

STEM-based Static Fluid Learning: A quasi-experimental comparison of students' 4C skills and learning outcomes

Mardia Hi Rahman¹, Sukarsi Talib², Fatma Hamid¹, Hutri Handayani Isra¹, Asmita Basir¹

¹Physics Education Study Program, Faculty of Teacher Training and Education, Universitas Khairun, Ternate, Indonesia

²SMA Negeri 7 Halmahera Selatan, South Halmahera, Indonesia

Email: mardiah.rahman1@gmail.com

ARTICLE INFO

Keywords:

4C skills;
Learning outcomes;
Physics instruction;
STEM;
21st-century skills

Article history:

Received 2026-08-17
Revised 2026-09-26
Accepted 2026-10-27

ABSTRACT

This study aimed to analyse differences in 21st-century skills (4C) and physics learning outcomes between students receiving Science, Technology, Engineering, and Mathematics (STEM)-based instruction and those receiving conventional instruction. A quantitative approach with a quasi-experimental, nonequivalent control group design involved 60 Grade XI students at SMA Negeri 7 Halmahera Selatan in the 2026/2027 academic year, comprising two intact classes of 30 students each selected through purposive sampling. Critical thinking, creativity, communication, and collaboration were observed by two observers over three instructional sessions using a structured observation instrument, and learning outcomes on static fluids were measured with a pretest and a posttest. Data were analysed using descriptive statistics, independent-samples t-tests, normalised gain (N-gain), and analysis of covariance (ANCOVA). The experimental group achieved a higher mean overall 4C score ($M = 89.15$, $SD = 2.52$) than the control group ($M = 73.82$, $SD = 2.48$), $t(58) = 23.760$, $p < .001$, and a higher posttest mean ($M = 84.73$, $SD = 3.99$) than the control group ($M = 70.07$, $SD = 3.10$). The difference in posttest scores remained significant after controlling for pretest scores, $F(1, 57) = 249.77$, $p < .001$. N-gain values of 0.70 and 0.401, respectively, indicated greater improvement in the experimental group. These findings suggest that STEM-based physics instruction was associated with higher 4C scores and better learning outcomes, although the quasi-experimental design limits causal inference.

This is an open access article under the [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



Corresponding Author:

Mardia Hi Rahman
Physics Education Study Program, Faculty of Teacher Training and Education, Universitas Khairun,
Ternate; mardiah.rahman1@gmail.com

INTRODUCTION

Paradigms in modern education are increasingly shifting from students' mastery of subject content toward problem-solving skills needed to address increasingly complex and diverse changes and challenges. At the senior high school level, information-processing skills, ideation skills, teamwork skills, and data-driven decision-making skills are essential competencies that need to be developed. These competencies are reflected in 21st-century skills, which encompass critical thinking, creativity, communication, and collaboration (the 4Cs), widely regarded as core components of 21st-century learning. Therefore, meaningful learning experiences need to be designed to help students apply the skills they acquire in relevant contexts (Kain et al., 2024; Lavi et al., 2021)

Physics knowledge is closely related to phenomena encountered in everyday life. Physics learning should not be limited to students' ability to manipulate mathematical equations; rather, it should provide opportunities for them to interpret natural phenomena, identify relationships between variables, apply reasoning, and construct arguments. In this way, physics knowledge can serve as a tool for explaining phenomena rather than merely as a collection of mathematical formulas. Physics instruction should therefore provide students with opportunities to learn in ways that foster conceptual understanding and enable them to apply their knowledge to problems requiring logical reasoning (Bao & Koenig, 2019).

One instructional approach that can provide such learning contexts is STEM education, which integrates Science, Technology, Engineering, and Mathematics. STEM-based learning connects knowledge from different disciplines through authentic problems and problem-solving activities, which are characteristic features of STEM education. Students can engage in STEM learning to identify problems, analyze information, develop alternative solutions, test ideas, refine solutions, and communicate their findings (Kelley & Knowles, 2016). STEM integration also requires meaningful connections between disciplinary concepts and learning activities. Thus, STEM should not be reduced to merely adding technology or engineering activities to instruction (Roehrig et al., 2021).

Previous studies have reported positive effects of STEM-based learning on the development of 21st-century skills. Damayanti et al., (2023) reported that the use of STEM-based physics modules could improve students' critical thinking skills. Similarly, Asrizal et al., (2024) found that the STEM-Smart Physics E-Module supported conceptual understanding and several components of 21st-century skills, particularly critical thinking, creative thinking, and communication. Yulianti et al., (2022) also reported that STEM-based physics learning supported by Scratch could be implemented to develop 21st-century learning skills. Furthermore, Wulandari & Pujianto, (2025) concluded that integrating STEM into physics instruction was associated with the development of critical thinking, creativity, communication, and collaboration.

The development of 21st-century skills does not necessarily follow a uniform pattern. Each skill may respond differently to learning experiences. Students may demonstrate strong collaboration skills while still needing further support in analyzing information or evaluating alternative solutions. Similarly, students may communicate actively without necessarily being able to generate diverse ideas. Therefore, analyzing each 21st-century skill separately is necessary to obtain a more accurate understanding of its development. Research on STEM implementation also indicates that adopting a STEM approach does not automatically lead to

improvements in all 21st-century skills. The quality of integration, activity design, teachers' understanding, and opportunities for students to engage in problem-solving activities are important factors in STEM implementation (Margot & Kettler, 2019; Roehrig et al., 2021). Dare et al., (2021) likewise emphasized the importance of real-world problems and 21st-century skills in conceptualizing integrated STEM education.

