JPF | Volume 14 | Number 2 | 2026 | 494 - 513
p - ISSN: 2302-8939
e - ISSN: 2527-4015

Jurnal Pendidikan Fisika
https://journal.unismuh.ac.id/index.php/jpf
DOI: 10.26618/m5aeba30

Integrating the STEM Approach into Discovery Learning to Improve Students' Scientific Literacy in Temperature and Heat

Arini Rosa Sinensis*, Roja Septiani, Febrianto Putra, Thoha Firdaus

Department of Physics Education, Universitas Nurul Huda, Ogan Komering Ulu Timur, South Sumatra,32161, Indonesia

**Corresponding author: arini@unuha.ac.id*

Received: February 04, 2026; Accepted: April 28, 2026; Published: May 18, 2026

Abstract *- Scientific literacy is an essential skill in physics education, as students must understand scientific principles, interpret evidence, and apply knowledge to real-world issues. In Indonesia, however, scientific literacy among students remains a concern, evidenced by low PISA science scores and predominantly teacher-centered classroom practices. This study aimed to analyze the effect of integrating the STEM approach into the Discovery Learning model on eleventh-grade students' scientific literacy in the topic of temperature and heat. A quasi-experimental method with a non-equivalent control group design was employed at MA Nurul A'la Jatimulyo II during the odd semester of the 2025/2026 academic year. The participants were 58 eleventh-grade science students. The experimental class ($n = 29$) learned through STEM-integrated Discovery Learning (DL-STEM), whereas the control class ($n = 29$) received conventional instruction. Scientific literacy was measured using a 25-item multiple-choice test developed from five indicators: explaining scientific phenomena, using scientific evidence, identifying scientific statements, understanding phenomena, and solving problems. Data were analyzed using descriptive statistics, assumption tests, the Mann–Whitney test for initial equivalence, an independent-samples t-test for posttest comparison, effect size r, and N-Gain analysis. The results indicated that both groups had comparable baseline abilities. Post-intervention, the experimental group significantly outperformed the control group, with mean scores of 64.69 versus 53.93 ($p < 0.001$), and demonstrated a moderate effect size ($r = 0.47$). Moreover, N-Gain analysis revealed improvements across all scientific literacy indicators in the experimental group, categorized as moderate and surpassing those of the control group. The novelty of this study lies in mapping STEM components onto the Discovery Learning syntax through a simple thermos engineering task in temperature-and-heat instruction. These findings indicate that DL-STEM can support students in connecting physics concepts, scientific evidence, and engineering design, thereby contributing to scientific literacy-oriented physics education.*

Keywords: *discovery learning; heat transfer; physics learning; scientific literacy; STEM education.*

© 2026 The Author(s). Licensed under CC BY-SA 4.0 International.

I. INTRODUCTION

In 21st-century education, scientific literacy has become an essential competence that students need to develop. Students are not only expected to understand scientific theories and concepts but also to use scientific knowledge to explain phenomena, evaluate the validity of information, and make decisions based on scientific evidence (Osborne & Allchin 2025). Scientific literacy also prepares students to engage in scientific reasoning, interpret evidence, and make informed decisions in real-life situations (Putri et al., 2026; Roy et al., 2025). In physics learning, this competence is particularly important because physics concepts are closely related to everyday issues, such as energy use, heat transfer, household technology, and safety in learning environments. Therefore, science learning in schools should be directed toward learning processes that help students connect concepts, scientific practices, and contextual problem solving (Rahmi et al., 2024; Hardianti et al., 2021; Asriadi & Lazulva, 2021; Mellyzar et al., 2022; Usta et al., 2025; Wulandari et al., 2023).

According to the PISA framework, scientific literacy refers to an individual's ability to engage with science-related issues, explain phenomena using scientific principles, evaluate and design scientific investigations, and critically interpret data and evidence (OECD, 2019, 2023a). Thus, scientific literacy cannot be understood merely as the memorization of concepts. Rather, it involves the ability to think and act scientifically. Scientifically literate students are expected to connect scientific ideas with technology, evaluate scientific claims, use data to support arguments, and apply concepts to real-life problems (Coppi et al., 2024; De Loof et al., 2022).

The urgency of improving scientific literacy in Indonesia is reflected in the PISA 2022 results. Indonesia's mean science score was 383, while the OECD average was 485 points (OECD, 2023b, 2023c). Although Indonesia's ranking improved compared with PISA 2018, its science score remains below the international average. This result indicates that many students have not yet reached a basic level of proficiency in using scientific knowledge to explain phenomena and draw evidence-based conclusions. This condition highlights the need for science learning that places greater emphasis on inquiry activities, data-based argumentation, and authentic problem solving (Çalik & Wiyarsi, 2024; Kumar & Choudhary, 2025).

Low scientific literacy is influenced by various factors, including the selection of learning resources, misconceptions, teacher-centered instruction, limited use of learning models, unsupportive learning environments, quality of learning implementation, laboratory facilities, student motivation, and socioeconomic background (Fuadi et al., 2020; Sudarto, 2026). Preliminary observations and interviews at MA Nurul A'la Jatimulyo II in East Ogan Komering Ulu Regency revealed a similar pattern. Students had difficulty understanding physics content and were not accustomed to connecting concepts of temperature and heat to their daily experiences. Learning was still frequently conducted through lectures, question-and-answer sessions, and assignments. As a result, students' active participation, confidence in asking questions, and creative thinking practices had not developed optimally. The preliminary scientific literacy test also yielded a score of 36.78%, placing students' scientific literacy in the low category.

One relevant approach to addressing this issue is STEM, which integrates science, technology, engineering, and mathematics. The STEM approach emphasizes integrating scientific concepts, technological applications, engineering processes, and mathematical reasoning to solve real-world problems. Effective STEM integration does not treat the four disciplines as separate elements (Portillo-Blanco et al., 2024). Instead, it connects them through learning activities that encourage students to observe phenomena, design solutions, test evidence, perform calculations, and revise ideas based on their observations (Abdi et al., 2024; Bybee, 2013; Kelley & Knowles, 2016; Roehrig et al., 2021).

