

Exercise or Problem? An Epistemological Analysis of Primary and Secondary Teachers' Conceptions Through Task Design

Orlando García-Marimón¹, Marcos Campo-Nava², Agustin Torres-Rodriguez², Luisa Morales-Maure^{1,*}

¹ Faculty of Natural Sciences, Exact Sciences and Technology, Universidad de Panamá, Member of the National System of Researchers (SNI), Panama, Panama

² Autonomous University of the State of Hidalgo, Hidalgo, Mexico

<https://orcid.org/0000-0002-0334-6133> | <https://orcid.org/0000-0001-9112-3070> | <https://orcid.org/0000-0002-7534-3193> | <https://orcid.org/0000-0003-3905-9002>

* Corresponding author: luisa.morales@up.ac.pa

HOW TO CITE:

Orlando García-Marimón, Marcos Campo-Nava, Agustin Torres-Rodriguez, Luisa Morales-Maure (2026). Exercise or Problem? An Epistemological Analysis of Primary and Secondary Teachers' Conceptions Through Task Design. International Journal of Special Education, 41(12s), 508-524.

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 study examines how practicing Panamanian primary teachers distinguish between a routine mathematical exercise and a genuine problem when asked to justify their decision in writing. The analysis is framed through the Ontosemiotic Approach to Mathematical Knowledge and Instruction, specifically the epistemic suitability criterion and its four components: errors, ambiguities, richness of processes, and representativeness of complexity. The data set comprises 25 written responses drawn from a diagnostic activity administered within the EDEM diploma programme for in-service teachers. Teachers classified a contextualized subtraction task involving money and were required to explain whether the task should be considered an exercise or a problem. Responses were examined with a rubric scored from 0 to 3 on each epistemic component. The overall distribution reveals heterogeneous performance: 40% of participants reached the high level, 24% the medium level, 32% the low level, and only one teacher achieved the excellent level. At component level, representativeness of complexity showed the strongest performance ($M = 2.44$), whereas richness of processes was the weakest dimension ($M = 1.28$). These findings suggest that teachers frequently identify the task correctly but do not always justify their decisions with conceptually precise or sufficiently elaborated arguments. The study contributes a fine-grained account of teacher discourse about mathematical tasks and highlights the need for teacher education that connects epistemic analysis, task design, and reflective pedagogical reasoning.

Keywords: Epistemic suitability, exercise versus problem, Ontosemiotic Approach, teacher knowledge, mathematics education, pedagogical argumentation.

INTRODUCTION

The difference between exercise and problem is a crucial aspect of mathematics education that allows for the design of effective teaching strategies. Understanding this distinction goes beyond a simple classification; it involves recognizing the cognitive demands of each type of task. Some authors agree that an exercise is an activity where the student solves directly by applying an algorithm and knows in advance how to solve it; whereas a non-routine problem or a true problem is a challenge to the intellect and requires a degree of creativity to find the solution. Furthermore, it demands complex cognitive processes, which can be likened to what mathematicians develop in their common activity, contributing to the development of mathematical ways of thinking. (Mason, et al. 2010; NCTM, 2000; Pólya, 1945; Schoenfeld, 1985).

In this regard, as Jessup (1998) pointed out in the scientific context, a similar problem can be observed in the field of mathematics, where many teachers, after explaining a topic, methodically resort to solving application exercises as a means of reinforcement. However, this practice falls short of fostering a deep understanding because it tends to promote the rote memorization of knowledge and the automatic application of previously taught rules. Furthermore, Jessup (1998) identifies the existence of a common confusion between what constitutes a “problem” and what is an exercise; this confusion is evident in school mathematical discourse and practice, where both notions are often used interchangeably. This situation can lead to the development of “functional fixedness¹” in students, which limits their ability to tackle tasks that require flexibility, analysis, and creativity.

Strictly speaking, an exercise corresponds to a mechanical and repetitive activity, allowing for the direct application of previously learned processes. In contrast, a problem requires an

uncertain analysis, where knowledge is mobilized and strategies are generated to find the solution(s). In this line of thought, Mancera and Escareño (1993) pointed out the following:

Despite the recognized importance of problem-solving in the development of mathematics, its presence in classroom instruction has diminished over the past two decades. Mathematical content is often reduced to sets of rules and procedures to be memorized rather than explored through meaningful problem situations (p. 79).

However, despite this conceptual distinction, its practical application in the classroom is not always straightforward or obvious. Given this situation, this study proposes to analyze epistemologically how a group of teachers classify and justify a contextualized mathematical situation as an exercise or a problem. From the perspective of teacher education, this issue becomes especially important. Teachers do not simply deliver tasks; they interpret them, classify them, justify them, and transform them into classroom opportunities. Their discourse about what counts as a problem reveals the didactic-mathematical knowledge they mobilize. This study examines how a group of Panamanian primary teachers classified a contextualized arithmetic task and how they justified their decision. Three research questions guide the paper: (RQ1) How do teachers classify the proposed task as an exercise or as a problem? (RQ2) What is the epistemic quality of their written justifications? and (RQ3) Which dimensions of epistemic suitability reveal the greatest strengths and weaknesses?

The contribution of the study is twofold. First, it documents a teacher-knowledge issue that is pedagogically consequential but often taken for granted in school mathematics. Second, it applies epistemic rubric to teacher discourse in order to move beyond dichotomous classification and

¹ According to Duncker (1945), functional fixedness is a cognitive bias that limits an individual's ability to use an object beyond its functional purpose.

toward a richer account of conceptual precision, ambiguity, argumentative density, and representational adequacy.

THEORETICAL FRAMEWORK

The exercise–problem distinction has been examined from cognitive, didactic, and curricular perspectives. A routine exercise is typically associated with familiarity, algorithmic immediacy, and procedural consolidation. A problem, by contrast, involves uncertainty, requires strategy generation, and cannot be resolved by a straightforward application of an already rehearsed algorithm. Importantly, a task may function as an exercise for one learner and

as a problem for another, depending on prior knowledge and available heuristics (Arnaiz-Barrios et al., 2020; Morales Jaimes, 2025).

Contemporary discussions also emphasise that contextualization alone does not transform an exercise into a problem. A monetary, geometric, or everyday setting may still hide a highly routinized demand if the operation to be used is transparent from the wording. Research on integrated work with exercises and problems suggests that both are pedagogically useful, but they support different learning functions and should therefore be intentionally distinguished.

