Research Article

Tracing teachers' engagement in counterintuitive mathematics from intuition to reasoning, calculation, and peer discussion

Arkhat Baimukhanov [1*]
 0009-0001-1782-954X

Sara D. Sony [2]
 0000-0001-5203-1958

Gulash Kochshanova [3]
 0000-0001-5716-3415

Yessengali Smagulov [1]
 0009-0001-9324-5095

Nuri Balta [4]
 0000-0002-6364-5346

Raquel Fernández Cézar [5]
 0000-0002-9013-7734

[1] Faculty of Physics and Mathematics, Zhetysu University named after I. Zhansugurov, Taldykorgan, KAZAKHSTAN
[2] School of Mathematics and Statistics, Northwest Missouri State University, Maryville, MO, USA
[3] Department of Basic Sciences, Yessenov University, Aktau, KAZAKHSTAN
[4] SDU University, Almaty, KAZAKHSTAN
[5] University of Castilla-La Mancha, Ciudad Real, SPAIN
* Corresponding author: arkhatbaimukhanov@gmail.com

ARTICLE INFO	ABSTRACT
	This study examines mathematics teachers' engagement with counterintuitive problems through a structured four-stage process: intuitive responses, reasoning, calculation, and peer discussion. Using convenience sampling method, 60 mathematics teachers from Kazakhstan's Zhetysu Region were selected for the study. Data were collected using an adapted set of four counterintuitive mathematics tasks. Results revealed that initial intuitive responses frequently led to systematic errors. Reasoning explanations were often only partially correct, indicating that intuitive responses were grounded in incomplete conceptual models. Formal calculations improved accuracy across most tasks, though the average speed problem remained especially resistant to resolution. Peer discussion proved to be the most effective intervention, leading to substantial gains in correctness, particularly for the most counterintuitive tasks. Comparative analysis across the four stages revealed distinct learning trajectories, with collaborative reasoning consistently yielding the largest improvements. Regression analysis demonstrated a moderate positive relationship between reasoning quality and solution accuracy. These findings show the pedagogical value of counterintuitive problems for improving reasoning, conceptual understanding, and collaborative learning in mathematics education. **Keywords:** counterintuitive problems, mathematics education, reasoning, problem-solving, peer discussion, teacher learning

INTRODUCTION

Problem-solving is widely recognized as a cornerstone of mathematics education, as it increases both procedural fluency and higher-order thinking skills essential for real-life application of knowledge (Santos-Trigo, 2024). While traditional mathematics problems often rely on straightforward application of formulas or algorithms, counterintuitive problems present a unique challenge by confronting learners with situations where their immediate, intuitive responses are frequently misleading (Balta & Eryılmaz, 2017; Brookman-Byrne et al., 2018). Engaging with counterintuitive problems compels teachers and students alike to move beyond rote computation toward deeper reasoning, reflection, and justification. Within broader educational priorities, this focus is significant as modern pedagogy increasingly emphasizes the cultivation of critical thinking, conceptual understanding, and collaborative learning practices. By integrating counterintuitive problems into teaching and professional development, educators can provide learners with opportunities to reconstruct conceptions while simultaneously strengthening reasoning and problem-solving skills (Shahbari, 2021).

Although mathematics is often perceived as a discipline governed by logical rules and precise calculations, both students and teachers frequently rely on intuitive judgments when solving problems. These intuitive responses, while efficient in everyday contexts, often lead to systematic errors in mathematical and science domains such as proportional reasoning, scale, and averages (Balta & Asikainen, 2019; Stavy & Tirosh, 2000). For example, learners may assume that increasing one quantity will linearly increase another, or that averaging speeds can be accomplished by a simple arithmetic mean, when in fact such reasoning produces incorrect conclusions. Research consistently shows that these misconceptions are highly resistant to change, persisting even among advanced learners unless instructional approaches explicitly confront and address them (Brookman-Byrne et al., 2018; Shahbari, 2021). Without targeted pedagogical strategies, intuitive but flawed mental models continue to dominate problem-solving processes, limiting the development of deeper conceptual understanding and robust reasoning skills.

Searching mathematics teachers' responses through the structured sequence of intuition, reasoning, calculation, and discussion provides a valuable lens for understanding how mathematical thinking develops and how counterintuitive questions are gradually solved. In this study, counterintuitive refers to situations in which an individual's immediate, intuitive judgment, typically based on a superficial approach, leads to an incorrect conclusion that conflicts with the mathematically correct solution. Such tasks require the suppression of an initial intuitive response and engagement with formal reasoning or calculation to arrive at the correct answer.

By capturing teachers' initial intuitive responses, the study reveals the nature and persistence of common errors. The subsequent reasoning stage allows researchers to examine the underlying justifications and partial conceptual frameworks that shape these intuitive responses. When participants then engage in formal calculations, it becomes possible to evaluate the extent to which procedural problem-solving can override or reinforce their prior reasoning. Finally, the peer discussion stage highlights the powerful role of social interaction, as collaborative dialogue often prompts participants to revisit their assumptions, confront inconsistencies, and converge on more accurate solutions.

This staged design thus makes it possible to trace the trajectory of conceptual change, from intuitive errors, through emerging reasoning, to refined understanding, while also identifying which types of incorrect intuitions are most resistant to change. Importantly, the inclusion of peer discussion reflects a growing body of research showing that collaborative learning is one of the most effective strategies for overcoming deeply rooted intuitive errors (Smith III et al., 1994).

The purpose of this study is to examine, in a systematic and observable way, how mathematics teachers' problem-solving performance changes when they engage with counterintuitive tasks under different instructional conditions. Specifically, the study analyzes teachers' responses across four sequential stages: initial intuitive answers, written reasoning, formal mathematical calculation, and peer discussion, to determine

(a) the types of intuitive errors teachers commonly produce,

(b) the extent to which formal reasoning and calculation reduce these errors, and

(c) how collaborative dialogue supports the resolution of counterintuitive conflicts.

In addition, the study investigates whether the quality of teachers' written reasoning statistically predicts the accuracy of their mathematical solutions. Thus, following research questions (RQs) were determined to guide our study.

1. **RQ1.** How do teachers' intuitive responses influence their initial responses to counterintuitive mathematics problems?
2. **RQ2.** What is the quality of mathematics teachers' reasoning when justifying their intuitive responses to counterintuitive problems?
3. **RQ3.** To what extent does the quality of teachers' reasoning predict the accuracy of their mathematical solutions?
4. **RQ4.** In what ways do formal mathematical calculations and peer discussion each contribute to improving teachers' performance on counterintuitive problems?
5. **RQ5.** How do teachers' performance trajectories differ across the four stages of problem-solving, intuition, reasoning, calculation, and discussion, when addressing counterintuitive tasks?

THEORETICAL FRAMEWORK

Intuition has had a somewhat negative reputation in the history of mathematics, unlike in other sciences. This is evident in philosophical traditions from Platonism (where mathematical truth is seen as existing independently of human perception) through formalism, to constructivism, which is more aligned with experiential learning (Benacerraf & Putnam, 1983; Parsons, 2018). Counterintuitive results, however, help deepen our understanding of mathematical structures and reasoning.

Counter-intuitiveness has long been contemplated by both religion and philosophy (Berlyne, 1960). It emerges from human experience and impacts not only mathematics but all natural and experimental sciences, as it is rooted in our experiential understanding of the world. In mathematics, counter-intuitiveness occurs when facts or concepts conflict with commonsense expectations (Everet & Pennathur, 2007). Mathematical intuition can be considered a pre-theoretical way of explaining the world through mathematical concepts. While experiential events have historically pushed the advancement of knowledge, intuition remains limited when facts contradict expectations.

Russell and Gobet (2013) noted that although there is no universal definition, scientific community agrees that counter intuitiveness involves the violation of prior knowledge, triggering a cognitive conflict. Nonetheless, rather than being detrimental to learning, such conflict can enhance student motivation through the element of surprise (Balta & Eryilmaz, 2017), a view now supported by brain science (Brookman-Byrne et al., 2018; Dumontheil et al., 2022; Everett & Pennathur, 2007; Wessel, 2018).

Surprise, triggered when outcomes defy expectations, plays a key role in student motivation and engagement (Pekrun, 2006). From a cognitive psychology standpoint, dissonance not only creates emotional discomfort but also activates mechanisms of resolution that support more sophisticated learning (D'Mello & Graesser, 2012). In mathematics, this dissonance becomes productive when it drives students to revise and refine their previous assumptions.

