true or false. Let x represent the number of additional days and y represent the total cost. $y - 10,000x = 120,000$ $y - 20,000x = 200,000$ Substitution: $y = 120,000 + 10,000x, \dots, x = 7, \dots$" (OC-PT-2a.1.4).	Progressed from a one-word response without evidence of process to sequential algebraic reasoning. The participant's intentions related to modelling and verification expressed in the reflection questionnaire were reflected in the post-test response.
LRT	Pre-test Response: "Median calculates the average temperature that occurs each day. Meanwhile, mode is to find the dominant temperature without knowing or affecting the data. Both are important to know the measure of temperature or changes in sea temperature." (K-2b-03). Reflection Questionnaire: "I am confused about the concept of median; sometimes I think it is the mean. I am not yet used to ordering	"Mean temperature of Flores Sea = $\frac{28.9+29.0+29.1+29.3+29.4+29.5+30.1}{7} = 29.32^\circ\text{C}$. Mean temperature of Sumba Strait = $\frac{28.5+28.6+28.7+28.8+28.9+29.1+29.3}{7} = 28.80^\circ\text{C}$. Median of Flores Sea = 29.3°C and Median of Sumba Strait = 28.8°C." (OC-PT-1b.2).	Demonstrated conceptual improvement in distinguishing between mean and median. On parallel item 1.b, the participant correctly calculated both measures and differentiated their functions.

Student	Pre-test Response and Reflection Questionnaire	Post-test Response	Description of Change
	data first before taking the middle value, so I misinterpret it." (M-2b-02); "I think I will try to read more carefully, write the mathematical model more clearly, calculate completely, and verify the result." (A-09).		
NRA	Pre-test Response: "Raka's savings if he is sick for 5 days within 30 days amount to Rp2,125,000 and are still Rp475,000 short. An alternative is to increase savings from pocket money." (J-1c-01). Reflection Questionnaire: "I have not checked in detail whether that strategy is truly sufficient because I did not substitute the values into the model completely." (M-1c-01, Part B Item 1.c); "I want to write the mathematical model first, calculate with more complete steps, and verify the result again before concluding the answer." (A-08); "To make it easier to solve similar items in the future, I need to get used to writing the model first clearly, reading the item more carefully, and always taking time to recheck calculation results before concluding the answer." (C-02).	"Disruption cost for 4 days is calculated as $4 \times \text{Rp}30,000 = \text{Rp}120,000$. For Days 1–4, the Gojek package cost is Rp120,000 and the Grab package cost is $\text{Rp}60,000 \times 2 = \text{Rp}120,000$. Remaining budget for Days 1–4, so that the total cost does not exceed Rp220,000, is $\text{Rp}220,000 - \text{Rp}120,000 = \text{Rp}100,000$, equivalent to Rp25,000 per day. Based on this calculation, total cost over 8 days is not below the Rp220,000 budget." (OC-PT-2c.3).	Progressed from savings-shortfall calculations without model verification to disruption-cost analysis and partial package-scenario comparison. However, the post-test response remained incomplete because the participant did not fully integrate the package conditions across the entire eight-day period. Reflections M-1c-01 and C-02 appear to have been reflected in the more systematic analysis observed in the post-test response.
BLW	Pre-test Response: "Because with the graph the extreme values are already visible. I remember it; it is just that from the graph the answer is already visible." (M-2a.2-01). Reflection Questionnaire: "This is where the role of verification is very important in providing the final answer to the item." (A-06); "Modelling is possible, but I still doubt this." (A-08).	"Daily temperature range in Flores Sea is larger than in Sumba Strait, so temperature fluctuation in Flores Sea is higher. Range = $X_{\max} - X_{\min} = 30.1 - 28.9 = 1.2$. For Sumba Strait, Range = $29.3 - 28.5 = 0.8$. Verification." (OC-PT-1a.2.2).	Shifted from graph-based interpretation to explicit calculation of range and verification of results. The reflection on item 2.a.2 was consistent with the more explicit calculation and verification

Student	Pre-test Response and Reflection Questionnaire	Post-test Response	Description of Change
			observed in post- test item 1.a.2.

Note. Codes in parentheses refer to pre-test open-coding categories (J/K/M, A/C) and post-test codes (OC-PT). Reflection questionnaires were administered after the numeracy pre-test, whereas evidence of subsequent change was identified retrospectively through responses to parallel items on the post-test.