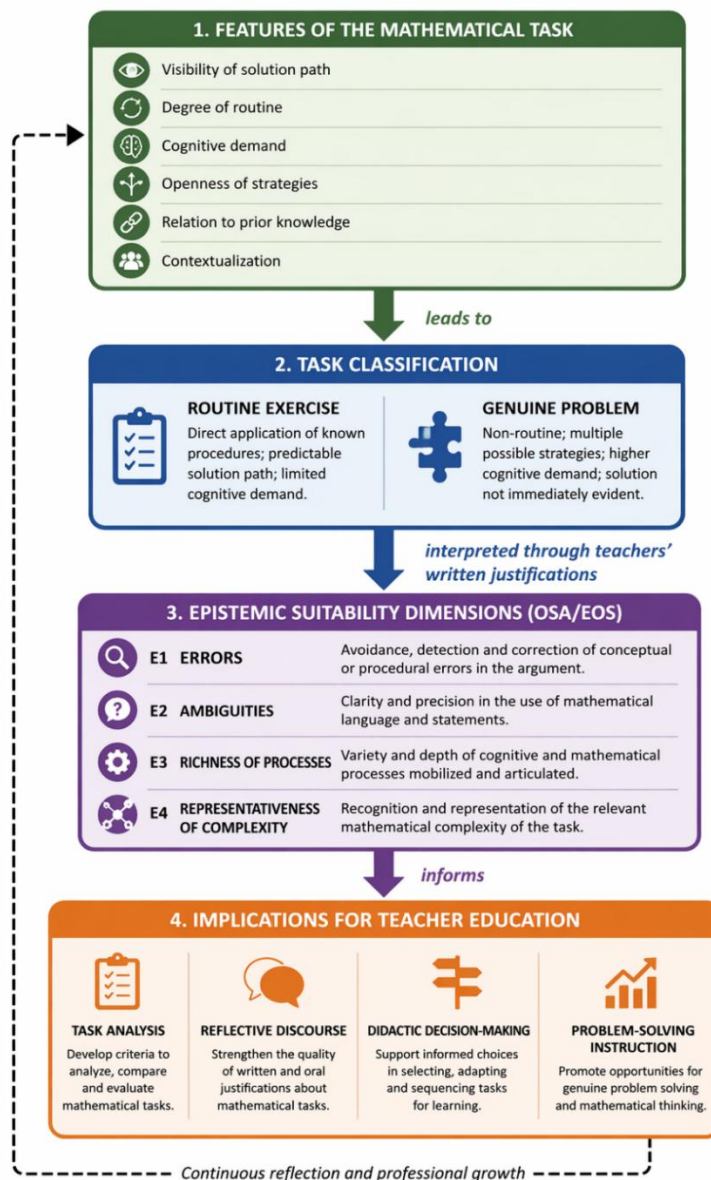


Figure 1. Theoretical articulation between task type and epistemic suitability

Exercises, problems, and cognitive demand

The distinction between exercise and problem has been addressed by multiple authors, who have highlighted the difference both in terms of cognitive complexity and in the pedagogical purposes that each type of task fulfills in the school environment.

Several didactic approaches have proposed classifications to distinguish types of mathematical tasks according to their level of cognitive demand and the function they perform in mathematical learning (Stein et al., 2009; Stein & Smith, 2018). From the Ontosemiotic Approach to Mathematical Knowledge and Instruction (OSA), the distinction is not only related to the level of difficulty, but also to the analysis of the mathematical objects involved, the processes at play, and the type of practice promoted in the classroom. In this framework, exercises tend to focus on the direct application of previously established procedures, which are geared toward reinforcement, such as the automation of techniques. In contrast, problems are cognitively demanding tasks that require interpretation, decision-making, and the generation of non-obvious solutions, so their analysis requires the teacher to have a deep understanding of the didactic variables involved (Godino et al., 2007).

Krulik and Rudnik (1980) point out that a problem is a situation, quantitative or otherwise, that requires a solution, for which the individuals involved do not know the obvious means or ways to obtain it. Perales Palacios (1993) indicates that a problem is any anticipated or spontaneous situation that produces, on the one hand, a certain degree of uncertainty and, on the other, behavior aimed at finding a solution. Along the same lines, Míguez (2003) argues that a problem, whether real or fictitious, should represent a challenge to the student's conceptual understanding and not be limited solely to knowledge of content previously covered in class. From a cognitive perspective, Schoenfeld (1985) emphasizes that the nature of a problem lies not only in the task itself, but also in the relationship between the person solving it and their available

knowledge; a task may be a problem for one student and an exercise for another, depending on their experience and understanding. Villalobos Fuentes (2008) confirms this difference when he points out that solving an exercise involves applying a routine procedure, while solving a problem requires reflection and even the execution of original steps not anticipated in advance in routine activities. In the case of Mazzilli et al. (2016), the authors argue that exercises are activities primarily oriented toward systematic repetition aimed at consolidating knowledge and skills, whereas problems require the active application of knowledge to arrive at a solution that is not immediately apparent.

This comparison has been highlighted by Jessup (1998), although in the context of science teaching, he warns of a constant tendency among teachers to use application exercises immediately after theoretical explanations, with the aim of reinforcing content. However, he points out that this often leads to a mechanization of knowledge rather than true understanding. Jessup (1998) also identifies a conceptual confusion between exercises and problems, which is reflected in the way they are presented in textbooks and in everyday teaching. This confusion generates a “functional fixation” in students that limits their ability to deal with situations that require creativity or non-routine reasoning. Extrapolating this reflection to the field of mathematics, it is possible to recognize that many of the activities presented as “problems” are disguised exercises, which hinders the development of genuine problem-solving skills.

In this regard, it is also worth mentioning that Mancera and Escareño (1993) point out that people (not only teachers) have different conceptions of what a problem is. Some consider exercises to be application problems, while others consider problems to be riddles. In general, there are a variety of conceptions, which is why they suggest trying to define what a problem is. In terms of first identifying what is not a problem. Table 1 summarizes six fundamental differences between exercises and mathematical problems from an epistemological perspective. This

comparison allows us to observe how aspects such as cognitive demand, task structure, pedagogical purpose, and relationship to knowledge vary, providing a useful tool for reflecting on the design and selection of classroom activities.

Epistemic suitability in the Ontosemiotic Approach

To analyse teachers' written justifications, this paper adopts the Ontosemiotic Approach to Mathematical Knowledge and Instruction (OSA). Within this framework, didactical suitability refers to the degree to which a teaching-learning process can be considered appropriate according to several interrelated criteria. Among them, epistemic suitability is particularly relevant here because it concerns the quality, coherence, and representativeness of the mathematical meanings activated in instructional practice.