Counterintuitive Tasks as Tools for Deep Understanding

Counterintuitive tasks are powerful pedagogical tools that foster deep understanding in mathematics education by challenging learners to confront and revise their intuitive conceptions. These tasks often present scenarios in which initial, instinctive responses lead to incorrect conclusions, thereby creating productive cognitive conflict (Fischbein, 1987). When guided effectively, this dissonance prompts learners to reflect, analyze, and reconstruct their understanding in a more robust and conceptual manner. Tasks that contradict intuition, such as paradoxes or unexpected results, can foster robust mathematical understanding by compelling students to reexamine their reasoning (Kapur, 2016). This "productive struggle" allows cognitive conflict to act as a driver of learning, promoting deeper conceptual restructuring (Hiebert & Grouws, 2007). Unlike disorganized confusion, well-structured cognitive conflict arising from counterintuitive tasks leads to enhanced critical thinking and understanding (Ran, 2022).

Far from the current moment, research shows that learners tend to rely on heuristics and everyday experiences to interpret mathematical problems (Tversky & Kahneman, 1974). While such intuition can be helpful in some contexts, it frequently leads to errors in situations involving probability, geometry, or algebraic reasoning (OBrien & NíRíordáin, 2025). For instance, tasks such as the Monty Hall problem or the "missing dollar riddle" highlight how intuitive reasoning can conflict with formal logic or statistical thinking. These tasks require students to suspend automatic judgments and engage in critical analysis, which is essential for developing mathematical sophistication (Borovcnik & Kapadia, 2009).

Engagement with counterintuitive tasks facilitates the transition from surface-level procedural knowledge to deeper conceptual understanding. According to Hershkowitz and Vinner (1980), when learners are confronted with contradictions between their concept image and formal definitions, they are more likely to revise and refine their mental models. This process strengthens metacognitive skills and encourages learners to adopt a more reflective and analytical stance toward mathematical problems.

However, the effectiveness of counterintuitive tasks is closely tied to instructional context. Without appropriate scaffolding and discussion, these tasks can lead to frustration (Stylianides & Stylianides, 2007). Teachers play a critical role in helping students articulate their reasoning, confront their errors, and co-construct meaning through dialogic interactions. Structured reflection and comparison between intuitive and formal approaches are key to transforming confusion into insight.

In conclusion, counterintuitive tasks are not merely puzzles or curiosities; they are essential instruments for cultivating deep mathematical understanding. By destabilizing naïve conceptions and promoting inquiry, these tasks help learners build more coherent and transferable knowledge. When thoughtfully integrated into instruction, they can enhance both cognitive engagement and conceptual clarity (Gennari & Valentini, 2024).

Counter Intuitiveness in Mathematics Education and Neuroscience

Since the 1980s, mathematics didactics have examined this conflict. In Brousseau's (2011) theory of didactic situations, counter intuitiveness emerges in a-didactic situations, i.e., contexts not explicitly designed for direct instruction, where students interact autonomously with the environment and receive feedback. The resulting cognitive conflict or disequilibrium becomes a mechanism for meaningful learning.

Currently, research in cognitive neuroscience has shown that cognitive disequilibrium activates brain regions associated with inhibitory control and working memory (Foisy et al., 2015). Such conflict is particularly relevant to mathematics education, where resilience facing the error is essential. Emotions like curiosity, frustration, and awe also play a pivotal role in mathematical problem-solving (Daniel & DiNapoli, 2024: Immordino-Yang & Damasio, 2007). The theory of productive failure proposes that initial failure can be valuable when learners reflect on their mistakes, allowing them to build deeper understanding through conflict resolution (Kapur, 2016).

Besides that, recent neuroscience research has also investigated how counterintuitive thinking in STEM disciplines is linked to executive functions, especially inhibitory control and visuospatial reasoning. These studies indicate that when students solve counterintuitive problems, intuitive responses are suppressed and cognitive resources are activated, enhancing learning outcomes (Dumontheil et al., 2022). In a study with adolescents aged 11-15, brain magnetic resonance discovered brain regions associated with both counterintuitive reasoning and inhibitory control were activated, though the neural patterns were not identical (Brookman-Byrne et al., 2018). They further found that higher engagement of response inhibition (e.g., go/no-go tasks) and semantic inhibition (e.g., suppressing intuitive strategies) correlated with improved performance on counterintuitive tasks. Another study by Wessel (2018) used electroencephalography to explore how surprise affects visuospatial working memory, which is crucial for geometry. Surprising stimuli activated motor inhibition mechanisms but did not disrupt visuospatial working memory accuracy, suggesting distinct cognitive mechanisms from those affecting verbal working memory.

Ultimately, brain science shows that counterintuitive reasoning in mathematics and science involves overlapping but distinct neural systems related to inhibitory control and visuospatial processing. Understanding these mechanisms is essential for developing instructional strategies that promote critical thinking and problem-solving skills in mathematics (Lanting, 2025).

Distinguishing Misconceptions from Counterintuitive Reasoning

In mathematics education, it is essential to distinguish between misconceptions and counterintuitive reasoning, as each presents distinct challenges and opportunities for instruction. While both involve errors in thinking, misconceptions typically arise from incorrect or incomplete prior knowledge that is resistant to change, whereas counterintuitive reasoning reflects the tension between intuitive responses and formal mathematical logic (Fischbein, 1987; Smith III et al., 1994).

Misconceptions are often stable, deeply rooted understandings that are internally coherent but incompatible with formal mathematics. Many misconceptions in mathematics stem from overreliance on sensory-based or heuristic intuition (Clements & Battista, 1992; Clements et al., 2013), or that the sum of angles in all triangles is 180 degrees regardless of context for geometry (Euclidean vs non-Euclidean geometries), which reflects overgeneralization of rules (Vosniadou, 2013), that can be driven with an appropriate teaching approach that connects sensory-based information with abstract representation. These erroneous beliefs are usually persistent and may not change unless explicitly challenged through targeted instruction and conceptual conflict (Hamza & Wickman, 2008).

In contrast, counterintuitive reasoning emerges in situations where learners' intuitive heuristics conflict with mathematically correct conclusions. These tasks do not necessarily reflect a misunderstanding, but rather a natural reliance on experiential or perceptual reasoning (Tversky & Kahneman, 1974). For instance, students may predict that a longer object will fall faster or that a non-square rectangle cannot be a parallelogram, errors arising not from ignorance but from intuitive inferences based on prior experiences. When learners are presented with counterintuitive tasks in a supportive environment, they often revise their reasoning once given time and tools to reflect (Borovcnik & Kapadia, 2009).

Recognizing the difference between misconceptions and intuitive reasoning is crucial for instructional decision-making. Treating intuitive errors as misconceptions may lead to unnecessary remediation, while overlooking true misconceptions could allow flawed reasoning to persist. This distinction has major educational implications: misconceptions require corrective instruction, while counterintuitive phenomena offer opportunities to promote critical thinking (Chi, 2005). Teachers must therefore adopt diagnostic approaches that examine not only student answers but also the reasoning behind them. Open-ended questions, think-aloud protocols, and classroom discussions can provide insight into whether students are applying flawed concepts or grappling with the surprise of an unfamiliar, yet ultimately accessible, idea (Stylianides & Stylianides, 2007).

Ultimately, distinguishing misconceptions from counterintuitive reasoning enables educators to tailor interventions appropriately, offering clarification and restructuring for misconceptions, and cognitive support and reflection for counterintuitive conflicts. Both phenomena can serve as productive entry points for deeper learning when afforded with pedagogical insight.

Counterintuitive Tasks in Mathematics

Frederick's (2005) cognitive reflection test is one of the most widely cited demonstrations of counterintuitive reasoning. The classic "bat and ball" question illustrates how intuitive answers can override logical analysis. Recent neuroscience and developmental psychology studies emphasize that inhibitory control, the ability to suppress immediate but incorrect responses, plays a central role in solving counterintuitive tasks (Palmer et al., 2025). Similarly, research shows that students who perform better on tasks requiring the inhibition of misleading intuitions often demonstrate stronger mathematical problem-solving skills (Dumontheil et al., 2023).

Several problems in mathematics are frequently cited as counterintuitive. The Monty Hall problem demonstrates how switching doors after one option is revealed increases the probability of winning from one-third to two-thirds (Gill, 2025). The birthday Paradox shows that in a group of only 23 people, there is a greater than 50% chance that two individuals share the same birthday (Sun, 2021). Other examples include the "rope around the earth" problem, in which adding a small amount of length to a rope encircling the Earth creates a surprisingly large gap (Campanario, 1998), and Gabriel's horn, which has finite volume but infinite surface area (Kaufman, 2023). Hilbert's hotel illustrates the paradoxical properties of infinite sets, revealing how a "full" hotel with infinitely many rooms can still accommodate new guests (Chen et al., 2022).

Counterintuitive tasks have been studied as instructional tools for enhancing mathematical learning. Collection of classroom tasks (Pruner, 2016) demonstrates how such problems engage students in productive struggle, encouraging them to question assumptions and refine reasoning. Large-scale evaluations, such as the National Foundation for Educational Research (2019) study of the "stop and think" intervention, show that structured engagement with counterintuitive tasks can improve reasoning outcomes in primary education. Earlier work, such as *That can't be right!* (Maylone, 2000), also provides curricular resources that embed counterintuitive problems in classroom practice

METHOD

Research Design

This study employed a quantitative, within-subjects research design combining descriptive, comparative, and correlational analyses. Each participant responded to the same set of counterintuitive problems across four sequential stages; intuition, reasoning, calculation, and peer discussion, allowing for direct comparison of performance changes within individuals.