Research on 21st-century skills in physics learning needs to consider not only students' achievements at the end of instruction but also the patterns of skill development throughout the learning process. End-of-instruction assessments provide information about students' achievement levels, whereas repeated observations can reveal changes in critical thinking, creativity, communication, and collaboration from one lesson to the next. In STEM-based physics learning research, these aspects warrant investigation to determine whether the development patterns of the 4Cs differ between students receiving STEM-based instruction and those receiving conventional instruction. Accordingly, research combining observations of 4C skills across several lessons with measurements of learning outcomes before and after instruction is needed. Based on these considerations, this study compared the development patterns of the 4Cs observed over three instructional sessions and learning outcomes in static fluids between students participating in STEM-based learning and those receiving conventional instruction in Grade XI at SMA Negeri 7 Halmahera Selatan.

This study aimed to analyze: (1) differences in the development of 4C skills between students participating in STEM-based physics learning and those receiving conventional instruction over three sessions; (2) which 4C skill dimension showed the greatest improvement in each group over the three sessions; and (3) differences in physics learning outcomes between students participating in STEM-based learning and those receiving conventional instruction.

RESEARCH METHODS

This study employed a quantitative approach with a quasi-experimental, nonequivalent control group design (pretest–posttest with intact classes) and was conducted at SMA Negeri 7 Halmahera Selatan during the 2026/2027 academic year. The study population comprised 165 Grade XI students enrolled in the science specialization across five classes. The sample was selected using purposive sampling and consisted of Class Peminatan IPA Fisika-1, comprising 30 students (14 male and 16 female), and Class Peminatan IPA Fisika-2, also comprising 30 students (12 male and 18 female). The two classes were selected because they had been established before the study was conducted. Based on students' daily test scores on the previously taught material, the mean learning outcomes for Class Peminatan IPA Fisika-1 and Class Peminatan IPA Fisika-2 were 72.5 and 71.9, respectively. Both mean scores exceeded the school's established Criteria for Learning Objective Attainment, which was set at 70. This information was one of the considerations in selecting the two classes as research participants, while acknowledging the limitations of a quasi-experimental design that does not involve individual student randomization. Furthermore, both classes had the same number of students and studied static fluids, making them suitable for comparing learning outcomes under different instructional conditions. Class Peminatan IPA Fisika-1 was designated as the experimental group and received instruction based on Science, Technology,

Engineering, and Mathematics (STEM), whereas Class Peminatan IPA Fisika-2 served as the control group and received conventional instruction.

The experimental group received integrated STEM instruction implemented through Project-Based Learning (STEM-PjBL), in which students completed a project designed to address a problem involving static fluids. The learning process followed six stages of PjBL: formulating fundamental questions, designing the project, developing a schedule, monitoring project implementation, testing the results, and evaluating and disseminating the project outcomes. STEM integration involved applying physics concepts, using technology, designing and testing solutions, and conducting quantitative measurements and analyses. In contrast, the control group received conventional instruction involving teacher explanations, the presentation of example problems, question-and-answer sessions, and discussions. The same teacher taught both groups. Students' 4C skills, namely critical thinking, creativity, communication, and collaboration, were assessed by two observers using the same observation instrument during each instructional session.

Students' 21st-century skills were measured using a structured observation instrument comprising four dimensions: critical thinking, creativity, communication, and collaboration. Each dimension consisted of five indicators, yielding a total of 20 indicators. Each indicator was rated on a four-point scale (1–4). A score of 4 indicated comprehensive and consistent performance, a score of 3 indicated good performance with minor shortcomings, a score of 2 indicated partial performance with several shortcomings, and a score of 1 indicated very limited performance or the absence of observable performance. The content validity of the questionnaire was assessed by two experts using Aiken's V index on a four-point scale (1–4). The Aiken's V values ranged from 0.667 to 1.000, with a mean of 0.850. All items were deemed acceptable and suitable for use in the study. The maximum possible score was 80, which was converted to a 0–100 scale using the following formula:

$$Score_{0-100} = \frac{\text{Obtained Score}}{\text{maximum Score}} \times 100 \quad (1)$$

Students' learning outcomes in static fluids were measured using seven essay questions that met the validity and reliability criteria. The test was administered before and after instruction (pretest and posttest) to assess conceptual mastery at cognitive levels C3–C6. The assessed topics included hydrostatic pressure, Pascal's law, hydraulic devices, Archimedes' principle, and buoyant force. Before implementation, the original 10-item instrument was piloted with 22 Grade XII students. Analysis of the pilot data using SPSS version 24 indicated that seven items met the validity criteria, whereas the remaining three items were excluded. The item validity analysis is presented in Table 1

Table 1. Results of the Item Validity Analysis

Item	r _{calculated}	Sig.	r _{table}	Decision
1	0.457	0.032	0.423	Valid
2	0.743	0.000	0.423	Valid
3	0.809	0.000	0.423	Valid
4	0.233	0.297	0.423	invalid
5	0.743	0.000	0.423	Valid
6	0.140	0.535	0.423	invalid
7	0.743	0.000	0.423	Valid
8	0.452	0.035	0.423	Valid

9	0.331	0.133	0.423	invalid
10	0.809	0.000	0.423	Valid

The reliability analysis yielded a coefficient of 0.763, indicating adequate internal consistency of the instrument. Inter-observer reliability was assessed using the unweighted Cohen's Kappa coefficient based on ratings provided by two observers for a subsample of 30 students, with each student serving as one unit of analysis. The analysis yielded a Kappa coefficient of 0.808, indicating substantial agreement between the two observers..

Learning outcome data were collected through a written test consisting of seven essay questions on static fluids, administered before and after instruction (pretest and posttest) to measure students' learning achievement. Each item had a maximum score of 10, yielding a total score ranging from 0 to 100. The scoring criteria covered conceptual accuracy, appropriate formula application, calculation procedures, correct use of units, and the consistency of conclusions with the questions posed. Pretest and posttest scores were used to describe changes in students' learning outcomes.

