STEM Education

STEM Education, 6(4): 697–715
DOI: 10.3934/steme.2026028
Received: 10 February 2026
Revised: 30 April 2026
Accepted: 24 June 2026
Published: 06 July 2026

https://www.aimspress.com/journal/steme

Research article

Empirical evaluation of the effectiveness of the STEAM approach on the development of technical thinking in primary school students

Guldana A. Totikova[1], Aidarbek A. Yessaliyev[2,*], Nurgul N. Medetbekova[2,*], Laura T. Iskakova[3] and Zhanar Sh. Zhiyasheva[2]

[1] Department of Pedagogy, Central Asian Innovation University, Kazakhstan; totikovag@gmail.com
[2] M. Auezov South Kazakhstan University, Kazakhstan; aidar.esali@mail, ru, nnnsssbbb@mail.ru, zhiyasheva88@mail.ru
[3] Academician A. Kuatbekov Peoples' Friendship University, Kazakhstan; laura_uko70@mail.ru

* **Correspondence:** Email: aidar.esali@mail.ru, nnnsssbbb@mail.ru; Tel: +77470162092.

Academic Editor: William Guo

Abstract: In this study, we aimed to empirically evaluate the effectiveness of a specialized science, technology, engineering, arts, and mathematics (STEAM) program that purposefully utilizes 2D and 3D modeling tools to develop technical, logical, and spatial thinking in 1st-4th-grade students.

To test the hypothesis of an association between the STEAM approach and the development of thinking, a quasi-experiment with pre- and post-testing in parallel groups was conducted (N = 172). The experimental group (EG) (n = 85) participated in project-based activities involving the creation of physical and digital models throughout the academic year, whereas the control group (CG) (n = 87) followed a standard curriculum for the same period.

Adapted versions of the Bennett Mechanical Comprehension Test (technical thinking), Raven's Progressive Matrices (logical thinking), and the Yakimanskaya–Zarkhin–Kadayas Spatial Representations Test (spatial thinking) were used, with grade-level differentiation and score normalization to ensure cross-grade comparability of the results.

The results revealed a statistically significant superiority of the EG over the CG in the post-test across all measured indicators ($p < 0.001$, Mann-Whitney U test) with large effect sizes ($r = 0.72–0.76$). The greatest improvement was recorded in spatial-thinking skills. Intra-group analysis (Wilcoxon signed-rank test) confirmed significant dynamics within the EG ($r > 0.80$) and no significant change in the CG. The results also suggested the potential importance of the teacher's role

as a facilitator in implementing STEAM-based activities.
The findings provide empirical support for the potential effectiveness of integrating a modeling-based STEAM approach into primary education to foster the development of technical, logical, and spatial thinking in children.

Keywords: STEAM education, primary school, spatial thinking, technical thinking, logical thinking, 3D modeling, educational experiment, project-based learning

1. Introduction

The dynamic development of modern society and the knowledge-based economy generate a consistent demand for specialists with interdisciplinary competencies capable of innovative activities and solving complex, practice-oriented problems. In this context, science, technology, engineering, arts, and mathematics (STEAM) education has established itself as one of the most promising approaches to address the challenges of the 21st century [1,2]. Its integrative nature aims to foster a holistic worldview in students and develop their critical thinking and a range of soft skills (communication, collaboration, and problem-solving) through the application of knowledge from various disciplines to implement projects [3,4].

Despite the growing interest in the STEAM approach and its recognized potential, particularly at the primary school level, where the foundations of cognitive activity are laid [5,6], a significant gap persists in the scientific and pedagogical literature. First, there is a lack of consensus on specific pedagogical strategies and didactic tools for effectively integrating STEAM into the educational process [7,8]. Second, there is a scarcity of empirically grounded research examining the effectiveness of specific STEAM practices in developing strictly defined cognitive abilities in primary school students, such as technical and spatial thinking [9,10]. Third, the role of digital and physical modeling tools (2D/3D) as catalysts for development remains unclear.

Consequently, the scientific novelty of this study lies in the comprehensive empirical assessment of the impact of a structured STEAM program, which purposefully utilizes 2D and 3D modeling tools, on the development of specific domains of thinking in primary school students, namely, technical, logical, and spatial thinking.

We aimed to evaluate the effectiveness of a specially designed STEAM-oriented educational program employing 2D and 3D modeling tools to develop technical, logical, and spatial thinking in 1st–to 4th-grade students.

Research hypothesis: Implementing a structured STEAM program based on the systematic use of 2D and 3D modeling in the primary school curriculum will lead to a statistically significant improvement in the indicators of technical, logical, and spatial thinking among students compared to traditional instruction.

1.1. Literature review

The theoretical foundation of this research is based on the principles of competency-based and integrative approaches to education. STEAM is not a simple sum of disciplines but a meta-approach fostering a holistic worldview and the formation of key competencies [11,12]. An analysis of the

literature reveals several key vectors in STEAM research at the primary school level, the first of which is associated with the motivational and cognitive effects of learning. Studies have confirmed that project-based activities within STEAM contribute to increasing student interest in the natural sciences and technical disciplines, thereby reducing academic anxiety and fostering a positive attitude toward learning [13,14]. The potential of STEAM to overcome gender stereotypes and engage girls in STEM fields is particularly emphasized [15].

The second direction focuses on developing soft skills. Integrating the arts into STEM disciplines is crucial for developing creativity, divergent thinking, and emotional intelligence [16,17]. Collaborative work on projects fosters communication and collaboration skills [18].

The third and least developed direction concerns the impact of STEAM on specific cognitive functions and thinking styles. Research indicates a positive influence of integrated approaches on computational thinking [19] and general scientific skills. However, there is a lack of studies investigating the relationship between STEAM practices, particularly those utilizing modeling tools, and the development of technical (engineering) and spatial thinking skills in primary school students. This identifies the research gap addressed in this study.

Despite the growing popularity of the STEAM approach, a consensus regarding its operationalization and a unified definition are lacking in the academic community [20,21]. STEAM is interpreted as either the development of creative abilities or as a tool for solving practical problems, leading to significant variability in the content of educational programs, assessment methods, and pedagogical strategies [6]. This methodological ambiguity creates serious difficulties for researchers and practitioners, hindering the comparison of results from different studies and the development of universal criteria for effective STEAM education. We bridge this gap by operationalizing STEAM through the lens of developing cognitive constructs, technical and spatial thinking, and applying standardized diagnostic tools to measure them.