Interpretation of Table 4 indicates that, in several paired cases, improvement intentions expressed in participants' reflection questionnaires were subsequently reflected in their post-test responses on parallel items. The case of AMD illustrates a shift from visual estimation to complete and systematic mean calculation. DSW progressed from brief responses lacking procedural evidence to more structured algebraic reasoning. LRT demonstrated conceptual improvement in distinguishing between mean and median. NRA showed development from unverified calculations to model-based analysis and scenario comparison, whereas BLW moved from graph-based interpretation to explicit use of the range formula accompanied by verification of results.

Nevertheless, these developments were not observed uniformly across all participants or across all items. Improvements were most evident in tasks requiring modelling, calculation, and procedural verification, whereas items demanding higher levels of contextual argumentation continued to exhibit variation in response quality. These findings reinforce the results presented in Figure 2, indicating that post-intervention changes did not occur as a uniform transformation but rather as gradual shifts in the quality of problem-solving processes.

To illustrate the relationship between patterns of numeracy difficulties identified in the pre-test and response patterns observed in the post-test, the synthesis of findings from both phases is presented in Figure 3. The figure organizes the developmental pathway into five linked components: initial numeracy difficulties, participants' reflections after the pre-test, the implementation of DIFUNDES-PROBLEMING, post-test response patterns and achievement, and overall developmental synthesis. Across these components, the pattern shows a shift from predominantly incomplete problem-solving processes and Basic-level achievement toward more structured modelling, calculation, and verification, accompanied by statistically significant improvement in numeracy achievement. Nevertheless, difficulties in integrative contextual reasoning and evidence-based argumentation remained evident, particularly on items 1.d, 2.c, and 2.d. Thus, Figure 3 represents numeracy development as a gradual process involving reflection, instructional support, and increasingly organized problem-solving processes rather than the complete elimination of difficulties.

Interpretation of Figure 3 further suggests that numeracy development in this study occurred through the interaction of difficulty awareness, instructional support, and increasingly accountable problem-solving practices. Improvements were most evident in the organization of modelling, calculation, and verification, whereas integrative contextual reasoning remained the least stable dimension of performance. This pattern aligns with the paired cases in Table 4, in which pre-test reflection intentions were reflected in post-test procedural improvements but did not uniformly extend to complete contextual argumentation. Taken together, Figure 3 integrates the quantitative evidence of achievement change presented in Tables 1–3 with the qualitative patterns of difficulty, reflection, and problem-solving development represented in Figures 1–2 and Table 4.