Learning outcome data were analyzed descriptively and inferentially using SPSS version 24. Means and standard deviations were calculated to summarize physics learning outcomes and the four dimensions of 21st-century skills. Data normality was assessed using the Shapiro-Wilk test, while homogeneity of variance was examined using Levene's test (Field, 2018). Independent-samples t-tests were subsequently conducted to compare pretest and posttest scores between the experimental and control groups. Analysis of covariance (ANCOVA) was also performed, with pretest scores as the covariate and instructional group as the independent variable, to examine differences in posttest scores after controlling for students' initial abilities (Field, 2018; Sudjana, 2002). Learning gains were calculated using the normalized gain (N-gain) formula:

$$\left[g = \frac{Posttest - Pretest}{Maximum\ Score - Pretest} \right] \quad (2)$$

Cohen's d was calculated to estimate the magnitude of between-group differences using a standardized effect size. Reporting effect sizes alongside statistical significance provides information about the practical magnitude of the observed differences (Field, 2018). The analysis of 21st-century skills included descriptive statistics and independent-samples t-tests for critical thinking, creativity, communication, and collaboration. Cohen's d was calculated for each dimension to estimate the magnitude of the differences between groups. Thus, the analysis covered two aspects: physics learning outcomes based on pretest and posttest scores, and 4C skills based on structured observation results.

RESULTS AND DISCUSSION

Results

The observations of students' 21st-century skills, comprising critical thinking, creativity, communication, and collaboration (4C), were conducted over three learning sessions. The mean observation scores for both groups are presented in Table 2.

Table 2. Mean Scores of Students' 21st-Century Skills

<i>Skill</i>	<i>Class</i>	<i>Session 1</i>	<i>Session 2</i>	<i>Session 3</i>	<i>Improvement</i>
Critical thinking	Experimental	83.33	89.00	91.67	8.34

Creativity	Control	69.67	70.33	76.17	6.50
	Experimental	84.50	89.00	91.83	7.33
Communication	Control	68.67	73.50	75.33	6.66
	Experimental	85.17	89.83	92.00	6.83
Collaboration	Control	69.67	74.50	77.33	7.66
	Experimental	87.50	92.50	94.00	6.50
	Control	70.00	77.83	82.50	12.50

The research data were subsequently analyzed descriptively and presented in tabular form to provide an overview of students' 21st-century skill (4C) scores in the experimental and control classes throughout the learning process.

Table 3. Descriptive Statistics of 21st-Century Skills

Dimension	Experimental Group Mean $\pm$ SD	Control Group Mean $\pm$ SD	Mean Difference
<i>Critical thinking</i>	88.00 $\pm$ 3.32	72.07 $\pm$ 3.30	15.93
<i>Creativity</i>	88.07 $\pm$ 2.95	72.67 $\pm$ 2.75	15.40
<i>Communication</i>	88.93 $\pm$ 2.96	73.87 $\pm$ 3.19	15.07
<i>Collaboration</i>	91.60 $\pm$ 2.70	76.67 $\pm$ 2.59	14.93
Overall 4C skills	89.15 $\pm$ 2.52	73.82 $\pm$ 2.48	15.33

The descriptive analysis indicated that the experimental group obtained higher mean scores than the control group across all four dimensions of 21st-century skills. Within the experimental group, collaboration had the highest mean score (91.60 $\pm$ 2.70), followed by communication (88.93 $\pm$ 2.96), creativity (88.07 $\pm$ 2.95), and critical thinking (88.00 $\pm$ 3.32). In contrast, the control group recorded lower mean scores across all four dimensions.

Table 2 illustrates the development of students' 21st-century skills across the three learning sessions, while Table 3 summarizes the mean scores and score variability for each group. Overall, the mean scores for all four dimensions increased from the first to the third session in both the experimental and control groups. However, the experimental group achieved higher scores across all dimensions than the control group. As shown in Table 3, the experimental group obtained an overall mean score of 89.15 (SD = 2.52) for 21st-century skills, whereas the control group obtained a mean score of 73.82 (SD = 2.48), representing a difference of 15.33 points. For critical thinking, creativity, communication, and collaboration, the experimental group achieved mean scores of 88.00 (SD = 3.32), 88.07 (SD = 2.95), 88.93 (SD = 2.96), and 91.60 (SD = 2.70), respectively. In contrast, the control group obtained corresponding mean scores of 72.07 (SD = 3.30), 72.67 (SD = 2.75), 73.87 (SD = 3.19), and 76.67 (SD = 2.59). The largest mean difference was observed in critical thinking, at 15.93 points, while the smallest was found in collaboration, at 14.93 points. Nevertheless, collaboration was the highest-scoring dimension in the experimental group.

Taken together, the two tables provide complementary insights into the development and achievement of students' 21st-century skills. Table 2 shows the increase in scores over the three learning sessions, whereas Table 3 indicates that the experimental group achieved higher mean scores across all four dimensions than the control group. This pattern suggests that Science, Technology, Engineering, and Mathematics (STEM)-based learning has the potential to support the development of critical thinking, creativity, communication, and collaboration

through problem-solving activities and project completion. These differences are consistent with the characteristics of integrated STEM learning. In integrated STEM instruction, students analyze real-world problems, explore solutions, participate in discussions to address problems, collaborate in teams, and present their findings based on evidence. These activities are closely associated with the competencies emphasized in 21st-century learning and frameworks for integrated STEM education (Dare et al., 2021; Kain et al., 2024; Kelley & Knowles, 2016).