A methodological challenge for modern STEAM pedagogy is the development of valid and reliable diagnostic tools to measure the acquisition of specific competencies [10,22]. To overcome this challenge, a set of tools adapted for primary school-age children was applied, which constitutes an element of its novelty.

2. Methodology

To test this hypothesis, a pedagogical experiment was conducted following a pre-test/post-test design with parallel groups (control group [CG] and experimental group [EG]).

2.1. Participants

The study involved 1st–4th grade students (N = 172) from general education schools in Southern Kazakhstan. The sample was divided into an EG and a CG using a matched group assignment procedure based on initial academic performance levels.

Matching was conducted using school-provided academic achievement indicators, including overall academic performance and subject-level results (where available), to improve the baseline comparability between groups prior to the intervention.

The EG (n = 85) received instruction according to the developed STEAM program, while the CG (n = 87) followed the standard curriculum without STEAM.

To ensure age comparability between the groups, the participants were proportionally distributed across grade levels (Table 1).

Table 1. Distribution of participants by grade and study group.

Grade	Experimental Group	Control Group
Grade 1	20	21
Grade 2	22	23
Grade 3	20	21
Grade 4	23	22
Total	85	87

This grade-balanced distribution supported the methodological validity of the age-based diagnostic adaptation and ensured a reliable comparison of outcomes between the EG and CG.

2.2. Intervention program

The program was implemented over an academic year (34 weeks). Sessions were held once a week, with each session lasting 40 minutes for grades 1-2 and 45 minutes for grades 3-4, aligning with the schedule of the partner institutions.

In addition, the EG participated in structured STEAM-based learning activities, including 2D and 3D modeling tasks, engineering-oriented problem solving, and project-based learning modules. The CG followed a standard curriculum without STEAM integration.

The program content was designed for the targeted development of engineering-technical and spatial thinking, as well as related cognitive processes (e.g., working memory, attention, and logical reasoning) in primary school students. The sessions were organized as group project activities (4-5 students per group), which fostered the development of communication skills and working memory through task distribution [23,24]. Each project was carried out over 3-4 weeks and included the following stages, aimed at the gradual formation of cognitive strategies: analysis of the technical task (logical thinking), creation and mental manipulation of images of the future object (spatial thinking), prototype creation (technical thinking), testing and problem-solving (logical and critical thinking), presentation of results, and reflection.

2.3. Examples of STEAM projects targeting key competencies

–"Bridge Across the Gorge" (using paper and cardboard): This project aimed to develop technical thinking through calculating strength and load and spatial thinking through creating and mentally rotating 2D nets of the future 3D structure [10]. Group work required the simultaneous holding of multiple parameters in working memory (length, weight, and strength).

–"Eco-House of the Future" (using wooden construction kits, counting sticks): This project stimulated spatial thinking and visual memory through the need to create a 3D model based on a pre-designed sketch plan. Logical thinking was developed by justifying the choice of materials and structures from the perspective of their efficiency.

–"Symmetrical World" (digital 3D modeling in Tinkercad): This project targeted the profound development of spatial thinking through mental rotation, scaling, and projection of virtual objects along three coordinate axes [9,25]. Creating symmetrical models requires a high level of abstract, logical thinking and concentration.

For younger students, especially those in grades 1–2, digital modeling tasks were scaffolded through step-by-step teacher demonstrations, simplified object manipulation, the use of predefined geometric shapes, and guided practice with basic navigation tools. Older students were given more autonomy in modifying, rotating, and combining the objects. This differentiation was used to ensure age-appropriate access to the same spatial reasoning tasks across grades.

2.4. Teacher qualifications

The program was delivered by primary school teachers (n = 5) who underwent specialized training in the fundamentals of STEAM education and the development of cognitive functions in primary school students through technical creativity (a 36-hour course), focusing on the implementation of modeling-based, project-oriented instructional practices. All teachers had over 10 years of experience and the highest qualification category, which ensured not only a high degree of implementation fidelity for the experimental program but also accurate tracking of progress in the development of students' target cognitive abilities.

2.5. Instruments and data collection

A battery of validated diagnostic instruments was used to assess changes in key cognitive components relevant to STEAM learning in the participants. The instruments were adapted for each grade level through the structured selection of developmentally appropriate items from the full test pools.

All diagnostic instruments were applied in age-adapted formats with grade-level differentiation. The adaptation was based on cognitive load regulation and the developmental characteristics of primary school students, including attention span, processing speed, and working memory capacity.

The number of tasks differed according to the grade level.

Raven’s Progressive Matrices — 6 / 8 / 10 / 10 tasks (Grades 1–4)

Bennett Mechanical Comprehension Test: 13 / 13 / 17 / 17 tasks

Yakimanskaya–Zarkhin–Kadayas Spatial Representations Test — 8 / 8 / 10 / 10 tasks.

Score Normalization Procedure

To ensure the comparability of results across grade levels, where the number of test items differed, raw scores were transformed into normalized proportional scores.

For each participant, the normalized score was calculated as follows:

normalized score = (number of correct responses) / (total number of items in the grade-specific test version)

This transformation resulted in a continuous scale ranging from 0 to 1, which represents the proportion of correctly solved tasks.

The normalization procedure was selected based on the following considerations:

- It preserves the relative performance of students within each grade level.
- It ensures comparability across tests with different numbers of items.
- It avoids distortion associated with unequal test lengths.

This is consistent with the diagnostic and developmental orientation of the adapted instruments.

The diagnostic battery included the following:

– Technical Thinking: Adapted version of the Bennett Mechanical Comprehension Test.

– Logical and Abstract Thinking: Raven's Progressive Matrices in an age-adapted format.

– Spatial Thinking: Yakimanskaya–Zarkhin–Kadayas Spatial Representations Test, which assesses mental rotation and object manipulation skills.

Instrument Adaptation and Construct Validity

All diagnostic instruments were adapted for primary school students using a structured and theoretically grounded item-selection approach, rather than the full-scale administration of the original tests.