Sample

The sample of this study consisted of mathematics teachers from the Zhetysu Region of Kazakhstan. A total of 60 teachers were selected using convenience sampling. The first group of 22 respondents completed the survey on August 13 at Secondary School No. 25 in Taldykorgan during a professional development course on "Teaching mathematics: Current trends and methods" organized by the "Orleu" Institute for Professional Development. This group included teachers not only from Taldykorgan but also from nearby districts such as Koksu, Alakol, Karatal, Aksu, and Kerbulak.

The second group of 25 teachers completed the survey on August 20 as part of the second cohort of the same course. An additional 13 teachers participated from the Nazarbayev Intellectual School (NIS) of natural sciences and mathematics in Taldykorgan, where the survey was administered on August 27. These teachers were colleagues of the first author, but only those who had not participated in the prior piloting of the instrument were included. Teachers were teaching from 9th to 12th grades, and they have teaching experiences varying between 1 to 26 years.

The teachers varied in age and professional background. Older teachers tended to be less actively engaged in the survey tasks, focusing more on the workshop activities of the course, while younger teachers were highly active, often presenting problem solutions on the board and leading discussions after completing the survey. Randomly chosen group sizes during the administration ranged from 3 to 4 teachers.

Across all three survey administrations, participants were assured that the study aimed solely at gathering data for research purposes and was not intended to evaluate their teaching skills or knowledge. This sample therefore reflects a broad cross-section of mathematics teachers from urban and rural areas of the Zhetysu Region, representing a diversity of ages, teaching experiences, and institutional contexts.

Instrument

The instrument consisted of four counterintuitive multiple-choice problems, each designed to elicit counter intuitions in mathematics. It is a task-based performance assessment, not a questionnaire or survey. The first problem, *string around the earth*, addressed intuitions about scale effects (Campanario, 1998). The second, *potato problem*, focused on proportional reasoning when water content changes (Burton, 1978). The third, *algae doubling*, targeted counter intuition about exponential growth (Illustrative Mathematics, 2026). The fourth, *average-speed trap*, explored difficulties with averages and impossibility reasoning (GMAT Club Forum, n.d.). All items were compiled from publicly available sources known in mathematics education.

Each problem followed the same structured format, consisting of four parts. First, participants provided a "quick answer," in which they selected the option that immediately came to mind without calculation, thus capturing their intuitive response. Second, they were asked to explain the reasoning that led to this initial answer. Third, they solved the task formally using mathematical calculations to verify or revise their earlier responses. Finally, after engaging in small-group peer discussion, they recorded whether they changed their

answer and, if so, why. In addition, teachers were given the opportunity to write about their own counterintuitive mathematics problem at the end.

The questions were administered in printed form, allowing participants to record their choices, explanations, and calculations in writing. Thirty minutes were allocated for the test stages (quick answer, reasoning, and calculation), and with the addition of peer discussion, the total administration time was approximately thirty-five to forty minutes. Group discussions were conducted in small groups of two to three participants.

Responses were coded using a four-level rubric for both reasoning and mathematical calculations. For the Reasoning dimension, a score of 4 indicated an explanation that was logical, thorough, and connected intuition to mathematics; a score of 3 reflected mostly logical reasoning with some gaps; a score of 2 was assigned to vague or partially incorrect reasoning; and a score of 1 indicated little or no reasoning. For the Mathematical Calculations dimension, a score of 4 was given for complete, correct, and well-organized calculations; a score of 3 for mostly correct work with minor errors; a score of 2 for partially correct work with major steps missing or serious errors; and a score of 1 for incorrect or missing calculations. Quick answers and post-discussion answers were scored dichotomously as correct or incorrect. Inter-rater reliability was established using a two-way random-effects intraclass correlation coefficient (ICC), appropriate for rubric-based ratings with raters treated as random effects. The resulting ICC indicated satisfactory agreement (.82).

Data Collection

The questions were administered to three groups of mathematics teachers from the Zhetysu Region. The first administration involved 22 teachers from the first stream, who completed the test at the end of their professional development course on August 13. The second administration involved 25 teachers from the second stream, who completed the test at the beginning of their course on August 20. A third administration was carried out with 13 teachers from the NIS of natural sciences and mathematics in Taldykorgan on August 27. For each session, teachers were given 30 minutes to complete the test. With the addition of peer discussion at the end, the total time required was approximately 35-40 minutes. The tasks were printed out and distributed, ensuring that all participants provided written responses. Before beginning, participants were briefed on the purpose of the study and were assured that their responses would be used strictly for research, not for evaluating their teaching performance.

Data were collected during a single session using a structured, four-stage procedure. Each stage was designed to capture a distinct phase of teachers' problem-solving processes.

1. **Stage 1.** Participants first selected a solution option for each problem without performing calculations. This stage was intended to capture their immediate, intuitive judgments.

2. **Stage 2.** Participants were then asked to provide a written explanation justifying their initial choice. This stage focused on eliciting the conceptual reasoning underlying intuitive responses.

3. **Stage 3.** Next, participants solved each problem using formal mathematical procedures to verify or revise their earlier answers. All calculations were recorded in writing.

4. **Stage 4.** Finally, participants discussed their solutions in small groups of two to three teachers. Following discussion, they were allowed to revise their final answers and briefly note whether and why their responses changed.

Data Analyses

Descriptive statistics were first used to summarize teachers' performance across the four stages of the task sequence. To test for differences across stages, the Friedman test was conducted, followed by Wilcoxon signed-rank post-hoc comparisons with Bonferroni correction. The effect of peer discussion was examined using Chi-square tests to compare correct versus incorrect responses before and after discussion. Finally, a simple linear regression analysis was performed to determine whether reasoning quality significantly predicted mathematical solution accuracy. Prior to interpreting the regression results, standard assumptions of linear regression were examined. No substantial deviations from these assumptions were detected.
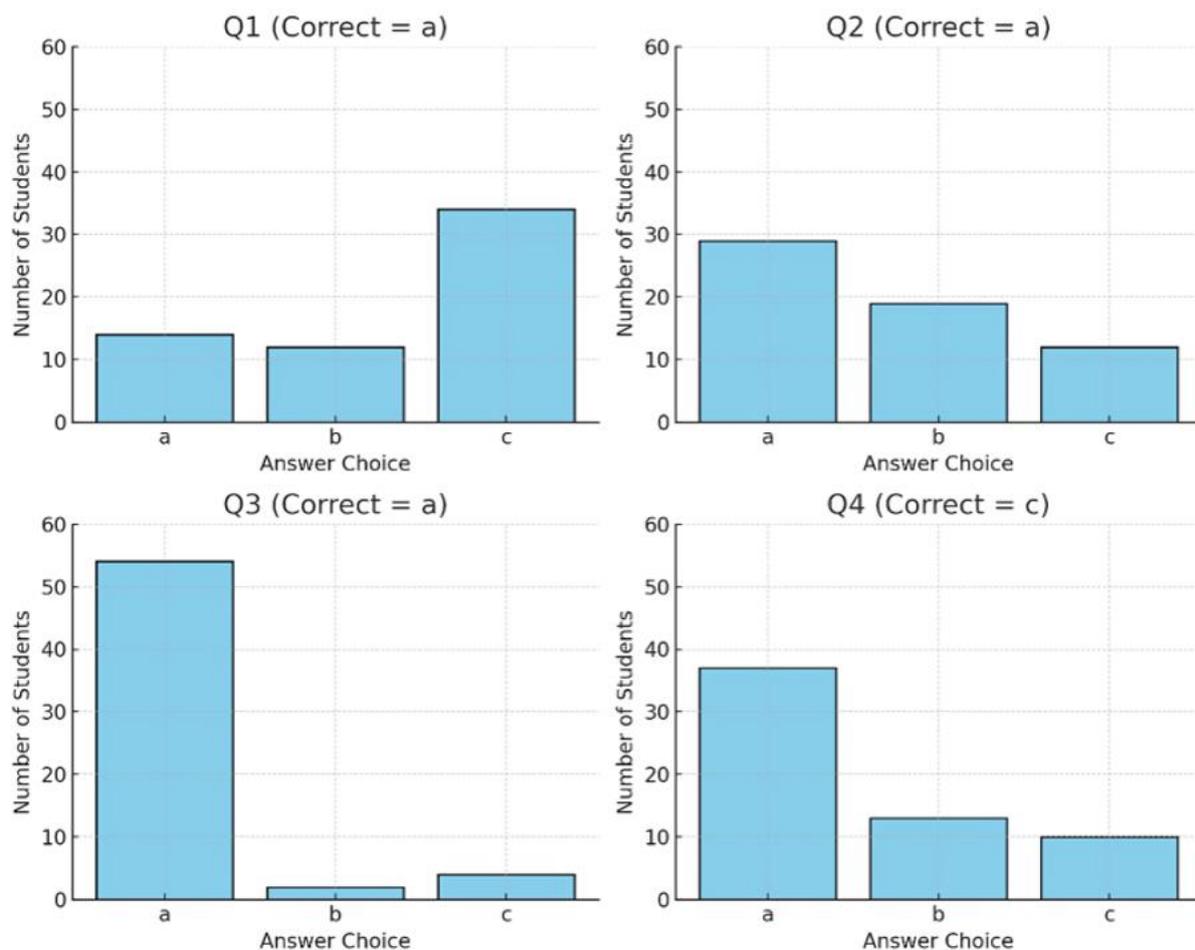
Figure 1. Distribution of initial intuitive responses to the four counterintuitive questions (Source: Authors)

RESULTS

Results are organized according to the four sequential stages of the problem-solving process rather than by RQs, because this structure reflects the study design and enables a direct examination of changes in teachers' responses across stages. Moreover, each stage directly corresponds to specific RQs.