Following previous OSA-based work, epistemic suitability was operationalised through four dimensions: errors, ambiguities, richness of processes, and representativeness of complexity. Errors refer to mathematically or didactically incorrect ideas; ambiguities concern imprecise or confusing formulations; richness of processes addresses the extent to which a response mobilises explicit reasoning and justification; and representativeness of complexity refers to whether the explanation captures multiple relevant features of the object under discussion rather than reducing it to a simplistic trait.

Table 1. Analytical distinctions between routine exercises and mathematical problems

Dimension	Routine exercise	Problem
Nature of the task	Structured and familiar	Non-routine or strategically open
Solution path	Known in advance	Not fully evident at the outset

² Stein et al. (2009) define the "Cognitive Demand" of a learning task as the level of complexity of thought

Dominant demand	Execution and reinforcement	Analysis, exploration, and decision-making
Relationship with prior knowledge	Direct application of rehearsed procedures	Reorganization and coordination of knowledge
Expected reasoning	Procedural efficiency	Interpretive and strategic reasoning
Pedagogical function	Practice and consolidation	Inquiry, modelling, and meaning-making

Didactical Suitability Criteria and the epistemic approach.

Distinguishing between exercise and a problem not only has practical implications for teaching mathematics but also allows for analysis from theoretical frameworks that enable the quality of didactic-mathematical knowledge present in these classifications to be assessed. In this regard, the OSA, developed by Godino et al. (2007), offers special conceptual tools for such analysis.

Table 2: Exercise vs. Problem

	Exercise	Problem
Type of task	Structured, routine	New, non-routine
Problem-solving procedure	Familiar and direct	Initially unknown; requires exploration
Cognitive demand ²	Low or medium; requires technical execution	High; requires analysis, reflection, and decision-making

required to perform a mathematical task, classifying it into four levels.

	Exercise	Problem
Relationship to knowledge	Mechanical application of concepts	Active construction of knowledge ³
Context ⁴	Artificial or decontextualized	Realistic or meaningful
Expected result	Unique, predictable	Possibly multiple or with diverse strategies ⁵

Source: own elaboration

Within the OSA are the Didactical Suitability Criteria (DSC) as a key construct. The DSC are a set of dimensions that assess the quality of mathematics teaching-learning processes. The dimensions they include are epistemic, cognitive, affective, mediational, ecological, and interactional, and these allow for a comprehensive analysis of educational praxis. Each dimension represents a series of requirements that must be met for a didactic practice to be considered suitable. In this study, we have chosen to focus the analysis on epistemic suitability, due to its relevance for examining the mathematical knowledge that teachers mobilize when classifying a task as an exercise or problem, as well as the way in which this knowledge is conceptualized, argued, and communicated. This perspective is consistent with the approach proposed by Godino et al. (2006), who highlights the role of representativeness analysis as a key element in assessing the adequacy of the mathematical knowledge involved in tasks.

Delving deeper into the epistemic suitability, and in accordance with the proposal by Breda et al. (2017), it is evaluated since four specific components (see Table 2):

Table 3: Epistemic suitability and its components

Criteria	Components
Epistemic	• Errors • Ambiguities • Richness of Processes • Representativeness of Complexity

Source: Breda et al. (2017)

- Errors (E1): These are understood as incorrect conceptions of the mathematical object under study. These errors reflect a lack of validity from a formal mathematical perspective, and their identification allows for the evaluation of the accuracy of the knowledge activated by the participants. In this sense, research such as that of Mellone et al. (2020) on teachers' interpretive knowledge of errors and non-standard reasoning highlights how the detection and analysis of errors reveal fundamental epistemological aspects of teachers' understanding.
- Ambiguities (E2): These refer to unclear or confusing expressions or formulations that can give rise to multiple interpretations. When ambiguities are detected, conceptual and linguistic clarity in the construction of definitions such as didactic classifications is assessed.
- Richness of processes (E3): This focuses on the degree of argumentation involved in the construction or justification of a given definition. Defining whether an activity is a problem or an exercise involves a complex argumentative process, where the subject mobilizes ideas, examples, and counterexamples. In this sense, Kublikowski (2009)

³ Mazur (1997) indicates that the concept of "active learning" is a process in which students cease to be passive recipients of information and become cognitively involved in constructing their own knowledge.

⁴ Barrera and Santos (2000) define that the problems used in mathematical learning tasks can arise from at least three

contexts: purely mathematical, hypothetical, or real-world.

⁵ Becker & Shimada (1997) indicate that a problem that has multiple correct answers or multiple methods for obtaining a correct answer allows the diversity of mathematical thinking to be revealed.

indicates that the development of a definition involves a process of rigorous argumentation, which enriches the understanding of the mathematical object studied.

- Representation of complexity (E4): This component assesses the extent to which the ideas proposed by participants capture the complexity of the mathematical object under review. An ideal classification should not be a reduction of excessive simplifications, but rather a reflection of multiple properties or aspects of the concept, allowing for a deep and nuanced understanding. Likewise, D'Amore and Radford (2017) indicate the importance of simultaneously considering epistemological and semiotic dimensions in mathematics education, highlighting how recognizing complexity in discourse contributes to meaningful and reflective teaching.

With this proposed structure, it is possible to accurately assess the mathematical knowledge activated by teachers, as well as their ability to communicate, justify, and relate it to other mathematical concepts. By analyzing these components, it is possible to assess whether didactic decisions regarding exercises and

problems are based on solid and articulated mathematical knowledge.

METHODOLOGY

Using the EOS framework, this work focuses on the epistemic suitability for assessing the quality of mathematical knowledge mobilized by elementary school teachers when classifying a task as an exercise or problem. Specifically, it considers the four components proposed by Breda et al. (2017): errors, ambiguities, richness of processes, and representativeness of complexity.

To operate these components in the analysis, the epistemological evaluation matrix developed by García Marimón et al. (2021) is adopted, which allows for the systematic evaluation of the mathematical reasoning indicated by teachers. This matrix translates the theoretical components into observable and quantifiable dimensions, facilitating their application to the analysis of qualitative data.

Epistemological evaluation matrix

The matrix considers four dimensions (E1 to E4), each corresponding to one of the components of epistemic suitability, and establishes four levels of performance, assigning a score from 0 (lowest level) to 3 (highest level) in each component. The maximum possible score per response is 12 points, as shown in Table 3.