The adaptation process was guided by the following principles.

- Preservation of the core diagnostic constructs (technical, logical, and spatial thinking)
- Developmental appropriateness and cognitive load regulation
- Progressive increase in task complexity across grade levels
- Use of perceptually clear, contextually meaningful, and age-relevant items
- Equivalence and comparability of the pre- and post-test measurements.

Importantly, the adaptation did not involve arbitrary simplification but rather a systematic selection of tasks that retained the underlying cognitive processes measured by the original instruments.

For Raven's Progressive Matrices, selected items preserved key processes such as pattern recognition, analogical reasoning, and structural completion, which constitute the core of nonverbal logical intelligence.

In the adapted Bennett Mechanical Comprehension Test, the focus was maintained on causal reasoning in mechanical systems, functional understanding, and spatial-mechanical relationships, while excluding developmentally inappropriate complexity.

The spatial thinking test adaptation targeted essential spatial processes, including mental rotation, projection, and object manipulation, which are foundational to early engineering cognition.

Scoring across all instruments was based on the proportion of correct responses relative to the total number of items, ensuring comparability across grade-specific versions and alignment with developmental diagnostic practices.

This structured adaptation framework supports the construct validity and methodological consistency of the instruments in primary education research.

Furthermore, the adapted instruments were used as developmentally oriented diagnostic tools rather than fully standardized psychometric scales. Therefore, this study emphasizes construct preservation, developmental appropriateness, and interpretive consistency across measurement points.

Detailed descriptions of the adaptation procedures, item selection rationale, grade-specific test structures, and additional statistical indicators are provided in the Supplementary Material (S1–S2).

2.6. Procedure

Testing was conducted twice: at the diagnostic (pre-intervention) and post-intervention stages of the study.

Detailed descriptions of the age-adaptation procedures, task selection rationale, and grade-specific structure of the diagnostic instruments are provided in the Supplementary Materials to ensure methodological transparency and reproducibility of the study.

Data Analysis

Statistical data processing was performed using STATISTICA (version 10.0). Owing to the

ordinal nature of the underlying test scores and deviations from the normal distribution identified using the Shapiro–Wilk test, non-parametric statistical methods were applied. Although normalized scores were expressed on a continuous 0–1 scale, they were derived from discrete item-level responses with grade-specific test lengths; therefore, non-parametric methods were considered more appropriate for conservative inferences.

– The Wilcoxon signed-rank test (W) was used to compare the results before and after the intervention within groups.

– The Mann–Whitney U test was used to compare the results between the EG and CG at both measurement stages.

– The Kruskal–Wallis H test was used to assess the influence of the grade factor, followed by the η^2 calculation.

– Effect sizes (r) were calculated for paired and unpaired comparisons to assess the practical significance of differences. Additional non-parametric effect size measures (rank-biserial correlation and common language effect size) are provided in the Supplementary Materials for interpretive completeness of the results.

The applied statistical approach was primarily descriptive and aimed to identify consistent patterns of change within groups and differences between groups at each measurement stage.

It should be noted that we did not directly model interaction effects (group × time), which would provide a formal test of differential change between groups. Therefore, the findings should be interpreted as converging evidence from complementary within- and between-group comparisons rather than as a formal test of interaction effects.

Exact p-values are reported where available; otherwise, threshold values are indicated (e.g., $p < 0.001$) in accordance with statistical reporting standards.

Given the strong significance levels observed in the main comparisons, no formal correction for multiple testing was applied; however, this should be considered when interpreting the marginal results.

Ethical Considerations

The study was conducted in accordance with the institutional ethical guidelines. Informed consent was obtained from the parents and school administration. Participation was voluntary, and data confidentiality and anonymity were guaranteed.

3. Results

To evaluate the effectiveness of the STEAM program, a comparative analysis of the dynamics of technical (Bennett test), logical-abstract (Raven's Progressive Matrices), and spatial thinking (Yakimanskaya-Zarkhin-Kadayas test) indicators was conducted in the CG (n = 87) and EG (n = 85) before (Pre) and after (Post) the experiment.

At the pre-intervention stage (Pre), statistical analysis using the Mann-Whitney U test revealed no significant differences between the groups on all three measured indicators ($p > 0.05$ for all tests, Table 2), confirming the initial equivalence of the groups and supporting the validity of the applied matching procedure.

Additional effect size measures (rank-biserial correlation and common language effect size) confirmed the magnitude and direction of the observed differences (see Supplementary Table S1 for details).

After the implementation of the STEAM program at the post-intervention stage, statistically significant differences were observed between the EG and CG. The descriptive statistics for the post-stage are shown in Table 2. These differences are visually demonstrated by the box plots in Figure 1.

Table 2. Descriptive statistics results at the post-intervention stage for the Bennett, Raven, and Yakimanskaya-Zarkhin-Kadayas tests.

Test	Group	N	Mean	SD	Min	Q1	Median	Q3	Max
Bennett	Experimental	85	9.4	2.4	6	7	9	12	13
	Control	87	6.6	1.8	4	5	7	8	9
Raven	Experimental	85	7.1	2.2	4	5	7	8	11
	Control	87	4.8	1.5	3	4	5	5	8
Spatial thinking	Experimental	85	6.8	1.7	4	5	7	8	10
	Control	87	4.2	1.5	2	3	4	5	7

Note: N — number of participants; Mean — arithmetic mean; SD — standard deviation; Min — minimum value; Q1 — first quartile; Median — median; Q3 — third quartile; Max — maximum value.

Because non-parametric methods were used, medians and interquartile ranges were considered the primary descriptive indicators, whereas means and standard deviations were reported for additional descriptive context.

The distribution of normalized pre- and post-test scores across groups, together with descriptive mean trends, is presented in Figure 1.