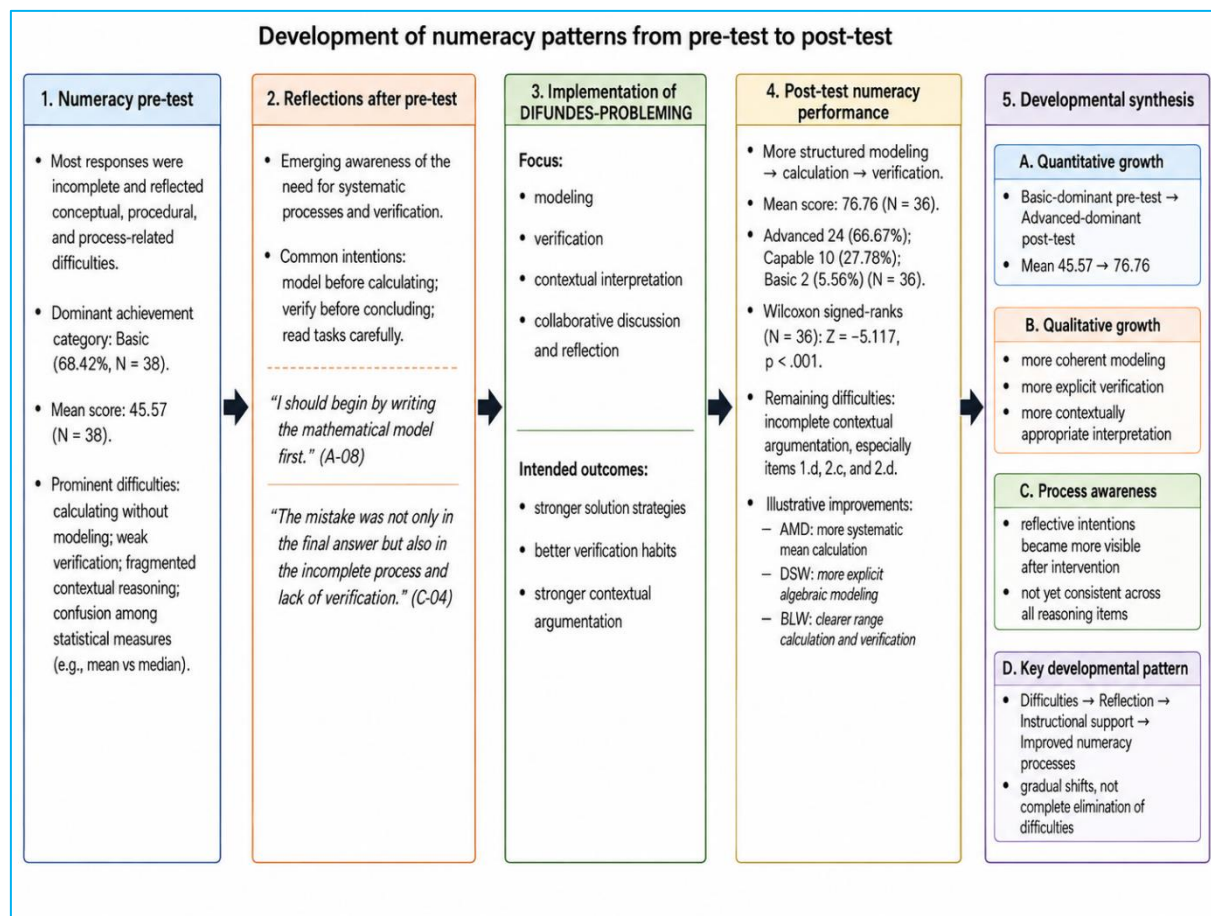


Figure 3. Development of Numeracy Patterns from Pre-test to Post-test

DISCUSSION

The findings indicate that the implementation of DIFUNDES-PROBLEMING contributed to changes in numeracy achievement and, more importantly, to shifts in the problem-solving processes used by pre-service mathematics teachers. The statistically significant improvement identified through Wilcoxon signed-rank tests (Tables 1–3) was accompanied by qualitative changes in how participants understood contextual problems, constructed mathematical models, carried out calculations, verified results, and concluded (Figures 1–3; Table 4). The main contribution of this study therefore lies not merely in showing that scores improved, but in demonstrating that numeracy development involved a gradual reorganization of problem-solving processes. This process-oriented interpretation is important because numeracy among pre-service mathematics teachers should not be reduced to procedural accuracy alone; it also involves the capacity to interpret quantitative information meaningfully in authentic situations (Geiger & Schmid, 2024).

At the same time, the findings show that numeracy development was uneven across components of problem solving. Improvement was more evident in modelling, calculation, and procedural verification than in integrative contextual reasoning and evidence-based argumentation. Post-test performance was not distributed evenly across all items: responses were generally more complete on tasks emphasizing modelling and calculation, whereas items 1.d, 2.c, and 2.d, which required learners to coordinate multiple quantitative results with contextual constraints, remained the most persistent challenges, as already reported in the Results section. This pattern suggests that participants became more systematic in executing solution procedures, while their ability to produce complete contextual arguments remained less consistent. Procedural readiness appeared to improve more readily than the capacity to integrate calculations, contextual information, and evidence-based justification in a fully coherent manner (Geiger et al., 2015; Geiger & Schmid, 2024).

The persistence of these difficulties was also evident in individual post-test responses. On item 1.d, CKD attempted to reconsider the effects of a temperature decline on several statistical measures but produced inconsistent recalculations and did not fully explain why the measures were affected differently (OC-PT-1d.3). On item 2.c, NRA progressed to disruption-cost analysis but did not fully integrate the package conditions across the entire eight-day period (OC-PT-2c.3). Similarly, on item 2.d, RNG compared voucher and package alternatives but did not completely coordinate voucher use, package costs, daily costs, and the overall budget constraints (OC-PT-2d.3). These examples reinforce the finding that participants increasingly organized their calculations and models systematically, while integrating multiple pieces of contextual information into evidence-based conclusions remained more difficult.

