

Think, Retrieve, Master: An AI-Assisted Intervention for Thermodynamics Conceptual Mastery in Physics 2

Mariane L. Alleta

Master Teacher II, Schools Division Office of Antipolo DLC I-A Antipolo City Senior High School

DOI: <https://doi.org/10.47772/IJRISS.2026EDU0562>

Received: 27 August 2026; Accepted: 01 September 2026; Published: 10 September 2026

ABSTRACT

This study examined changes in thermodynamics conceptual mastery following the implementation of the Artificial Intelligence (AI)-Assisted Retrieval Model (AI-ARM) among Grade 11 learners in Physics 2. Specifically, it examined learners' conceptual understanding, quantitative reasoning, and application of thermodynamic principles before and after the intervention and documented learners' perceptions of the learning experience. The study employed a mixed-method explanatory sequential design, with a one-group pretest–posttest quantitative component and thematic analysis of qualitative reflections. Nineteen (19) purposively selected learners participated in a four-week intervention conducted during homeroom sessions, integrating structured retrieval practice and AI-assisted explanatory feedback. Data were collected using a 40-item Thermodynamics Conceptual Mastery Test, reflection logs, and a perception survey. Mean performance increased from 24.58 (SD = 5.52) in the pretest to 33.16 (SD = 2.21) in the posttest, a mean increase of 8.58 points. A paired-samples t-test indicated a statistically significant pretest–posttest difference, $t(18) = 6.67$, $p < .001$. For the paired design, Cohen's d_z was estimated as $t/\sqrt{n} = 1.53$, indicating a large standardized mean difference; the 95% confidence interval for the mean difference, based on the reported t statistic, was approximately 5.88 to 11.28 points. Qualitative findings highlighted simplification of complex concepts, conceptual clarity, guided error correction, increased confidence, and improved application. Because the study had no control or comparison group and used a small purposive sample, the observed gains cannot be interpreted as definitive evidence that AI-ARM alone caused the improvement. Repeated testing, normal classroom learning, learner motivation, and other concurrent learning experiences may also have contributed. The findings therefore provide preliminary classroom-based evidence that AI-assisted retrieval is a promising remediation approach that warrants further evaluation using controlled or comparison-group designs.

Keywords: Artificial intelligence (AI), AI-assisted learning, retrieval practice, thermodynamics, conceptual mastery, physics education, remediation

CONTEXT AND RATIONALE

Action research is a systematic and reflective process that enables teachers to examine classroom challenges while implementing evidence-based solutions to improve learner outcomes. Recent scholarship emphasizes that classroom-based action research strengthens instructional responsiveness and promotes data-driven innovation in secondary education settings (Morales & Santos, 2022). Within the Strengthened Senior High School Curriculum, educators are encouraged to design targeted interventions that address measurable learning gaps, particularly in Science, Technology, Engineering, and Mathematics (STEM) disciplines where conceptual mastery is foundational for academic progression and workforce readiness.

In Fourth Quarter Physics 2, learners are expected to demonstrate integrated understanding of thermodynamics across eight competencies, including recalling and applying methods of heat transfer in real-world contexts, explaining temperature measurement in varied systems, interpreting tabular representations of energy flow, calculating heat during temperature and phase changes, investigating thermal expansion quantitatively, describing engine and refrigeration processes using secondary sources, analyzing factors affecting engine efficiency, and applying thermodynamic laws to sustainable energy technologies. These competencies require not only procedural fluency but systems thinking, conceptual coherence, and applied reasoning.

However, classroom assessment data in SY 2025–2026 indicate persistent conceptual difficulties. Students often recall formulas such as $Q = mc\Delta T$ yet struggle to explain the physical meaning of heat transfer processes or distinguish clearly between heat and temperature in applied contexts. Similar patterns have been documented in recent physics education research, which shows that misconceptions in thermodynamics remain prevalent even after instruction and require deliberate conceptual reinforcement (Singh & Marshman, 2021). Learners also demonstrate difficulty interpreting energy movement across environmental and technological systems, particularly when presented in tabular or data-driven formats. Quantitative problems involving phase changes and thermal expansion frequently reveal computational errors rooted in fragmented conceptual understanding.

This situation is significant because thermodynamics principles underpin modern technological systems, environmental sustainability initiatives, and energy efficiency innovations. As global emphasis on renewable energy and sustainable technologies increases, learners must be able to apply thermodynamic laws to real-world systems such as heat engines, refrigeration cycles, and solar thermal technologies. Recent research in science education highlights the importance of strengthening systems-based reasoning to support students' ability to connect scientific principles with sustainability contexts (Kelley & Knowles, 2021). Without strong conceptual foundations, students may struggle to transfer classroom knowledge to higher-level STEM studies and emerging technological fields.

