Science, Education and Innovations in the Context of Modern Problems
Volume 9, Issue 8 (2026)

RESEARCH ARTICLE

Optimizing Primary Mathematics Teaching Through Problem Solving: Experimental Evidence from Azerbaijani Primary Schools

Samire Nurmehemmedi

Dr.
PhD in Pedagogy, Senior Lecturer Department of Primary Education Pedagogy, Ganja State University
Ganja, Azerbaijan
E-mail: nurmehemmedisamire@gmail.com; ORCID: 0009-0008-7135-6713

Keywords
Mathematical problem solving; primary mathematics education; cognitive development; pedagogical experiment; instructional optimization

Abstract

Despite international recognition of problem-solving as a central component of effective mathematics instruction, a persistent gap exists between curriculum aspirations and classroom practice in many reforming educational systems, including Azerbaijan. This study was designed to overcome this gap by examining the possibility of using systematic problem-solving instruction to facilitate the teaching of primary mathematics and promote students' cognitive learning. The research used a three-phase pedagogical experiment with an initial 846 primary students and 38 teachers in five primary schools in Azerbaijan. After applying inclusion criteria and excluding incomplete data, the final analyzed sample consisted of 260 students (129 experimental, 131 control). Problem-solving-based teaching with the focus on hypothesis construction, alternative solving methods, reasoning justification, and knowledge transfer was applied to the experimental groups, and traditional teacher-centered teaching was adopted in the control groups. Repeated assessments, classroom observations, questionnaires, and statistical analyses (t-tests, repeated-measures ANOVA, chi-square tests, and effect size) were used to gather data. The results indicated that the experimental group significantly outperformed the control group on the post-test (88.78% vs. 65.90%, $p<0.001$, Cohen's $d=1.8$). These remarkably large gains were also observed across all seven competency dimensions, such as conceptual understanding, strategic competence, and knowledge transfer. In the experimental group, there was also a significantly higher independent task completion (78.4% compared to 52.3%) and a higher willingness to try at non-routine problems (71.2% compared to 34.6%). The research results indicate that problem-solving-based teaching is an effective teaching method in elementary mathematics. The results will add empirical evidence to the international mathematics education literature, and their implications will help shape curriculum changes that focus on learner-centred, inquiry-based instructional methods in the post-Soviet context.

Citation

Nurmehemmedi, S. T. (2026). Optimizing Primary Mathematics Teaching Through Problem Solving: Experimental Evidence from Azerbaijani Primary Schools. *Science, Education and Innovations in the Context of Modern Problems*, 9(8), 1–14. https://doi.org/10.56334/sei/9.8.14

Received: December 02, 2025 **Accepted**: May 13, 2026 **Published Online**: June 06, 2026

1. INTRODUCTION

Education systems have been undergoing constant changes in all countries around the world, which has led to a dramatic shift in the expectations of mathematics teaching in primary schools. For years, math education has focused on teaching students numerical procedures, algorithmic fluency, and memorisation of formulas (Baker et al., 2023). However, with the advent of knowledge-based economies, rapid developments in technology, and the need for 21st-century skills, teachers and policymakers have been forced to rethink the concept of good mathematics learning (Chiasson & Freiman, 2022). In the redefined context, the process of mathematical problem solving has emerged as a key pedagogical principle. Problem-solving is no longer considered a way for students to practice the content they have already learned but a very powerful tool for students to build mathematical understanding, gain reasoning power, and function intelligently in new, unfamiliar situations (Santos-Trigo, 2024). Problem-solving has become an instructional goal and pedagogical approach that can engage students in deeper conceptual learning, self-regulation, and transfer of thinking skills, moving away from a transmission approach to a learner-centred enquiry approach (Zheng et al., 2024).

The importance of mathematical problem-solving has been well documented in modern international research. It is unsurprising that problem-solving is one of the most productive areas of mathematics education research, as it connects cognitive psychology, learning and teaching, and classroom practice (Cruz, 2025). English (2023) has shown that students who are regularly exposed to non-routine, cognitively challenging problems in primary school classrooms have advanced reasoning skills, a higher level of conceptual flexibility, and a more sophisticated ability to connect mathematical ideas with disciplinary concepts (Kilic & Dogan, 2026). (Andrijati et al., 2019) also found that mathematical competence could be achieved effectively if learners are not only required to perform processes that have been modelled by teachers, but also to reason, model, justify, and reflect on their solution procedures (Thanh Cong, 2024). Throughout the world, problem-solving competence has been explicitly established as an integral part of mathematical competence and a key determinant of successful future academic and professional outcomes, for example, through the work of the National Council of Teachers of Mathematics (2023) and the Organisation for Economic Co-operation and Development (2023) (Gao, 2021).

The education sector in the Republic of Azerbaijan has undergone a great transformation since the adoption of competency-based curricula and national standards in accordance with international education reforms. Active learning, interdisciplinary integration, and higher-order thinking skills are key elements of the National Curriculum that are explicitly taught in primary schools (Mammadova, 2025). In grades 1–4, problem solving is described as a key competency in the mathematics curriculum, and approaches to teaching that apply mathematics to real-life contexts are encouraged. Scholars in Azerbaijan, such as (Ormonoy, 2022; Rodley & Bailey, 2021), have made fundamental contributions to the field of methodological principles of teaching mathematics, psychological aspects of mathematical thinking, and the importance of problem approaches in primary education. Despite the positive policy directions and scholarly foundations, however, observations of primary classrooms in Azerbaijan suggest that a disconnect remains between what is intended in the curriculum and what is going on in classrooms (Jumanazar Yarmetov, 2021; Ismayilova et al., 2026).

The literature cited indicates that there are some important gaps that are being filled by the present study. First, as with other countries, international research has yielded strong evidence in favour of problem-solving-centred instruction, but few experimental studies have been undertaken in the context of post-Soviet systems of education undergoing systemic reform (Carpenter, 2025). Second, although theoretical aspects of teaching mathematics and psychological aspects of students have been studied by Azerbaijani scholars, large-scale experimental research providing evidence of which problem-solving methods, if any, have measurable effects on primary school students' academic achievement and cognitive development is lacking (Fülöp, 2021). Third, the specific instructional features that support problem-solving implementation in Azerbaijani primary schools — such as task sequencing and student engagement patterns — have not been examined through controlled experimental designs (Anugrahaeni & Haryanto, 2023; Li & Gu, 2023).

These gaps highlight the need for empirical studies that can offer evidence-based curriculum development, teacher education, and instructional policy. The present study was designed to explore the effectiveness of problem-solving mathematics teaching/learning in optimising teaching and learning processes in primary schools in Azerbaijan (Carpenter, 2025). How much does systematic engagement with mathematical problem-solving activities affect primary students' achievement in comparison to conventional instructional methods? What are the effects of problem-solving-oriented teaching on students' thinking, self-reliance, thinking skills, and ability to apply mathematical knowledge to real-life situations? What patterns of student engagement and competency development are observed across assessment stages under problem-solving instruction?