Integrated STEM learning requires teachers to connect disciplinary concepts with inquiry practices, engineering design, technological use, and mathematical reasoning, rather than presenting STEM as an additional activity at the end of a lesson (Mansour et al., 2024; Halawa et al., 2024). Previous studies have reported that integrated STEM can support students' cognitive performance, interest, and motivation when teachers structure learning around meaningful problems and guide students to use evidence in decision-making (De Loof et al., 2022; Demirkol et al., 2022). Therefore, the instructional value of STEM depends not only on the presence of projects but also on how projects are embedded within the sequence of concept construction, investigation, and reflection.

Several studies have implemented STEM and Discovery Learning in science education. Fadlina et al. (2021) showed that STEM-based Discovery Learning can improve critical thinking skills. Other studies found that Discovery Learning contributes to scientific literacy (Maulana et al., 2024), Discovery Learning contributes positively to student achievement (Kumar & Choudhary, 2025), the STEM approach supports data literacy (Satriana, 2023), and STEM integration in Learning Cycle and Project-Based Learning can strengthen scientific literacy and critical thinking skills (Batubara et al., 2025; Ibrahim et al., 2025; Nisah et al., 2024). However, three gaps remain. First, many studies report achievement gains without explaining the mechanism through which Discovery Learning and STEM complement each other. Second, the contribution of engineering design tasks to scientific literacy is often asserted but not clearly described in terms of learning syntax. Third, empirical work on temperature and heat still rarely connects everyday heat-transfer phenomena, data interpretation, and product design within a single Discovery Learning sequence (Zhan & Niu, 2023).

Based on this gap, the present study focuses on the implementation of STEM-based Discovery Learning (DL-STEM) to improve eleventh-grade students' scientific literacy in the topic of temperature and heat. The added value of combining Discovery Learning and STEM lies in their complementary mechanisms. Discovery Learning structures the process through which students move from observing phenomena to constructing concepts, whereas STEM provides an interdisciplinary problem context in which those concepts are used to design, test, and justify a solution. Temperature and heat are suitable topics for this integration because the concepts of conduction, convection, radiation, temperature change, and insulation can be directly observed in household technologies, such as a thermos, and quantified through simple measurements.

Theoretically, DL-STEM is expected to enhance scientific literacy, particularly in the indicators of explaining scientific phenomena and using scientific evidence. The stimulation and problem-statement stages encourage students to observe heat-transfer phenomena and formulate investigable questions. The data collection, data processing, and verification stages require students to compare measurements and evaluate claims. The generalization stage then helps students connect empirical evidence with physics principles. The engineering task extends this process by asking students to design a simple thermos, select appropriate materials, measure temperature change, and justify design decisions. Thus, this study does not claim to introduce a wholly new model. Instead, it offers empirical evidence on how STEM can be mapped onto the Discovery Learning syntax for temperature-and-heat instruction and how this mapping relates to students' scientific literacy.

II. METHODS

This study was conducted at MA Nurul A'la Jatimulyo II and involved eleventh-grade science students during the odd semester of the 2025/2026 academic year. A quasi-experimental method with a non-equivalent control group design was used because the classes had already been established, and individual randomization could not be applied. The sampling technique was non-probability sampling with total sampling. The entire population, consisting of 58 students, was used as the research sample. Class XI Science 1, consisting of 29 students, was assigned as the experimental class, whereas Class XI Science 2, also consisting of 29 students, was assigned as the control class. The same physics teacher implemented both learning conditions using lesson plans developed in collaboration with the researchers to minimize teacher effects. Permission from the school was obtained before data collection. Students were informed that the test data would be used anonymously for research purposes and that their participation would not affect their academic grades.

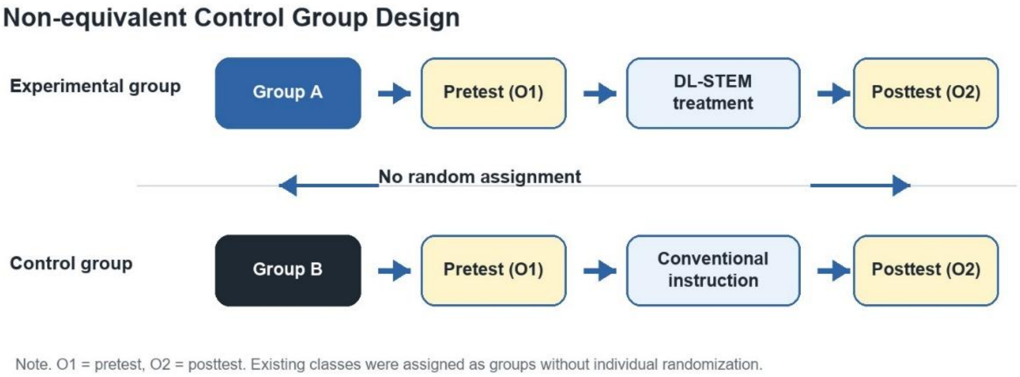

Figure 1. Non-equivalent control group design

The intervention was carried out during the scheduled topic of temperature and heat. It consisted of four instructional meetings conducted after the pretest and before the posttest. In the experimental class, each meeting followed the Discovery Learning syntax while embedding STEM components (Mansour et al., 2024). In the control class, the same topic coverage and learning duration were maintained. However, instruction was delivered through teacher explanation, question-and-answer activities, worked examples, and individual assignments without an explicit engineering-design task.