This uneven pattern was also reflected in comparisons between pre-test and post-test responses. Pre-test difficulties were dominated by incomplete modelling, weak verification, and limited contextual reasoning (Figure 1). Following the intervention, post-test responses showed increasingly structured sequences of modelling, calculation, and verification (Figure 2), and several paired cases in Table 4 indicated that improvement intentions expressed in reflection questionnaires were subsequently reflected in post-test responses on parallel items. Nevertheless, verification and contextual argumentation remained variable across participants and items (Figure 3). These findings indicate that post-intervention changes did not occur as a uniform transformation but rather as gradual shifts in the quality of problem-solving processes.

This pattern is especially significant because the participants were pre-service mathematics teachers. As future teachers, they are expected not only to solve contextual numeracy problems themselves but also to facilitate learning experiences in which school students interpret contexts, construct models, verify results, and justify conclusions. Previous research has similarly emphasized that pre-service mathematics teachers' understanding of mathematical problem solving and its teaching constitutes an important component of their professional preparation (Jiang et al., 2022). From this perspective, the persistence of difficulty in contextual reasoning extends beyond performance in a university course and may have implications for how these future teachers later facilitate numeracy-oriented instruction in schools (Forgasz & Hall, 2019; Getenet, 2022; O'Sullivan, 2022).

The observed shifts may be interpreted in relation to the instructional mechanisms embedded in DIFUNDES-PROBLEMING. Problem- and project-based activities provided structured opportunities for participants to engage with contextual mathematical tasks. Project work in mathematics has been shown to support multiple dimensions of mathematical proficiency (Viro & Joutsenlahti, 2020), while evidence from sustained problem-based mathematics learning indicates improvements in problem-solving performance, reasoning, and the use of multiple representations (Alashwal & Barham, 2025). Differentiated Instruction may have helped accommodate variation in participants' readiness (Tomlinson, 2017). The coherence emphasized through Understanding by Design, including opportunities to rethink, revise, and evaluate solutions, appears to have strengthened attention to evidence, verification, and alignment between learning goals, assessment, and explanation (Wiggins & McTighe, 2005). Within the Deep Learning approach in education, these integrated components were expected to situate mathematical concepts in authentic decision-making contexts and support meaningful transfer of understanding (Anggraena et al., 2025). However, because contextual reasoning requires the simultaneous coordination of models, calculations, contextual information, and justification, improvement in this area appeared less consistent than improvement in more procedural aspects of performance.

This interpretation is consistent with previous research showing that mathematical modelling requires learners to coordinate multiple processes rather than merely execute mathematical procedures. Guerrero-ortiz & Ferri (2022) identified challenges among pre-service mathematics teachers in translating between representations and using mathematical knowledge meaningfully when engaging with modelling tasks. Mathematical modelling activities have also been positioned as a means of engaging pre-service mathematics teachers with complex and authentic interdisciplinary situations as part of their professional preparation (Wiegand & Borromeo Ferri, 2023). Zuo et al. (2024) further showed that developing pre-service teachers' competencies for noticing and supporting mathematical modelling requires sustained, context-specific learning opportunities. Similarly, Wirth & Greefrath

(2024) emphasized the need to coordinate multiple stages of the modelling process and reported that transferring an approach to a related modelling problem may remain challenging even when learners are supported by a structured solution process. The present study extends these perspectives by showing that, among pre-service mathematics teachers, improvements in modelling and calculation may become evident while contextual evaluation and evidence-based argumentation remain less consistently developed.

These findings carry several practical implications. For mathematics education lecturers, instruction should move beyond emphasizing correct final answers and explicitly scaffold modelling, calculation, verification, and contextual explanation through differentiated tasks, structured reflection, and systematic opportunities to recheck solutions. For curriculum developers, numeracy curricula and assessments should sequence learning experiences from relatively accessible modelling and verification toward tasks requiring learners to coordinate multiple pieces of quantitative and contextual information, evaluate alternative solutions, and construct evidence-based arguments. For developers of numeracy learning programs, interventions such as DIFUNDES-PROBLEMING should provide repeated opportunities for learners to justify decisions, evaluate the reasonableness of results, compare alternative solutions, and revise incomplete arguments rather than focusing primarily on achievement-score improvement. These experiences are particularly important for pre-service mathematics teachers, who are expected eventually to facilitate similar numeracy processes in school classrooms.