To examine differences in 21st-century skills between students receiving STEM-based instruction and those receiving conventional instruction, inferential statistical analysis was conducted using independent-samples t-tests. The analysis covered each of the four 4C dimensions—critical thinking, creativity, communication, and collaboration—as well as the overall 4C score. In addition to the t-statistics and p-values, the analysis included mean differences, 95% confidence intervals (95% CIs), and Cohen’s d effect sizes to quantify the magnitude of the differences between the groups.

Table 4. Independent-Samples t-Test Results and Effect Sizes for 21st-Century Skills

Dimension	Mean Difference	t	df	p-value	95% CI	Cohen’s d
Critical thinking	15.93	18.637	58	<0.001	[14.22; 17.64]	4.81
Creativity	15.40	20.940	58	<0.001	[13.93; 16.87]	5.41
Communication	15.07	18.961	58	<0.001	[13.48; 16.66]	4.90
Collaboration	14.93	21.864	58	<0.001	[13.56; 16.30]	5.65
Overall 4C skills	15.33	23.760	58	<0.001	[14.04; 16.62]	6.13

The independent-samples t-test results revealed statistically significant differences between the experimental and control groups across all four dimensions of 21st-century skills. Critical thinking differed significantly between the two groups, $t(58) = 18.637$, $p < .001$, with a mean difference of 15.93 points and a very large effect size ($d = 4.81$).

Creativity also showed a statistically significant difference, $t(58) = 20.940$, $p < .001$. The experimental group scored 15.40 points higher than the control group, with a very large effect size ($d = 5.41$). This finding is consistent with the characteristics of engineering design, which require students to generate, test, evaluate, and refine alternative solutions (Khalil et al., 2023).

Communication differed significantly between the groups by 15.07 points, $t(58) = 18.961$, $p < .001$, with a very large effect size ($d = 4.90$). STEM activities require students to explain procedures, discuss evidence, present findings, and communicate decisions in collaborative learning settings. These activities provide repeated opportunities for students to develop scientific communication and argumentation skills (Patriot et al., 2025; Rizakhojayeve et al., 2025).

Collaboration yielded the highest mean score in the experimental group, at 91.60 ± 2.70 , compared with 76.67 ± 2.59 in the control group. This difference was statistically significant, $t(58) = 21.864$, $p < .001$, with a very large effect size ($d = 5.65$). This finding is consistent with

the collaborative structure of STEM-based problem-solving, in which students work together to investigate problems, distribute responsibilities, evaluate alternatives, and develop solutions.

For the overall 4C score, the experimental group obtained a mean of 89.15 ± 2.52 , compared with 73.82 ± 2.48 in the control group. This difference was statistically significant, $t(58) = 23.760$, $p < .001$, with a very large standardized effect size ($d = 6.13$). The consistently higher scores across all four dimensions indicate that STEM-based learning was associated with better observed performance in the four 21st-century competencies assessed in this study.

The analysis revealed a very large and statistically significant difference in 21st-century skills (4C) scores between the experimental and control groups ($p < 0.001$), with Cohen's d values ranging from 4 to 6. The magnitude of these effect sizes is mathematically associated with substantial differences in group means and relatively small pooled standard deviations, indicating that the between-group differences were large relative to the variability within each group. Descriptively, the largest effect size was observed for collaboration, which may be associated with the characteristics of STEM-based physics instruction, which provides students with opportunities to work together, share responsibilities, discuss alternative solutions, and design problem-solving strategies. These activities may also support the development of creativity, communication, and critical thinking.

The study involved 60 Grade XI students, equally divided between the experimental and control groups, with 30 students in each group. The experimental group received STEM-based instruction, whereas the control group received conventional instruction. Learning outcomes were assessed before and after instruction using a test instrument administered as a pretest and posttest.

Table 5. Descriptive Statistics of Students' Learning Outcomes

Group	n	Pretest Mean $\pm$ SD	Posttest Mean $\pm$ SD	Mean Gain
Experimental	30	50.97 ± 2.59	84.73 ± 3.99	33.76
Control	30	49.90 ± 3.37	70.07 ± 3.10	20.17

Based on Table 5, the standard deviation (SD) values for both the experimental and control groups were relatively small at both the pretest and posttest stages. At pretest, the experimental group had an SD of 2.59, whereas the control group had an SD of 3.37. Following instruction, the SD of the experimental group increased to 3.99, while that of the control group decreased to 3.10. These SD values indicate that students' scores within each group were relatively concentrated around their respective means, although the degree of score variability differed between groups and across measurement occasions. The increase in SD in the experimental group indicates greater variability in posttest scores than in pretest scores, whereas the decrease in SD in the control group indicates that score variability became smaller following instruction.

The data analysis showed that the two groups had relatively similar baseline performance, as indicated by their pretest scores. The experimental group obtained a mean pretest score of 50.97, whereas the control group obtained a mean score of 49.90, representing a difference of 1.07 points. Following the instructional intervention, the mean posttest score increased to 84.73 in the experimental group, while the control group achieved a mean score of 70.07. These results indicate score increases of 33.76 points in the experimental group and 20.17 points in

the control group. Overall, the experimental group demonstrated a greater improvement in learning outcomes than the control group.

The post-intervention difference in learning outcomes is consistent with the characteristics of the Science, Technology, Engineering, and Mathematics (STEM) approach, which integrates science, technology, engineering, and mathematics into instructional activities. This integration provides students with opportunities to explore physics concepts through inquiry, experimentation, quantitative analysis, and the design of solutions to relevant problems. These processes may help students connect theoretical concepts with their applications in real-world situations (Kelley & Knowles, 2016; Thibaut et al., 2018). Research in physics education has also emphasized the importance of instruction that connects conceptual understanding with the application of concepts in meaningful contexts (Bao & Koenig, 2019). Nevertheless, these descriptive differences alone are insufficient to establish that the STEM approach caused greater improvements in learning outcomes. Therefore, the interpretation of these findings should also consider the results of inferential statistical analyses and the characteristics of the research design (Asrizal et al., 2024; Kelley & Knowles, 2016).