Table 4: Evaluation of Components of Epistemic Suitability

Component	Level 0	Level 1	Level 2	Level 3
<i>E1. Errors</i>	All ideas are incorrect	Some ideas are incorrect	Most ideas are correct	All ideas are correct
<i>E2. Ambiguities</i>	The response is entirely ambiguous	Most of the wording is unclear	Some wording causes confusion	No relevant ambiguity is observed
<i>E3. Richness of processes</i>	No argumentation	Minimal or weak argumentation	Moderately developed argumentation	Rigorous and well-structured argumentation
<i>E4. Representativeness of complexity</i>	No relevant features are represented	Few relevant mathematical elements are used	The task is represented in a broadly adequate way	The explanation captures diverse and relevant features of the task

Source: García Marimón et al. (2021)

This tool allows key aspects of mathematical knowledge used by teachers to justify their didactic decisions to be identified and quantified.

Methodological design

This study adopts a qualitative approach supported by quantitative analysis, aimed at assessing the epistemological reasoning of practicing teachers when faced with a simple mathematical task.

The activity analyzed is part of Module 1 of the EDEM⁶, Diploma Course, which is developed in Panama and aimed at practicing teachers in basic education. From a population of 150 participating teachers, an intentional sample of 25 teachers was selected, chosen based on criteria of diversity in training and professional experience. The task administered was diagnostic in nature and was specifically designed to examine teachers' initial epistemological conceptions regarding the distinction between a routine exercise and a mathematical problem. Its purpose was to elicit participants' underlying understandings about the nature of mathematical tasks and the cognitive demands associated with each, thereby providing insight into how teachers conceptualize and differentiate procedural practice from genuine problem-solving situations

Situation presented: Each teacher was given the following task:

"You are going to buy groceries for \$/.100 and you spend \$/.67.15. How much money do you have left? Is this an exercise or a problem?"

Participants were asked to classify the task as an exercise or a problem and justify their choice in writing. These justifications served as the basis for the analysis.

Coding of responses: The responses were coded and evaluated according to the four components in Table 3, assigning each component a score from 0 to 3. Based on the averages obtained, the

general epistemological level of each teacher was classified and is shown in Table 5.

Table 5: Epistemic Level.

Levels	Average range	Brief description
Excellent	2.5 – 3.0	Clear reasoning, without errors or ambiguities
High	2.0 – 2.49	Good level of understanding, correct classifications
Medium	1.5 – 1.99	Correct classification with simple or ambiguous reasoning
Low	< 1.5	Frequent confusion, poor or incorrect reasoning.

Source: own elaboration

Table 6 summarizes the distribution of teachers according to epistemic suitability levels. Of the 25 participants, the majority are at the "High" level (40%), followed by the "Low" level (32%) and "Medium" level (24%). Only one teacher (4%) reached the "Excellent" level. This distribution reflects a heterogeneous epistemological understanding, where, although solid performance is observed in a significant group, significant difficulties in conceptual argumentation and clarity of ideas persist.

Table 6: Overall evaluation of teacher responses

Epistemic level	Number of teachers	Percentage	Main characteristics
Excellent	1	4%	Clear classification and strong justification
High	10	40%	Generally correct and sufficiently coherent reasoning
Medium	6	24%	Acceptable classification with partial or

⁶ The EDEM Diploma is part of the project "Competencies and Knowledge of Primary and Secondary School Teachers for Teaching Mathematics," which was approved by the National Secretariat for Science, Technology, and

Innovation (SENACYT) and has been taught in a hybrid format to primary school teachers in various regions of Panama.

Epistemic level	Number of teachers	Percentage	Main characteristics
Low	8	32%	superficial justification Frequent conceptual imprecision or weak explanation

Source: own elaboration

To complement the above results, Figure 1 was created, which provides a graphical representation of the epistemic levels achieved by teachers. This facilitates the identification of general patterns in the sample, complementing the detailed information presented in Table 5 and providing an intuitive perspective on the data.

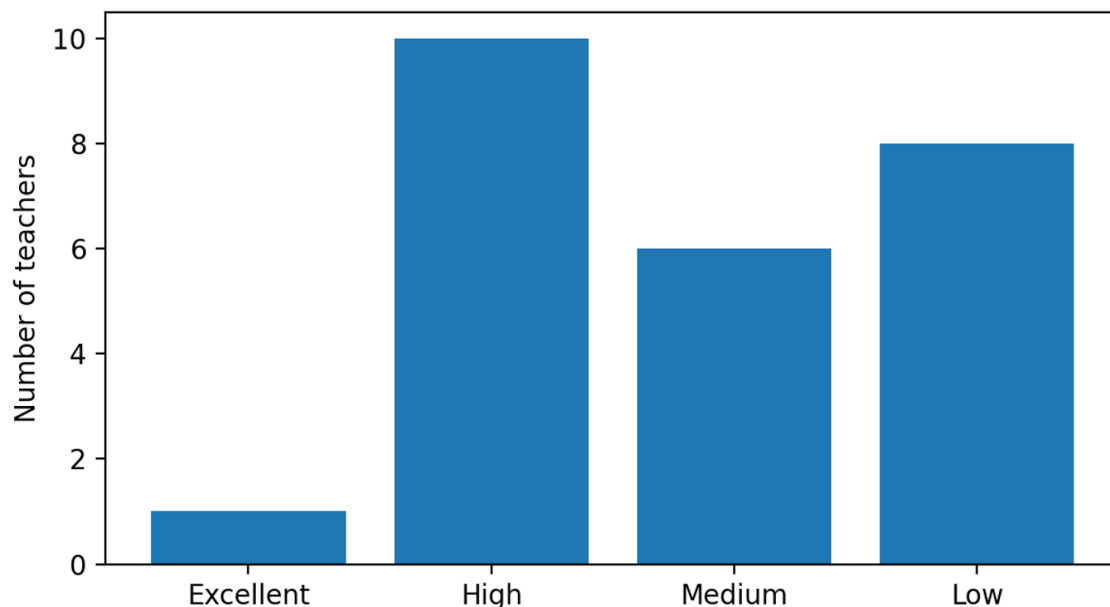


Figure 2: Distribution of overall epistemic levels among participating teachers.

Source: own elaboration

Below are three representative responses from different levels (see Tables 6, 7, and 8) given by teachers in the study. Qualitative analysis identifies notable differences in conceptual accuracy, such as clarity of language and soundness of argumentation.