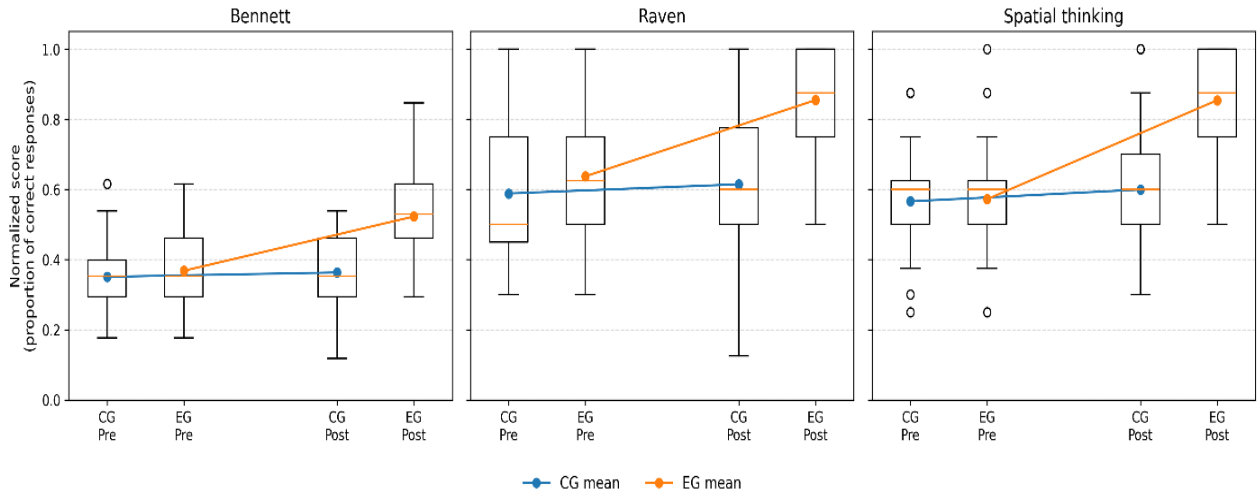

Figure 1. Distribution of normalized pre- and post-test scores for Bennett, Raven, and Spatial measures in the control (CG) and experimental (EG) groups.

Note: Box plots represent the median, interquartile range (IQR), data range and outliers. The lines indicate the descriptive mean trends within each group and are provided for visual interpretation only. The scores were normalized to the proportion of correct responses. The y-axis starts at zero to ensure an accurate visual representation.

Statistical analysis using the Mann-Whitney U test confirmed that the differences between the groups in the post-stage were statistically significant for all three tests ($p < 0.001$), with a large effect

size (Table 3).

Table 3. Results of the Mann-Whitney U test for intergroup comparisons.

Test	Stage	n (EG)	n (CG)	p	r	Interpretation
Bennett	Pre	85	87	0.519	0.048	Negligible
	Post	85	87	< 0.001	0.721	Large
Raven	Pre	85	87	0.211	0.092	Negligible
	Post	85	87	< 0.001	0.743	Large
Spatial thinking	Pre	85	87	0.23	0.088	Negligible
	Post	85	87	< 0.001	0.759	Large

Intra-group analysis using the Wilcoxon signed-rank test showed that statistically significant changes occurred in the EG for all three indicators between the pre- and post-stages ($p < 0.001$ for all tests) with large effect sizes. No significant dynamics were identified in the CG ($p > 0.05$), and the effect size was negligible (Table 4).

Taken together, the presence of statistically significant within-group improvement in the EG, combined with the absence of meaningful change in the CG and the emergence of strong between-group differences at the post-test stage, indicates a consistent pattern of differential development associated with the intervention.

Table 4. Results of the Wilcoxon test and effect size (r) for intra-group comparisons (pre → post).

Test	Group	N pairs	p	r	Interpretation
Bennett	Control	87	0.502	0.072	Negligible
	Experimental	85	< 0.001	0.814	Large
Raven	Control	87	0.427	0.085	Negligible
	Experimental	85	< 0.001	0.807	Large
Spatial thinking	Control	87	0.51	0.071	Negligible
	Experimental	85	< 0.001	0.826	Large

The effect size calculation (Figure 2) confirmed that the changes in the EG were not only statistically significant but also of high practical importance (large effect size, $r > 0.80$), in contrast to the negligible changes in the CG.

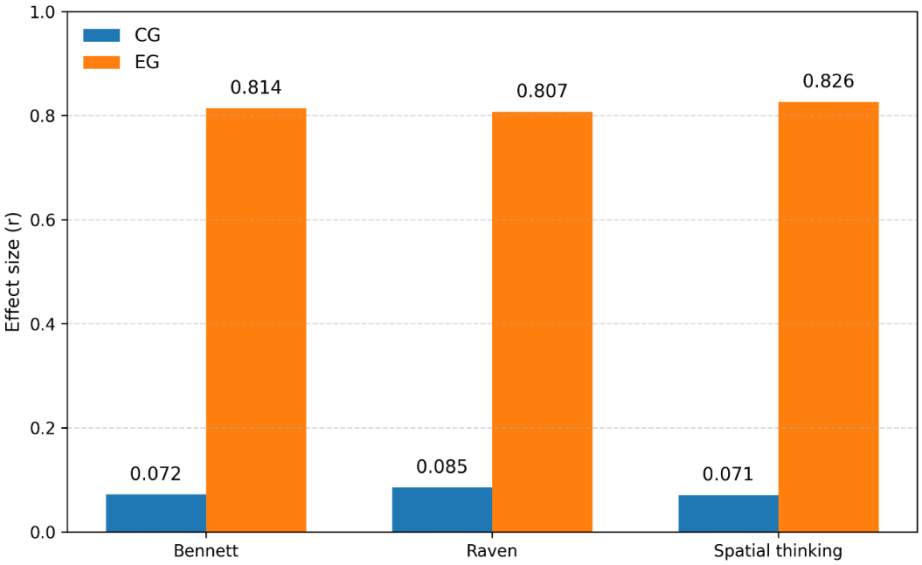

Figure 2. Effect sizes (r) for within-group pre→post changes.

Note: Effect sizes were computed using the Wilcoxon signed-rank test. The values on the bars indicate the effect size r magnitudes. The experimental group shows large effects across all tests, whereas the control group shows negligible effects across all tests. The y-axis starts at zero to ensure an accurate visual interpretation.

Additional analysis using the Kruskal–Wallis test showed no statistically significant differences in the results between students of different grades (1–4) either at the pre- or post-stage ($p > 0.05$ for all tests, Table 5). Effect size estimates (η^2) were negligible in all comparisons ($\eta^2 \approx 0$), indicating that grade level did not meaningfully influence the test performance. This suggests a comparable pattern of results across age groups within primary school; however, the formal equivalence of effects was not directly tested, and this should be interpreted with caution.