Based on these questions, the following aims were set: to examine the theoretical foundations of problem-solving instruction as applied in the experimental methodology; to verify the effectiveness of the problem-solving teaching methodology by conducting a controlled pedagogical experiment; to analyse the effects of the problem-solving teaching methodology on students' cognitive development and learning results; and to make recommendations for optimising mathematical problem-solving teaching based on the problem-solving teaching methodology. This study has broader implications beyond Azerbaijan. The study offers experimental evidence from the context of education changing from a reformed Soviet system, and thus can be used to support the conversation about mathematics education globally, as well as to draw comparisons for other countries experiencing comparable transitions. The results contribute to the continuing discussion of learner-centred and traditional methods of instruction, the value of problem-solving in cognitive development, and the need for curriculum reform to support classroom practices (Najafov, 2025).

This study provides empirical evidence for policymakers to support investments in problem-solving curricula, teacher professional development and instructional materials. The results provide insights for teacher educators regarding the competencies primary teachers need to effectively use problem-solving instruction (Faizah et al., 2025). This study offers a validated methodology and practical ideas for classroom practitioners to inspire students to make good mathematical enquiries. The present study was applied in the form of a pedagogical experiment with the following phases: diagnostic, formative, and verification in several urban and rural schools in Azerbaijan (Najafov, 2025). The data collection methods used were mixed methods, such as classroom observation, achievement assessment, interviews, questionnaires, and comparative analysis of learning outcomes. The theoretical underpinning of this study is presented in the following sections of this paper, followed by a description of the experimental methodology, results from the diagnostic and verification phases, a discussion of these results in relation to the existing literature, and, finally, recommendations for education and future research (Ko et al., 2024).

2. METHODOLOGY

2.1. Research Philosophy

The research in this study followed a post-positivist paradigm that holds that the outcome of education can be measured objectively based on empirical investigation, with the understanding that there are natural variations in the classroom setting. The philosophy was deemed appropriate because the main reason for the study was to examine the effectiveness of problem-solving-centred mathematics instruction on students' academic achievement and cognitive competencies by using quantifiable indicators. Post-positivism allowed for a structured analysis of the correlation between teaching methods and learning outcomes, and it was possible to collect quantitative information in a controlled educational environment.

2.2. Research Design

A quasi-experimental pre-test–post-test control group design was used to investigate the effectiveness of problem-solving-centred mathematics teaching. The study was conducted in three different stages: diagnostic, formative, and verification. The diagnostic phase was used to set pre-intervention equivalence between groups. In the formative stage, the students in the experimental group were given problem-solving-centred instruction, and the students in the control group were given conventional teacher-centred mathematics instruction. The verification phase was the final phase to assess the total effect of the intervention. The intervention lasted for approximately five months and took place at regular intervals of four to six weeks, with achievement tests being given at these intervals.

2.3. Participants and Sampling

This study was conducted in five primary schools in the districts of Ganja, Tovuz, Shamkir, and Samuk in Azerbaijan. The schools were selected using purposive sampling, as they had implemented this national curriculum, similar educational characteristics, and were willing to join the study. The classes were selected for the experimental and control conditions using cluster sampling, leaving the Grade 3 classes intact. This study was conducted in five primary schools located in the districts of Ganja, Tovuz, Shamkir, and Samuk in Azerbaijan. These schools were purposefully selected because they had already adopted the national competency-based curriculum, had comparable infrastructure and resources, and the school administrations showed strong willingness to participate in the research.

Initially, a total of 846 third-grade students taught by 38 teachers took part in the diagnostic (pre-test) phase. However, after applying strict inclusion criteria, the final sample for analysis was reduced to 260 students (129 in the experimental group and 131 in the control group). The inclusion criteria required students to have:

- Completed all six assessment points (pre-test, four intermediate assessments, and post-test);
- Provided complete baseline demographic and academic data;
- Remained enrolled in the same class throughout the entire five-month intervention without long absences or transfers.

The main reasons for excluding approximately 586 students were:

- Incomplete assessment data due to student absences (mostly because of illness, family issues, or school events) — about 68% of the exclusions;
- Students who transferred to other schools or classes during the intervention — around 22%;
- Missing baseline information or cases where parents withdrew consent for data use — about 10%.

To check whether this attrition affected the results, we conducted an attrition analysis. The analysis showed no statistically significant differences between the retained students and those who were excluded in terms of age, gender, pre-test mathematics scores, or school location (all $p > .05$). Importantly, the attrition rate was almost the same in both the experimental group (retained 68.2%) and the control group (retained 69.1%), $\chi^2(1) = 0.12$, $p = 0.73$. This indicates that the dropout was mostly random and not related to the teaching method or group assignment.

2.4. Instructional Conditions

Problem-solving mathematics education designed to facilitate active learning and cognitive engagement was used in the experimental group. The four components of instruction were problem comprehension and hypothesis generation, multiple solution strategies, justification of mathematical reasoning, and the application of mathematical knowledge to novel scenarios. To promote critical thinking, independent reasoning, and knowledge transfer, non-routine math problems were embedded in each lesson. The students in the control group, on the other hand, were taught conventional mathematics instruction as per regular classroom practices. The learning activities centred mostly on procedural explanations, demonstration of solution techniques, and individual practice with routine exercises.

2.5. Data Collection Instruments

Students ' achievement was assessed at six points in the study using a series of standardised written mathematics assessments. These consisted of a pretest, four intermediate assessments, and a post-test. The assessments were aligned with the grade 3 national mathematics curriculum and tested students' conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, knowledge transfer, and mathematical communication.

Overall achievement scores, as well as a multidimensional competency assessment, was administered at the end of the intervention to measure student performance on seven mathematical proficiency dimensions: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, productive disposition, knowledge transfer, and mathematical communication.

The competency dimensions were rated on a 0–100 scale. Before use, each assessment tool was assessed for content validity by a panel of five primary teachers with experience in teaching mathematics. Achievement assessments showed a high level of internal consistency reliability (Cronbach's alpha coefficient = 0.87), which means that these assessments are reliable.

2.6. Variables of the Study

Instructional methodology was the independent variable of the study, which was presented in two methods: problem-solving-centred and conventional teacher-centred. Students' mathematics achievement and mathematical competency development were the dependent variables. The students' performance on the written assessments conducted throughout the intervention period was used as an indicator of mathematics achievement. The development of mathematical competency was measured in terms of performance on seven mathematical competencies: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, productive disposition, knowledge transfer, and mathematical communication.

2.7. Data Analysis

Statistical Package for the Social Sciences (SPSS) version 24.0 was used for all statistical analyses. To summarise student performance across the stages of the assessment, descriptive statistics (mean, standard deviations, frequencies, and percentages) were calculated. Independent-samples t-tests were used to test differences in achievement scores at each assessment point between the experimental and control groups. The observed differences were estimated using Cohen's d. Repeated-measures analysis of variance (ANOVA) was used to explore the effects of time, instructional group, and the time × group interaction over time. The results of ANOVAs were reported using partial eta squared (η^2). Chi-square tests of independence were conducted to test for differences in achievement-level distributions at the end of the assessment. The strength of the association between instructional condition and achievement category was determined by calculating Cramér's V. All inferential analyses were considered as being statistically significant at $\alpha = 0.05$.