Contemporary cognitive science continues to identify retrieval practice as a high-impact strategy for strengthening durable learning and conceptual organization. A recent meta-analysis confirms that structured retrieval significantly improves retention and transfer compared to passive review strategies (Yang, Luo, Vadillo, Yu, & Shanks, 2021). Retrieval practice enhances long-term consolidation of knowledge and promotes deeper conceptual integration when combined with feedback. Moreover, feedback that is timely, explanatory, and corrective has been shown to significantly improve student learning outcomes, particularly in complex scientific domains (Wisniewski, Zierer, & Hattie, 2021).

The integration of artificial intelligence in education provides new opportunities to operationalize retrieval practice with immediate and adaptive feedback. AI-assisted retrieval intervention strategies can generate instant explanations, identify misconception patterns, and scaffold student reasoning in real time. Recent studies suggest that AI-supported instructional tools, when intentionally integrated into pedagogical design, can enhance engagement, personalization, and conceptual clarification (Zawacki-Richter et al., 2022). However, despite increasing global research on AI in education, there remains limited classroom-based action research within public secondary school contexts examining how AI-assisted retrieval strategies specifically improve thermodynamics conceptual mastery.

In response to these identified gaps, this study implemented an AI-assisted retrieval intervention designed to support thermodynamics conceptual mastery in Physics 2. The intervention integrated structured retrieval tasks, AI-generated explanatory feedback, cumulative mastery monitoring, and reflective correction processes aligned with the targeted competencies. The study was designed as classroom-based action research; consequently, the purpose was to examine whether measurable changes occurred in the participating learners and to understand their reported learning experiences, rather than to establish a definitive causal effect of AI-ARM independent of other learning influences.

The significance of this research lies in its potential to provide an evidence-informed instructional model that Physics teachers can examine and adapt when addressing conceptual gaps in thermodynamics and related STEM domains. The study contributes classroom-based evidence on the combined use of retrieval practice and AI-assisted explanatory feedback. However, the findings should be interpreted within the limits of the one-group design, purposive sampling, small sample size, and concurrent classroom learning. These limitations are explicitly considered in the interpretation and recommendations.

Action Research Questions

This study aimed to examine changes in thermodynamics conceptual mastery following the implementation of the AI-Assisted Retrieval Model (AI-ARM) among Grade 11 learners in Physics 2. Specifically, it sought to answer the following questions:

1. What is the level of learners' thermodynamics conceptual mastery before and after the intervention in terms of:
 - 1.1 conceptual understanding of heat transfer and temperature measurement;
 - 1.2 quantitative reasoning in heat calculations, phase change, and thermal expansion; and
 - 1.3 application of thermodynamic principles to engines, refrigeration systems, and sustainable energy technologies?
2. Is there a significant difference between the pretest and posttest scores of learners after the implementation of the AI-Assisted Retrieval Model?
3. What is the magnitude of effect of the AI-Assisted Retrieval Model on learners' thermodynamics conceptual mastery?

Innovation, Intervention, and Strategy

To address the identified gaps in thermodynamics conceptual mastery, this study implemented the AI-Assisted Retrieval Model (AI-ARM) as a targeted remedial intervention. The intervention was designed as a focused academic support mechanism rather than a replacement for regular classroom instruction and was conducted during scheduled homeroom periods. It provided additional opportunities for learners who required scaffolding beyond regular instructional activities.

The intervention responded directly to persistent learning difficulties observed in Physics 2, particularly learners' inability to articulate conceptual reasoning, their tendency to rely on formula memorization without understanding, and their difficulty applying thermodynamic principles in real-world contexts. Common misconceptions included confusion between heat and temperature, challenges in interpreting energy transfer across systems, errors in quantitative problem-solving involving phase change and thermal expansion, and limited ability to connect thermodynamic laws to engines, refrigeration systems, and sustainable energy technologies. These patterns highlighted the need for structured, small-group academic reinforcement beyond regular instructional time.

The AI-Assisted Retrieval Model (AI-ARM) integrated systematic retrieval practice with AI-assisted explanatory feedback to strengthen conceptual understanding, quantitative reasoning, and applied analysis. During each homeroom session over a four-week period, learners engaged in 30–40 minutes of structured retrieval activities. These included conceptual multiple-choice questions, short explanatory prompts, guided table-completion tasks for energy systems analysis, and scaffolded quantitative problem-solving exercises. The retrieval process required learners to actively recall prior knowledge, confront misconceptions, and apply principles across varied contexts.