First Stage. Intuitive Responses

Figure 1 presents the distribution of teachers' responses to the four counterintuitive mathematics questions. The bars represent teachers' response counts, showing both correct answers and the relative strength of distractors.

In the *string around the earth* problem, only 23% of teachers identified the correct solution (≈10 meters), while a majority (57%) selected the option of 1000 kilometers, reflecting a strong scale counterintuitive in which teachers overestimated the effect of raising the string due to the earth's size. The *potato problem* produced a near-even split, with 48% correct and the remainder distributed between 98 kg and 99 kg, suggesting that many teachers held proportional reasoning errors, assuming a small percentage change in water content implied a small overall weight change. The *algae growth* question proved least effective as a counterintuitive task, with 90% answering correctly that the lake would be fully covered on Day 31, indicating that most teachers recognized the exponential growth principle without difficulty. Finally, *the average speed* trap was the most challenging, with only 17% of teachers recognizing that achieving a 60 mph average was impossible, while 62% incorrectly chose 90 mph.
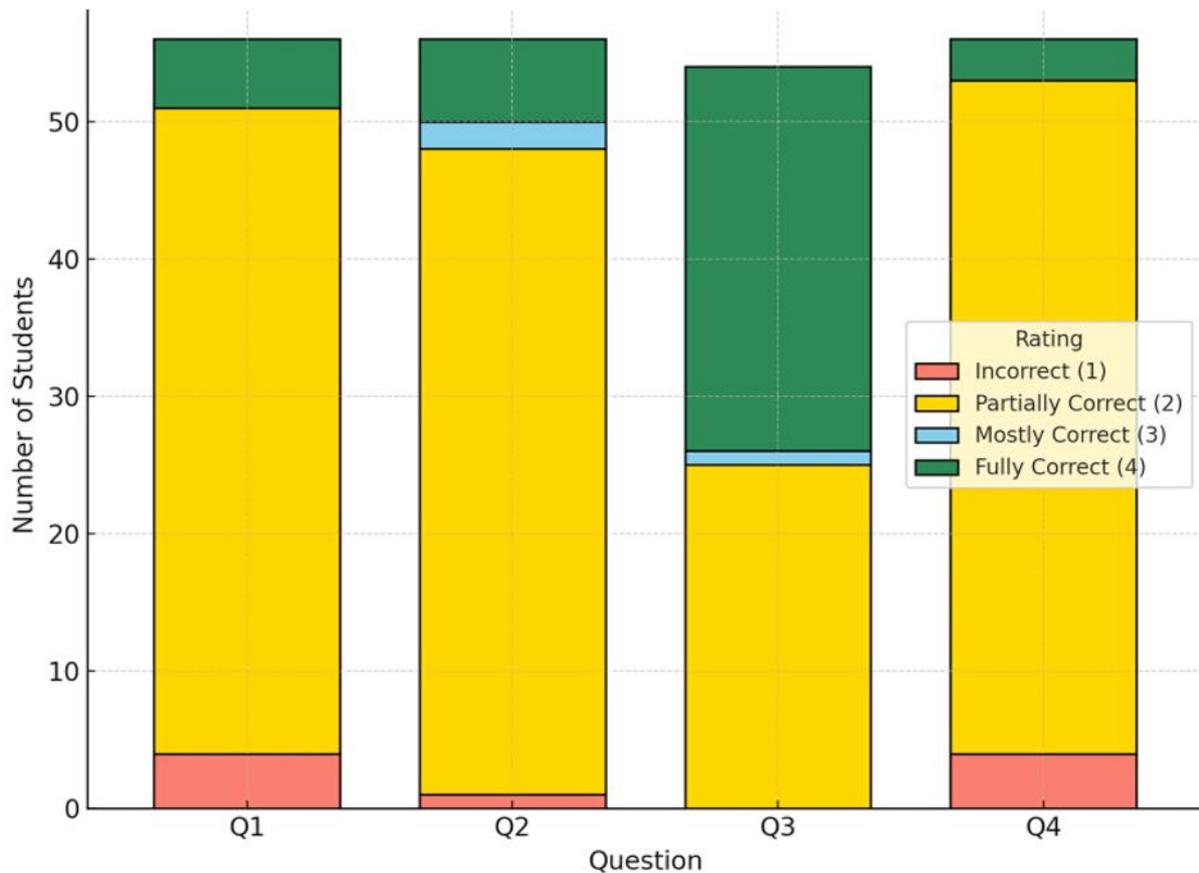
Figure 2. Distribution of teachers' reasoning quality across the four counterintuitive problems (Source: Authors)

Second Stage. Reasoning

Figure 2 presents the distribution of teachers' reasoning quality across the four counterintuitive problems. The stacked bars display the number of teachers whose solutions were rated as incorrect, partially correct, mostly correct, or fully correct.

The distribution of reasoning quality shows that most teachers' explanations were rated as only partially correct. In Q1 and Q4, the overwhelming majority of responses fell into this category, with very few fully correct explanations, indicating persistent counterintuitive cases about scale effects and average speed. Q2 produced a similar profile, though with slightly more fully developed responses, showing that teachers were often on the right track but lacked precision in proportional reasoning. In contrast, Q3 stood out as the strongest, with nearly half of the teachers providing fully correct reasoning and the remainder offering partially correct justifications.

The reasoning quality ratings align closely with the answer-choice patterns reported in **Figure 1**. For the *string around the earth* (Q1) and *average speed trap* (Q4) problems, where the majority of teachers initially selected dominant distractors, reasoning was overwhelmingly rated as only partially correct, reflecting incomplete or flawed conceptual models. These findings confirm that counterintuitive case was not random guess but were supported by partially coherent, yet ultimately incorrect, explanations. In the *potato problem* (Q2), a large proportion of teachers also demonstrated partial reasoning, paralleling the split distribution of answers across the correct and distractor options. In contrast, the *algae growth* problem (Q3) showed both high correctness in answer choice and strong reasoning quality, indicating that this problem was not as counterintuitive as intended. Taken together, the alignment between response distributions and reasoning ratings demonstrates that teachers counter intuitions were deeply embedded in their reasoning processes.
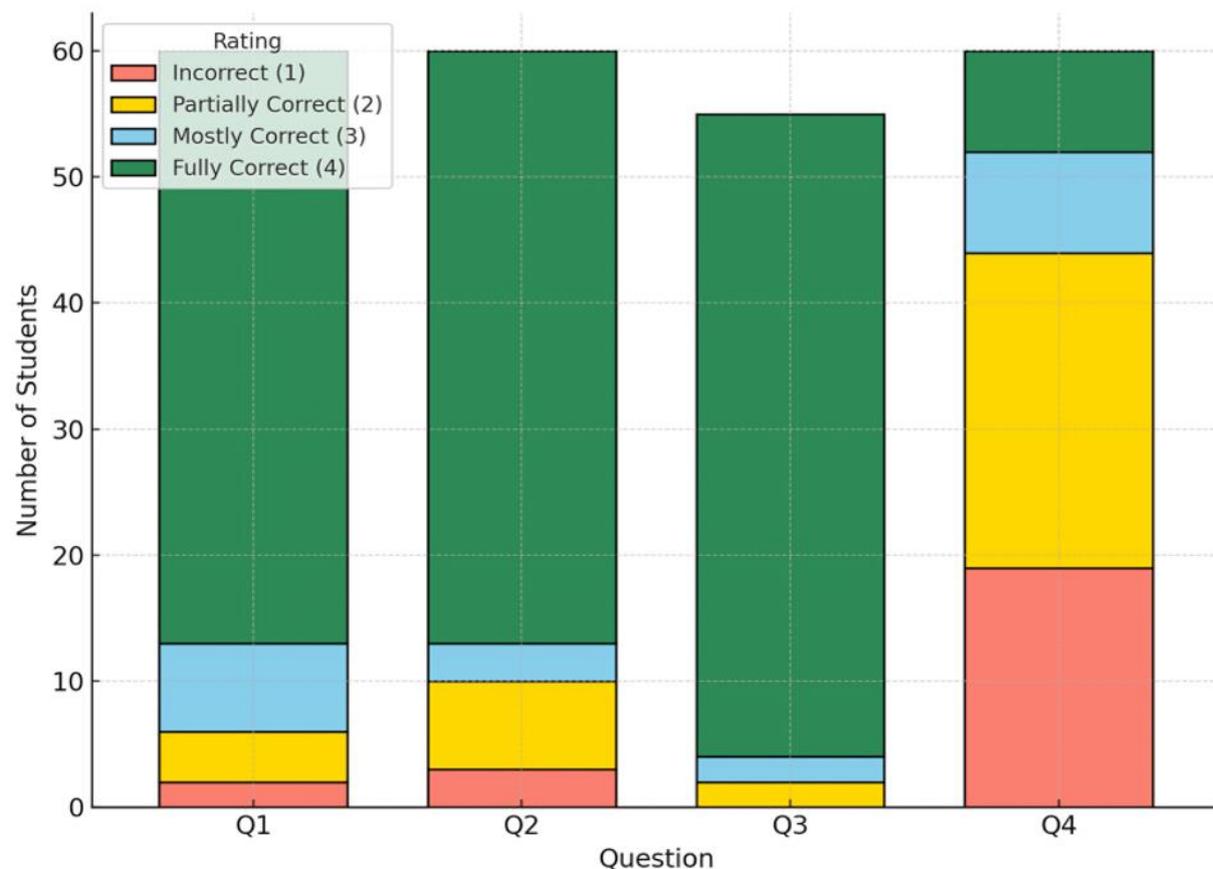
Figure 3. Distribution of teachers' mathematical solution quality across the four counterintuitive problems (Source: Authors)