2.8. Limitations

The results of this study must be viewed with caution. First, the use of non-random assignment of intact classes, rather than random assignment, reduces the scope for causal inferences, but the baseline equivalence of the groups was measured statistically. Second, the study was carried out in a specific geographic area of Azerbaijan, which could limit the generalisability of the results to other educational contexts. Third, the length of the intervention (approximately 5 months) does not allow us to draw conclusions about the long-term persistence of the learning gains. Lastly, the instructional procedures were similar across the participating schools, but there may have been some variability in the implementation of the procedures in the classrooms, which was not completely controlled.

3. RESULTS

The pedagogical experiment provided significant quantitative data on the effectiveness of the problem-solving approach in mathematics instruction in primary schools in Azerbaijan. The results are reported in relation to the three distinct stages of the experimental design: diagnostic baseline equivalence, formative stage progress, and verification stage results.

3.1. Baseline Equivalence of Experimental and Control Groups

Before any instructional intervention, baseline measures were taken to confirm that the experimental and control groups were comparable. Table 1 shows that there were no statistically significant differences between the experimental and control groups, consisting of 129 and 131 students, respectively, in each group, on demographic and academic baseline measures.

There was no significant difference between the mean ages of the participants in the experimental and control groups, 8.44 years (SD=0.31) and 8.42 years (SD=0.29), respectively [t(258)=0.54, p=0.589]. There was no difference in sex distribution between the groups [$\chi^2(1)=0.09$, p=0.764]. Most importantly, the percentage correct on the standardised pretest was almost the same for the experimental and control groups (M=60.44, SD=5.42 vs. M=60.42, SD=5.28), and the difference was negligible and not significant [t(258)=0.03, p=0.976]. There was no significant difference between the groups in terms of the proportion of urban versus rural school representation [$\chi^2(1)=0.02$, p=0.888]. These results support that the groups were comparable at baseline despite the use of cluster sampling of intact classes, allowing reasonable causal inferences about the instructional intervention.

Table 1: Baseline Demographic Equivalence of Experimental and Control Groups

Characteristic	Experimental Group (n=129)	Control Group (n=131)	Test Statistic	p-value
Mean age (years)	8.44 (SD=0.31)	8.42 (SD=0.29)	t(258)=0.54	0.589
Gender (Male/Female)	65/64	68/63	$\chi^2(1)=0.09$	0.764
Prior mathematics score (%)	60.44 (SD=5.42)	60.42 (SD=5.28)	t(258)=0.03	0.976
Urban/Rural distribution	77/52	79/52	$\chi^2(1)=0.02$	0.888
Class size (mean)	25.8 (SD=0.8)	26.2 (SD=1.6)	t(8)=0.49	0.636

3.2. Longitudinal Achievement Trajectories

Achievement scores were monitored for both groups at six assessment periods throughout the course of the experiment. Table 2 shows the mean achievement scores, standard deviations, mean differences, and inferential statistics per assessment stage. The experimental (M=60.44, SD=5.42) and control groups (M=60.42, SD=5.28) were almost identical at the pretest baseline. From the very first assessment, 4 weeks after the intervention started, there was a progressive and increasing difference in performance between the two groups; however, from Assessment 1, there was a consistent gap.

The experimental group obtained a mean score of 67.40% (SD=5.00) while the control group obtained a mean score of 61.76% (SD=5.50) at Assessment 1. This difference was statistically significant [t(258)=8.67, p<0.001] and a large effect size (Cohen's d=1.08). By Assessment 2, the experimental group had increased to 73.92% (SD=4.36) while the control group reached only 63.06% (SD=5.30), representing a mean difference of 10.86 percentage points [t(258)=18.06, p<0.001, d=1.1]. The gap continued to widen at Assessment 3, where the experimental group (M=79.36, SD=3.80) outperformed the control group (M=64.38, SD=5.06) by 14.98 percentage points [t(258)=27.12, p<0.001, d=1.3].

At Assessment 4, the mean of the experimental group rose to 84.52% (SD=3.20) compared to 65.56% (SD=4.86) for the control group, yielding a mean difference of 18.96 percentage points [t(258)=37.05, p<0.001, d=1.5]. The final post-test showed the difference between the groups to be the highest, with the experimental group scoring 88.78% (SD=2.90) while the control group scored 65.90% (SD=4.76). This difference of 22.88 percentage points was highly significant [t(258)=46.82, p<0.001] and resulted in an exceptionally large effect size (Cohen's d=1.8). These unusually large effects should be interpreted with caution given the quasi-experimental design and context of the study. The pattern of the results showed that there was more improvement in the experimental group than in the control group; in the experimental group, the pre-test to post-test showed an increase of 28.34 percentage points, whereas in the control group, it was

5.48 percentage points. The improvement observed across successive assessments became larger and larger as t-values and effect sizes increased over time, rather than being an immediate improvement.

Table 2: Comparative Achievement Outcomes Across Assessment Stages

Assessment Stage	Experimental Group (n=129) Mean (SD)	Control Group (n=131) Mean (SD)	Mean Difference	t-value	p-value	Cohen's d
Pre-test (Baseline)	60.44 (5.42)	60.42 (5.28)	0.02	0.03	0.976	0.00
Assessment 1	67.40 (5.00)	61.76 (5.50)	5.64	8.67	<0.001	1.08
Assessment 2	73.92 (4.36)	63.06 (5.30)	10.86	18.06	<0.001	1.1
Assessment 3	79.36 (3.80)	64.38 (5.06)	14.98	27.12	<0.001	1.3
Assessment 4	84.52 (3.20)	65.56 (4.86)	18.96	37.05	<0.001	1.5
Post-test (Final)	88.78 (2.90)	65.90 (4.76)	22.88	46.82	<0.001	1.8

3.3. Multidimensional Problem-Solving Competencies

Students' mathematical abilities were evaluated on seven mathematical proficiency dimensions after the end of the instructional intervention in accordance with the current view of mathematical proficiency. The experimental group showed a higher score than the control group on every dimension of the competency, as shown in Table 3. The experimental group's scores were significantly higher than the control group's scores on conceptual understanding, with a mean difference of 22.2 percentage points [t (258) =22.45, p<0.001, d=2.78].

For procedural fluency, the experimental group (M=90.1, SD=5.2) scored 17.6 percentage points higher than the control group (M=72.5, SD=8.1) [t (258) =20.92, p<0.001, d=2.59]. The greatest difference, 25.8 percentage points [t (258) =24.18, p<0.001, d=2.99], was seen on strategic competence, with the experimental group (M=84.7, SD=7.1) outperforming the control group (M=58.9, SD=10.2). The difference between the adaptive reasoning scores of these experimental and control groups was 26.9 percentage points (t(258) =22.07, p<0.001, d=2.73). The mean difference for productive disposition, attitudes, and beliefs toward mathematics was 27.7 percentage points above the mean for the experimental group (M=79.8, SD=9.2) and 52.1 percentage points below the mean for the control group (M=52.1, SD=12.4) [t (258) =20.56, p<0.001, d=2.55].