A component-level analysis reveals a more differentiated pattern. The strongest dimension was representativeness of complexity (E4, $M = 2.44$), suggesting that many teachers were able, at least in broad terms, to recognize relevant features of the task as a routine activity. However, this comparatively stronger performance did not automatically translate into precise or elaborate argumentation. Richness of processes (E3) was clearly the weakest component ($M = 1.28$), followed by ambiguities (E2, $M = 1.92$) and errors (E1, $M = 2.04$). In practice, many teachers appeared able to sense what the correct label should be, but not to explain why with sufficient clarity and depth.

Individual evaluation of some responses given by teachers.

Teacher's response #18: *"In my opinion, this is a problem. It requires analysis and reading comprehension."*

Table 7: Evaluation of teacher #18's response

Component	Qualitative evaluation	Score
E1	Incorrect classification. This exercise does not require complex analysis	1
E2	Vague expression. Does not specify what analysis or reading comprehension is required	1
E3	Poor argumentation, without clear support	0
E4	Does not correctly represent the nature of the task	1

Source: own elaboration

The evaluation of teacher #18's response shows a total of 3 points and an average of 0.75, indicating a low epistemic level. This reflects that the response presents unclear ideas, does not

specify the type of analysis well, and does not adequately justify their opinion on the task.

Teacher #13's response: *"I consider it an exercise; we can solve it immediately."*

Table 8: Evaluation of teacher #13's response

Component	Qualitative evaluation	Score
E1	Correct classification without conceptual errors	3
E2	Brief expression; "immediately" somewhat ambiguous	2
E3	Limited argumentation does not go into depth	1
E4	Adequately represents a routine exercise	3

Source: own elaboration

The evaluation shows a total of 9 points and an average of 2.25, indicating an average epistemic level. This means that the answer shows acceptable understanding and some arguments but could still be improved in terms of clarity and depth.

Teacher's response #7: *"It's an exercise. It can be solved through a sequence of steps, and prior knowledge of these steps is required..."*

Table 9: Evaluation of teacher #7's response

Component	Qualitative evaluation	Score
E1	Correct classification without errors	3
E2	Clear language, without ambiguities	3
E3	Solid and well-founded argumentation	3
E4	Clearly represents the nature of the exercise	3

Source: own elaboration

The evaluation indicates a total of 12 points and an average of 3.0, reflecting an excellent epistemic level. This means that the response is clear, well-argued, and demonstrates a good understanding of the subject. To further evaluate the epistemic levels achieved by teachers, Figure 2 breaks down the distribution of their responses according to

four specific component (E1, E2, E3, and E4). This visualization shows in detail how performance levels—from "Low" to "Excellent"—are distributed in each dimension evaluated, which facilitates the identification of strengths and areas for improvement in teachers' analysis and argumentation.

Table 10. Distribution by epistemic component

Component	Excellent	High	Medium	Low	Average
E1. Errors	0	11	5	9	2.04
E2. Ambiguities	0	7	11	7	1.92
E3. Richness of processes	0	1	10	14	1.28
E4. Representativeness	1	17	1	6	2.44

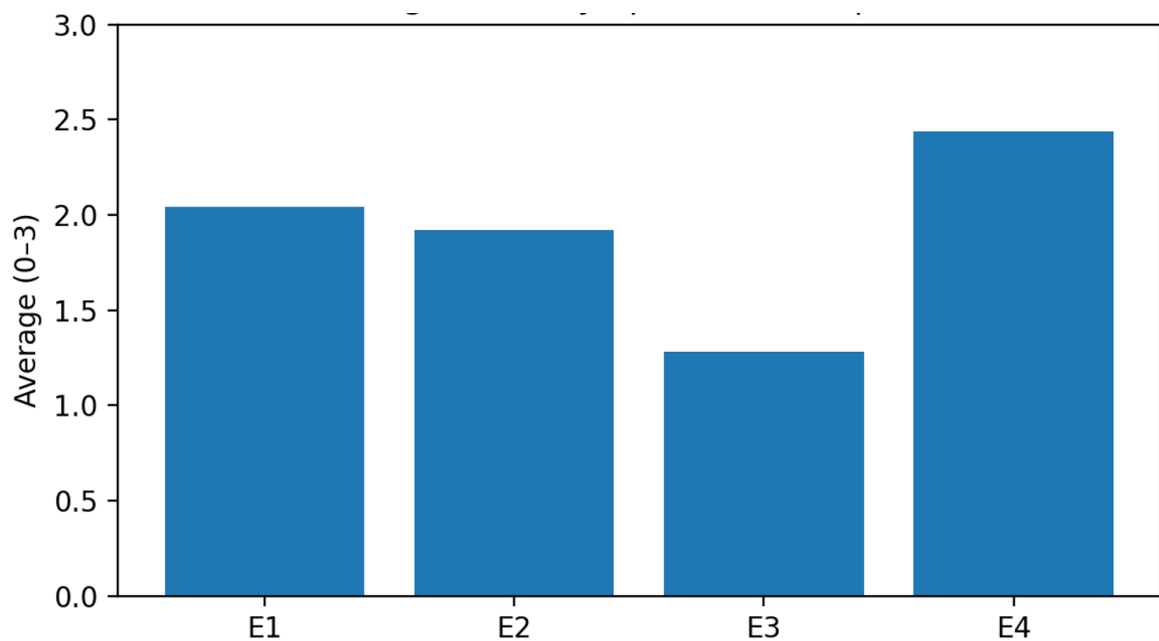


Figure 3. Average score by epistemic component.

A component-level analysis reveals a more differentiated pattern. The strongest dimension was representativeness of complexity (E4, $M = 2.44$), suggesting that many teachers were able, at least in broad terms, to recognize relevant features of the task as a routine activity. However, this comparatively stronger performance did not automatically translate into precise or elaborate argumentation. Richness of processes (E3) was clearly the weakest component ($M = 1.28$), followed by ambiguities (E2, $M = 1.92$) and errors (E1, $M = 2.04$). In practice, many teachers appeared able to sense what the correct label should be, but not to explain why with sufficient clarity and depth.

The distribution of E1 indicates that 44% of participants reached the high level, but 36% remained in the low category. This means that

conceptual correctness was far from uniform across the sample. In E2, no teacher reached the excellent level and 72% were concentrated in the medium or low categories, which points to a persistent problem of linguistic and conceptual imprecision. E3 shows the most critical weakness: 56% of teachers were at the low level and only one teacher reached the high category. By contrast, E4 was comparatively stronger, with 68% of teachers at the high level and one teacher at the excellent level. Taken together, these patterns suggest that the main challenge is not merely recognising whether a task is routine, but articulating a robust didactic-mathematical justification for that judgement.