Third Stage. Mathematical Solution

Figure 3 presents the distribution of teachers' mathematical solution quality across the four counterintuitive problems. The stacked bars display the number of teachers whose solutions were rated as incorrect, partially correct, mostly correct, or fully correct.

The distribution of ratings for teachers' mathematical solutions shows a marked improvement compared to their earlier reasoning. In Q1 and Q2, the majority of teachers achieved fully correct solutions, with only a small minority remaining at the partially correct or incorrect levels. Q3 displayed the strongest performance overall, with nearly all teachers providing fully correct solutions, confirming that the exponential growth task was conceptually straightforward once formal calculations were applied. By contrast, Q4 continued to pose difficulties: almost three-quarters of teachers produced either partially correct or incorrect solutions, and only a small group succeeded in solving it fully.

Fourth Stage. Peer Discussion

Figure 4 displays the distribution of teacher responses to the four counterintuitive mathematics questions following peer discussion. After peer discussion, teachers' performance improved substantially across all four counterintuitive problems. In the *string around the earth* problem (Q1), over 50 teachers identified the correct solution compared to only 14 in the initial round, indicating that discussion effectively dispelled the common scale counterintuitive. The *potato problem* (Q2) showed an even more dramatic shift, with nearly all teachers (≈58 out of 60) arriving at the correct answer after discussion, suggesting that collaborative reasoning helped clarify counterintuitive situation about proportional reasoning. The *algae growth* problem (Q3), which was already answered correctly by most teachers initially, became almost unanimous, reinforcing its limited diagnostic value as a counterintuitive item. The greatest conceptual change occurred in the *average speed problem* (Q4): while only 17% were initially correct, nearly 50 teachers selected the correct answer after discussion, although a small minority persisted with the intuitive but incorrect choice.
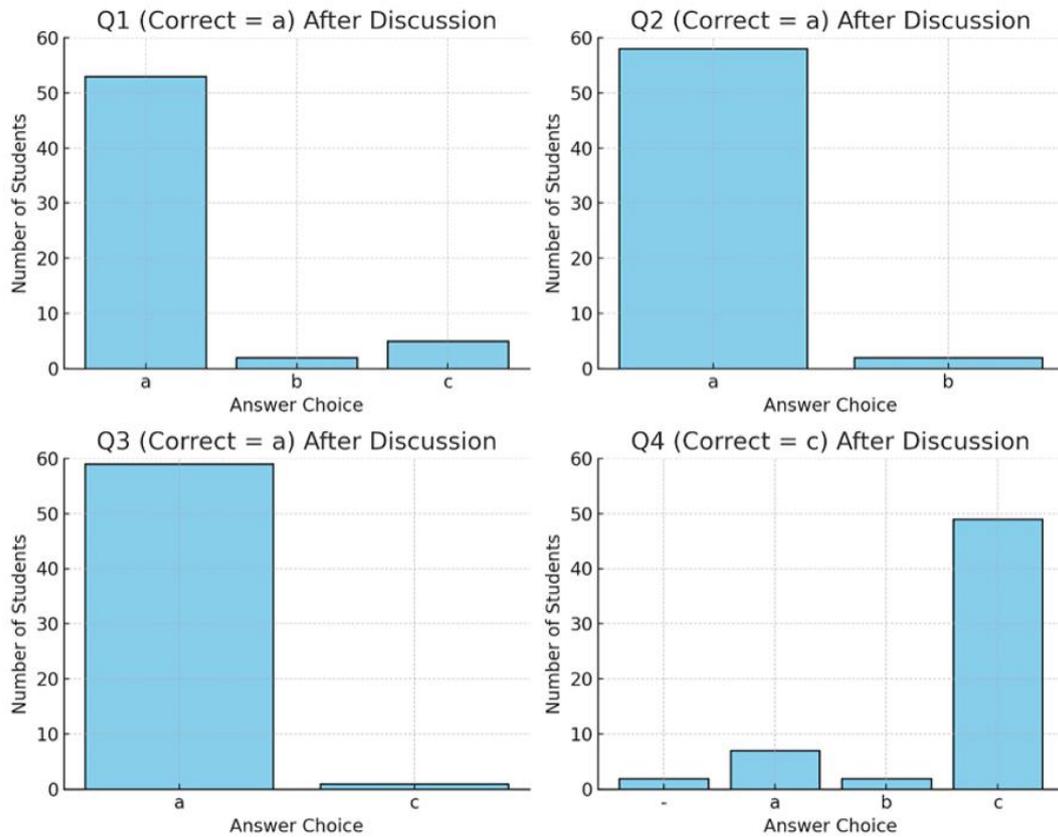
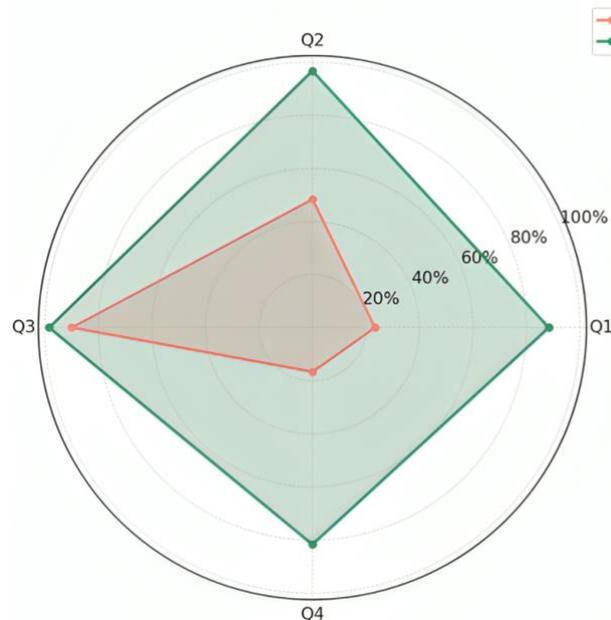**Figure 4.** Distribution of responses following peer discussion (Source: Authors)

Figure 5. The percentage of correct responses before and after peer discussion using a spider chart (Source: Authors)

Effect of Peer Discussion

Figure 5 illustrates the percentage of correct responses to each counterintuitive mathematics question before and after peer discussion using a spider chart. This format highlights the relative strengths and weaknesses across the four items and makes visible the substantial gains achieved through collaborative reasoning, particularly in Q1 and Q4, where initial counterintuitive cases were most pronounced.