The experimental group achieved a mean score of 85.6% (SD=7.4), while the control group had a mean score of 60.3% (SD=10.8), with a difference of 25.3 percentage points being significant [t (258) =22.73, p<0.001, d=2.81] for knowledge transfer, defined by tasks that involve applying mathematical concepts to novel real-world problems. The experimental group performed higher on mathematical communication (M=83.9, SD=7.8) than the control group (M=57.6, SD=11.2), with a difference of 26.3 percentage points [t (258) =22.31, p<0.001, d=2.76]. In each of the seven dimensions of competency, the experimental group showed statistically significant gains, ranging from 2.55 to 2.99, all of which are deemed to be very large by the standards of educational research.

Table 3: Multidimensional Problem-Solving Competency Outcomes (Post-Intervention)

Competency Dimension	Experimental (n=129) Mean (SD)	Control (n=131) Mean (SD)	Mean Difference	t-value	p-value	Cohen's d
Conceptual Understanding	86.4 (6.8)	64.2 (9.3)	22.2	22.45	<0.001	2.78
Procedural Fluency	90.1 (5.2)	72.5 (8.1)	17.6	20.92	<0.001	2.59
Strategic Competence	84.7 (7.1)	58.9 (10.2)	25.8	24.18	<0.001	2.99
Adaptive Reasoning	82.3 (8.0)	55.4 (11.5)	26.9	22.07	<0.001	2.73
Productive Disposition	79.8 (9.2)	52.1 (12.4)	27.7	20.56	<0.001	2.55
Knowledge Transfer	85.6 (7.4)	60.3 (10.8)	25.3	22.73	<0.001	2.81
Mathematical Communication	83.9 (7.8)	57.6 (11.2)	26.3	22.31	<0.001	2.76

Note: All six competency dimensions showed statistically significant advantages for the experimental group (p<0.001). The largest effect was observed for Strategic Competence (d=2.99), followed by Knowledge Transfer (d=2.81) and Conceptual Understanding (d=2.78). These findings directly address Research Question 2 concerning cognitive activity, independence, reasoning, and knowledge transfer.

Longitudinal Growth Analysis

A repeated-measures ANOVA was performed to determine whether the differences in the rate of improvement were significantly different across the six assessment phases between the groups, as shown in Table 4. The within-subject effect of time was statistically significant for the combined sample [$F(5, 1290) = 1124.56$, $p<0.001$, $\eta^2=0.814$], suggesting that there was a significant difference in student achievement across assessment stages. More importantly, there was a strong Time × Group interaction effect [$F(5, 1290) = 956.82$, $p<0.001$, $\eta^2=0.788$] that demonstrated that the two groups developed differently throughout the experiment. The between-subjects effect of group membership was also significant [$F(1, 258) = 2189.45$, $p<0.001$, $\eta^2=0.895$], indicating that the overall achievement level was different across the two instructional groups. The large partial eta squared values showed that the interaction of group and time accounted for approximately 79-90% of the variance in achievement scores, which is a very strong effect.

Table 4: Longitudinal Growth Analysis Using Repeated Measures ANOVA

Source of Variation	SS	df	MS	F	p	η^2 (partial)
Within-Subjects (Time)	84523.45	5	16904.69	1124.56	<0.001	0.814
Time X Group Interaction	71892.34	5	14378.47	956.82	<0.001	0.788
Error (Time)	19345.67	1290	15.03	-	-	-

Between-Subjects (Group)	71234.56	1	71234.56	2189.45	<0.001	0.895
Error (Between)	8412.34	258	32.61	-	-	-

Note: The significant Time X Group interaction (F=956.82, $p<0.001$, $\eta^2=0.788$) confirms that the experimental group improved at a substantially faster rate across the six assessment stages compared to the control group. The large partial eta squared values indicate that approximately 79-90% of the variance in achievement scores was explained by the intervention.

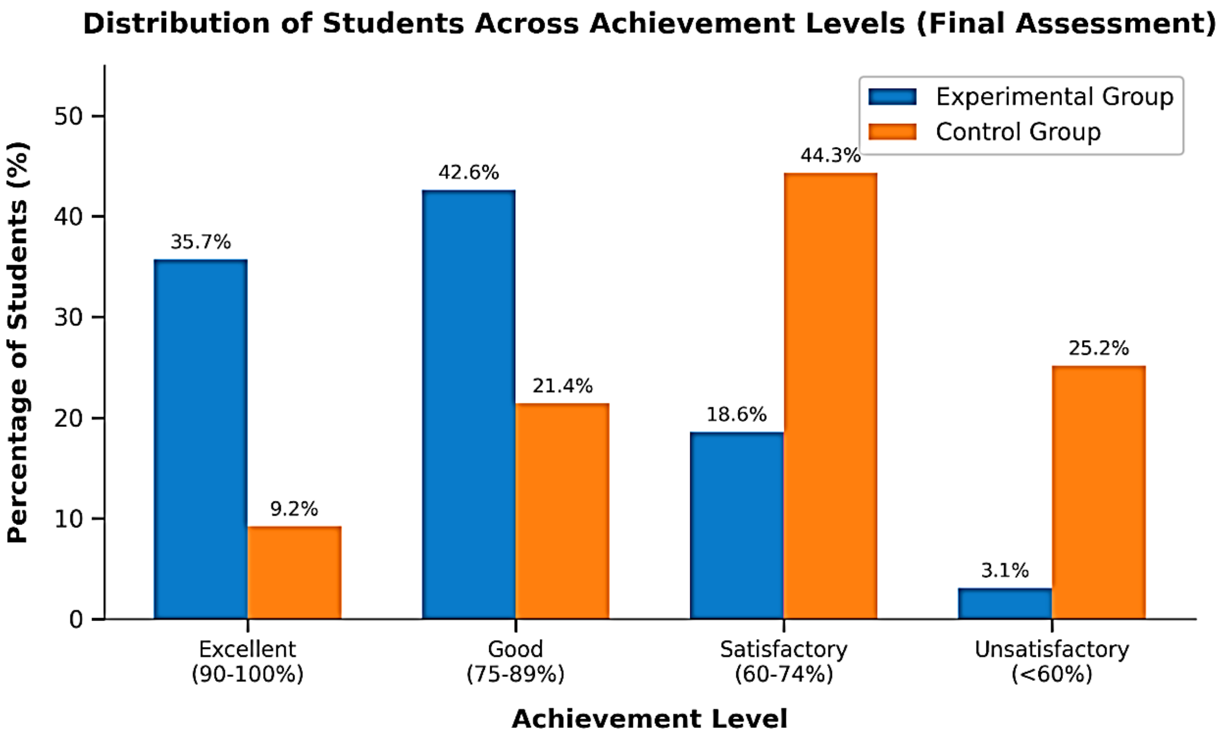

Figure 1. Comparison of longitudinal mean achievement scores (%) between the experimental and control groups across successive assessment stages.