Although a considerable number of teachers demonstrate adequate control with minimal conceptual errors (E1), 50% show deficiencies in

correcting these errors in their arguments. No teacher achieved an “excellent” level when analyzing ambiguities. More than 70% were found to be at medium and low levels, which shows the persistence of ambiguities in arguments, limiting the clarity and precision of epistemic school discourse.

Four percent of teachers achieved a “high” level in process richness, indicating that the rest were between the medium and low levels (96%). The high concentration in the medium and low levels revealed significant limitations for the integration and articulation of complex cognitive processes.

In the case of E4, better performance was revealed, with more than 70% of teachers ranking at the high and excellent levels. In contrast, 24% showed total failures in this dimension, indicating areas for improvement.

DISCUSSION

The findings indicate that correct classification and epistemic quality are not equivalent constructs. Although a substantial proportion of the participating teachers correctly identified the proposed task as an exercise, the quality of their written justifications revealed important weaknesses in conceptual precision, argumentative clarity, and epistemic depth. This pattern is particularly visible in the overall distribution of epistemic levels: 40% of participants were located at the high level, 24% at the medium level, 32% at the low level, and only 4% at the excellent level. These results suggest that recognizing a task as routine does not necessarily imply that teachers possess sufficiently elaborate criteria to explain why it should be considered an exercise rather than a genuine mathematical problem. From an ontosemiotic perspective, this distinction is highly relevant, since the quality of teachers’ discourse about mathematical tasks affects not only classification itself, but also subsequent instructional decisions about task selection, implementation, and adaptation (Breda et al., 2017; García-Marimón et al., 2021).

A particularly significant result is the weak performance in richness of processes (E3), where

96% of participants were concentrated at medium or low levels. This finding suggests that many teachers classify tasks without making explicit reference to central analytical criteria such as procedural familiarity, visibility of the solution path, degree of strategic uncertainty, or cognitive demand. In this sense, the results are consistent with recent literature showing that effective teaching for mathematical problem solving requires more than procedural competence; it demands professional knowledge capable of sustaining task complexity and eliciting students’ mathematical thinking in analytically grounded ways (Lit et al., 2026; Parrish & Bryd, 2022). Likewise, recent reviews on teacher knowledge for mathematical problem-solving stress that maintaining or productively adjusting cognitive demand is one of the most challenging dimensions of instructional practice, especially when teachers tend to reduce tasks to solvable routines during explanation or classroom mediation (Lit et al., 2026; Parrish & Bryd, 2022). The present study extends this discussion by showing that such reduction may begin even earlier, at the level of teachers’ own epistemic discourse about what counts as an exercise or a problem.

The results obtained for ambiguities (E2) reinforce this interpretation. More than 70% of the teachers were located at medium or low levels in this component, which points to persistent difficulties in communicating mathematically and didactically precise judgments. This is not a minor issue of wording; rather, it reveals a fragile professional metalanguage for analyzing task structure. Recent work on task design and task transformation in teacher education has shown that teachers need explicit conceptual tools to evaluate, modify, and justify mathematical tasks according to pedagogical intention and mathematical meaning (Markle & McGarvey, 2025; Montes et al., 2024). From this perspective, the ambiguity observed in the present study may help explain why many apparently contextualized word tasks continue to be treated as “problems” in school practice, even when they demand only direct application of a known procedure. In line with Arnaiz-Barrios et al. (2020), the mere

presence of a real-world context should not be confused with authentic problem solving; what matters is whether the task requires integrated mathematical activity beyond routine execution.

By contrast, the comparatively stronger performance in representativeness of complexity (E4) should be interpreted with caution. Although more than 70% of participants reached high or excellent levels in this component, 24% still showed serious shortcomings, and the overall pattern suggests that some teachers recognized isolated features of the task without being able to organize those features into a coherent epistemic explanation. In other words, part of the group seems to possess a practical intuition about routine tasks, but not yet a stable meta-didactic language for articulating that intuition. This interpretation is coherent with recent work emphasizing that mathematical tasks are multifaceted objects whose educational value depends on how teachers select, redesign, and sequence them in relation to learning goals (Markle & McGarvey, 2025). It also resonates with Morales Jaimes (2026), who argues that contextualized didactics can promote meaningful learning only when contexts are pedagogically mediated through reflective analysis rather than reduced to superficial realism. Similarly, Pérez Castillo et al. (2026) argue for problematizing didactics inspired by Pólya and Freire, in which the educational value of a task lies in its capacity to provoke inquiry, interpretation, and critical reasoning rather than merely activate a memorized algorithm.

From a theoretical standpoint, the study contributes to the literature by showing that the distinction between exercise and problem should not be examined only as a matter of taxonomic correctness, but as a matter of epistemic suitability of teachers' discourse. This contribution is important because it operationalizes that distinction through four analytically differentiated components: errors, ambiguities, richness of processes, and representativeness of complexity thus offering a more nuanced account of teachers' didactic-mathematical knowledge. In that sense, the

findings support the usefulness of the Didactical Suitability Criteria and the broader ontosemiotic framework for identifying not only whether teachers "know" a classification, but also whether they can justify it with conceptual clarity and communicative precision (Breda et al., 2017; García-Marimón et al., 2021). This is especially relevant for Q1-level discussions on teacher knowledge, because it shifts attention from declarative correctness to the quality of professional reasoning that underpins pedagogical decisions.

The practical implications for teacher education are equally significant. The results suggest that teacher preparation programs should explicitly train participants to analyze and justify mathematical tasks using criteria such as routine versus non-routine structure, openness of strategy, degree of procedural predictability, and cognitive demand. This recommendation is supported by recent reviews showing that task design, task modification, and problem posing are central dimensions of mathematics teacher education and require deliberate pedagogical attention rather than incidental exposure (Markle & McGarvey, 2025; Montes et al., 2024). In addition, within the broader post-pandemic debate on higher education quality and social responsibility, the findings also point to the need for teacher education institutions to strengthen intellectual rigor in professional discourse, not only technological adaptation or access conditions (Parrish & Bryd, 2022; Lit et al., 2026).