Nevertheless, this study should be interpreted within the scope of the article. The analysis focused on changes in numeracy difficulties and problem-solving processes during the retrospective phase of a broader DDR project and did not directly examine participants' subsequent ability to design numeracy instruction. The persistence of difficulty on items requiring integrative contextual reasoning suggests that exposure to contextual tasks alone may not be sufficient to ensure fully developed argumentation; sustained opportunities for reflection, explanation, and evidence-based justification may be required. Therefore, the implications proposed here remain interpretive rather than demonstrative, and further research is needed to investigate how improvements in pre-service teachers' own numeracy processes may translate into their capacity to design, implement, and reflect on contextual numeracy learning for students.

CONCLUSION

The implementation of DIFUNDES-PROBLEMING contributed to changes in both numeracy achievement and numeracy problem-solving processes among pre-service mathematics teachers. The statistically significant improvement identified through the Wilcoxon signed-ranks test was accompanied by qualitative shifts in how participants understood contextual problems, constructed mathematical models, carried out calculations, verified results, and concluded. Compared with the pre-test, participants demonstrated more structured approaches to solving contextual problems through more explicit modelling, more systematic calculation, and more consistent verification of results. However, development was not equally evident across all dimensions of numeracy problem solving. Modelling, calculation, and procedural verification showed clearer progress, whereas integrative contextual reasoning and evidence-based argumentation remained the most challenging aspects, particularly on items 1.d, 2.c, and 2.d, which required learners to coordinate multiple quantitative results with contextual constraints. These findings indicate that numeracy development occurred through the gradual reorganization of interconnected problem-solving processes rather than through simultaneous growth across all dimensions. Because the participants were pre-service mathematics teachers, the persistence of difficulty in contextual reasoning is pedagogically significant and may have implications for the numeracy-oriented instruction they may later facilitate in schools.

Further research is needed to examine these changes over longer instructional periods and across more diverse participants and mathematical topics, with particular attention to integrative contextual reasoning and evidence-based argumentation. Longitudinal studies and research examining the contributions of different components of DIFUNDES-PROBLEMING may further clarify how numeracy problem-solving processes develop over time and how improvements in pre-service teachers'

own numeracy processes may translate into their capacity to design, implement, and reflect on contextual numeracy learning for students.

ACKNOWLEDGMENTS

The authors would like to thank the 38 students from the Department of Mathematics Education, Universitas Jambi, who participated in this study. The authors also acknowledge the academic support provided by the Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Jambi, and the valuable contributions of the validators involved in the development and validation of the research instruments.

DECLARATIONS

- Author : **Sri Winarni**: Conceptualization, Investigation, Data Curation, Formal Analysis, Contribution Visualization, Writing – Original Draft. **Kamid**: Conceptualization, Methodology, Supervision, Writing – Review & Editing. **Jefri Marzal**: Methodology, Formal Analysis, Writing – Review & Editing. **Asrial**: Validation, Writing – Review & Editing.
- Funding Statement : This research received no external funding.
- Conflict of Interest : The authors declare no conflict of interest.
- AI Use Statement : ChatGPT and Cursor AI were used in a limited and supervised manner to assist with the organization of qualitative data, preliminary coding, initial categorization, and the preparation of analytical summaries. These tools served solely as research-support instruments and were not used to make scientific, methodological, interpretive, or authorship decisions. All essay-test scoring, code verification, category refinement, data interpretation, and conclusion development were conducted and independently validated by the authors through repeated comparison with the original data, research records, and study context. The authors take full responsibility for the accuracy, integrity, and interpretation of all reported findings.
- Additional Information : -