Table 1. Chi-square tests comparing responses before and after peer discussion

Question	χ² (Chi-square)	df	p-value
Q1	48.80	1	< .001 ***
Q2	32.77	1	< .001 ***
Q3	2.43	1	.119 (ns)
Q4	48.15	1	< .001 ***
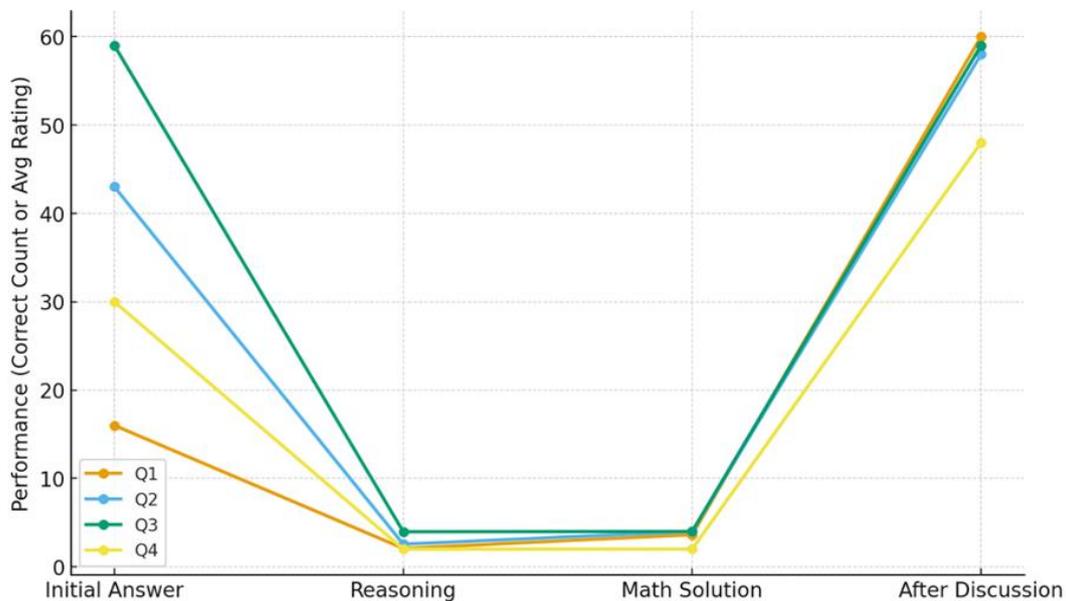
Figure 6. Progression of teacher performance across four stages (Source: Authors)

The red area illustrates the limited initial success rates, particularly in Q1 (23%) and Q4 (17%), where strong counter intuitions dominated teachers' reasoning. In contrast, the green area shows the dramatic gains after discussion, with correct responses rising to 89% in Q1 and 82% in Q4. Q2 and Q3 also showed increases, from 48% to 97% and 90% to 98%, respectively, though the relative improvement was less pronounced given their higher starting points.

Table 1 presents the results of Chi-square tests comparing correct and incorrect responses before and after peer discussion for each of the four counterintuitive problems.

The Chi-square analyses compared the distribution of correct and incorrect responses before and after discussion for each problem (Reasoning and mathematical solutions are not included). Results showed a significant improvement for Q1 ($\chi^2 = 48.80$, df = 1, p < .001), Q2 ($\chi^2 = 32.77$, df = 1, p < .001), and Q4 ($\chi^2 = 48.15$, df = 1, p < .001), indicating that peer discussion substantially increased the number of correct answers. In contrast, Q3 did not yield a significant effect ($\chi^2 = 2.43$, df = 1, p = .119), reflecting the already high rate of correct responses before discussion.

Comparison of Four Stages

Figure 6 displays the progression of teacher performance across four stages: initial answers, reasoning, mathematical solutions, and peer discussion, for each of the counterintuitive problems. The progression lines across all four problems illustrate distinct learning trajectories from initial answers through reasoning, mathematical solutions, and peer discussion. For Q1 and Q2, the number of correct answers was initially low to moderate, and reasoning quality averaged only around the "partially correct" level; however, mathematical problem-solving substantially improved performance, and peer discussion ultimately led to near-universal correctness. Q3 followed a different pattern, beginning with a high rate of correct responses and strong reasoning, and showing only incremental gains thereafter, indicating that this item was less counterintuitive. By contrast, Q4 started with very few correct answers, was characterized by weak reasoning, and showed only limited gains from mathematical problem-solving. Although discussion improved performance substantially, nearly one-fifth of teachers still failed to resolve the counter intuitions.

Table 2. Pairwise comparisons of teachers' performance across the four predefined data-collection stages

Stage 1	Stage 2	Wilcoxon Z	p-value	Adjusted Sig. (Bonferroni)
Initial	Reasoning	0.0	< .001 ***	< .001 ***
Initial	Math	0.0	< .001 ***	< .001 ***
Initial	Discussion	97.5	< .001 ***	< .001 ***
Reasoning	Math	492.5	.025 *	.149 (ns)
Reasoning	Discussion	0.0	< .001 ***	< .001 ***
Math	Discussion	0.0	< .001 ***	< .001 ***

Table 3. The regression table for the effect of reasoning quality on mathematical solutions

Predictor	Unstandardized B	Standard error	Standardized beta	t	Sig.	95% CI lower	95% CI upper
(Constant)	2.38	0.15	-	16.16	< .001	2.09	2.67
Reasoning rating	0.39	0.05	.433	7.47	< .001	0.29	0.49

Note. [a]Dependent variable: Mathematical solution rating (1-4); [b]Predictor: Reasoning rating (1-4); & CI: Confidence interval

The Friedman test was run to locate the statistically significant differences between the four stages. The Friedman test revealed a significant effect across the four stages (χ^2 = 137.20, df = 3, p < .001). This indicates that teachers' performance changed significantly from initial answers to reasoning, then to mathematical solutions, and finally to after discussion. **Table 2** shows the pairwise Wilcoxon signed-rank post-hoc results with Bonferroni correction.

The analysis revealed a statistically significant effect of stage on performance, χ^2 (3, N = 60) = 137.20, p < .001, showing that teachers' scores varied substantially depending on the stage. Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed that performance improved significantly from the initial answers to reasoning, mathematical solutions, and peer discussion (all p < .001). Although the increase from reasoning to mathematical solutions was marginal and did not remain significant after correction, the transition from both reasoning and mathematics to peer discussion yielded significant gains.

Finally, we conducted a regression analysis to examine whether reasoning quality predicted teachers' subsequent mathematical solutions. The regression model revealed a moderate positive relationship (R = .433) between reasoning and mathematical solution accuracy, explaining about 19% of the variance (R^2 = .187). The model fit was acceptable, with a standard error of .818. **Table 3** displays the regression output for the effect of reasoning quality on mathematical solutions.

Results indicated a moderate positive correlation between reasoning and solution accuracy (r = .43, p < .001). The regression model was significant, $F (1,242)$ = 55.77, p < .001, with an R^2 of .19, showing that reasoning quality explained approximately 19% of the variance in mathematical solution ratings. The regression coefficient was positive (B = 0.39, p < .001), showing that a one-point increase in reasoning rating is associated with an expected 0.39-point increase in mathematical solution rating (1-4 scale).

Main Findings

1. At the intuitive stage, most teachers relied on their initial judgments, which frequently led to incorrect answers, particularly in problems involving scale and averages, while only the exponential growth task was widely understood correctly

2. When asked to justify their answers, most teachers provided only partially correct explanations, indicating that the reasoning behind their intuitive choices was incomplete or flawed.

3. Engaging in formal calculations led to substantial improvements in accuracy for most problems, demonstrating the power of structured problem-solving, although the average-speed task remained troublesome for many.

4. Collaborative discussion markedly enhanced performance across all problems, as teachers clarified their thinking, corrected mistakes, and converged on correct solutions.

5. A comparison of correct-response rates before and after discussion highlighted dramatic gains on the most counter-intuitive items and smaller, but still positive, gains on tasks that were already widely understood.

6. Examining performance across the four stages showed distinct learning trajectories—initial answers and reasoning were often incorrect, calculations improved understanding, and peer discussion produced the largest gains, though some counter intuitions persisted in certain tasks.

7. A regression of solution accuracy on reasoning quality revealed a moderate positive relationship, suggesting that better reasoning contributes to better solutions but does not fully account for performance.

DISCUSSION

The findings of this study show the critical importance of incorporating counterintuitive problems in mathematics classrooms. Such problems examine how students' initial intuitive answers often lead to unexpected errors. However, structured interventions such as guided reasoning, mathematical solution, and peer discussion can effectively bridge the gap between intuition and formal mathematical understanding. Therefore, these results consistent with Kapur's (2016) theory of productive failure, which shows those initial errors, when addressed through reflective processes, can promote deeper conceptual learning. Similarly, Liljedahl (2020) notes that nonroutine tasks in thinking classrooms spark cognitive conflict, increasing deeper mathematical understanding through collaboration and reflection, thus reinforcing the efficacy of instructional strategies. By integrating such problems with targeted scaffolding, teachers can enhance their students' mathematical proficiency and instructional effectiveness in their mathematics classrooms. This directly addresses the first and third RQs by demonstrating the influence of intuitive responses and the corrective potential of structured interventions.