These findings should nevertheless be interpreted in light of several limitations. First, the study is based on an intentional sample of 25 Panamanian elementary teachers, which restricts statistical generalization and calls for caution in extrapolating the results to other educational levels or national contexts. Second, the analysis relies on a single diagnostic task and on brief written justifications, which are highly valuable for capturing initial epistemic conceptions but may not fully reflect the complexity of teachers' reasoning in interviews, lesson planning, or classroom interaction. Third, because the study is cross-sectional, it does not allow examination of

how these conceptions evolve through sustained professional development. Finally, the manuscript itself includes a discrepancy between the average ranges reported for epistemic levels in Table 4 and Table 5, so those ranges should be standardized before submission in order to strengthen internal consistency and methodological transparency (Markle & McGarvey, 2025; Montes et al., 2024).

Future research could extend this line of inquiry in several directions. First, it would be valuable to include larger and more diverse samples of pre-service and in-service teachers in order to explore whether the same epistemic patterns persist across educational stages and institutional settings. Second, future studies should incorporate multiple tasks varying in contextualization, openness, mathematical structure, and cognitive demand, so that researchers can better identify which task features most strongly influence teachers' classifications and justifications. Third, combining written responses with interviews, think-aloud protocols, lesson plans, and classroom observations would make it possible to examine the relationship between teachers' explicit discourse and their actual instructional decisions. Finally, intervention studies are needed to test whether professional development centered on task analysis, problem posing, and epistemic justification can improve teachers' ability to distinguish between exercises and problems in theoretically grounded ways. Such a research agenda would be consistent with current international calls to treat task design and problem solving as core domains of mathematics teacher education rather than peripheral methodological concerns (Lit et al., 2026; Markle & McGarvey, 2025; Montes et al., 2024).

CONCLUSIONS

This study examined how a group of Panamanian primary teachers classified a contextualized arithmetic task and how they justified that classification in writing. Regarding RQ1, the findings show that most participants identified the task as an exercise rather than as a problem, which suggests a basic awareness of the

distinction between routine procedural activity and genuine problem solving. However, this recognition was not uniform, and the presence of incorrect classifications indicates that the exercise–problem distinction is not yet fully consolidated across the group.

Concerning RQ2, the epistemic quality of the teachers' written justifications was markedly uneven. Although 40% of the participants reached a high epistemic level, 24% were located at the medium level, 32% at the low level, and only one teacher (4%) achieved an excellent level. These results confirm that correct classification alone is not sufficient evidence of robust didactic-mathematical understanding. In many cases, teachers were able to identify the nature of the task, but they were not able to justify their decision with the degree of conceptual precision, argumentative clarity, and epistemic depth expected in reflective professional discourse.

With respect to RQ3, the component-level analysis revealed a differentiated epistemic profile. The strongest dimension was representativeness of complexity (E4), where 72% of the participants reached high or excellent levels, suggesting that many teachers recognized at least some relevant features of a routine exercise. Nevertheless, this relative strength should be interpreted cautiously, since a non-negligible proportion still failed to integrate those features coherently into their pedagogical reasoning. By contrast, the greatest weaknesses were found in richness of processes (E3), where 96% of participants remained at medium or low levels, and in ambiguities (E2), where more than 70% were also concentrated in medium and low levels. In addition, around 50% showed difficulties related to conceptual errors (E1). Taken together, these results indicate that the main challenge is not merely identifying task type, but constructing clear, rigorous, and theoretically grounded justifications for that identification.

These findings have important implications for teacher education. First, they suggest that initial and continuing professional development should place greater emphasis on the explicit analysis of mathematical tasks, particularly on criteria such as

routine versus non-routine structure, procedural predictability, cognitive demand, and openness of solution strategies. Second, they point to the need to strengthen teachers' professional discourse so that pedagogical decisions are supported not by intuition alone, but by explicit and communicable epistemic criteria. Third, they highlight the importance of critically reviewing curricular materials and textbooks, since overreliance on routine exercises may limit the development of mathematical thinking and reduce opportunities for authentic problem solving.

Overall, the study shows that the participating teachers possess an incipient but uneven understanding of what distinguishes an exercise from a mathematical problem. While there is evidence of partial recognition of this distinction, substantial limitations remain in how it is justified and communicated. In this sense, the Didactical

Suitability Criteria, and particularly the dimension of epistemic suitability, proved to be a valuable theoretical and methodological lens for identifying the strengths and weaknesses of teachers' conceptions. The study therefore contributes to the field by showing that what matters is not only whether teachers classify tasks correctly, but whether they can do so through coherent, precise, and pedagogically meaningful reasoning.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the Office of the Vice President for Research and Graduate Studies (VIP) of the University of Panama through the University Call for Research Funding (CUFI 2024), under project VIP-01-90-2024-01. This support made the development and completion of this study possible.

Bibliography

- Arnaiz-Barrios, I., García-Rodríguez, J. A., & Díaz-Gómez, M. (2020). Concepción didáctica para aplicar integralmente las habilidades matemáticas en la solución de ejercicios y problemas. *Educación y Sociedad*, 18(3), 16–29.
- Barrera-Mora, F. y Santos-Trigo, M. (2001). Cualidades y procesos matemáticos importantes en la resolución de problemas: un caso hipotético de suministro de medicamentos. En Ministerio de Educación Nacional (ed.) *Memorias del Seminario Nacional de Formación de Docentes: Uso de Nuevas Tecnologías en el Aula de Matemáticas*, p.166-185. Bogotá: Ministerio de Educación. <https://surl.li/vbmuge>
- Becker, J.P. & Shimada, S. (1997). *The Open-Ended Approach: A New Proposal for Teaching Mathematics*. Reston, VA: NCTM.
- Guo, X. (2026). Construction and Empirical Validation of The "Faculty-Student-Ai-Environment-Culture" Five-Element Synergistic Model In Higher Education: Evidence From The Integration of Sdg 4 and Artificial Intelligence. *International Journal of Engineering Sciences & Research Technology*, 15(1), 16–32. Available at:<https://doi.org/10.64149/j.ijesrt.15.1.16-32>
- Breda, A., Pino-Fan, L. R., & Font, V. (2017). Meta Didactic-Mathematical Knowledge of Teachers: Criteria for The Reflection and Assessment on Teaching Practice. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 1893-1918. <https://doi.org/10.12973/eurasia.2017.01207a>
- D'Amore, B., & Radford, L. (2017). Enseñanza y aprendizaje de las matemáticas: Problemas semióticos, epistemológicos y prácticos. Universidad Distrital Francisco José de Caldas. <https://repository.udistrital.edu.co/handle/11349/5884>
- Duncker, K. (1945). On problem-solving (L.S. Lees, Trans). *Psychological Monographs*, 58(5), 11-23. <http://doi.org/10.1037/h0093599>
- García-Marimón, O., Morales-Maure, L., Díez-Palomar, J. y Durán-González, R. E. (2021). Evaluación de secuencias de aprendizaje de matemáticas usando la herramienta de los Criterios de Idoneidad Didáctica. [Using the Didactical Suitability Criteria tool when assessing mathematics learning sequences]. *Bolema* 35(70), 1047-1072. <http://dx.doi.org/10.1590/1980-4415v35n70a23>