Following individual responses, the AI-assisted tool generated immediate explanatory feedback. The feedback protocol was intended to move beyond simply indicating whether an answer was correct or incorrect by clarifying reasoning, identifying possible misconceptions, modelling solution processes, and reinforcing conceptual connections. Because AI-generated content can contain inaccuracies or misleading explanations, learner responses and AI feedback were subject to teacher supervision and correction.

The intervention was organized into weekly competency clusters aligned with identified areas of difficulty. The first phase focused on heat transfer, temperature measurement, and interpretation of energy movement across systems. The second phase emphasized quantitative reasoning in heat calculations and phase change. The third phase targeted thermal expansion and factors affecting engine efficiency. The final phase integrated higher-order applications of thermodynamic laws in engines, refrigeration systems, and sustainable energy technologies. Throughout the program, learners maintained structured reflection logs to document misconceptions, corrections, and evolving understanding, thereby promoting metacognitive awareness and academic confidence.

The intervention was examined using a one-group pretest–posttest design. A validated Thermodynamics Conceptual Mastery Test was administered before and after the four-week intervention period. Mean scores, proficiency distributions, paired-samples t-test results, and an appropriate paired-sample effect size were used to describe the magnitude and statistical significance of the observed change. These analyses identify change within the participating group.

Ultimately, the AI-Assisted Retrieval Model (AI-ARM) functioned as a structured academic remediation program embedded within homeroom support time. By combining retrieval practice, explanatory feedback, and reflective learning, the model was intended to provide additional opportunities for learners to retrieve, explain, correct, and apply thermodynamics concepts.

Action Research Methods

This study employed a mixed-method explanatory sequential design. The quantitative component used a one-group pretest–posttest approach to examine changes in thermodynamics conceptual mastery, followed by qualitative analysis of learners' reflections and perceptions to contextualize the observed changes. The design was appropriate for a classroom-based action research inquiry in which the researcher sought to examine an instructional response to an identified learning need.

The quantitative component utilized a one-group pretest–posttest design. Nineteen learners completed a pretest before the four-week intervention and a posttest after the intervention. A paired-samples t-test was used because the analysis compared scores from the same learners. The unit of inferential analysis was the within-learner difference between posttest and pretest scores. Random assignment and a comparison group were not feasible because the intervention was implemented as targeted remedial support for learners who had identified learning gaps. Accordingly, the statistical findings are interpreted as evidence of within-group change rather than proof of causal attribution.

Following the quantitative phase, qualitative data were collected through structured reflection logs and open-ended perception surveys. These instruments examined how retrieval practice cycles influenced conceptual understanding and how AI-assisted feedback supported misconception correction and confidence in quantitative problem solving. Thematic analysis was employed to identify recurring patterns that helped explain and contextualize the quantitative findings.

Integration of findings occurred during interpretation. Quantitative results described the magnitude and statistical significance of within-group change, while qualitative evidence provided learner-reported explanations and experiences that may help contextualize the observed gains. The qualitative findings were therefore treated as complementary evidence rather than independent proof that AI-ARM caused the quantitative improvement.

A. Participants and/or other Sources of Data/Information

Participants consisted of Grade 11 learners enrolled in Fourth Quarter Physics 2 at Antipolo City Senior High School who met the inclusion criteria for the remedial intervention. Learners who obtained below the 75% mastery benchmark in thermodynamics-related summative assessments and demonstrated recurring misconceptions based on item analysis were invited to participate. Informed consent was secured from all participants, with parental consent obtained for minors. Learners who had already achieved mastery or required specialized interventions beyond the scope of the program were excluded.

The study employed purposive criterion-based sampling because participants were selected using predefined performance indicators aligned with the remedial objective. Nineteen (19) learners met the inclusion criteria and completed the study. The small, non-random sample limits the generalizability of the findings beyond the participating learners and school context. The sample should therefore be understood as an action-research group rather than a statistically representative sample of Grade 11 Physics 2 learners.

For the qualitative phase, all participants completed structured reflection logs and perception surveys.

This sampling approach supported the equity-oriented purpose of the intervention and ensured alignment between the research design and the instructional problem identified through classroom assessment data.