Before hypothesis testing, normality and homogeneity of variance were examined to assess the suitability of parametric statistical tests. Normality was assessed using the Shapiro–Wilk test, whereas homogeneity of variance between groups was examined using Levene’s test. The results of these tests were used to determine the appropriateness of the independent-samples t-test for examining differences in mean physics learning outcomes between the experimental and control groups.

Table 6. Results of the Shapiro–Wilk Normality Test

Group	Statistic	Sig.	Interpretation
Experimental	0.967	0.472	Normally distributed
Control	0.948	0.145	Normally distributed

The Shapiro–Wilk test yielded p-values of .472 for the experimental group and .145 for the control group. Because both values exceeded .05, the posttest score distributions in both groups did not show statistically significant departures from normality, allowing parametric analyses to proceed. Levene’s test of homogeneity of variance yielded a p-value of .133, which was also greater than .05. Therefore, there was insufficient evidence of a statistically significant difference in variances between the two groups. These results indicate that the homogeneity of variance assumption was satisfied and that the parametric analysis could proceed.

Table 7. Results of the Homogeneity of Variance Test

Test	Statistic	Sig.	Interpretation
Levene’s test	2.326	0.133	Homogeneity of variance assumed

After these assumptions were examined, an independent-samples t-test was conducted to compare the pretest scores of the experimental group ($M = 50.97$; $SD = 2.59$) and the control group ($M = 49.90$, $SD = 3.37$). The test was performed to determine whether the groups differed in their initial abilities before the instructional intervention. The mean difference of 1.07 points had a 95% confidence interval ranging from -0.49 to 2.63 , with

Cohen's $d = 0.36$. The results indicated that no statistically significant difference in initial pretest performance was detected between the two classes. However, the equivalence of the groups in other characteristics cannot be established because students were not individually randomized.

Table 8. Independent-Samples t-Test Results for Pretest and Posttest Scores

Outcome	Experimental Group, Mean $\pm$ SD	Control Group, Mean $\pm$ SD	Mean Difference	t	df	p-value	95% CI	Cohen's d
Pretest	50.97 $\pm$ 2.59	49.90 $\pm$ 3.37	1.07	1.375	58	0.174	[-0.49; 2.63]	0.36
Posttest	84.73 $\pm$ 3.99	70.07 $\pm$ 3.10	14.66	15.906	58	<0.001	[12.81; 16.51]	4.11

The experimental group obtained a mean posttest score of 84.73 (SD = 3.99), whereas the control group achieved a mean score of 70.07 (SD = 3.10). The independent-samples t-test revealed a statistically significant difference between the groups, $t(58) = 15.906$, $p < .001$, with a mean difference of 14.66 points and a 95% confidence interval ranging from 12.81 to 16.51. Cohen's $d = 4.11$ indicated an exceptionally large standardized difference. These findings suggest that students receiving STEM-based instruction achieved higher physics learning outcomes than those receiving conventional instruction. This result is consistent with the STEM integration framework, which emphasizes the application of interdisciplinary knowledge to authentic problem-solving (Kelley & Knowles, 2016; Thibaut et al., 2018), as well as previous findings in physics education (Asrizal et al., 2024; Patriot et al., 2025; Yulianti et al., 2025). However, this difference should be interpreted in light of the quasi-experimental design.

To compare learning gains between the experimental and control groups, normalized gain (N-gain) scores were calculated. The results are presented in Table 9.

Table 9. Normalized Gain in Physics Learning Outcomes

Group	Mean N-Gain	Interpretation
Experimental	0.70	High
Control	0.401	Moderate

As shown in Table 9, the mean N-gain in physics learning outcomes was 0.70 in the experimental group and 0.401 in the control group. According to the N-gain interpretation criteria, the experimental group was at the threshold of the high-gain category, whereas the control group fell within the moderate-gain category. Descriptively, these results indicate that the experimental group achieved greater gains in physics learning outcomes than the control group. Nevertheless, the observed difference in N-gain scores does not, by itself, establish a statistically significant difference in learning gains without supporting inferential analysis. These findings are consistent with the characteristics of STEM-based learning, which integrates investigation, experimentation, engineering design, and problem-solving. Interdisciplinary activities and authentic problems are important components of the STEM approach (Dare et al., 2021; Thibaut et al., 2018). However, descriptive differences in N-gain

scores alone do not demonstrate that the differences in learning gains are statistically significant.

The subsequent analysis employed analysis of covariance (ANCOVA) to compare learning outcomes between the groups while accounting for students' pretest scores. In this analysis, posttest scores were specified as the dependent variable, instructional approach as the independent variable, and pretest scores as the covariate.

Table 10. ANCOVA Results for Posttest Physics Learning Outcomes

Source of Variation	F	df	p-Value	Adjusted Mean Difference	95% CI
Pretest	5.20	1, 57	0.026	—	—
Instructional group	249.77	1, 57	<0.001	14.30	[12.49; 16.11]

After controlling for initial ability based on pretest scores, the effect of instructional group remained statistically significant, $F(1, 57) = 249.77$, $p < .001$. The adjusted mean difference was approximately 14.30 points, with a 95% confidence interval ranging from 12.49 to 16.11 points. The ANCOVA results reinforce the findings of the independent-samples t-test, as the difference in posttest scores remained statistically significant after accounting for students' initial abilities. Nevertheless, because the participants were drawn from pre-existing classes and were not randomly assigned individually, these findings should be interpreted as evidence supporting the effectiveness of the intervention under the conditions examined, rather than as definitive proof of causality.