REFERENCES

- Alashwal, H. A., & Barham, A. I. (2025). Sustaining problem-based learning: A mixed-methods exploration of its long-term effects on primary students' mathematical problem solving. *Social Sciences and Humanities Open*, 12. <https://doi.org/10.1016/j.ssaho.2025.101717>
- Anggraena, Y., Ginanto, D. E., Kesuma, A. T., & Setiyowati, D. (2025). Pembelajaran dan Asesmen Edisi Revisi Tahun 2025. *Kementerian Pendidikan Dasar Dan Menengah Republik Indonesia*.
- Aprilianti, W., Ardellea, F., & Hamdu, G. (2022). Pentingnya Kemampuan Guru Sekolah Dasar dalam Mengembangkan Soal Tes Literasi dan Numerasi Berbasis Education for Sustainable Development (ESD). *Edu Cendikia: Jurnal Ilmiah Kependidikan*, 2(02), 220–227. <https://doi.org/10.47709/educendikia.v2i02.1587>
- Bolstad, O. H. (2020). *Operasionalisasi literasi matematika guru sekolah menengah*. 8(3), 115–135.
- Botha, H., & van Putten, S. (2018). How Mathematical Literacy Teachers Facilitate Mathematisation in Modelling Situations. *African Journal of Research in Mathematics, Science and Technology*

- Education, 22(1), 93–102. <https://doi.org/10.1080/18117295.2018.1437337>
- Brousseau, G. (1997). *Theory of Didactical Situations in Mathematics*. In Kluwer Academic Publishers. <https://doi.org/10.1007/0-306-47211-2>
- Corbin, J., & Strauss, A. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. SAGE Publications.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.)*. SAGE Publications, Inc.
- Dole, S., & Geiger, V. (2018). *Numeracy Across the Curriculum: Research-based Strategies for Enhancing Teaching and Learning*. taylorfrancis.com. <https://doi.org/10.4324/9781003116585>
- Efriani, A., Putri, R. I. I., & Hapizah. (2019). Sailing context in pisa-like mathematics problems. *Journal on Mathematics Education*, 10(2), 265–276. <https://doi.org/10.22342/jme.10.2.5245.265-276>
- Fanani, A., Hanindita, A. W., Rosidah, C. T., & Susiloningsih, W. (2022). Peningkatan Kompetensi Guru dalam Penyusunan Instrumen Soal AKM Literasi Teks SD. *Jurnal Abdi Masyarakat Indonesia*, 2(4), 1333–1338. <https://doi.org/10.54082/jamsi.422>
- Forgasz, H. J., & Hall, J. (2019). Learning about Numeracy : The Impact of a Compulsory Unit on Pre-service Teachers ' Understandings and Beliefs. *Australian Journal of Teacher Education*, 44(2). <https://doi.org/10.14221/ajte.2018v44n2.2>
- Geiger, V., Goos, M., & Forgasz, H. (2015). A Rich Interpretation of Numeracy for the 21st Century: a Survey of the State of the Field. *ZDM - International Journal on Mathematics Education*, 47(4), 531–548. <https://doi.org/10.1007/s11858-015-0708-1>
- Geiger, V., & Schmid, M. (2024). A critical turn in numeracy education and practice. *Frontiers in Education*, 9(1363566). <https://doi.org/10.3389/educ.2024.1363566>
- Getenet, S. T. (2022). Teachers ' Knowledge Framework for Designing Numeracy Rich Tasks across Non-Mathematics Curriculum Areas. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, 10(3663–680). <https://doi.org/10.46328/ijemst.2137>
- Guerrero-ortiz, C., & Ferri, R. B. (2022). Pre-service teachers' challenges in implementing mathematical modelling: Insights into reality. *PNA*, 16(4), 309–341. <https://doi.org/10.30827/pna.v16i4.21329>
- Herizal, H., Marhami, M., Fonna, M., & Rohantizani, R. (2022). Preservice Mathematics Teachers' Reasoning in Solving Critical Thinking Problem. *Barekeng Jurnal Ilmu Matematika Dan Terapan*, 16(1), 1–6. <https://doi.org/10.30598/barekengvol16iss1pp001-006>
- Jiang, P., Zhang, Y., Jiang, Y., & Xiong, B. (2022). Preservice mathematics teachers' perceptions of mathematical problem solving and its teaching: A case from China. *Frontiers in Psychology*, 13, 1–16. <https://doi.org/10.3389/fpsyg.2022.998586>
- Kemendikbud. (2020). AKM dan Implikasinya pada Pembelajaran. *Pusmenjar Balitbang Kemendikbud*, 1–37.
- Marhami, Rohantizani, Muhammad, I., Samsidar, & Anggraini, I. (2023). *Pre-service mathematics teachers ' numeracy in Acehnese culture-based minimum competence assessment*. 9(January), 109–118.
- Marx, M., & Winter, J. (2014). *Pre-Service Teacher Learning and Practice for Mathematical Literacy*. University of the Witwatersrand, Johannesburg,.