Further analysis of four problems revealed that teachers experienced varying levels of intuitive conflict. The *string around the earth* (Q1) and *average speed trap* (Q4) generated the strongest counter intuitions, while the *algae growth* problem (Q3) was predominantly answered correctly from the outset (see **Figure 1**). This pattern further illustrates how counterintuitive problems expose the disconnect between everyday reasoning and formal mathematics. This result aligns with Russell and Gobet's (2013) perspective on how counterintuitive tasks expose gaps between everyday reasoning and formal mathematics, triggering cognitive conflict that drives learning. Similarly, Schoenfeld (2020) emphasizes that tasks challenging intuitive assumptions promote reflective reasoning essential for conceptual development. Guided reasoning, mathematical solutions, and peer discussions facilitated resolution of these conflicts, aligning with Kapur's (2016) productive failure framework and addressing the **RQ3**. Specifically, these interventions enabled teachers to achieve conceptual views by systematically confronting intuitive errors.

In terms of reasoning quality, teachers' initial errors typically stemmed from coherent but incomplete conceptual models rather than random guessing, addressing the **RQ2**. For example, common misconceptions such as believing that heavier objects fall faster or that non-square rectangles cannot be parallelograms reflect misapplied everyday experiences (Hershkowitz & Vinner, 1980). In Q1 and Q4, such reasoning was prevalent, consistent with Tversky and Kahneman's (1974) findings on intuition-driven errors.

A regression analysis showed a moderate correlation between reasoning quality and solution accuracy (see **Table 3**), suggesting that stronger justifications tended to predict better outcomes, although a considerable amount of variance remained unexplained. This modest effect size, compared to more robust findings in probability-based research, points to the differing cognitive demands of geometric and probabilistic reasoning (Borovcnik & Kapadia, 2009), further informing the **RQ2**.

Furthermore, mathematical computations significantly improved accuracy in Q1, Q2, and Q3, particularly in proportional and geometric tasks, indicating their role in correcting intuitive errors. This aligns with Hiebert and Grouws' (2007) focus on structured problem-solving as a pathway to meaningful learning. In contrast, the limited improvement in Q4 through computations alone suggests that some counter intuitions, such as those involving average speed, resist procedural approaches, and requiring conceptual restructuring (Vosniadou, 2013). Peer discussion proved highly effective; in Q4, the correct response rate increased from 17% to 82% following discussion (see **Figure 5**), addressing **RQ3**. This supports Stylianides and Stylianides' (2008) perspective on collaborative discourse as means of co-constructing mathematical meaning, as well as Brousseau's (2011) theory of didactic situations, where peer interactions facilities the resolution of cognitive disequilibrium. Similarly, Boaler (2022) point out those collaborative activities, combined with innovative

scaffolding, enable teachers to reshape misconceptions and collaboratively build mathematical understanding. However, the minimal gains from peer discussion in Q3, likely due to its intuitive accessibility, show the influence of task difficulty on productive struggle (Balta & Eryilmaz, 2017).

Building on the impact of peer discussion, teachers' performance trajectories across the four instructional stages (intuition, reasoning, mathematical solution, and peer discussion) varied significantly by problem type, thus addressing to the **RQ4**. Statistical analysis using the Friedman test ($p < 0.05$; see **Figure 6**) confirmed distinct developmental patterns. Q1 and Q2 followed a classic cognitive conflict trajectory, beginning with low initial accuracy, showing moderate gains through reasoning, significant improvement after solution, and achieving near-complete accuracy following discussion. This trajectory reflects D'Mello and Graesser's (2012) model of dissonance-driven insight. Conversely, Q3 exhibited a flat trajectory, likely due to its intuitive simplicity, supporting Fischbein's (1987) distinction between counterintuitive and complex but accessible tasks. Meanwhile, Q4 presented a more resistant pattern; despite instructional scaffolding, 20% of teachers persisted in holding incorrect conceptions. This suggests the presence of deeply rooted misconceptions, consistent with Hamza and Wickman's (2008) argument that stable misconceptions often require targeted and conceptually focused interventions.

Collectively, these findings show the value of well-designed counterintuitive tasks in mathematics instruction, addressing all four RQs. Problems such as the string around the earth (Campanario, 1998) illustrate how cognitive conflict can lead to conceptual breakthroughs when supported by structured reasoning and collaborative dialogue (Stylianides & Stylianides, 2008). Furthermore, the study aligns with research indicating that such problems engage key cognitive mechanisms, including inhibition and working memory (Brookman-Byrne et al., 2018; Dumontheil et al., 2022). Emotional factors such as frustration likely influenced students' engagement and learning processes (Immordino-Yang & Damasio, 2007). However, the persistence of counter intuition in problems like Q4 challenges claims of universal insight following cognitive conflict (Maylone, 2000; Pruner, 2016).

CONCLUSIONS, IMPLICATIONS, AND LIMITATIONS

This study contributes to research on counterintuitive problems by providing a stage-based analysis of how mathematics teachers' understanding develops across intuition, reasoning, calculation, and peer discussion. Unlike prior studies that primarily compare pre- and post-performance, this study traces within-session changes, revealing which stages are most effective in resolving counterintuitive conflicts. In addition, the study demonstrates that reasoning quality predicts, but does not fully determine, solution performance, highlighting the distinct instructional roles of formal calculation and peer discussion. By focusing on in-service teachers, the findings offer new insights relevant to teacher professional development that have been largely absent from previous research.

These results show the necessity of careful task design and robust scaffolding to navigate intuitive conflicts and promote lasting mathematical understanding. By showing how counterintuitive problems can reveal and resolve intuitive conceptual challenges, these findings highlight the importance of rethinking current instructional practices. Specifically, these problems serve as powerful diagnostic tools to reveal gaps in students' conceptual understanding, which often arise from a dependence on everyday intuition instead of formal mathematical principles. This necessitates the integration of such tasks into professional development programs, where educators can practice navigating cognitive conflicts in a supportive environment. For instance, teacher education curricula could incorporate modules that simulate these scenarios, encouraging pre-service (Reyes, 2025; Sony, 2025) and in-service teachers to engage in reflective practices that mirror the study's interventions with guided reasoning, mathematical solutions, and peer discussions. This approach not only enhances individual mathematical proficiency but also equips teachers to foster similar persistence in their students, promoting a classroom culture where errors are viewed as opportunities for growth rather than failures.

Furthermore, the implications extend to broader educational policy and resource planning. Schools and policymakers should prioritize invest resources that promote the use of counterintuitive problems, such as digital tools and instructional materials featuring nonroutine tasks across grade levels. Encouraging collaborative teacher communities and peer discussions can help address persistent conceptual difficulties.

Additionally, supporting teachers with training in metacognitive and emotional regulation strategies may strengthen their ability to manage cognitive challenges and model effective problem-solving for students. These efforts can foster deeper learning environments in mathematics classrooms, ultimately contributing to a more equitable and effective mathematics education system.

However, while these implications are encouraging, the study has few limitations that qualify its conclusions and highlight important directions for future research. The small sample size (N = 60) limits the generalizability of findings, necessitating larger, more diverse samples to enhance external validity. Additionally, the algae growth problem (Q3) appeared to be less counterintuitive than the other tasks, which may have influenced the comparative analyses. During peer discussions, 25 teachers mentioned encountering a similar problem in preparing students for the university entrance exam. This prior familiarity with the task may have reduced the cognitive conflict intended by the problem. In contrast, the remaining 35 teachers either did not recall the problem or had no prior familiarity with it, introducing potential variability in responses that was unrelated to the instructional interventions. These limitations suggest that future research should include larger samples and ensure all tasks are novel and equally counterintuitive to participants. Doing so would help isolate the effects of instructional strategies more reliably and provide a clearer understanding of how teachers engage with conceptual conflict in mathematics learning.

REFERENCES

Balta, N., & Asikainen, M. A. (2019). A comparison of Olympians' and regular students' approaches and successes in solving counterintuitive dynamics problems. *International Journal of Science Education, 41*(12), 1644-1666. https://doi.org/10.1080/09500693.2019.1624990

Balta, N., & Eryılmaz, A. (2017). Counterintuitive dynamics test. *International Journal of Science and Mathematics Education, 15*, 411-431. https://doi.org/10.1007/s10763-015-9694-6

Benacerraf, P., & Putnam, H. (Eds.). (1983). *Philosophy of mathematics: Selected readings*. Cambridge University Press. https://doi.org/10.1086/289270

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. McGraw Hill Book Company. https://doi.org/10.1037/11164-000

Boaler, J. (2022). *Mathematical mindsets: Unleashing student' potential through creative mathematics, inspiring messages and innovative teaching*. John Wiley & Sons.