Discussion

The results showed that students who participated in physics instruction based on *Science, Technology, Engineering, and Mathematics* (STEM) demonstrated an increase in 21st-century skills scores over the three instructional sessions. The mean 4C skills score in the experimental group increased from 85.13 in the first session to 90.08 in the second session and 92.38 in the third session. The mean score in the control group also increased from 69.50 to 74.04 and 77.83, respectively. Descriptively, the experimental group achieved higher scores than the control group in every session. The between-group score differences were 15.63 points in the first session, 16.04 points in the second session, and 14.55 points in the third session. This pattern indicates that the experimental group maintained higher scores than the control group throughout the three instructional sessions.

The development of critical thinking skills can be understood through the characteristics of STEM-based learning, which position students as active participants in scientific inquiry and problem-solving. In physics learning, students need to identify relevant information, connect concepts with observed phenomena, evaluate evidence, and formulate explanations based on their analyses. These activities provide opportunities for students to develop reasoning that extends beyond factual knowledge and the mechanical application of formulas. Bao & Koenig (2019) emphasize the importance of physics teaching approaches that focus on scientific practices and competency development, whereas Dare et al. (2021) highlight the implementation of integrated STEM education by considering the characteristics and

practices that support integration across disciplines. Taken together, these studies suggest that STEM-based physics learning can be understood as an approach that integrates scientific practices, inquiry, and problem-solving through the application of concepts from multiple disciplines. These characteristics provide students with opportunities to analyze information, evaluate evidence, connect concepts with phenomena, and formulate logical explanations. Such activities may support the development of critical thinking skills because students not only acquire physics concepts but also apply them to understand and solve problems systematically.

Creativity is associated with opportunities for students to explore alternative solutions to problems and develop ideas based on physics principles. In STEM-based learning, the engineering design process may involve identifying needs, developing designs, testing solutions, and making improvements based on evaluation results. This iterative process allows students to consider more than one approach when completing tasks. The findings of Khalil et al., (2023) indicate that STEM-based learning can support the development of several aspects of students' creative thinking, particularly fluency, flexibility, and originality. These findings are consistent with the meta-analysis by Novianti et al., (2023), which indicated that the implementation of STEM in science education has a relatively large impact on 21st-century skills, including creative thinking. Together, these studies provide empirical support for the present study, particularly regarding the development of students' creativity through STEM-based learning. STEM learning activities can provide students with opportunities to generate ideas, explore alternative solutions, and apply physics concepts to problem-solving. When learning involves designing, testing, and refining products, these activities give students opportunities to develop and evaluate their ideas.

Communication is an essential skill in project-based STEM learning, particularly when students engage in investigation, discussion, and presentation activities. Through these activities, students have opportunities to express their ideas, explain their investigative findings, and communicate the solutions they develop. In line with this, Patriot et al., (2025) reported that the implementation of STEM-Project-Based Learning can support the development of 21st-century skills, including communication and creative thinking. Project-based learning activities provide opportunities for students to exchange ideas, discuss their work, and articulate their thoughts throughout the problem-solving process. These findings are relevant to explaining the importance of learning activities in providing students with opportunities to practice and develop their communication skills.

In the collaboration dimension, the experimental group recorded the highest score among the four 4C skill dimensions assessed. This finding is consistent with the characteristics of STEM-based learning, which may require students to work together to identify problems, distribute responsibilities, reach collective decisions, and develop group solutions. In this context, collaboration involves both individual contributions and the ability to coordinate efforts toward a shared goal. Rizakhojayeve et al., (2025); Thibaut et al., (2018) concluded in their respective studies that STEM learning can support the development of students' skills through integrated, problem-centered, inquiry-based learning activities that incorporate design processes and promote collaboration. Nevertheless, a higher collaboration score than those of the other dimensions does not, in itself, indicate that collaboration experienced the greatest improvement.

Conceptually, the development of 4C skills can be understood as part of a learning process that integrates knowledge, scientific practices, and problem-solving. Kelley & Knowles, (2016) describe a conceptual framework for integrated STEM education that connects different disciplines through learning contexts and practices. Meanwhile, Kain et al., (2024) provide a broader perspective on the development of 21st-century skills in secondary education. Their findings emphasize the importance of systematically integrating these skills into the learning process. In the context of STEM-based physics learning, critical thinking, creativity, communication, and collaboration can be developed through activities involving problem analysis, idea exploration, the presentation of investigative findings, and group coordination. These activities provide students with opportunities to apply their physics knowledge while developing skills relevant to the demands of 21st-century learning.

Methodologically, the findings should be interpreted in light of the characteristics of the quasi-experimental design. The study involved two intact classes at a single school; therefore, differences in students' initial characteristics and the influence of classroom context cannot be completely ruled out. In addition, 4C skill scores were obtained through observations conducted over three sessions. Accordingly, the measurements represent skill-related behaviors observed during the learning activities investigated and do not directly capture every aspect of students' competencies outside those situations. The higher scores recorded by the experimental group from the first session also highlight the importance of distinguishing differences observed at the beginning of the instructional period from score changes across the three sessions. If 4C skills were not measured before the intervention, the initial difference cannot be attributed with certainty to the implementation of STEM-based learning. Although between-group differences were tested inferentially, changes across the three sessions were analysed only descriptively; repeated-measures analyses that account for group and session effects are needed to determine whether these patterns of change are statistically supported.