Table 1. Instructional sequence and comparison between groups

Meeting	Experimental class: DL-STEM	Control class: conventional instruction
Pretest	Scientific literacy pretest on temperature and heat.	Scientific literacy pretest on temperature and heat.
1	Stimulation and problem statement through everyday heat-transfer phenomena, including hot drinks and insulating materials.	Class questions followed the teacher's explanation of temperature, heat, and heat transfer.
2	Data collection through observation / experiment: students measured temperature changes and discussed material properties.	Worked examples and exercises on heat, temperature change, and heat transfer.
3	Data processing and verification; students compared measurements and evaluated claims about insulation effectiveness.	Guided discussion and practice questions using textbook-based contexts.
4	Generalization and engineering task; groups designed, tested, and reflected on a simple thermos based on heat-transfer principles.	Review, summary, and assignments on temperature-and-heat concepts.
Posttest	Scientific literacy posttest using the same blueprint as the pretest.	Scientific literacy posttest using the same blueprint as the pretest.

The scientific literacy instrument consisted of a 25-item multiple-choice test developed based on five scientific literacy indicators. Each item was scored 1 for a correct answer and 0 for an incorrect answer. The raw scores were then converted to a 0–100 scale using the formula: obtained score/25 × 100. The five indicators were derived from a PISA-oriented scientific literacy construct, which includes the ability to explain phenomena, use evidence, evaluate scientific statements, understand phenomena, and solve contextual problems (Zhang et al., 2023).

Because the same test design was used for both the pretest and posttest, several procedures were implemented to minimize the possibility of a testing effect, namely, the influence of participants having previously seen the questions. First, the test was administered before and after the instructional process. Second, students were not given feedback or correct answers after the pretest. Third, the posttest results were primarily used to compare outcomes between classes rather than to measure students' memorization of the test items.

Content validity was examined by three validators across material, construct, and language aspects. The validation sheets used relevant/not-relevant judgments, which were summarized using the content validity ratio (CVR) and content validity index (CVI). The validation evidence showed that all 25 items were rated as relevant by the three validators across the material, construct, and language aspects. Therefore, the CVR for each item and the overall CVI were 1.00. The instrument also showed that the items were contextualized within temperature and heat phenomena, such as thermos insulation, cooking, heatwaves, condensation, material selection, and radiative cooling. Additional validation evidence was obtained from the supplied expert validation sheets and is reported to enhance the transparency of the research instrument.

Table 2. Validation, scoring, and item contexts

Aspect	Evidence reported in the revised manuscript
Number and scoring	25 multiple-choice items; correct = 1, incorrect = 0; converted to a 0–100 scale.
Content validity	Three validators reviewed material, construct, and language aspects; all items were judged relevant (CVR = 1.00; CVI = 1.00).
Indicator representation	Five indicators were represented: explaining phenomena, using evidence, identifying scientific statements, understanding phenomena, and solving problems.
Example contexts	Thermos insulation, cooking and heating, heatwave mitigation, condensation on car glass, material choice, and radiative cooling.
Remaining limitation	Distractor analysis and internal consistency were not available in the supplied dataset and are acknowledged as limitations.

The five indicators presented in Table 3 served as the operational basis for developing the test instrument. In this study, scientific literacy was based on the PISA 2018 framework (OECD), which emphasizes competencies in explaining scientific phenomena, using and interpreting scientific evidence, and applying scientific knowledge to solve contextual problems. To fit the learning context and the needs of the research instrument, these competencies were operationalized into five indicators: (1) Explaining Scientific Phenomena, (2) Using Scientific Evidence, (3) Identifying Scientific Statements, (4) Understanding Phenomena, and (5) Solving Problems. These indicators are adaptations of PISA scientific literacy competencies, tailored to the learning objectives and characteristics of the topic studied.

Table 3. Scientific literacy indicators

No.	Indicator	Description
1	Explaining Scientific Phenomena	Students' ability to describe and explain natural phenomena accurately based on relevant scientific concepts, principles, or theories.
2	Using Scientific Evidence	The ability to use data, facts, and scientific information to support arguments, draw conclusions, and evaluate a scientific claim.
3	Identifying Scientific Statements	The ability to distinguish scientific and non-scientific statements and assess their validity and relevance based on testable scientific procedures.
4	Understanding Phenomena	The ability to comprehend fundamental scientific concepts in order to interpret, analyze, and predict phenomena or relationships among variables in natural events.
5	Solving Problems	The ability to apply scientific knowledge and skills to formulate problems, design solutions, make decisions, and evaluate results.

The data were analyzed through several procedures. First, descriptive statistics were used to summarize the data, including the minimum, maximum, mean, median, and standard deviation. Second, assumption testing was conducted using the Kolmogorov–Smirnov and Shapiro–Wilk tests for normality, while homogeneity of variance was examined using Levene's test. The Shapiro–Wilk p-values were 0.175 for the control pretest, 0.074 for the control posttest, 0.025 for the experimental pretest, and 0.300 for the experimental posttest. Levene's test for the posttest comparison yielded F = 1.491, p = 0.227.

Because the experimental-class pretest did not meet the Shapiro–Wilk normality assumption, the equivalence of initial ability between groups was analyzed using the Mann–Whitney test. Third, differences in posttest scores between the control and experimental groups were examined using an independent-samples t-test. Fourth, the magnitude of the treatment effect was calculated using the following formula: $r = \sqrt{\frac{t^2}{t^2+df}}$. Fifth, the improvement of scientific literacy for each indicator was analyzed using N-Gain.

Table 4. Normality and homogeneity tests

Test	Dataset/comparison	Statistic	df	p-value	Interpretation
Shapiro–Wilk	Control pretest	0.949	29	0.175	Normal
Shapiro–Wilk	Control posttest	0.935	29	0.074	Normal
Shapiro–Wilk	Experimental pretest	0.917	29	0.025	Not normal
Shapiro–Wilk	Experimental posttest	0.958	29	0.300	Normal
Levene	Posttest comparison	F = 1.491	-	0.227	Homogeneous variance

The hypotheses in this study were formulated as follows:

- H_0 (Null Hypothesis): The scientific literacy of eleventh-grade science students taught using Discovery Learning with a STEM approach is not higher than, or is equal to, that of students taught using conventional instruction.
- H_1 (Alternative Hypothesis): The scientific literacy of eleventh-grade science students taught using Discovery Learning with a STEM approach is significantly higher than that of students taught using conventional instruction.