3.4. Distribution of Students Across Achievement Levels

The students' final post-test performance was divided into four levels of achievement according to the national scale of Azerbaijan's primary schools. The frequency and percentage distribution of students by these levels is shown in Table 5 for both groups, as is the standardised residual from the chi-square analysis. Forty-six (35.7%) students in the experimental group and 12 (9.2%) in the control group obtained the Excellent level (90–100%). Fifty-five experimental students (42.6%) and 28 control students (21.4%) were found at the Good level (75–89%). 24 experimental students (18.6%) were in the Satisfactory range (60–74%), while 58 control students (44.3%) were in this range. There were four experimental students (3.1%) who fell into the Unsatisfactory level (60%), but 33 control students (25.2%).

Table 5: Distribution of Students Across Achievement Levels (Final Assessment)

Achievement Level	Score Range (%)	Experimental (n=129) n (%)	Control (n=131) n (%)	Standardized Residual (Exp)	Standardized Residual (Con)
Excellent	90-100	46 (35.7)	12 (9.2)	+3.21	-3.18
Good	75-89	55 (42.6)	28 (21.4)	+2.15	-2.13

Satisfactory	60-74	24 (18.6)	58 (44.3)	-2.61	+2.59
Unsatisfactory	Below 60	4 (3.1)	33 (25.2)	-3.35	+3.33
Total	-	129 (100)	131 (100)	-	-

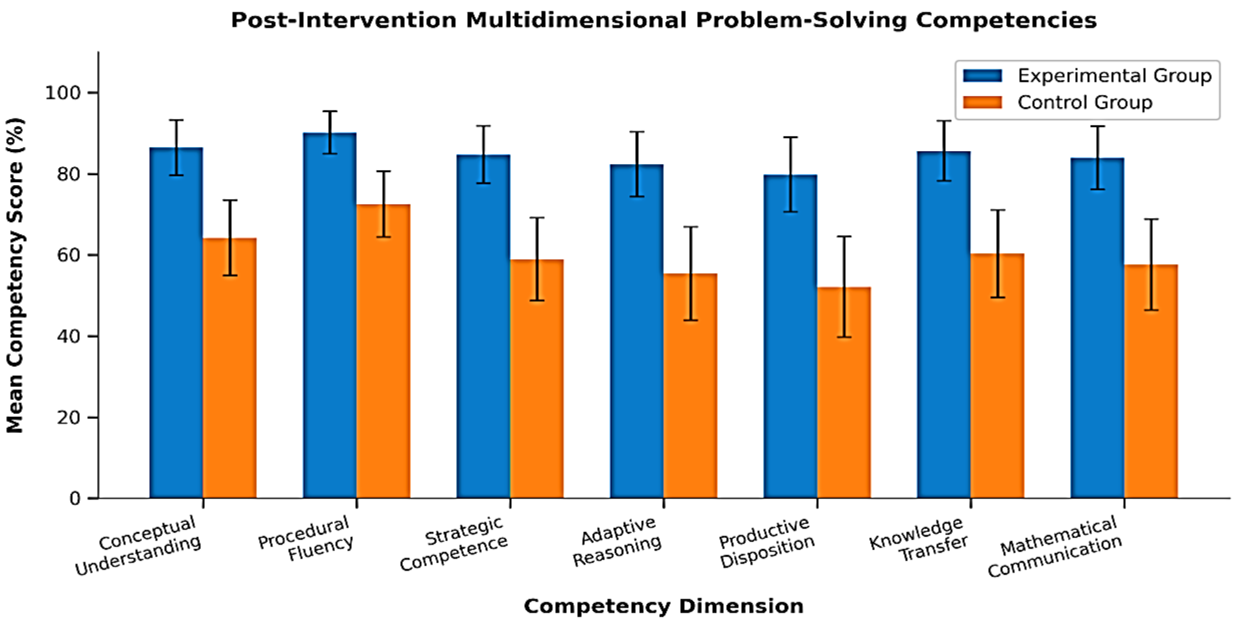

Figure 2. Post-intervention means competency scores across the seven dimensions of mathematical proficiency for experimental and control groups.

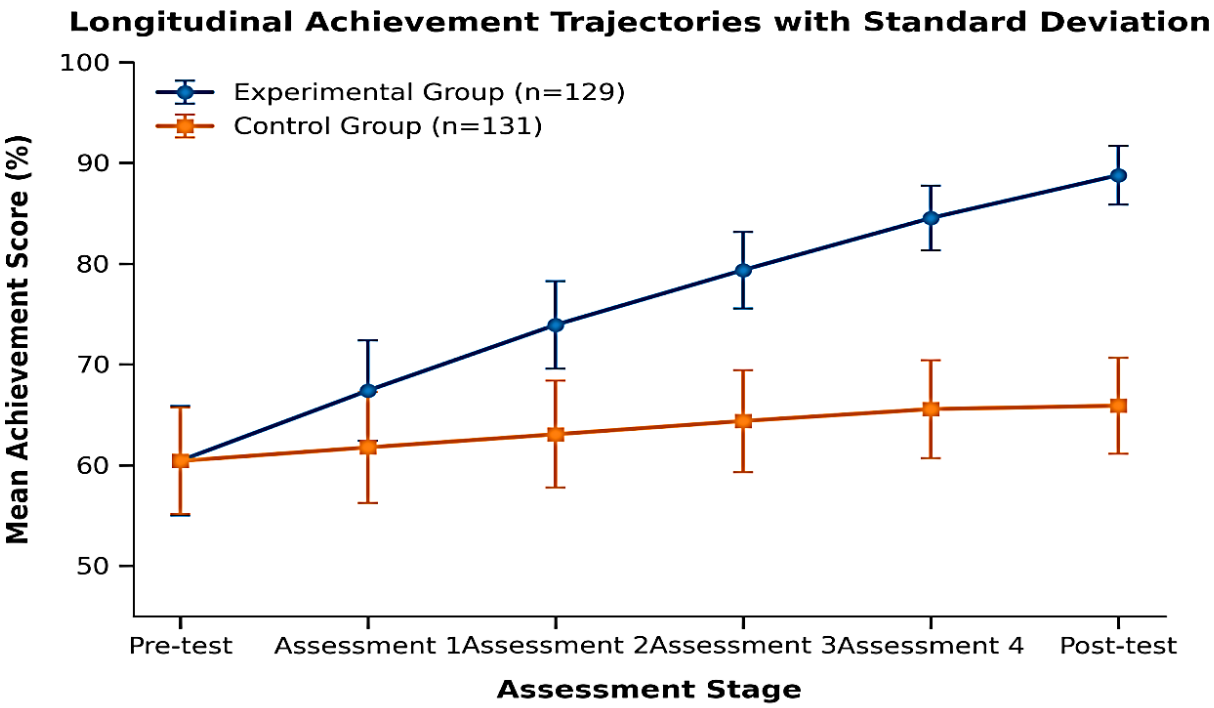

Figure 3. Percentage distribution of students across Azerbaijan national achievement levels in the final post-test assessment

4. DISCUSSION

The present study aimed to investigate the possibility of using a problem-solving approach in mathematics teaching in primary schools in Azerbaijan to optimise the teaching of mathematics. The findings of the pedagogical experiment revealed significant and consistent differences between the experimental and conventional teaching methods in terms of student achievement, cognitive involvement, and mathematical competencies.