Borovcnik, M., & Kapadia, R. (2009). Research and developments in probability education. *International Electronic Journal of Mathematics Education, 4*(3), 111-130. https://doi.org/10.29333/iejme/233

Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS ONE, 13*(6), Article e0198973. https://doi.org/10.1371/journal.pone.0198973

Brousseau, G. (2011). *La théorie des situations didactiques en mathématiques* [The theory of didactic situations in mathematics]. Presses Universitaires de Rennes. https://doi.org/10.4000/educationdidactique.1005

Burton, W. G. (1978). The physics and physiology of storage. In P. M. Harris (Ed.), *The potato crop* (pp. 545-606). Springer. https://doi.org/10.1007/978-1-4899-7210-1_15

Campanario, J. M. (1998). Using counterintuitive problems in teaching physics. *The Physics Teacher, 36*(7), 439-441. https://doi.org/10.1119/1.879917

Chen, X., Wang, S., You, C., Magaña-Loaiza, O. S., & Jin, R. B. (2022). Experimental implementation of the fractional-vortex Hilbert hotel. *Physical Review A, 106*(3), Article 033521. https://doi.org/10.1103/PhysRevA.106.033521

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*(2), 161-199. https://doi.org/10.1207/s15327809jls1402_1

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 420-464). Macmillan. https://doi.org/10.1108/978-1-60752-874-620251022

Clements, P., Buffington, P., & Tobey, C. (2013). *Middle-grade students' misconceptions about the graphical representation of simple fractions: An assessment from the eliciting mathematical misconceptions project (EM 2)*. Society for Research on Educational Effectiveness.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Daniel, A., & DiNapoli, J. (2024). Investigating pedagogies in undergraduate precalculus and their relationships to students' attitudes towards mathematics and perseverance in problem-solving. *Journal of Research in Science, Mathematics and Technology Education, 7*(SI), 61-83. https://doi.org/10.31756/jrsmte.314si

Dumontheil, I., Brookman-Byrne, A., Tolmie, A. K., & Mareschal, D. (2022). Neural and cognitive underpinnings of counterintuitive science and math reasoning in adolescence. *Journal of Cognitive Neuroscience, 34*(7), 1205-1229. https://doi.org/10.1162/jocn_a_01854

Dumontheil, I., Wilkinson, H. R., Farran, E. K., Smid, C., Modhvadia, R., Mareschal, D., & UnLocke Team. (2023). How do executive functions influence children's reasoning about counterintuitive concepts in mathematics and science? *Journal of Cognitive Enhancement, 7*, 257-275. https://doi.org/10.1007/s41465-023-00271-0

Everett, L., & Pennathur, A. (2007, June). A design process for conceptually based, counterintuitive problems. In 2007 Annual Conference & Exposition (pp. 12-37). https://doi.org/10.18260/1-2--1586

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach*. Springer. https://doi.org/10.2307/3619981

Foisy, L. M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics?. *Trends in Neuroscience and Education, 4*(1-2), 26-36. https://doi.org/10.1016/j.tine.2015.03.001

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42. https://doi.org/10.1257/089533005775196732

Gennari, A. S., & Valentini, M. (2024). The effects of physical activity on cognitive and learning abilities in childhood. *European Educational Researcher, 7*(1), 1-30. https://doi.org/10.31757/euer.711

Gill, R. D. (2025). Monty hall problem: Solution. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1547-1553). Springer. https://doi.org/10.1007/978-3-662-69359-9_380

GMAT Club Forum (2026, February 23). *Distance, speed, & time word problems made easy*. GMAT Club. https://gmatclub.com/forum/distance-speed-time-word-problems-made-easy-87481.html

Hamza, K. M., & Wickman, P. O. (2008). Describing and analyzing learning in action: An empirical study of the importance of misconceptions in learning science. *Science Education, 92*(1), 141-164. https://doi.org/10.1002/sce.20233

Hershkowitz, R., &Vinner, S. (1980). Concept images and some common cognitive paths in the development of some simple geometric concepts. In R. Karplus (Ed.), *Proceedings of the 4th International Conference for the Psychology of Mathematics Education* (pp. 177-184). University of California.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 371-404). Information Age. https://doi.org/10.1108/978-1-28141-200-320251013

Illustrative Mathematics. (2026, February 23). Lake algae. *Illustrative Mathematics*. https://tasks.illustrativemathematics.org/content-standards/tasks/533

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education, 1*(1), 3-10. https://doi.org/10.1111/j.1751-228X.2007.00004.x

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289-299. https://doi.org/10.1080/00461520.2016.1155457

Kaufman, R. (2023). Gabriel's horn and the painter's paradox in perspective. *Ohio Journal of School Mathematics, 94*, 1-5. https://doi.org/10.18061/ojsm.4192

Lanting, L. S. (2025). Fostering the development of young students' analytical thinking by use of a problem-solving method. *Journal of Research in Science, Mathematics and Technology Education, 8*(SI), 377-401. https://doi.org/10.31756/jrsmte.4117si

Liljedahl, P. (2020). *Building thinking classrooms in mathematics, grades K-12: 14 teaching practices for enhancing learning*. Corwin Press.

Maylone, N. J. (2000). Using counterintuitive problems to promote student discussion. *Mathematics Teaching in the Middle School, 5*(8), 542-546. https://doi.org/10.5951/mtms.5.8.0542

National Foundation for Educational Research. (2019). Learning counterintuitive concepts evaluation report. *NFER*. https://www.nfer.ac.uk/media/uqapkjif/learning_counterintuitive_concepts_evaluation_report_-final.pdf

OBrien, A., & NíRíordáin, M. (2025). Irish 14-year-old students' knowledge of initial algebra. *The European Educational Researcher, 8*(2), 7-29. https://doi.org/10.31757/euer.822

Palmer, L. R., Sumanapala, D. K., Mareschal, D., Dumontheil, I., & UnLocke Team. (2025). Neural associations between inhibitory control and counterintuitive reasoning in science and maths in primary school children. *Journal of Cognitive Neuroscience, 37*(5), 915-940. https://doi.org/10.1162/jocn_a_02303

Parsons, C. D. (2018). *Mathematics in philosophy: Selected essays*. Cornell University Press. https://doi.org/10.1086/524822

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315-341. https://doi.org/10.1007/s10648-006-9029-9

Pruner, M. (2016, February 23). *BCAMT weekly math tasks*. https://peterliljedahl.com. https://peterliljedahl.com/wp-content/uploads/Pruners-Math-Tasks.pdf

Ran, H. (2022). *Making struggle productive: The effects of cognitively guided instruction (CGI) professional development on struggling students' problem-solving* [PhD thesis, University of Miami].

Reyes, K. (2025). A systematic review of culturally relevant pedagogy in mathematics teacher education: Methods, successes, and room for improvement. *International Educational Review, 3*(1), 21-40. https://doi.org/10.58693/ier.312

Russell, Y. I., & Gobet, F. (2013). What is counterintuitive? Religious cognition and natural expectation. *Review of Philosophy and Psychology, 4*(4), 715-749. https://doi.org/10.1007/s13164-013-0160-5

Santos-Trigo, M. (2024). Problem solving in mathematics education: Tracing its foundations and current research-practice trends. *ZDM Mathematics Education, 56*, 211-222. https://doi.org/10.1007/s11858-024-01578-8

Schoenfeld, A. H. (2020). Mathematical practices, in theory and practice. *ZDM Mathematics Education, 52*, 1163-1175. https://doi.org/10.1007/s11858-020-01162-w

Shahbari, J. A. (2021). Cognitive conflict in technological environments: Cognitive process and emotions through intuitive errors in area, perimeter, and volume. *Mathematics, 9*(14), Article 1672. https://doi.org/10.3390/math9141672

Smith III, J. P., DiSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163. https://doi.org/10.1207/s15327809jls0302_1

Sony, S. D. (2025). Preservice teachers' perspectives on connections within mathematical concepts: A qualitative study. *Mathematics Teacher Education and Development, 27*(2), Article 2. https://mted.merga.net.au/index.php/mted/article/view/1240

Stavy, R., & Tirosh, D. (2000). *How students (mis-) understand science and mathematics: Intuitive rules*. Teachers College Press.

Stylianides, A. J., & Stylianides, G. J. (2007). Learning mathematics with understanding: A critical consideration of the learning principle in the principles and standards for school mathematics. *The Mathematics Enthusiast, 4*(1), 103-114. https://doi.org/10.54870/1551-3440.1063

Stylianides, A. J., & Stylianides, G. J. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical Thinking and Learning, 10*(2), 103-133. https://doi.org/10.1080/10986060701854425

Sun, Q. (2021). The probability principle of the birthday paradox and extended applications. *The Frontiers of Society, Science and Technology, 3*(3), 110-112. https://doi.org/10.25236/fsst.2021.030320

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vosniadou, S. (2013). *International handbook of research on conceptual change (2nd ed.).* Routledge. https://doi.org/10.4324/9780203154472

Wessel, J. R. (2018). A neural mechanism for surprise-related interruptions of visuospatial working memory. *Cerebral Cortex, 28*(1), 199-212. https://doi.org/10.1093/cercor/bhw367

❖❖❖