The decision criterion used a significance level of 0.05. If $p < 0.05$, there is a significant difference in scientific literacy improvement between students in the experimental class and those in the control class.

III. RESULTS

Students' scientific literacy in the control and experimental groups was examined using descriptive statistical analysis to identify changes in learning outcomes before and after the instructional process. The experimental group participated in DL-STEM, whereas the control group received conventional instruction. The research data were obtained from students' pretest and posttest scores. A summary of the descriptive statistics, including the mean, standard deviation, and distribution of scientific literacy scores, is presented in Table 5.

Table 5. Descriptive statistical results of the control and experimental classes

	Pretest		Posttest	
	Control	Experimental	Control	Experimental
Minimum score	28.00	28.00	40.00	48.00
Maximum score	60.00	64.00	84.00	92.00
Mean	41.72	41.86	53.93	64.69
Standard deviation	8.464	8.749	9.584	11.468
Median	40.00	40.00	52.00	64.00

Table 5 presents the descriptive comparison between the control and experimental groups. In the pretest phase, the mean scores of the two groups were nearly identical. The control group obtained a mean score of 41.72, while the experimental group obtained a mean score of 41.86. The median score was also the same in both groups, namely 40.00. These results indicate that students' initial scientific literacy levels in the two classes were relatively similar before the treatment. However, because this study used intact classes, group equivalence should be interpreted cautiously and not in the same way as equivalence obtained through random assignment.

After the instructional treatment, both groups showed an increase in scientific literacy scores. In the control group, the mean score increased from 41.72 in the pretest to 53.93 in the posttest, indicating a gain of 12.21 points. In the experimental group, which was taught using DL-STEM, the mean score increased from 41.86 to 64.69, indicating a gain of 22.83 points. The increase in the experimental group was higher than that in the control group. This result suggests that students who participated in DL-STEM experienced greater comparative improvement in scientific literacy. Nevertheless, this finding should be interpreted within the limitations of a quasi-experimental design involving two intact classes.

Hypothesis test results

Assumption testing showed that most of the research data were normally distributed. However, the pretest scores of the experimental group did not meet the assumption of normality, as indicated by the Shapiro–Wilk test. Therefore, the equivalence of students' initial abilities between the control and experimental groups was analyzed using the Mann–Whitney test. As shown in Table 6, the test yielded a p-value of 0.987 ($p > 0.05$). This result indicates that H0 was accepted, meaning that there was no statistically significant difference in initial scientific literacy between the two groups before the treatment.

Table 6. Mann-Whitney test results for control and experimental pretest data

Mann-Whitney test	Median (IQR)	U	Sig.	Decision
Control pretest	40.00 (14.00)	419.500	0.987	H_0 accepted
Experimental pretest	40.00 (12.00)			

Following treatment implementation, the posttest scores of the control and experimental groups were analyzed using an independent-samples t-test. This analysis was conducted to determine whether there was a statistically significant difference in scientific literacy between the two groups after the instructional process. As shown in Table 8, the t-test yielded a p-value < 0.001 ($p < 0.05$). Therefore, H0 was rejected, and H1 was accepted.

This result indicates a statistically significant difference between the posttest scores of the two groups. Students in the experimental group who experienced DL-STEM instruction achieved a higher mean posttest score ($M = 64.69$; $SD = 11.468$) than students in the control group who received conventional instruction (M = 53.93; SD = 9.584). These results suggest that DL-STEM was associated with better scientific literacy outcomes than conventional instruction in this study.

Table 7. Independent-samples t-test results for control and experimental posttest data

t-test (df = 56.08)	Mean (SD)	t	Sig.	Decision
Control posttest	53.93 (9.584)	3.992	<0.001	H_0 rejected
Experimental posttest	64.69 (11.468)			

The significant posttest difference indicates that the implementation of DL-STEM was related to an improvement in students' scientific literacy. The difference was not limited to score improvement; it also reflected the contributions of learning activities involving concept discovery, data processing, solution design, and the application of temperature and heat concepts to develop a simple product. These learning experiences may have helped students connect physics concepts to real-world contexts. However, the findings should be interpreted carefully. The improvement cannot be attributed solely to the DL-STEM model because the study used intact classes. Several factors may have influenced the results, including class characteristics, students' motivation, group interactions, and the consistency of treatment implementation during the learning process.

Effect size

Effect size analysis was conducted to determine the practical magnitude of the difference between the control and experimental groups. The effect size was calculated using the r value. The results are presented in Table 8.

Table 8. Effect size r results

Effect size (r)	Mean (SD)	t / (df = 56.085)	r	Category
Control posttest	53.93 (9.584)	3.992	0.47	Moderate
Experimental posttest	64.69 (11.468)			

The effect size value of r = 0.47 was classified as a moderate effect (Cao et al., 2025). This finding indicates that DL-STEM produced a practically meaningful, although not large, difference in students' scientific literacy. The moderate category should be interpreted proportionally. The learning sequence appears promising, but its effect may still have been influenced by the quality of syntax implementation, the depth of STEM problems, student readiness, group activity management, and the availability of simple experimental facilities (Ammar et al., 2024).

One factor that may explain the moderate effect is the quality of the stimulation stage. In the topic of temperature and heat, the stimulation stage should present authentic problems, such as heat loss in hot drinks, the selection of insulating materials, thermos efficiency, or heat transfer in household devices. If the initial phenomenon is not sufficiently challenging and does not explicitly connect science, technology, engineering, and mathematics, the discovery process may remain at the level of simple observation and may not fully encourage deeper scientific argumentation.

In addition, scientific literacy requires students to apply scientific concepts, data, and evidence to explain phenomena and make informed decisions (OECD, 2019, 2023a). In integrated STEM learning, the quality of the initial problem and the clarity of interdisciplinary connections play an important role in determining whether students merely perform practical activities or genuinely develop scientific reasoning and solution-design skills (Bybee, 2013; Kelley & Knowles, 2016; Roehrig et al., 2021). Therefore, the moderate improvement identified in this study can be considered a positive outcome. At the same time, it suggests that further improvements are needed, particularly in designing more authentic real-world problems, strengthening investigative questioning, and encouraging evidence-based reflection at the end of the learning process.