Table 5. Results of the Kruskal-Wallis test for assessing the influence of the "grade" factor.

Test	Stage	H	p	η^2	Interpretation
Bennett	Pre	2.64	0.45	0.000	No differences
	Post	1.37	0.713	0.000	No differences
Raven	Pre	1.88	0.597	0.000	No differences
	Post	1.26	0.739	0.000	No differences
Spatial thinking	Pre	1.93	0.586	0.000	No differences
	Post	0.13	0.988	0.000	No differences

Note: H — Kruskal-Wallis H statistic; p — significance level; η^2 — eta-squared effect size measure.

4. Discussion

We aimed to comprehensively evaluate the effectiveness of a specialized STEAM program utilizing 2D and 3D modeling to develop key domains of thinking in primary school students, namely technical, logical, and spatial thinking. The results support the main hypothesis and provide a deeper understanding of the mechanisms through which the integrated approach influences students' technical, logical, and spatial thinking and identifies the critical conditions for its successful

implementation.

The statistically significant superiority of the EG over the CG in the post-test, supported by large effect sizes (r = 0.72–0.76), indicates a strong association between the application of the STEAM approach and the development of the targeted types of thinking skills.

The large within-group effect sizes observed in the EG should also be interpreted in relation to the alignment between the intervention activities and the measured cognitive domains of interest. Although the STEAM tasks were not designed to directly replicate the diagnostic test items, the intervention and assessments involved overlapping cognitive processes, such as spatial manipulation, mechanical reasoning, and pattern recognition. Therefore, part of the observed effect may reflect domain-specific training sensitivity rather than broad transfer.

The contrast between the substantial progress in the EG and the minimal changes observed in the CG suggests the effectiveness of the intervention under the given conditions.

This contrast should also be interpreted in light of the differential sensitivity of the instructional approaches involved, where structured STEAM-based activities may produce more measurable gains in the assessed domains than traditional instruction over the same period.

Furthermore, these effects should be interpreted as resulting from the combined influence of the instructional approach and its pedagogical implementation, rather than from the intervention content alone.

In addition, the use of adapted and shortened diagnostic instruments should be considered when interpreting the magnitude of the observed effects, as the measurement sensitivity and reliability of such tools may differ from those of fully standardized psychometric tests.

In addition, students' prior familiarity with digital tools and interfaces may have influenced their engagement with the modeling tasks, particularly in activities involving spatial reasoning and virtual object manipulation. This factor was not explicitly measured in this study and should be considered in future research.

In addition, the analytical approach relied on complementary within- and between-group comparisons rather than a unified model of change, which should be considered when interpreting the magnitude and generalizability of the observed effects.

This is consistent with the structure of the intervention, which relied on the systematic use of 2D and 3D modeling tools that required students to constantly perform mental rotations, projections, and object manipulations. This result aligns with research demonstrating the plasticity of spatial thinking in primary school-age children and its high sensitivity to targeted interventions [19]. Furthermore, it finds methodological support in studies that also showed that the development of complex cognitive structures (in their case, spatial thinking) requires the application of specific tools and progresses unevenly, depending on the design of the learning activities.

The observed improvement in logical thinking can be explained by the structure of the project-based activities, which require students to systematically analyze task conditions, identify patterns, and construct causal relationships.

Engagement in modeling tasks and engineering problem-solving promotes the transition from concrete actions to more abstract reasoning, thereby supporting the development of analytical and nonverbal logical thinking, as measured by Raven's Progressive Matrices.

Qualitative analysis of the implementation process suggests that the teacher's role may be an important mediating factor in the observed outcomes. In particular, differences in teacher preparation, familiarity with STEAM pedagogy, instructional practices, and classroom-level conditions may have

contributed to the magnitude of observed effects.

Observations indicate that more active facilitation, including guidance, support of student autonomy, and structured organization of project-based activities, may enhance learning outcomes compared to procedural or instruction-driven approaches.

This directly relates to the thesis of Herro et al. [26] and Ortiz-Revilla et al. [18] regarding the need to shift the pedagogical role for STEAM success, which is strongly supported by empirical evidence from related fields. Researchers investigating the effectiveness of didactic games for developing spatial thinking have found that the effectiveness of a methodology directly depends on the teacher's readiness to organize activities rather than transmit knowledge. Their conclusion that "the teacher becomes the organizer and consultant of cognitive activity" aligns with our concept of a "facilitator" and confirms that investments in teachers' professional development are not an additional option but a necessary condition for the successful implementation of STEAM.

On an operational level, our study offers a way to overcome the methodological gap caused by the vague definition of STEAM. We avoided the lack of operational transparency through a detailed description of the independent variables: specific pedagogical practices (maintaining engineering journals, design thinking, and project defenses) [20]. This approach to strict operationalization, similar to the criterion-referenced framework for assessing spatial thinking [27], enables a clear demonstration of which specific practices led to which results, contributing to the refinement and operationalization of the STEAM concept.

This approach reflects a pragmatic balance between methodological rigor and developmental feasibility, which is particularly important in research involving primary school populations in the Netherlands.

An important observation was the minimal change identified in the CG over the study duration.

This pattern may indicate that traditional instructional approaches have a limited impact on the development of the targeted types of thinking (technical, logical, and spatial) within the same time frame, particularly when these skills are not explicitly addressed in the curriculum.

Furthermore, this result may also be influenced by the characteristics of the measurement instruments and the applied normalization procedure. The use of proportional scores and cross-grade comparable scales may reduce the visibility of gradual developmental changes, especially when such changes are relatively small or nonlinear.

In addition, the diagnostic instruments used in this study may be more sensitive to structured, model-based learning gains than to incremental changes associated with routine classroom instruction.

These factors should be considered when interpreting the observed lack of significant dynamics in the CG and require further investigation in future studies.

Despite these convincing results, this study has several limitations.

First, a quasi-experimental design with class-level group allocation was used, which does not exclude the influence of unobserved variables on the results. Although a matching procedure based on academic performance indicators was applied, other potentially relevant factors (e.g., prior cognitive development, motivation, or socio-economic background) were not controlled.