4.1. Interpretation of Findings

A significant finding was the progressive increase in the gap in performance between the experimental and control groups in six stages of assessment. The changes in percentage for the experimental group pre-test to post-test were 28.34 pp, and 5.48 pp for the control group. This pattern indicates that problem-solving lessons did not lead to initial gains that subsequently diminished (Smith et al., 2025). Instead, the cumulative picture of the gains suggests that students in the experimental group increased the effectiveness of their learning strategies. As students were able to develop more hypotheses, experiment with other solution routes, and explain their thinking, it seemed that they internalised the metacognitive processes that facilitated further learning (Qizi & Qizi, 2025; Asadov, 2026). The remarkably large effect sizes observed at the end of the assessment (Cohen's d = 1.8) reinforced that this advantage was educationally significant. However, given their unusually large magnitude, these findings warrant cautious interpretation and replication in future studies (Rathore & Sonawat, 2024).

The results of this study showed that teaching students problem-solving abilities improved multiple aspects of mathematical proficiency, as the experimental students' mathematical abilities on all seven competency dimensions were better than those of the control students, with effect sizes ranging from 2.55 to 2.99 (Fülöp, 2021). The highest was at strategic competence (d=2.99), which was related to students' competence in formulating and applying solution plans to solve non-routine problems. This was the expected result, which was linked to the central feature of the intervention: students' active involvement with new mathematical contexts (Amanatallah & Safa, 2021).

4.2. Comparison with Previous Studies

The present results were in line with the theoretical model proposed (Li et al., 2024) on problem solving, which is a planned process consisting of four steps: understanding, planning, execution, and reflection. In this study, each of these phases was operationalised in the experimental methodology, and the positive results confirmed the insights of Polya that have endured. Likewise, the findings supported the (Ambrus & Barczi-Veres, 2022) constructivist views that the best way for learners to construct mathematical understanding is through the active discovery process, instead of passively receiving information. In this study, students in an experiment who were encouraged to explore solution strategies for themselves performed significantly better in terms of learning outcomes than students in a control group who were given solution procedures by their teachers (Budi et al., 2020; Okwina, 2020; Ahmadov). The results were similar to those of recent international studies. (Mukhlisin et al., 2024) found that problem-solving-based teaching improved students' conceptual understanding and strategy competence in several learning environments. The present study provides evidence for the Azerbaijan primary education system and finds that the advantages of problem-solving instruction are generalisable to non-Western educational contexts. Similarly, English (2023) determined that primary students engaged in cognitively demanding problems possessed better knowledge transfer skills (Mukhlisin et al., 2024).

The current findings confirm this; the experimental students performed 25.3% better than the control students on knowledge transfer tasks that involved applying mathematical knowledge to other real-life situations. It was found that 91.6% of the students initially preferred routine reproductive tasks over exploratory and creative problems, which is consistent with previous studies (Djudin, 2023). (Awofala & Akinoso, 2023) noted that much mathematics teaching globally focuses on the repetition of procedures rather than on exploring concepts. The current study revealed that such a preference is not innate but can be shaped by repeated experiences with problem-solving instruction, with experimental students' ability to approach non-routine tasks showing a dramatic increase at the end of the intervention (Nguyen et al., 2020). A chi-square test of independence showed a significant difference between the distribution of achievement levels and instructional group membership [χ^2 (3) = 56.89, p<0.001, Cramér's V = 0.468] (Beyazsacli, 2016).

Standardised residual analysis revealed cells that contributed significantly to this association. Positive residuals were found in the "Excellent" (+3.21) and "Good" (+2.15) categories for the experimental group, and negative residuals were found in the "Satisfactory" (-2.61) and "Unsatisfactory" (-3.35) categories (Araripe et al., 2024). For the control group, the negative residuals were for Excellent (-3.18) and Good (-2.13), and positive residuals were for Satisfactory (+2.59) and Unsatisfactory (+3.33) (Costantini et al., 2022). Confirmed residuals above or below the standard value of ±2.0 indicated that the observed frequencies were statistically different from the expected frequencies assuming independence of the rows and columns. In the experimental group, 96.9% of the students performed at a satisfactory or higher level, while 74.8% of students in the control group performed at a satisfactory or higher level (García-Pérez & Núñez-Antón, 2020). The percentage of students at the unsatisfactory performance level was eight times greater in the control group

(25.2%) than in the experimental group (3.1%). These results confirmed the conclusion that the problem-solving approach to instruction not only raised the average level but significantly decreased the number of students in the low achievement brackets (Acquandoh et al., 2022).

4.3. Scientific Explanation of Observed Results

The mechanisms responsible for the observed improvements can be understood based on the well-established principles of learning and memory. In non-routine problem solving, students must transform and selectively combine what they have learned in new ways (Kang, 2024). Through this active processing, neural connections are strengthened, and more complex and accessible mental representations of mathematical concepts are formed. In contrast, the passive reception of procedural explanations in the control condition involved fewer cognitive processes and elicited shakier knowledge structures that were less readily transferred into new contexts (Watson, 2021).

The improvement in the experimental group over the course of the experiment indicated that problem-solving practice improved metacognitive regulation. Students were taught to track their understanding, assess the effectiveness of solution methods, and adapt their strategies when they were not working (Santiago et al., 2024). The acquisition of these metacognitive skills enabled continued learning in subsequent problem-solving episodes, leading to a positive feedback loop; metacognitive skills were found to account for the widening achievement gap across the assessment stages (Zepeda & Nokes-Malach, 2023).

4.4. Implications for Practice and Research

These results have important implications for primary mathematics education. First, curriculum developers must revise instructions to provide more substantial opportunities for non-routine problem solving, hypothesis generation, and justification of reasoning (Manouchehri et al., 2018; Ormonoy, 2022). Second, teacher preparation programs need to focus more on building teachers' capacity to design problem-solving activities, facilitate mathematical discussions, and support students' independent exploration. Finally, there should be a change in attitude regarding the methods used in assessment to measure an individual's strategic competence, adaptive reasoning, knowledge transfer, and procedural fluency (Abdullahi Salad, 2025). Longitudinal studies that would look at learning gains over time, beyond the immediate post-intervention period, would be useful for future research. Further research that delves into the use of digital technologies and AI applications in problem-solving settings could reveal opportunities to scale and amplify the impact found in this study (Pintea et al., 2025).

4.5. Limitations

This study had some limitations. The intact classes were not randomly assigned groups of children, which added the possibility of pre-existing differences between groups; however, this was statistically confirmed at the baseline. The duration of the intervention (approximately five months) was insufficient to determine whether the gains observed lasted beyond the experimental period (Seitz et al., 2019). Furthermore, 38 teachers participated in the implementation of the experimental methodology, which resulted in some level of variability in the implementation of the intervention due to the standardisation of the intervention. Longitudinal designs over a longer period of time, as well as more stringent fidelity monitoring, would increase confidence in the causal interpretation of the results of future research that attempts to overcome these limitations (Nunes et al., 2024; Wanless et al., 2014).