The implications of this study lie in the importance of designing physics instruction that goes beyond content delivery and provides students with structured opportunities to practice 21st-century skills. STEM implementation should be accompanied by clear learning objectives, tasks aligned with physics concepts, well-defined group roles, and observation rubrics that operationally specify the skill indicators. Teachers should also provide feedback during inquiry activities and encourage reflection on students' work so that skill development can be monitored systematically. As demonstrated in this study, observations across multiple sessions provide an overview of score patterns during the learning process. Future research could strengthen the evidence by measuring 4C skills before the intervention, involving multiple classes or schools, and conducting repeated-measures analyses of score changes across sessions. These steps are important for distinguishing observed development during instruction from pre-existing differences between groups.

The overall evaluation of the experimental and control groups showed that the experimental group consistently achieved higher 4C skills scores across all observations. These findings suggest that the integration of problem-solving activities, physics concepts, and group work may provide students with opportunities to develop critical thinking, creativity, communication, and collaboration skills. Therefore, the higher 21st-century skills scores achieved by the experimental group indicate that these students demonstrated consistently

higher performance throughout the learning process. Real-life experiences and 21st-century skills are important components of integrated STEM learning. Incorporating real-life experiences can motivate students to apply their knowledge, engage in reasoning, seek solutions to problems, and collaborate with group members to complete learning tasks (Dare et al., 2021).

The findings indicate that the dimension with the greatest improvement differed between the experimental and control groups. In the experimental group, which implemented STEM-based physics learning, the greatest improvement occurred in critical thinking, with an increase of 8.34 points. This result indicates that, during the implementation of STEM-based learning, critical thinking showed the largest increase in scores compared with the other three 4C dimensions. This improvement may be associated with the characteristics of STEM activities, which provide students with opportunities to analyze problems, consider information, evaluate alternative solutions, and test the solutions they develop. In contrast, the greatest improvement in the control group occurred in collaboration, with an increase of 12.50 points. This increase was greater than that observed in the experimental group, where collaboration improved by 6.50 points. This difference may be related to the initial scores of the two groups, as the experimental group had already demonstrated relatively high collaboration scores from the first session, leaving less room for further score increases than in the control group. Therefore, the greater increase in the control group does not necessarily indicate that students in this group had better collaboration skills than those in the experimental group. Statistical test results showed that the experimental group achieved higher scores than the control group across all four 4C dimensions.

On average, learning outcomes in the experimental group were better than those in the control group. This finding is consistent with the results of another study by Asrizal et al., (2024), which examined the role of integrated STEM-based physics learning in conceptual understanding. From a pedagogical perspective, the findings suggest that STEM-based physics instruction should be organized to provide students with opportunities to: (1) analyze problems and apply critical thinking skills in practice; (2) formulate and modify alternative solutions to exercise their creative potential during learning; (3) communicate and defend their ideas through argumentation and presentations; and (4) share responsibility for problem-solving by engaging in genuine collaboration rather than merely dividing tasks. Thus, the effectiveness of STEM-based learning cannot be assessed solely by the inclusion of Science, Technology, Engineering, and Mathematics in the design of instructional materials. More importantly, the question is whether this integration helps students use their knowledge, thinking skills, communication abilities, and collaborative skills to solve problems.

This study shows that students who participated in *Science, Technology, Engineering, and Mathematics* (STEM)-based physics instruction achieved higher learning outcomes than those who received conventional instruction on static fluids. This finding is reflected in the mean posttest score of 84.73 (SD = 3.99) for the experimental group and 70.07 (SD = 3.10) for the control group. The difference was statistically significant, $t(58) = 15.906$, $p < .001$, with a mean difference of 14.66 points and a 95% confidence interval (CI) of [12.81, 16.51]. These findings suggest that STEM-based instruction has the potential to support physics learning outcomes by enabling students to apply concepts to problem-solving.

The higher learning outcomes in the experimental group can be understood in terms of the characteristics of STEM-based instruction, which connects scientific concepts with technology, engineering, and mathematics in problem-solving activities. In the context of static fluids, students not only learn physics principles and equations but also have opportunities to apply them to explain phenomena, analyze problems, and design or test solutions. Consequently, learning extends beyond memorizing formulas and solving problems through procedural methods. These findings suggest that physics teachers may consider implementing STEM-based instruction by designing activities that integrate conceptual explanations, scientific inquiry, mathematical applications, and solution-design processes. Successful implementation depends not only on the use of projects or equipment but also on establishing clear connections between student activities and learning objectives.

This study has several limitations. The sample comprised only 60 students from two classes at a single school; therefore, the generalizability of the findings should be considered cautiously. In addition, 4C skills were measured through observations conducted over three sessions, without the use of standardized pretest and posttest instruments. Future studies could involve more schools and students, implement longer intervention periods, and combine classroom observations with performance tests or standardized 4C skills instruments.

CONCLUSION

This study indicates that STEM-based physics learning was associated with higher 21st-century skills (4C) scores than conventional instruction. Descriptively, the experimental group obtained an overall mean 4C score of 89.15 (SD = 2.52), whereas the control group obtained a mean score of 73.82 (SD = 2.48). The independent-samples t-test revealed a statistically significant difference between the two groups, with a mean difference of 15.33 points, $t(58) = 23.760$, $p < .001$, and Cohen's $d = 6.13$. Statistically significant differences were also found in critical thinking, creativity, communication, and collaboration. Collaboration had the highest mean score in the experimental group, at 91.60 (SD = 2.70). In addition, 4C scores increased descriptively over the three sessions in both groups, although this pattern was not tested using inferential analyses of between-session changes.

Regarding physics learning outcomes, the experimental group obtained a mean posttest score of 84.73 (SD = 3.99), higher than that of the control group, which was 70.07 (SD = 3.10). The mean posttest difference of 14.66 points was statistically significant, $t(58) = 15.906$, $p < .001$, with Cohen's $d = 4.11$. The ANCOVA results also indicated that the between-group difference remained statistically significant after controlling for pretest scores, $F(1, 57) = 249.77$, $p < .001$, with an adjusted mean difference of 14.30 points (95% CI [12.49, 16.11]). The normalized gain (N-gain) values of 0.70 for the experimental group and 0.401 for the control group descriptively indicated greater improvement in learning outcomes in the experimental group.