Second, the statistical approach did not include direct modeling of the interaction effects between group and time, which limits the ability to formally test differential change between groups.

Third, the data were collected within intact classes and implemented by a limited number of teachers, which may have introduced a degree of clustering at the class or teacher level. This was not

explicitly modeled and may have influenced the independence of the observations and the precision of the statistical estimates.

Fourth, although the instruments were carefully adapted using a structured and theoretically grounded approach, formal psychometric validation (e.g., reliability coefficients such as Cronbach's alpha) was not conducted for the adapted version.

This limitation is related to the use of shortened and developmentally adapted test forms, which were designed to ensure age-appropriateness and feasibility within the educational setting. Because these adapted forms were not administered as fully standardized psychometric scales, reliability coefficients such as Cronbach's alpha or McDonald's omega could not be robustly estimated in this study. In future studies, researchers should include full-scale psychometric validation of the adapted instruments, including reliability and internal consistency analyses.

Fifth, the sample was limited to Southern Kazakhstan, requiring caution when generalizing the conclusions to other cultural and educational contexts in the future.

The specific characteristics of the regional educational system, including curriculum structure, institutional organization, and the level of integration of digital and STEAM-based practices, may have influenced the baseline performance and responsiveness to the intervention. In Southern Kazakhstan, where STEM- and STEAM-oriented practices are gradually being integrated into school education, students' prior exposure to modeling-based and engineering-oriented tasks may vary considerably across schools. This context should be considered when interpreting the magnitude of the observed gains and when transferring the program to other educational systems in the future.

Despite efforts to standardize the program, the personal and professional characteristics of the teachers involved in the experiment may have acted as confounding variables.

In particular, the EG teachers received specialized STEAM-oriented training, whereas comparable training conditions for the CG were not explicitly controlled in the study design. Therefore, some of the observed effects may be associated not only with the intervention content but also with differences in teacher preparation, pedagogical style, and level of instructional engagement.

Additionally, the limited observable change in the CG may be partly related to the sensitivity of the measurement instruments and normalization procedure, which may not fully capture gradual developmental progress in the absence of targeted instructional interventions.

The theoretical significance of this study lies in the empirical confirmation of the link between specific STEAM activities (2D/3D modeling) and the development of specific, measurable components of thinking, as well as in identifying the critical role of the pedagogical position in this process. From a practical standpoint, the results justify the feasibility of broadly implementing similar programs in the future in the same way.

However, the successful implementation of such approaches depends on the availability of appropriate technical infrastructure, including access to digital devices, stable software environments, and institutional support for STEAM-oriented instruction.

The key success conditions are as follows: 1) systematic sessions; 2) appropriate training for teachers to work in the new role of facilitator; and 3) the use of clearly described and reproducible pedagogical practices, ensuring the program's scalability [28,29].

For future research, we recommend: 1) Conducting a longitudinal study to test the sustainability of the effects; 2) using cluster randomization, involving more teachers and schools to mitigate the "teacher effect"; and 3) supplementing quantitative methods with qualitative interviews for a better understanding of students' perceptions of the STEAM environment.

5. Conclusions

In this study, we aimed to achieve our objective of evaluating the impact of a STEAM-oriented program utilizing 2D and 3D modeling on the development of technical, logical, and spatial thinking in primary school students. Based on this analysis, the following conclusions can be drawn:

– Hypothesis Confirmation: The implementation of a structured STEAM program based on the systematic use of 2D and 3D modeling was associated with statistically significant and practically important improvements in technical, logical, and spatial thinking among students in the EG compared to the CG.

– Intervention Effectiveness: The effect size of the STEAM program implementation was large ($r > 0.72$ for intergroup comparisons), indicating its high practical significance and effectiveness in developing technical, logical, and spatial thinking in students.

– Approach Universality: The results showed a consistent pattern across students in grades 1–4, suggesting the program's potential age-related adaptability within primary education.

– Contribution to Theory and Practice: The results provide empirical support for integrating STEAM-based modeling activities into primary education and clarify their association with the development of technical, logical, and spatial thinking. A practical outcome of this study is the development and testing of a lesson model that may be considered for implementation in educational practice, subject to contextual adaptation.

The prospects for further research are as follows.

– A longitudinal study is recommended to assess the sustainability of the observed effects.

– Expanding the sample size and conducting randomized controlled trials (RCTs) in various sociocultural contexts is recommended.

– A deeper study of the program's influence on students' motivation and academic performance in related disciplines is needed.

– Investigating the role of digital and analog modeling tools in achieving educational outcomes.

Thus, this study makes a significant contribution to the justification and development of effective pedagogical strategies for fostering critically important forms of thinking, including technical, logical, and spatial thinking, among primary-school students.

Author contributions

Guldana A. Totikova: development of the general idea and goals of the research, development of the methodological framework for the study, leadership and coordination of the research activities, preparation of the initial draft of the manuscript. Aidarbek A. Yessaliyev: development and validation of psycho-pedagogical research methods, conducting research activities, data analysis, verification and confirmation of the obtained results, application of statistical and qualitative methods for data analysis, revision and editing of the text. Nurgul N. Medetbekova: organization and conduct of the empirical research, data collection, management, annotation, and maintenance of research data integrity, provision of necessary materials and tools for the research, participation in revising the manuscript. Laura T. Iskakova**:** development of pedagogical methodology and design of educational programs, ensuring access to educational platforms and institutional resources, significant contribution to the critical analysis and approval of the final version of the manuscript. Zhanar Sh.

Zhiyasheva: creation and preparation of graphs, diagrams, and illustrative materials, participation in the practical implementation of research methods, data collection, participation in the revision and preparation of the manuscript for publication.

All authors have read and approved the final version of the manuscript.

Use of Generative-AI tools declaration

The authors used AI-based tools (ChatGPT) for language editing and text refinement during the preparation of this manuscript.

Acknowledgements

The authors express their gratitude to the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan and the research teams of M. Auezov South Kazakhstan State University for their support in collecting and analyzing the data.

Conflict of interest

The authors declare that there are no actual or potential conflicts of interest, financial or otherwise, that could have inappropriately influenced the results or interpretation of this study. The study was conducted impartially, with no affiliations that could have biased the results.