Improvement of scientific literacy

The improvement in students' scientific literacy was evaluated using the N-Gain score to determine the level of progress in each indicator. The N-Gain results for each scientific literacy indicator are presented in Figure 2.

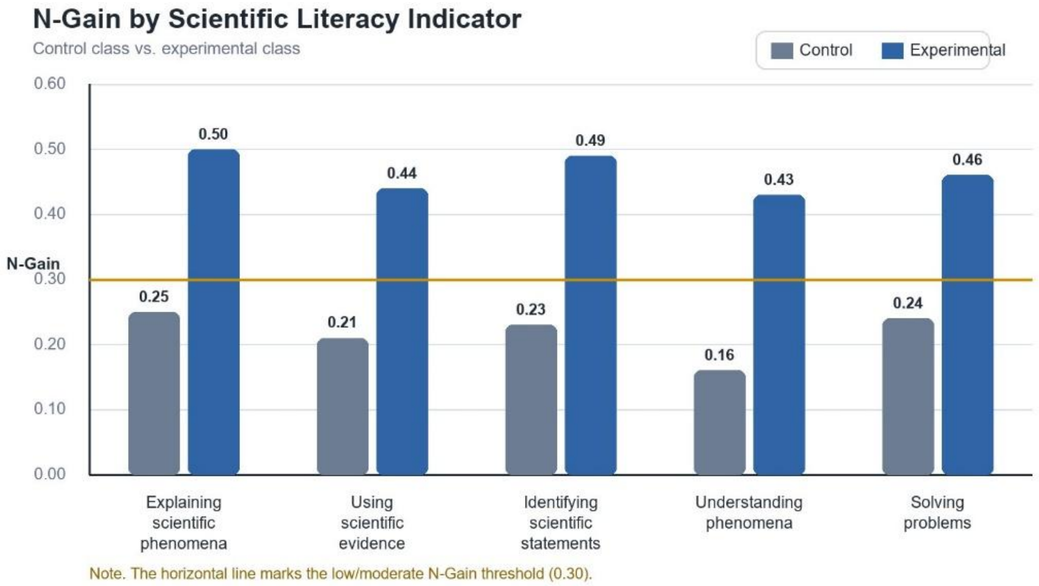

Figure 2. N-gain results for each scientific literacy indicator

Figure 2 shows that the N-Gain values of the experimental class were higher than those of the control class across all scientific literacy indicators, with improvement in the moderate category. The greatest improvement in the experimental class occurred in the indicator of explaining scientific phenomena. This pattern is reasonable because DL-STEM activities repeatedly required students to observe heat transfer, compare material behavior, and explain why the temperature changed in the simple thermos design. These activities were more directly aligned with explanation-oriented items than with indicators requiring more complex evaluation of evidence or claim validity.

IV. DISCUSSION

The improvement in indicators related to using scientific evidence and identifying scientific statements indicates that students began to develop the ability to interpret data, distinguish evidence-based claims from opinions, and draw conclusions from experimental results. This improvement is important because both indicators require students not only to remember concepts but also to use information obtained from observations and measurements. In the context of this study, students were required to examine temperature changes, compare the performance of insulating materials, and relate the obtained data to the concepts of conduction, convection, radiation, and heat transfer. These activities provided opportunities for students to practice using scientific evidence in a more contextual manner.

However, the moderate N-Gain scores suggest that these abilities had not yet developed to an optimal level. This condition may be related to the nature of evidence-based reasoning, which requires repeated and explicit practice in connecting claims, evidence, and reasoning. Although students were involved in observation, experimentation, and product testing, their reasoning process still needed stronger guidance. In this study, the learning activities were still largely centered on product testing and worksheet-based observations, while opportunities for deeper scientific argumentation and diagnostic analysis remained limited. Therefore, the improvement in evidence use and scientific statement identification should be understood as an initial development rather than a fully established competence.

In the learning activities, students were tasked with designing a simple thermos as a practical application of DL-STEM integration. This activity encouraged students to move beyond theoretical definitions and mathematical calculations of temperature and heat by applying these concepts to real-world problem solving. Through this task, students not only learned the meaning of heat transfer in an abstract way but also applied the concepts to determine how a container could maintain temperature more effectively. Such activities can actively engage students in investigating problems, designing solutions, developing models, and presenting their ideas (Urdanivia-Alarcon et al., 2023). Students were challenged to select appropriate materials, minimize heat transfer, monitor temperature changes, and evaluate the effectiveness of their designs.

The thermos design activity also helped students connect the topic's main concepts to observable phenomena. For example, students needed to consider how insulating materials could reduce heat loss, how temperature changes could be measured, and how design choices could influence the product's performance. In this process, students were encouraged to use scientific concepts as a basis for decision-making. The engineering component was assessed formatively by considering the rationale for the design, the selection of insulating materials, and students' ability to explain how their designs inhibited conduction, convection, and radiation. This means that the engineering task functioned as a learning activity that supported conceptual understanding and scientific literacy rather than as a separate product-performance assessment.

However, students' project outcomes were not analyzed as a separate research variable. The products created by students were used in learning activities and to support the assessment of creativity. Therefore, they were discussed as part of the instructional process rather than as independent outcomes analyzed statistically. This distinction is important because the study's main outcome was students' scientific literacy, measured using the scientific literacy test. Thus, the thermos project should be interpreted as an instructional context that helped students apply concepts and evidence, not as an additional measured variable.

During the stimulation stage, students were introduced to contextual problems related to everyday temperature and heat phenomena to encourage curiosity and activate prior knowledge. These contextual problems were important because the topic of temperature and heat is closely connected to daily experiences, such as hot drinks, insulating materials, and heat transfer in household objects. By beginning the lesson with familiar phenomena, students were expected to recognize that physics concepts are not isolated from real life. This stage also provided an entry point for students to observe problems before moving into a more structured investigation.