B. Data Gathering Methods

The primary quantitative instrument was a 40-item Thermodynamics Conceptual Mastery Test developed by the researcher to measure the eight targeted competencies. The test included conceptual multiple-choice items, scenario-based reasoning tasks, and structured quantitative problem-solving items. A parallel form was used for the pretest and posttest. The instrument's content validity was established through expert review by Physics teachers and a research coordinator.

Instrument validation included expert review for content alignment and pilot testing for item analysis.

Qualitative data were collected through structured reflection logs and a post-intervention perception survey. Reflection logs asked learners to identify misconceptions, describe corrected reasoning, and explain how feedback affected their understanding. The perception survey captured learners' experiences with retrieval activities, perceived changes in understanding and confidence, and remaining challenges. These responses were used as learner-reported evidence to contextualize the quantitative results.

Supplementary data sources included retrieval practice records and weekly mastery check summaries. These classroom artifacts were analyzed to monitor progression across competencies and to support data triangulation. Data collection followed a structured sequence: administration of the pretest, implementation of the four-week intervention, ongoing collection of reflection logs, administration of the posttest, and completion of the perception survey.

Ethical standards were observed throughout the study. Participation was voluntary, informed consent was secured, and parental consent was obtained for minors. Data were coded and stored securely, and personally identifiable information was excluded from reporting. Because participants were minors and an AI tool was used, responsible-use safeguards were necessary: learners were instructed not to enter personally identifying or sensitive information into the AI system; teacher supervision was maintained; AI-generated explanations were checked against established Physics content; and learners were reminded that AI output could be inaccurate and should not be treated as an unquestionable authority.

Quantitative data were analyzed using descriptive and inferential statistics. Descriptive measures included mean scores, standard deviations, MPS, and proficiency distributions. A paired-samples t-test was used to examine whether the mean within-learner difference between posttest and pretest scores differed significantly from zero. The normality assumption should be assessed on the difference scores. Based on the reported $t(18) = 6.67$ and mean difference of 8.58, the 95% confidence interval for the mean difference is approximately 5.88 to 11.28 points.

For the paired design, the standardized effect was expressed as Cohen's d_z , calculated as the mean paired difference divided by the standard deviation of the paired differences. An equivalent calculation is $d_z = t/\sqrt{n}$. Using the reported $t = 6.67$ and $n = 19$ gives $d_z = 1.53$, a large standardized within-group difference.

Qualitative data were analyzed using thematic analysis. Reflection-log and survey responses were organized, read repeatedly, and coded for recurring meanings related to conceptual understanding, error correction, confidence, and application. Codes were grouped into categories and refined into themes. Credibility was supported through analysis of the available learner reflections, repeated comparison of codes with the original responses, retention of illustrative excerpts, and triangulation with quantitative results. The qualitative component was intended to contextualize the quantitative findings rather than generate a standalone theory.

Mixed-method integration occurred by comparing the direction and content of quantitative changes with themes emerging from learner reflections. The qualitative evidence was used to contextualize possible learning processes and learner experiences.

DISCUSSION OF RESULTS AND REFLECTION

This section presents the quantitative and qualitative findings and interprets them in relation to the research questions and relevant literature. Because the study used a one-group pretest–posttest design with 19 purposively selected learners, the results are interpreted as observed within-group changes. The findings do not establish that AI-ARM alone caused the improvement.

Level of Learners' Thermodynamics Conceptual Mastery Before and After the AI-Assisted Retrieval Model

The level of learners' thermodynamics conceptual mastery was examined across three domains: conceptual understanding, quantitative reasoning, and application of principles.

Table 1 Conceptual Understanding of Heat Transfer and Temperature Measurement

Topic	Pretest			Posttest		
	Mean	MPS	Proficiency Level	Mean	MPS	Proficiency Level
Methods of Heat Transfer	6.16	76.97	Proficient	7.74	96.71	Highly Proficient
Temperature Measurement	5.05	63.16	Nearly Proficient	6.47	80.92	Proficient
Overall Mean	5.61	70.07	Nearly Proficient	7.11	88.82	Proficient

Table 1 presents learners' pretest and posttest performance in Methods of Heat Transfer and Temperature Measurement. Both areas showed higher posttest means than pretest means. Methods of Heat Transfer increased from 6.16 (76.97%) to 7.74 (96.71%), while Temperature Measurement increased from 5.05 (63.16%) to 6.47 (80.92%). The overall mean increased from 5.61 (70.07%) to 7.11 (88.82%). These descriptive changes indicate stronger performance after the four-week period, although the design does not permit the gain to be attributed uniquely to AI-ARM.