These findings support the use of STEM-based physics learning as an alternative approach to developing 4C skills while supporting students' understanding of physics concepts. Its implementation should be designed around inquiry, problem-solving, solution design, discussion, and presentation activities that provide students with meaningful opportunities to practice each skill. Nevertheless, the findings should be interpreted in light

of the quasi-experimental design, the involvement of two classes from a single school, and the measurement of 4C skills through observations conducted over three sessions. Future research should involve more schools, measure 4C skills before and after the intervention, and employ repeated-measures analyses to examine changes in skills over time. Thus, this study provides empirical evidence of between-group differences in the context investigated; however, the findings cannot yet be generalized broadly or regarded as definitive evidence of causality.

REFERENCES

- Asrizal, Nazifah, N., Effendi, H., & Helma. (2024). STEM-Smart Physics e-module to promote conceptual understanding and 4C skills of students. *International Journal of Information and Education Technology*, 14(2), 279–286. <https://doi.org/10.18178/ijiet.2024.14.2.2049>
- Bao, L., & Koenig, K. (2019). Physics education research for 21st century learning. *Disciplinary and Interdisciplinary Science Education Research*, 1, Article 2. <https://doi.org/10.1186/s43031-019-0007-8>
- Damayanti, E. D., Suwarma, I. R., & Muslim, M. (2023). The influence of STEM-based physics module to enhance critical thinking skills. *Unnes Science Education Journal*, 12(3), 110–115. <https://doi.org/10.15294/usej.v12i3.72443>
- Dare, E. A., Keratithamkul, K., Hiwatig, B. M., & Li, F. (2021). Beyond content: The role of STEM disciplines, real-world problems, 21st century skills, and STEM careers within science teachers' conceptions of integrated STEM education. *Education Sciences*, 11(11), Article 737. <https://doi.org/10.3390/educsci11110737>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. *Teaching and Teacher Education*, 151, Article 104739. <https://doi.org/10.1016/j.tate.2024.104739>
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, 3, Article 11. <https://doi.org/10.1186/s40594-016-0046-z>
- Khalil, R. Y., Tairab, H., Qablan, A., Alarabi, K., & Mansour, Y. (2023). STEM-based curriculum and creative thinking in high school students. *Education Sciences*, 13(12), Article 1195. <https://doi.org/10.3390/educsci13121195>
- Lavi, R., Tal, M., & Dori, Y. J. (2021). Perceptions of STEM alumni and students on developing 21st century skills through methods of teaching and learning. *Studies in Educational Evaluation*, 70, Article 101002. <https://doi.org/10.1016/j.stueduc.2021.101002>
- Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM Education*, 6, Article 2. <https://doi.org/10.1186/s40594-018-0151-2>
- Novianti, B. A., Lasmawan, I. W., & Suharta, I. G. P. (2023). Efek STEM dalam pembelajaran sains terhadap keterampilan abad 21: Meta analisis. *Kappa Journal*, 7(2), 307–312. <https://doi.org/10.29408/kpj.v7i2.18630>
- Patriot, E. A., Ismet, I., Wiyono, K., Rizka, M. R., & Melati, P. (2025). STEM-project based learning: Implementation to optimize the 21st century skills in pre-service teacher of physics education. In *Proceedings* (pp. 622–639). In M. Meilinda, et al. (Eds.), *Proceedings of the 6th Sriwijaya University Learning and Education International Conference 2024 (SULE-IC 2024)* Atlantis Press. https://doi.org/10.2991/978-2-38476-390-0_51



- Rizakhojayeva, G., Ramankulov, S., Akeshova, M., Nurizinova, M., Tuyakov, Y., & Abdrakhmanov, R. (2025). STEM-based approaches to soft skills development: A synthesis of meta-analytic findings and empirical evidence. *Frontiers in Education*, 10, Article 1663155. <https://doi.org/10.3389/feduc.2025.1663155>
- Roehrig, G. H., Dare, E. A., Ring-Whalen, E., & Wieselmann, J. R. (2021). Understanding coherence and integration in integrated STEM curriculum. *International Journal of STEM Education*, 8, Article 2. <https://doi.org/10.1186/s40594-020-00259-8>
- Sudjana. (2002). *Metoda statistika* (Ed. ke-6). Penerbit Tarsito. Bandung
- Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., Boeve-de Pauw, J., Dehaene, W., Deprez, J., De Cock, M., Hellinckx, L., Knipprath, H., Langie, G., Struyven, K., Van de Velde, D., Van Petegem, P., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education*, 3(1). <https://doi.org/10.20897/ejsteme/85525>
- Wulandari, N., & Pujiyanto. (2025). Integrating STEM in physics learning to improve 21st century skills: A systematic literature review. *Kasuari: Physics Education Journal*, 8(2), 360–369. <https://doi.org/10.30862/kpej.v8i2.952>
- Yulianti, D., Shafira, F. F., Marwoto, P., & Noviyatun, S. (2025). A systematic review of trend STEM education research on physics learning in Indonesia. *Jurnal Pendidikan Fisika Indonesia*, 21(2), 192–208. <https://doi.org/10.15294/jpfi.v21i2.23548>
- Yulianti, D., Sugianto, & Ngafidin, K. M. (2022). Scratch assisted physics learning with a STEM approach in the pandemic era to develop 21st century learning skills. *Jurnal Pendidikan IPA Indonesia*, 11(1), 185–194. <https://doi.org/10.15294/jpii.v11i1.32607>