In the problem statement stage, students formulated questions and problems related to heat concepts and the performance of insulating materials. This stage encouraged students to clarify what needed to be investigated and why the problem was scientifically relevant. Data collection was conducted through group observations and experiments, while data processing required students to classify data, calculate temperature changes, and compare results across designs. These activities helped students engage directly with data and relate it to the effectiveness of the simple thermos design. As a result, the learning process gave students opportunities to connect measurement, interpretation, and conceptual explanation.

The verification stage was used to examine the consistency between students' initial hypotheses and the data obtained during the activities. At this stage, students discussed why certain designs were more effective than others in maintaining temperature. This process was important because students had to compare their assumptions with empirical evidence. If the data did not fully support their initial ideas, students needed to reconsider their explanations and relate them to the principles of heat transfer again. The generalization stage was then used to formulate general principles of heat transfer and relate them to simple technological products. This sequence strengthened the connection between the Discovery Learning syntax and the characteristics of integrated STEM, which places inquiry, engineering design, technology, and mathematical reasoning within a single learning experience (Azizah & Setyawarno, 2025; Kelley & Knowles, 2016; Roehrig et al., 2021).

The results of this study show that the experimental class obtained greater improvement than the control class across all scientific literacy indicators. The experimental class achieved a posttest mean score of 64.69, while the control class achieved a posttest mean score of 53.93. The increase in the experimental class was also larger, from 41.86 in the pretest to 64.69 in the posttest, with a gain of 22.83 points. In comparison, the control class increased from 41.72 to 53.93, with a gain of 12.21 points. These results indicate that the DL-STEM sequence provided stronger support for students' development of scientific literacy than conventional instruction in this two-class setting.

The statistical test also confirmed the difference between the two groups after the intervention. The independent-samples t-test showed a significance value of $p < 0.001$, indicating that the posttest difference between the experimental and control groups was statistically significant. In addition, the effect size value of $r = 0.47$ was classified as moderate. This means that DL-STEM produced a meaningful practical effect on students' scientific literacy, although the effect was not large. Therefore, the findings should be interpreted proportionally. DL-STEM was effective in supporting better outcomes, but the learning implementation still had room for improvement.

Nevertheless, the N-Gain results, which remained in the moderate category, indicate that DL-STEM implementation still requires further strengthening. First, the stimulation stage should use phenomena that are closer to students' everyday experiences and that can create cognitive conflict. Cognitive stimulation can increase student engagement (Akdeniz, 2024; Lee & Lee, 2025). In this study, contextual phenomena were already used, but stronger cognitive conflict may help students question their initial understanding more deeply. This is especially important in the topic of temperature and heat, where students may have misconceptions about heat, temperature, insulation, and heat transfer processes.

Second, worksheets should guide students to explicitly write claims, evidence, and reasoning so that the indicator of using scientific evidence develops more strongly. Although students collected and processed data, the moderate N-Gain score suggests they still needed more structured support in explaining how the data could justify their conclusions. Explicit guidance on claim-evidence-reasoning would help students move from simply recording observations to constructing scientific explanations. This improvement is particularly relevant because the indicators of using scientific evidence and identifying scientific statements require students to evaluate information critically and support conclusions with appropriate data.

Third, the simple thermos design process should include a more explicit rubric that covers aspects such as temperature decrease, material efficiency, design explanation, and redesign decisions. A clearer rubric would help students understand the criteria for evaluating their designs and encourage them to connect design performance with scientific principles. With these improvements, STEM activities would not only lead students to produce a product but also train them in scientific argumentation and data-based decision-making. According to the National Academies of Sciences, Engineering, and Medicine (2025), STEM learning should help students understand natural processes and technological innovation.

Overall, the results show that DL-STEM produced greater improvements in scientific literacy than conventional instruction across all indicators in this two-class setting. The main contribution of this approach is that it provides learning experiences that enable students to discover concepts through phenomena, test ideas with data, and apply concepts to real-world problem-solving. This finding supports previous studies showing the usefulness of Discovery Learning and STEM integration in science learning (Amarlita & Saija, 2025; Azizah & Setyawarno, 2025; Fadlina et al., 2021; Karan, 2023; Maulana et al., 2024; Rahman et al., 2022). At the same time, this study adds a more specific account of how temperature-and-heat learning can connect scientific explanation, evidence use, and engineering design.

In the context of physics education, the findings suggest that learning about temperature and heat through DL-STEM can make abstract concepts more accessible by linking them to measurable, observable phenomena. The simple thermos task provided a bridge between physics concepts and practical application, while the Discovery Learning stages guided students from observation to generalization. Therefore, the strength of DL-STEM in this study lies not only in the use of an engineering product but also in the integration of scientific inquiry, data interpretation, and the application of concepts within a coherent instructional sequence. However, because the study used two intact classes from one school, the findings should be interpreted within the context of the research setting and should not be generalized too broadly without further studies.

V. CONCLUSION AND SUGGESTION

This study found that the DL-STEM was associated with higher students' scientific literacy in the topic of temperature and heat than conventional instruction in a quasi-experimental setting. The initial abilities of the control and experimental groups were relatively similar, but after the treatment, the experimental group achieved a higher posttest score than the control group. The hypothesis test showed a significant difference between the two classes ($p < 0.001$), with a moderate effect size ($r = 0.47$). The N-Gain results also showed that all scientific literacy indicators in the experimental class improved in the moderate category and were higher than those in the control class. These findings indicate that DL-STEM can help students connect temperature and heat concepts to real phenomena, use scientific evidence, and apply physics principles in simple problem-solving activities.