In the pretest, learners obtained a mean score of 6.16 with an MPS of 76.97% in Methods of Heat Transfer, interpreted as Proficient. This indicates that learners possessed a relatively adequate understanding of conduction, convection, and radiation prior to the intervention, although misconceptions and conceptual inconsistencies were still evident. In contrast, Temperature Measurement obtained a lower mean score of 5.05 and an MPS of 63.16%, interpreted as Nearly Proficient, suggesting that learners experienced greater difficulty distinguishing concepts related to heat and temperature, interpreting temperature scales, and explaining thermal behavior in different systems.

The pretest pattern suggests that temperature measurement was a greater area of difficulty than methods of heat transfer. The improvement in both areas is consistent with the intended focus of the retrieval activities, which repeatedly required learners to recall, explain, and apply core concepts. However, repeated exposure to the same content, regular Physics instruction, and familiarity with the assessment may also have contributed to the observed increase.

Following the intervention period, both competency areas showed higher mean scores. The descriptive increase is consistent with the possibility that structured retrieval and explanatory feedback supported learning, but the one-group design requires a cautious interpretation. The results show that learners performed better on the posttest; they do not isolate which component of the learning experience produced the change.

The observed pattern is compatible with research on retrieval practice, which emphasizes repeated active recall as a mechanism for strengthening retention and conceptual organization. The AI-assisted feedback may also have provided timely explanations that helped learners revisit errors. Nevertheless, these mechanisms remain plausible explanations rather than experimentally established causes in the present study.

Learner reflections also indicated that explanatory feedback helped some participants identify and correct errors. This qualitative evidence provides useful context for the quantitative gain, but it represents learners' perceptions and should not be interpreted as independent causal evidence.

Overall, the results show improved descriptive performance in heat transfer and temperature measurement after the intervention period. The findings are promising for classroom remediation.

Table 2 Quantitative Reasoning in Heat Calculations, Phase Change, and Thermal Expansion

Topic	Pretest			Posttest		
	Mean	MPS	Proficiency Level	Mean	MPS	Proficiency Level
Heat Calculations	5.11	63.82	Nearly Proficient	5.63	70.39	Nearly Proficient
Thermal Expansion	6.42	80.26	Proficient	6.58	82.24	Proficient
Overall Mean	5.77	72.04	Nearly Proficient	6.11	76.32	Proficient

Table 2 presents learners' pretest and posttest performance in Heat Calculations and Thermal Expansion. Heat Calculations increased from 5.11 (63.82%) to 5.63 (70.39%), while Thermal Expansion increased from 6.42 (80.26%) to 6.58 (82.24%). The overall mean increased from 5.77 (72.04%) to 6.11 (76.32%).

In the pretest, learners obtained a mean score of 5.11 with an MPS of 63.82% in Heat Calculations, interpreted as Nearly Proficient. This result indicates that learners experienced difficulty solving quantitative problems involving heat transfer, specific heat capacity, and phase change computations. Many learners demonstrated challenges in identifying the correct formula, substituting variables accurately, and interpreting relationships among thermodynamic quantities. These findings suggest that procedural problem-solving skills were not yet fully developed prior to the intervention.

In contrast, Thermal Expansion obtained a higher pretest mean score of 6.42 with an MPS of 80.26%, interpreted as Proficient. This indicates that learners showed relatively stronger understanding in solving problems related to linear expansion and changes in material dimensions due to temperature variation. However, despite achieving proficiency, some learners still committed computational and analytical errors, particularly in multi-step problem-solving tasks requiring careful interpretation of variables and units.

The overall pretest result indicates that quantitative reasoning was a relative area of need, particularly in heat calculations. Learners' difficulties with formula selection, variable substitution, units, and interpretation are consistent with the broader literature on conceptual and procedural difficulties in thermodynamics.

Following the intervention, improvement was observed across both competency areas. Heat Calculations increased to a mean score of 5.63 with an MPS of 70.39%, while Thermal Expansion improved to a mean score of 6.58 with an MPS of 82.24%. Although Heat Calculations remained within the Nearly Proficient level, the increase in mean performance indicates that learners developed better understanding of quantitative procedures and improved accuracy in solving computation-based problems. Meanwhile, Thermal Expansion maintained its Proficient classification and demonstrated further improvement in learner performance.

The overall posttest mean increased from 5.77 to 6.11, and the MPS increased from 72.04% to 76.32%, moving the descriptive proficiency classification from Nearly Proficient to Proficient. The gain is educationally relevant, but the modest increase compared with the conceptual domains also suggests that quantitative reasoning may require longer or more intensive practice.