Ethics declaration

This study was conducted in accordance with the institutional and national ethical guidelines for educational research involving minors. The research procedures were noninvasive and conducted within the standard educational process. Informed consent was obtained from the parents/legal guardians and school administration. Student participation was voluntary, and the data were anonymized.

According to institutional regulations, separate approval from an ethics committee was not required for this type of educational research conducted within the regular school curriculum.

The study involved a quasi-experimental design implemented in a natural classroom setting without invasive procedures or deviations from standard educational practices.

References

1. Belbase, S., Mainali, B., Kasemsukpipat, W., Tairab, H., Gochoo, M. and Jarrah, A.M., At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education: Prospects, priorities, processes, and problems. *International Journal of Mathematical Education in Science and Technology*, 2021, 53(11): 2919–2955. https://doi.org/10.1080/0020739x.2021.1922943
2. Taylor, J. and Hwang, J., Science, Technology, Engineering, Arts, and Mathematics remote instruction for students with disabilities. *Intervention in school and clinic*, 2021, 57(2): 111–118. https://doi.org/10.1177/10534512211001858
3. Bassachs, M., Ortí, D.C., Noguí, L., Serra, T., Bubnys, R. and Colomer, J., Fostering critical reflection in primary education through STEAM approaches. *Education Sciences*, 2020, 10(12): 384. https://doi.org/10.3390/educsci10120384

4. Panfilova, A. and Larchenko, L., Innovative tools for developing soft skills in the context of educational transformation. *SHS Web of Conferences*, 2021, 97: 01024. https://doi.org/10.1051/shsconf/20219701024
5. Istianah, F., Importance of STEAM Learning Implementing in Elementary Schools. *KnE Social Sciences*, 2023, 76–84. https://doi.org/10.18502/kss.v8i8.13287
6. Başaran, M. and Erol, M., Recognizing aesthetics in nature through STEM and STEAM education. *Research in Science & Technological Education*, 2021, 41(1): 326–342. https://doi.org/10.1080/02635143.2021.1908248
7. Perignat, E. and Katz-Buonincontro, J., STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 2019, 31: 31–43. https://doi.org/10.1016/j.tsc.2018.10.002
8. Yim, I.H.Y., Su, J. and Wegerif, R., STEAM in practice and research in primary schools: a systematic literature review. *Research in Science & Technological Education*, 2024, 43(4): 1065–1089. https://doi.org/10.1080/02635143.2024.2440424
9. Ortiz-Revilla, J., Greca, I.M. and Villagrá, J.Á.M., Effects of an integrated STEAM approach on the development of competence in primary education students (Efectos de una propuesta STEAM integrada en el desarrollo competencial del alumnado de Educación Primaria). *Journal for the Study of Education and Development Infancia y Aprendizaje*, 2021, 44(4), 838–870. https://doi.org/10.1080/02103702.2021.1925473
10. Totikova, G., Yessaliyev, A., Madiyarov, N. and Medetbekova, N., Effectiveness of the Development of spatial thinking in junior class schoolchildren by application of plane and spatial modeling of geometric figures in didactic games. *European Journal of Contemporary Education*, 2020, 9(4): 902–914. https://doi.org/10.13187/ejced.2020.4.902
11. Marín, J.A.M., Guerrero, A.J.M., Dúo-Terrón, P. and Belmonte, J.L., STEAM in education: a bibliometric analysis of performance and co-words in Web of Science (Review of STEAM in education: a bibliometric analysis of performance and co-words in Web of Science). *International Journal of STEM Education*, 2021, 8(1): 41. Springer Science+Business Media. https://doi.org/10.1186/s40594-021-00296-x
12. Leavy, A., Dick, L.K., Meletiou-Mavrotheris, M., Paparistodemou, E. and Stylianou, E., The prevalence and use of emerging technologies in STEAM education: A systematic Review of the literature (A systematic review of the prevalence and use of emerging technologies in STEAM education: A systematic review of the literature). *Journal of Computer Assisted Learning*, 2023, 39(4): 1061–1082. Wiley. https://doi.org/10.1111/jcal.12806
13. Bedar, R.W.A.H. and Al-Shboul, M., The Effect of Using the STEAM Approach on Motivation Towards Learning Among High School Students in Jordan. *International Education Studies*, 2020, 13(9): 48–57. https://doi.org/10.5539/ies.v13n9p48
14. Clements, D.H., Vinh, M., Lim, C.I. and Sarama, J., STEM for inclusive excellence and equity. *Early education and development*, 2020, 32(1): 148–171. https://doi.org/10.1080/10409289.2020.1755776
15. Ahmad, D.N., Astriani, M.M., Alfahnum, M. and Setyowati, L., Increasing students' creative thinking through learning organization with STEAM education. *Jurnal Pendidikan IPA Indonesia*, 2021, 10(1): 103. https://doi.org/10.15294/jpii.v10i1.27146
16. Taylor, P., Enriching STEM with the arts to better prepare 21st-century citizens. *AIP Conference*