This study has several limitations. It used two intact classes at one school, so the findings should not be generalized to other learning contexts without further investigation. The study also did not include random assignment, detailed treatment-fidelity observation, internal consistency analysis, distractor analysis, or separate statistical assessment of students' product-design outcomes. Future studies are recommended to involve broader, more diverse samples; use randomized or stronger quasi-experimental designs where possible; report instrument reliability and item analyses in greater detail; include treatment-fidelity measures; and examine product design performance as a separate outcome. Despite these limitations, this study contributes to physics education by providing empirical evidence on how STEM can be integrated into the Discovery Learning syntax for temperature-and-heat instruction. It also shows how physics learning can connect concept discovery, evidence-based reasoning, and simple engineering design to support students' scientific literacy.

REFERENCES

Abdi, A. I., Omar, A. M., Mahdi, A. O., Asiimwe, C., & Osman, M. A. (2024). Tracing the evolution of STEM education: A bibliometric analysis. *Frontiers in Education, 9*, 1-12. https://doi.org/10.3389/feduc.2024.1457938

Akdeniz, C. (2024). Effects of cognitive stimulation, physical arrangement of the learning environment, and instructional tendency on student engagement. *International Journal of Curriculum and Instruction, 16*(3), 492–516. https://ijci.net/index.php/IJCI/article/view/1502

Amarlita, D. M., & Saija, M. (2025). The effect of STEM with a discovery learning model on learning motivation on reaction rate concept. *AIP Conference Proceedings, 3206*(1). https://doi.org/10.1063/5.0259765

Ammar, M., Al-Thani, N. J., & Ahmad, Z. (2024). Role of pedagogical approaches in fostering innovation among K-12 students in STEM education. *Social Sciences & Humanities Open, 9*, 1-13. https://doi.org/10.1016/j.ssaho.2024.100839

Asriadi, A., & Lazulva, L. (2021). Desain dan uji coba video pembelajaran berbasis literasi sains dengan menggunakan Scratch pada materi kesetimbangan kimia. *Journal of Research and Education Chemistry, 3*(2), 143-156. https://doi.org/10.25299/jrec.2021.vol3(2).7921

Azizah, G. A. G., & Setyawarno, D. (2025). The effect of integrated STEM discovery learning model on computational thinking skills in heat and its conversion material for grade VII students at SMP Negeri 1 Kalasan. *Journal of Science Education Research, 9*(2), 107–117. https://doi.org/10.21831/jser.v9i2.88453

Batubara, Y., Marjanah, M., & Mahyuny, S. R. (2025). Peningkatan literasi sains melalui penerapan model pembelajaran learning cycle (5E) berbasis STEM (*Science, Technology, Engineering, and Mathematics*) di SMAN 3 Langsa. *Jurnal Ilmiah Profesi Pendidikan, 10*(2), 1032–1037. https://doi.org/10.29303/jipp.v10i2.3210

Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. NSTA Press.

Çalik, M., & Wiyarsi, A. (2024). The effect of socio-scientific issues-based intervention studies on scientific literacy: A meta-analysis study. *International Journal of Science Education, 47*(3), 399-421. https://doi.org/10.1080/09500693.2024.2325382

Cao, X., Lu, H., Wu, Q., & Hsu, Y. (2025). Systematic review and meta-analysis of the impact of STEM education on students learning outcomes. *Frontiers in Psychology, 16*, 1-15. https://doi.org/10.3389/fpsyg.2025.1579474

Coppi, M., Fialho, I., & Cid, M. (2024). Assessing scientific literacy: A study with 9th grade students in Portugal. *Frontiers in Education, 9*, 1-14. https://doi.org/10.3389/feduc.2024.1433919

De Loof, H., Boeve-de Pauw, J., & Van Petegem, P. (2022). Integrated STEM education: The effects of a long-term intervention on students' cognitive performance. *European Journal of STEM Education, 7*(1), 1-17. https://eric.ed.gov/?id=EJ1371646

Demirkol, K., Kartal, B., & Taşdemir, A. (2022). The effect of teachers' attitudes towards and self-efficacy beliefs regarding STEM education on students' STEM career interests. *Journal of Science Learning, 5*(2), 204–216. https://doi.org/10.17509/jsl.v5i2.43991

Fadlina, F., Artika, W., Khairil, K., Nurmaliah, C., & Abdullah, A. (2021). Penerapan model discovery learning berbasis STEM pada materi sistem gerak untuk meningkatkan keterampilan berpikir kritis. *Jurnal Pendidikan Sains Indonesia, 9*(1), 99–107. https://jurnal.usk.ac.id/JPSI/article/view/18591

Fuadi, H., Robbia, A. Z., Jamaluddin, J., & Jufri, A. W. (2020). Analisis faktor penyebab rendahnya kemampuan literasi sains peserta didik. *Jurnal Ilmiah Profesi Pendidikan, 5*(2), 108–116. https://jipp.unram.ac.id/index.php/jipp/article/view/122

Halawa, S., Lin, T. C., & Hsu, Y. S. (2024). Exploring instructional design in K–12 STEM education: A systematic literature review. *International Journal of STEM Education, 11*(43), 1-15. https://doi.org/10.1186/s40594-024-00503-5

Hardianti, F., Setiadi, D., Syukur, A., & Merta, I. W. (2021). Pengembangan bahan ajar berbasis SETS (*science, environment, technology, society*) untuk meningkatkan literasi sains peserta didik. *Jurnal Pijar Mipa, 16*(1), 68–74. https://doi.org/10.29303/jpm.v16i1.1636

Ibrahim, A., Supartin, S., & Samatowa, L. (2025). Meningkatkan literasi sains peserta didik dengan menggunakan pendekatan STEM (*science, technology, engineering, and mathematics*) pada materi gelombang bunyi. *JPFT: Jurnal Pendidikan Fisika Tadulako Online, 13*(2), 212–219. https://doi.org/10.22487/jpft.v13i2.4657

Karan, E. (2023). Discovery-based approach combined with active learning to improve student learning experiences for STEM students. *Journal of Education and Training Studies, 11*(4), 16–25. https://doi.org/10.11114/jets.v11i4.6205

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*,(11), 1-11. https://doi.org/10.1186/s40594-016-0046-z