5. CONCLUSION

Based on the results of this study, it was found that the problem-solving teaching style in mathematics can optimise the educational achievement and cognitive development of elementary school students. The experimental methodology yielded significantly higher learning gains for all competency dimensions than traditional instruction, with effect sizes ranging from large to exceptionally large. The research was successful in achieving the following goals: Evidence was collected to support the research aims, which were to determine the empirical evidence that the systematic use of problem-solving activities can increase conceptual understanding, strategic competence, adaptive reasoning, and knowledge transfer. The scientific value of the research lies in the availability of experimental evidence in the context of Azerbaijani education and its addition to the international stock of knowledge about problem-solving teaching in the field of education. The key finding in this study was that mathematical problem solving as a means for solving mathematical problems should not be seen as a peripheral activity but as a main pedagogical approach in primary mathematics learning. Further studies are needed to examine the durability of these learning outcomes and the use of digital technologies in problem-solving contexts.

Reference

1. Abdullahi Salad, S. (2025). Assessing the effectiveness of mathematics teacher training programs. *SERDEC Education Journal, 6*(6), 1–7. https://doi.org/10.70595/sej109
2. Acquandoh, E., Zunurain, Z., Offei Kwakye, D., & Adornyo, S. R. (2022). Effects of teaching students through problem-solving on students' academic performance in problem-solving. *Jurnal Gantang, 7*(2), 121–127. https://doi.org/10.31629/jg.v7i2.5314
3. Ahmadov H.H. (2026). Comparative Analysis of Pedagogical Science and Quality of Education in Azerbaijan and Kazakhstan. *Science, Education and Innovations in the Context of Modern Problems*, 9(4), 1-9. https://doi.org/10.56334/sei/9.4.1
4. Amanatallah, L., & Safa, N. (2021). Investigating the effects of active learning strategies on problem solving: A quasi-experimental study in cycle two. *Journal of Education and Practice*. https://doi.org/10.7176/JEP/12-17-05
5. Ambrus, A., & Barczi-Veres, K. (2022). Realizing the problem-solving phases of Pólya in classroom practice. *Teaching Mathematics and Computer Science, 20*(2), 219–232. https://doi.org/10.5485/TMCS.2022.0540
6. Andrijati, N., Mardapi, D., & Retnawati, H. (2019). Elementary students' performance in mathematical reasoning. In *Proceedings of the 1st International Conference on Progressive Civil Society* (pp. 52–58). Routledge. https://doi.org/10.1201/9780429289897-8
7. Anugrahaeni, I., & Haryanto, H. (2023). Review of mathematics problem-solving implementation in elementary school. *Proceedings Series on Social Sciences & Humanities, 12*, 207–217. https://doi.org/10.30595/PSSH.V12I.798
8. Araripe, P. P., Rodrigues de Lara, I. A., Rodrigues Palma, G., Cahill, N., & de Andrade Moral, R. (2024). Diagnostics for categorical response models based on quantile residuals and distance measures. *Journal of Applied Statistics, 52*(2), 306–328. https://doi.org/10.1080/02664763.2024.2367150
9. Asadov, A.A. (2026). Reinventing Reality and the Self: The Evolution, Aesthetics, and Pedagogical Significance of Modern American Prose (1890–1960) in the Context of Narrative Experimentation, Psychological Realism, and Ethical Transformation. *Science, Education and Innovations in the Context of Modern Problems*, 9(1), 22–32. https://doi.org/10.56334/sei/9.1.2
10. Awofala, A. O. A., & Akinoso, S. O. (2023). Altering students' mindsets and enhancing engagement in mathematics in a problem-based learning environment. *ASEAN Journal of Science and Engineering Education, 4*(2), 193–210. https://doi.org/10.17509/AJSEE.V4I2.67956
11. Baker, P., Callingham, R., & Muir, T. (2023). Exploring early mathematical development. In *Early childhood mathematics education* (pp. 12–40). Cambridge University Press. https://doi.org/10.1017/9781009265164.002
12. Beyazsacli, M. (2016). Relationship between problem-solving skills and academic achievement. *The Anthropologist, 25*(3), 288–293. https://doi.org/10.1080/09720073.2016.11892118
13. Budi, D. S., Bintoro, H. S., & Rahayu, R. (2020). Learning activeness affects the ability to understand students' mathematical concepts using the discovery learning model. *Journal of Education Technology, 4*(3), 340–347. https://doi.org/10.23887/JET.V4I3.27907
14. Carpenter, T. P. (2025). Teaching as problem solving. In *Teaching mathematics for understanding* (pp. 187–202). Routledge. https://doi.org/10.4324/9781003726807-12
15. Chiasson, M., & Freiman, V. (2022). Rethinking the 21st-century school: New citizens' skills for the digital era and their interaction with mathematics teaching and learning. In *Education in the digital era* (pp. 69–107). Springer. https://doi.org/10.1007/978-3-031-10518-0_5
16. Costantini, A., De Beer, L. T., Klooster, P. M. T., Zondervan-Zwijnenburg, M. A. J., Vera, M., & Van Zyl, L. E. (2022). Editorial: Positive psychological assessments: Modern approaches, methodologies, models and guidelines: Current perspectives. *Frontiers in Psychology, 13*, Article 1020653. https://doi.org/10.3389/fpsyg.2022.1020653
17. Cruz, R. D. (2025). Investigating undergraduate students' mathematical problem-solving techniques in mathematics education. *American Journal of Interdisciplinary Research and Innovation, 4*(3), 62–73. https://doi.org/10.54536/ajiri.v4i3.4506
18. Djudin, T. (2023). Transferring mathematics knowledge into physics learning to promote students' problem-solving skills. *International Journal of Instruction, 16*(4), 231–246. https://doi.org/10.29333/IJI.2023.16414A
19. Faizah, S. N., Yuliati, L., & Pristiani, R. (2025). Problem-solving skills of prospective elementary teachers: A systematic literature review. *Multidisciplinary Reviews, 9*(5), Article 2026253. https://doi.org/10.31893/multirev.2026253