Proceedings, 2018, 1923: 020002. https://doi.org/10.1063/1.5019491

17. Li, C., Wang, M., Chen, Y., Niu, W., Hong, M. and Zhu, Y., Design my music instrument: A project-based Science, Technology, Engineering, Arts, and Mathematics program for the development of creativity. *Frontiers in Psychology*, 2022, 12: 763948. https://doi.org/10.3389/fpsyg.2021.763948
18. Ortiz-Revilla, J., Ruiz-Mart ń, Á. and Greca, I.M., Conceptions and attitudes of pre-school and primary school teachers towards STEAM education in Spain. *Education Sciences*, 2023, 13(4): 377. https://doi.org/10.3390/educsci13040377
19. Kaewkamnerd, S. and Suwannarat, A., Enhancing deaf students' computational thinking using the STEAM approach. *Education Sciences*, 2025, 15(5): 627. https://doi.org/10.3390/educsci15050627
20. Perignat, E. and Katz-Buonincontro, J., STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 2019, 31: 31–43. https://doi.org/10.1016/j.tsc.2018.10.002
21. Yim, I.H.Y., Su, J. and Wegerif, R., STEAM in practice and research in primary schools: a systematic literature review. *Research in Science & Technological Education*, 2025, 43(4): 1065–1089. https://doi.org/10.1080/02635143.2024.2440424
22. Litovchin, Y.M., Avilova, N.L., Podvoiska, I.A., Valeyev, A.S., Yesina, E.A., Gataullina, R.V., et al., The Development of students' key professional competencies in the process of didactic tasks realization. *Journal of Sustainable Development*, 2015, 8(3): 285. https://doi.org/10.5539/jsd.v8n3p285
23. Rahmawati, Y., Ridwan, A., Hadinugrahaningsih, T. and Soeprijanto, S., Developing critical and creative thinking skills through STEAM integration in chemistry learning. *Journal of Physics Conference Series*, 2019, 1156: 012033. https://doi.org/10.1088/1742-6596/1156/1/012033
24. Anisimova, T., Sabirova, F. and Shatunova, O., Formation of design and research competencies in future teachers within the framework of STEAM education. *International Journal of Emerging Technologies in Learning (iJET)*, 2020, 15(2), 204–217. https://doi.org/10.3991/ijet.v15i02.11537
25. Liu, L., Instructional design of mathematical modeling in high school under the STEAM concept. *Journal of Education and Development*, 2022, 6(2), 17–23. https://doi.org/10.20849/jed.v6i2.1174
26. Herro, D., Quigley, C. and Jacques, L., Examining technology integration in middle school STEAM units. *Technology Pedagogy and Education*, 2018, 27(4): 485–498. https://doi.org/10.1080/1475939x.2018.1514322
27. Totikova G., Aldabergenov N., Salmirza J., Gurbanova, A., Madiyarov, N. and Yessaliyev, A., Criteria-based assessment of spatial representations in primary school students. *Ilkogretim Online*, 2019. 18(2): 796–809.
28. Juškevičienė, A., Dagienė, V. and Dolgopolovas, V., Integrated activities in STEM environment: Methodology and implementation practice. *Computer Applications in Engineering Education*, 2021, 29(1), 209–228. https://doi.org/10.1002/cae.22324
29. Spyropoulou, N. and Kameas, A., Augmenting the Impact of STEAM education by developing a competence framework for effective teaching and learning. *Education Sciences*, 2023, 14(1): 25. https://doi.org/10.3390/educsci14010025

Author's biography

Guldana A. Totikova – PhD, Associate professor at the Central Asian Innovation University. She has been conducting research in the field of STEAM education for primary school students since 2016. She leads a scientific project aimed at developing technical thinking in primary school students through the use of 2D and 3D modeling technologies. She is the author of a number of educational programs, teaching aids, and guidelines. Holds numerous certificates of registration for intellectual property objects.

Aidarbek A. Yessaliyev – Doctor of medical sciences, Professor at M. Auezov South Kazakhstan University. Since 2014, his area of scientific interest has included the formation of spatial concepts and the development of cognitive abilities in primary school children. He is a leading expert in the psychological and pedagogical support of STEAM interventions in the educational process.

Nurgul N. Medetbekova – Candidate of pedagogical sciences, Associate Professor at M. Auezov South Kazakhstan University. A leading specialist in the field of Pedagogy and Psychology. Her research area covers problems of intellectual development in early childhood, developmental psychology and interpersonal relationships, as well as the formation of spatial thinking in children and adolescents. For 14 years, she has been studying the development of spatial concepts in primary school students within the context of the STEAM approach.

Laura T. Iskakova – Doctor of pedagogical sciences, Professor at Academician A. Kuatbekov People's Friendship University. A leading specialist in the field of Pedagogy and Psychology. For 12 years, she headed an institute for advanced training of teaching staff. She is engaged in developing educational programs for integrating robotics and STEAM learning into the secondary school curriculum.

Zhanar Sh. Zhiyasheva – PhD, Associate professor at M. Auezov South Kazakhstan University. Her scientific and pedagogical interests focus on the methodology of teaching mathematics in secondary school using the principles and tools of STEAM education.

Appendix

Table S1. Mann–Whitney between-group comparisons: U, Z, p, r (Z/√N), rank-biserial r_rb, and common language effect size (CLES).

Test	Stage	n1 (EG)	n2 (CG)	U	Z	p (two-tailed)	r (Z/√N)	rank-biserial r_rb	CLES (EG > CG)
Bennett	Pre	85	87	3903	0.630	0.529	0.048	-0.056	0.472
	Post	85	87	6785	9.456	< 0.001	0.721	0.835	0.918
Raven	Pre	85	87	4091	1.207	0.228	0.092	-0.107	0.447
	Post	85	87	6879	9.744	< 0.001	0.743	0.860	0.930
Spatial thinking	Pre	85	87	4074	1.154	0.248	0.088	-0.102	0.449
	Post	85	87	6948	9.954	< 0.001	0.759	0.879	0.940

Note: The direction of rank-biserial correlation (r_rb) and CLES depends on the ordering of groups. In the present table, CLES is reported as the probability of superiority of the experimental group over the control group (EG > CG). Values of CLES greater than 0.5 indicate higher scores in the experimental group, whereas values below 0.5 indicate higher scores in the control group.

Table S2. Wilcoxon within-group changes: Z, exact p, r (Z/√N), rank-biserial r_rb, and common language effect size (CLES).

Test	Group	N pairs	Z	p (two-tailed)	r (Z/√N)	W+	W−	rank-biserial r_rb	CLES (Post>Pre)
Bennett	Control	87	0.672	0.502	0.072	2073	1755	0.083	0.542
	Experimental	85	7.505	< 0.001	0.814	3540	115	0.937	0.969
Raven	Control	87	0.793	0.428	0.085	2101	1727	0.098	0.549
	Experimental	85	7.440	< 0.001	0.807	3525	130	0.929	0.964
Spatial thinking	Control	87	0.662	0.508	0.071	2070	1758	0.082	0.541
	Experimental	85	7.615	< 0.001	0.826	3565	90	0.951	0.975

Note: For paired comparisons, rank-biserial correlation (r_rb) was computed from Wilcoxon signed-rank statistics as (W+ − W−) / [n(n+1)/2]. CLES is reported as the probability that post-test scores exceed pre-test scores (Post > Pre). Values greater than 0.5 indicate improvement from pre- to post-test.