- Nieveen, N. M., Mckenney, S., & Reeves, T. C. (2007). Educational Design Research. *Handbook of Information Technology in Education: Professional Learning and Development of Teachers*.
- Novita, N., Muliani, M., & Mellyzar, M. (2022). Pelatihan Pengembangan Soal Matematika Dan Sains Berbasis Numerasi Pada Guru Untuk Menunjang Asesmen Nasional. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, 6(1), 486. <https://doi.org/10.31764/jpmb.v6i1.7761>
- O'Sullivan, K. (2022). Investigating Pre-service Teachers' Knowledge of Numeracy and Their Ability to Teach Numeracy for Disciplinary Learning. *University of Limerick, January*. <https://doi.org/10.34961/researchrepository-ul.21591333.v1>
- OECD. (2022). *Pisa 2022 Mathematics Framework (Draft)*. OECD.
- Oktiningrum, W., Zulkardi, & Hartono, Y. (2016). Developing PISA-like mathematics task with Indonesia natural and cultural heritage as context to assess students' mathematical literacy. *Journal on Mathematics Education*, 7(1), 1–8. <https://doi.org/10.22342/jme.7.1.2812.1-8>
- Permatasari, B. D., Gunarhadi, & Riyadi. (2019). The Influence of Problem Based Learning Towards Social Science Learning Outcomes Viewed from Learning Interest. *International Journal of Evaluation and Research in Education*, 8(1), 39–46. <https://doi.org/10.11591/ijere.v8i1.15594>
- Suryadi, D. (2019). Penelitian Desain Didaktis (DDR) dan Implementasinya. In *Gapura Putra*.
- Tekin-Dede, A., & Bukova-Güzel, E. (2018). A Rubric Development Study for the Assessment of Modeling Skills. *The Mathematics Educator*, 27(2), 33–72.
- Toklu, D. A., & Hursen, C. (2021). Assessment of the Educational Needs of Preschool Teacher Candidates as Regards Teaching Practice. *Revista Romaneasca Pentru Educatie Multidimensionala*, 13(2), 356–381. <https://doi.org/10.18662/rrem/13.2/426>
- Tomlinson, C. A. (2017). *How to Differentiate Instruction in Academically Diverse Classrooms*. 3rd Ed. Alexandria, VA: ASCD.
- Vale, P., & Westaway, L. (2024). The development of pre-service teachers ' competence to teach mental calculation strategies. *South African Journal of Childhood Education*, 14(1), 1–9. <https://doi.org/10.4102/sajce.v14i1.1552>
- Viro, E., & Joutsenlahti, J. (2020). Learning Mathematics by Project Work in Secondary School. *Lumat*, 8(1), 107–132. <https://doi.org/10.31129/LUMAT.8.1.1372>
- Walsh, C., Bragg, L. A., Muir, T., & Oates, G. (2022). Unleashing Adult Learners' Numeracy Agency Through Self-Determined Online Professional Development. *The International Review of Research in Open and Distributed Learning*, 23(3), 240–258. <https://doi.org/10.19173/irrodl.v23i3.6046>
- Wiegand, S., & Borromeo Ferri, R. (2023). Promoting pre-service teachers' professionalism in steam education and education for sustainable development through mathematical modelling activities. *ZDM - Mathematics Education*, 55(7), 1269–1282. <https://doi.org/10.1007/s11858-023-01500-8>
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design*. Association for Supervision and Curriculum Development.
- Winarni, S., Kamid, K., Marzal, J., & Asrial, A. (2024). Numeracy Skills of Prospective Mathematics Teachers: An Investigation in the Context of Discounting. *AIP Conference Proceedings*, 3176(1). <https://doi.org/10.1063/5.0222947>

- Wirth, L., & Greefrath, G. (2024). Working with an instructional video on mathematical modeling : upper - secondary students ' perceived advantages and challenges. *ZDM – Mathematics Education*, 56(4), 573–587. <https://doi.org/10.1007/s11858-024-01546-2>
- Zuo, S., Liu, L., & Qi, C. (2024). Using video to develop pre-service teachers' noticing within a mathematical modelling context. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(5). <https://doi.org/10.29333/ejmste/14466>