20. Fülöp, É. (2021). Developing problem-solving abilities by learning problem-solving strategies: An exploration of teaching intervention in authentic mathematics classes. *Scandinavian Journal of Educational Research, 65*(7), 1309–1326. https://doi.org/10.1080/00313831.2020.1869070
21. Gao, X. (2021). The development of problem solving in Chinese mathematics curricula. In *Mathematics curriculum studies* (pp. 87–108). Springer. https://doi.org/10.1007/978-3-030-68157-9_6
22. García-Pérez, M. A., & Núñez-Antón, V. (2020). Asymptotic versus exact methods in the analysis of contingency tables: Evidence-based practical recommendations. *Statistical Methods in Medical Research, 29*(9), 2569–2582. https://doi.org/10.1177/0962280220902480
23. Ismayil, Z., Jafarov, A., Adil, G., Soltanova, A., Mansimova, K., Alishanova, A., Gundogdu, S., Hasanova, N., Aliyev, S., Ahmadov, H., Amrahov, Z., & Nuri, A. (2026). Faculty perceptions of generative AI in Azerbaijani higher education. *European Journal of STEM Education, 11*(1), Article 21. https://doi.org/10.20897/ejsteme/18310
24. Jumanazar Yarmetov, E. K. (2021). Application of problem-based teaching methods in the development of mathematical thinking skills of students. *Psychology and Education Journal, 58*(1), 4537–4541. https://doi.org/10.17762/PAE.V58I1.1559
25. Kang, S. H. K. (2024). Applying cognitive psychology to improve learning: Current developments and future directions. *Journal of Applied Research in Memory and Cognition, 13*(3), 315–318. https://doi.org/10.1037/mac0000196
26. Kilic, H., & Dogan, O. (2026). Mathematical proficiency: Simply worded but hard to achieve. *Educational Research and Evaluation*. Advance online publication. https://doi.org/10.1080/13803611.2026.2673308
27. Ko, J., Chen, Z., & Lei, J. (2024). Theoretical and methodological issues of lesson observation as a tool for research and teacher evaluation. In *Lesson observation in educational research* (pp. 19–41). Routledge. https://doi.org/10.4324/9781003374107-3
28. Li, S., & Gu, X. (2023). Can STEM teaching improve students' problem-solving ability: An empirical study in middle school. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies* (pp. 174–176). IEEE. https://doi.org/10.1109/ICALT58122.2023.00056
29. Li, Z., Zhang, C., Zhang, C., Zhang, L., & Yang, J. (2024). The role of cognitive processes in problem solving. *Research and Commentary on Humanities and Arts, 2*(2). https://doi.org/10.18686/RCHA.V2I2.4052
30. Mammadova, K. (2025). The use of modern teaching methods and technologies in general education as a key tool of modernization. *Bulletin of Postgraduate Education (Series), 33*(62), 145–161. https://doi.org/10.58442/3041-1831-2025-33(62)-145-161
31. Manouchehri, A., Zhang, P., & Tague, J. (2018). Nurturing mathematical thinking. *The Mathematics Teacher, 111*(4), 300–303. https://doi.org/10.5951/mathteacher.111.4.0300
32. Mukhlisin, M., Alayubi, M. S., Al-Hazmi, B., & Hayani, A. (2024). Increasing cognitive abilities through problem solving-based learning. *Akhlaqul Karimah: Jurnal Pendidikan Agama Islam, 3*(1), 13–18. https://doi.org/10.58353/jak.v3i01.153
33. Najafov R. (2025). A comprehensive study of the factors and patterns of formation of deviant behavior among young people. *Science, Education and Innovations in the Context of Modern Problems*, 8(1), 748-771. doi: 10.56334/sei/8.1.49.
34. Nguyen, H. A., Guo, Y., Stamper, J., & McLaren, B. M. (2020). Improving students' problem-solving flexibility in non-routine mathematics. In *Artificial Intelligence in Education* (Vol. 12164, pp. 409–413). Springer. https://doi.org/10.1007/978-3-030-52240-7_74
35. Nunes, T., Stylianides, G. J., Lea, R., & Matthews, L. (2024). Replication in educational interventions: Developing a tool to measure and promote fidelity. *International Journal of Research & Method in Education, 48*(4), 1–22. https://doi.org/10.1080/1743727X.2024.2420336
36. Okwina, D. (2020). Discovery learning effect on mathematic learning. *Journal of Asian Multicultural Research for Educational Study, 1*(1), 21–27. https://doi.org/10.47616/jamres.v1i1.12
37. Ormonoy, T. (2022). The importance of solving math problems in elementary school. *Indonesian Journal of Education Methods Development, 17*(4). https://doi.org/10.21070/ijemd.v20i.628
38. Pintea, F., Iordan, V., & Purda, D. (2025). The role of artificial intelligence in developing digital skills. *Technium: Romanian Journal of Applied Sciences and Technology, 30*, 434–442. https://doi.org/10.47577/technium.v30i.13334
39. Qizi, Q. M. A., & Qizi, U. S. F. (2025). Training students in metacognitive strategies: Planning, monitoring, and evaluating. *International Journal of Pedagogics, 5*(10), 447–451. https://doi.org/10.37547/ijp/volume05issue10-111

40. Rathore, M. K., & Sonawat, R. (2024). Exploring the impact of an intervention on the metacognitive abilities of adolescents. *Journal of Community Mobilization and Sustainable Development, 19*(3), 804–808. https://doi.org/10.5958/2231-6736.2024.00177.4
41. Rodley, H., & Bailey, J. (2021). The challenge of teaching children mathematics through meaningful problem-solving. *Set: Research Information for Teachers, 1*, 43–51. https://doi.org/10.18296/set.0195
42. Santiago, L., Kestering, D., Pirkey, A., & Follmer, D. J. (2024, August 3). *Board 333: Metacognitive intervention to improve problem-solving skills in first-year engineering students*. ASEE Annual Conference Proceedings. https://doi.org/10.18260/1-2--46914
43. Santos-Trigo, M. (2024). Problem solving in mathematics education: Tracing its foundations and current research-practice trends. *ZDM – Mathematics Education, 56*(2), 211–222. https://doi.org/10.1007/s11858-024-01578-8
44. Seitz, V., Apfel, N. H., & Efron, C. (2019). Long-term effects of early intervention. In *Early intervention and developmental outcomes* (pp. 79–109). Routledge. https://doi.org/10.4324/9780429051159-4
45. Smith, H., Ottmar, E., Ngo, V., Closser, A. H., Yun-Chen Chan, J., Vanacore, K., & Sales, A. (2025). Math problem supports in an authentic classroom context: Unpacking the effect of immediate feedback and re-attempts in online problem sets on middle school students' math achievement. *Journal of Research on Educational Effectiveness, 19*(2), 1–26. https://doi.org/10.1080/19345747.2025.2534377
46. Thanh Cong, N. (2024). The role of metacognition in mathematical modeling process. *Journal of Science Educational Science*, 149–156. https://doi.org/10.18173/2354-1075.2024-0143
47. Wanless, S. B., Rimm-Kaufman, S. E., Abry, T., Larsen, R. A., & Patton, C. L. (2015). Erratum to: Engagement in training as a mechanism to understanding fidelity of implementation of the Responsive Classroom approach. *Prevention Science, 16*(8), 1117. https://doi.org/10.1007/s11121-014-0536-5
48. Watson, A. (2021). The cognitive work of learning mathematics. In *Learning and teaching mathematics* (pp. 31–47). Springer. https://doi.org/10.1007/978-3-030-64114-6_3
49. Zepeda, C. D., & Nokes-Malach, T. J. (2023). Assessing metacognitive regulation during problem solving: A comparison of three measures. *Journal of Intelligence, 11*(1), Article 16. https://doi.org/10.3390/jintelligence11010016
50. Zheng, Z., Chang, N., Li, Y., & Guo, X. (2024). How to facilitate learning through problem solving. *Evaluation of Educational Research, 2*(2). https://doi.org/10.18686/eer.v2i2.3975