- Godino, J. D., Contreras, A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathematiques*, 26(1), 39-88. <https://dialnet.unirioja.es/servlet/articulo?codigo=2200862>
- Godino, J.D., Batanero, C. & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM Mathematics Education* 39, 127–135 <https://doi.org/10.1007/s11858-006-0004-1>
- Jessup C., M. N. J. C (1998). Resolución de Problemas y Enseñanza de las Ciencias Naturales. *Tecné, Episteme y Didaxis: TED*, (3). <https://doi.org/10.17227/ted.num3-5701>
- Krulik, S. & Rudnick, K. (1980). *Problem solving in school mathematics*. National council of teachers of mathematics. Yearbook: Reston, Virginia. <https://eric.ed.gov/?id=ED183420>
- Kublikowski, R. (2009). Definition within the structure of argumentation. *Studies in Logic, Grammar and Rhetoric*, 16(29), 229-244. [https://pbc.biaman.pl/Content/25585/PDF/Studies16\(29\)-srodek.pdf#page=229](https://pbc.biaman.pl/Content/25585/PDF/Studies16(29)-srodek.pdf#page=229)
- Tabase, F., Kumi-Boateng, B., Yakubu, I., & Ziggah, Y.Y., (2026). A Review of Models For Predicting Customer Satisfaction. *International Journal of Engineering Sciences & Research Technology*, 15(1), 1–15. Available at: <https://doi.org/10.64149/j.ijesrt.15.1.1-15>
- Lit, S., Kool, M., & Drijvers, P. (2026). Teacher knowledge and skills for mathematical problem solving: A systematic literature review. *Journal of Mathematics Teacher Education*. Advance online publication. <https://doi:10.1007/s10857-025-09737-8>
- Mancera, E. y Escareño, F. (1993). Problemas, Maestros, y la Resolución de Problemas. *Educación Matemática*, 5(3), 78-92. <https://doi.org/10.24844/EM0503.05>
- Markle, J. M., & McGarvey, L. M. (2025). Research on task design in pre-service mathematics teacher education: A scoping review. *Frontiers in Education*, 10, 1467482. <https://doi:10.3389/educ.2025.1467482>
- Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically* (2nd ed.). Harlow: Pearson.
- Mazur, E. (1997). *Peer Instruction: A User's Manual*. New Jersey: Prentice Hall.
- Mazzilli D.M., Hernández, De La Hoz S.I. (2016). Procedimiento para Desarrollar la Competencia Matemática Resolución de Problemas [Procedure to develop the competition mathematical resolution of problems]. *Escenarios*, 14 (2), 103-119. <http://dx.doi.org/10.15665/esc.v14i2.935>
- Mellone, M., Ribeiro, M., Jakobsen, A., Carotenuto, G., Romano, P., & Pacelli, T. (2020). Mathematics teachers' interpretative knowledge of students' errors and non-standard reasoning. *Research in Mathematics Education*, 22(2), 154-167. <https://doi.org/10.1080/14794802.2019.1710557>
- Míguez, Á. (2003). Los ejemplos, ejercicios, problemas y preguntas en las actividades de aprendizaje de matemática [The examples, exercises, problems, and questions in mathematics learning activities]. *Revista Educación y Pedagogía*, (35), 141-149. <https://revistas.udea.edu.co/index.php/revistaeyp/article/view/5949>
- Montes, M., Chico, J., Martín-Díaz, J. P., & Badillo, E. (2024). Mathematics teachers' specialized knowledge mobilized through problem transformation, *The Journal of Mathematical Behavior*, 73, 101132, ISSN 0732-3123, <https://doi.org/10.1016/j.jmathb.2024.101132>.
- Morales Jaimes, A. D. (2025). Resolución de problemas en el área de matemáticas desde la perspectiva de una didáctica contextualizada [Tesis doctoral, Universidad Pedagógica Experimental Libertador].
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*.
- Parrish, C. W., & Bryd, K. O. (2022). Cognitively demanding tasks: Supporting students and teachers during engagement and implementation. *International Electronic Journal of Mathematics Education*, 17(1), em0671. <https://doi:10.29333/iejme/11475>
- Perales Palacios, F. J. (1993). La resolución de problemas: una revisión estructurada. *Enseñanza De Las Ciencias. Revista de investigación y Experiencias Didácticas*, 11(2), 170–178. <https://doi.org/10.5565/rev/ensciencias.4533>
- Pérez Castillo, J. C., de Nacimiento Capado, R. M., & Calunga Gando, Y. A. (2026). Hacia una didáctica problematizadora en la resolución de problemas matemáticos desde Pólya y Freire. *Revista Conrado*, 22(108), e5118.

-
- Pólya, G. (1945). *How to solve it: Anew aspecto of mathematical method*. Princeton: Princeton University Press.
- Said, H. J. N. (2004). *Resolución de Problemas*. Maracaibo: Talleres de Formación Matemática. vol. 1, p. 3-10
- Schoenfeld, A. H. (1985) *Mathematical Problem Solving*. Orlando, Florida: Academic Press.
- Stanic, G. & Kilpatrick, J. (1989). Historical perspectives on problem solving in the mathematics curriculum. In R. Charles Silver (Eds.) *The teaching and assessing of mathematical problem solving*, 1-22 Reston, VA: NCTM.
- Stein, M.K. & Smith, M.S. (2018). *5 Practices for Orchestrating Productive Discussions* (2nd ed.). Reston, VA: NCTM. <https://pubs.nctm.org/display/book/9780873538015/9780873538015.xml>
- Stein, M.K., Smith, M.S., Henningsen, M.A., & Silver, E.A. (2009). *Implementing standards-based mathemtaics instruction: A casebook for profesional development* (2nd ed). Teachers College Press. <https://cir.nii.ac.jp/crid/1971430859776813580>
- Villalobos Fuentes, X. (2008). Resolución de problemas matemáticos: un cambio epistemológico con resultados metodológicos. *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 6(3), 36-58. <https://www.redalyc.org/articulo.oa?id=55160303>