Kumar, V., & Choudhary, S. K. (2025). Reimagining scientific literacy: A textbook framework for future-focused science education. *Research in Science Education, 55*, 1109–1127. https://doi.org/10.1007/s11165-025-10269-7

Lee, M. Y., & Lee, J. S. (2025). Project-based learning as a catalyst for integrated STEM education. *Education Sciences, 15*(7), 1-8. https://doi.org/10.3390/educsci15070871

Mansour, N., Said, Z., & Abu-Tineh, A. (2024). Factors impacting science and mathematics teachers' competencies and self-efficacy in TPACK for PBL and STEM. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(5), 1-17. https://doi.org/10.29333/ejmste/14467

Maulana, M., Rosmaiyadi, R., & Kariadi, D. (2024). Pengaruh model discovery learning untuk meningkatkan literasi sains siswa. *Journal of Educational Review and Research, 7*(1), 34–49. https://journal.stkipsingkawang.ac.id/index.php/JERR/article/view/5919

Mellyzar, M., Zahara, S. R., & Alvina, S. (2022). Literasi sains dalam pembelajaran sains siswa SMP. *Pendekar: Jurnal Pendidikan Berkarakter, 5*(2), 119-124. https://doi.org/10.31764/pendekar.v5i2.10097

National Academies of Sciences, Engineering, and Medicine. (2025). *Transforming undergraduate STEM education; A framework for institutions*. Educators and Disciplines. https://www.nationalacademies.org/projects/DBASSE-BOSE-22-04/publication/28268

Nisah, K., Saminan, S., Syukri, M., Elisa, E., & Markisni, M. (2024). Optimizing of physics learning through PjBL-STEM model to improve critical thinking skills and students responsibility attitudes. *Jurnal Penelitian Pendidikan IPA, 10*(4), 1770–1778. https://doi.org/10.29303/jppipa.v10i4.6795

OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

OECD. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023b). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2023c). *PISA 2022 (Volume I and II) - Country notes: Indonesia*. OECD Publishing. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html

Osborne, J., & Allchin, D. (2025). Science literacy in the twenty-first century: Informed trust and the competent outsider. *International Journal of Science Education, 47*(15–16), 2134–2155.https://doi.org/10.1080/09500693.2024.2331980

Portillo-Blanco, A., Deprez, H., De Cock, M., Guisasola, J., & Zuza, K. (2024). A systematic literature review of integrated STEM education: Uncovering consensus and diversity in principles and characteristics. *Education Sciences, 14*(9), 1-24. https://doi.org/10.3390/educsci14091028

Putri, M. D., Fitria, D., Nurlaini, N., & Berutu, N. J. (2026). Validity and reliability of science literacy assessment instruments for measuring science competencies in the context of PISA 2025 using the Rasch model. *Jurnal Pendidikan Fisika, 14*(1), 119–137. https://journal.unismuh.ac.id/index.php/jpf/article/view/19809/10351

Rahman, M. H., Latif, S., & Saban, M. (2022). Implementasi model discovery learning untuk meningkatkan kemampuan literasi sains siswa kelas XI MAN 2 Halmahera Utara. *Jurnal Pendidikan Fisika FKIP UM Metro, 10*(2), 259–270. https://doi.org/10.24127/jpf.v10i2.5660

Rahmi, A., Zahara, S. R., Alvina, S., Fadli, M. R., & Juliana, E. (2024). Analysis of students' scientific literacy abilities on science learning in high school. *Proceedings of Malikussaleh International Conference on Multidisciplinary Studies (MICoMS), 4*, 1-6. https://doi.org/10.29103/micoms.v4i.897

Roehrig, G. H., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research, 3*(11), 1-18. https://doi.org/10.1186/s43031-021-00041-y

Roy, G., Sikder, S., & Danaia, L. (2025). Adopting scientific literacy in early years from empirical studies on formal education: A systematic review of the literature. *International Journal of STEM Education, 12*(26), 1-24. https://doi.org/10.1186/s40594-025-00547-1

Satriana, A. (2023). Peningkatan literasi data melalui model discovery learning dengan pendekatan ilmu pengetahuan alam, teknologi, rekayasa dan matematika (STEM). *LOKAKARYA, 2*(1), 41–55. https://doi.org/10.30821/lokakarya.v2i1.2754

Sudarto, S. (2026). Factors causing low the scientific ability of Indonesian children: A systematic review. *Pendas: Jurnal Ilmiah Pendidikan Dasar, 11*(1), 145–160. https://journal.unpas.ac.id/index.php/pendas/article/view/43557

Urdanivia-Alarcon, D. A., Talavera-Mendoza, F., Rucano Paucar, F. H., Cayani Caceres, K. S., & Machaca Viza, R. (2023). Science and inquiry-based teaching and learning: A systematic review. *Frontiers in Education, 8*, 1-10. https://doi.org/10.3389/feduc.2023.1170487

Usta, Z. B., Mertoglu, H., & Akgül, E. (2025). Examining prospective teachers' scientific literacy and STEM efficacy beliefs. *Science Insights Education Frontiers, 30*(1), 4825–4840. https://doi.org/10.15354/sief.25.or824

Wulandari, S., Yuliani, H., & Azizah, N. (2023). Pengaruh e-modul berbasis discovery learning (DL) terhadap literasi sains siswa pada materi gelombang bunyi. *LAMBDA: Jurnal Ilmiah Pendidikan MIPA dan Aplikasinya, 3*(2), 72-77. https://doi.org/10.58218/lambda.v3i2.552

Zhan, Z., & Niu, S. (2023). Subject integration and theme evolution of STEM education in K–12 and higher education research. *Humanities and Social Sciences Communications, 10*, 1-13. https://doi.org/10.1057/s41599-023-02303-8

Zhang, L., Liu, X., & Feng, H. (2023). Development and validation of an instrument for assessing scientific literacy from junior to senior high school. *Disciplinary and Interdisciplinary Science Education Research, 5*(21), 1-15. https://doi.org/10.1186/s43031-023-00093-2