The pattern is consistent with the intended role of explanatory feedback: immediate clarification may help learners inspect solution steps and recognize procedural errors. The observed gain may also reflect regular classroom instruction, repeated practice, or growing familiarity with the tested skills.

Additionally, repeated retrieval practice likely reinforced procedural memory and strengthened learners' ability to transfer conceptual knowledge into quantitative applications. Retrieval-based learning encourages active recall and repeated application of information, which improves both retention and problem-solving performance. Research further suggests that retrieval practice combined with explanatory feedback enhances learners' capacity to apply knowledge flexibly across different contexts and supports deeper cognitive processing.

Overall, learners showed a descriptive improvement in quantitative reasoning, although heat calculations remained below the 75% benchmark. This result suggests that retrieval with explanatory feedback may be useful as supplementary support, while also indicating that additional instructional time may be needed for multi-step quantitative problem solving.

Table 3 Application of Thermodynamic Principles to Engines, Refrigeration Systems, and Sustainable Energy Technologies

Topic	Pretest			Posttest		
	Mean	MPS	Proficiency Level	Mean	MPS	Proficiency Level
Energy Movement in Systems	5.05	63.82	Nearly Proficient	6.74	84.21	Proficient
Overall Mean	5.05	63.82	Nearly Proficient	6.74	84.21	Proficient

Table 3 presents performance in applying thermodynamic principles to energy movement in systems. The mean increased from 5.05 (63.82%) in the pretest to 6.74 (84.21%) in the posttest, shifting the descriptive proficiency classification from Nearly Proficient to Proficient.

During the pretest, learners obtained a mean score of 5.05 with an MPS of 63.82%, interpreted as Nearly Proficient. This result indicates that learners initially experienced difficulty applying abstract thermodynamic principles to practical situations involving energy transfer, system interactions, and real-world thermodynamic processes. Many learners struggled to analyze how heat energy moves within and between systems, interpret energy transformations, and explain the behavior of thermodynamic systems using scientific reasoning. These findings suggest that although learners possessed some foundational conceptual knowledge, they encountered challenges when required to transfer that knowledge into authentic and application-based contexts.

The relatively low pretest performance supports existing literature emphasizing that application-oriented competencies are among the most difficult areas in science learning. According to Kelley and Knowles (2021), learners often struggle to apply scientific principles to real-world systems when instruction focuses heavily on isolated concepts rather than integrated problem-solving experiences. Without sufficient scaffolding and contextualized learning opportunities, students may fail to develop the systems thinking necessary for analyzing complex scientific phenomena.

The posttest result indicates a substantial descriptive increase in application performance. Learners performed better on tasks requiring interpretation of energy movement and thermodynamic systems after the four-week learning period.

One plausible explanation is the intervention's progression from foundational concepts to contextualized applications. Retrieval tasks requiring learners to explain relationships and apply principles may have provided repeated opportunities to connect abstract ideas with concrete systems. This interpretation is also consistent with learner reflections.

Moreover, the AI-assisted feedback component provided immediate clarification and scaffolded explanations that helped learners correct misconceptions and refine analytical reasoning. By receiving step-by-step guidance during retrieval activities, learners were better able to understand how thermodynamic principles operate within engines, refrigeration systems, and energy-transfer processes. This aligns with the argument of Kelley and

Knowles (2021), who emphasized that scaffolded learning experiences are essential for developing systems thinking, contextual understanding, and real-world problem-solving skills in science education.

Overall, the application domain showed a notable descriptive gain and moved above the 75% proficiency benchmark.

Table 4 summarizes the overall pretest and posttest proficiency distribution of the 19 participating learners. The corrected frequencies total 19 in both administrations. In the pretest, 12 learners were Nearly Proficient, 3 were Low Proficient, and 4 were Proficient. In the posttest, 16 were Proficient and 3 were Highly Proficient; no learner remained in the Nearly Proficient or Low Proficient categories.

Table 4 shows a shift toward higher proficiency categories. This descriptive movement is consistent with the increase in the overall mean from 24.58 to 33.16 and the MPS from 61.45% to 82.89%. Because the same learners completed both tests, the distributional shift should be interpreted as a within-group change.

Table 4 Pretest and Posttest Scores of Learners

Range	Proficiency Level	Pretest	Posttest	Increase/ Decrease
36 – 40	Highly Proficient	0	3	+3
30 – 35	Proficient	4	16	+12
20 – 29	Nearly Proficient	12	0	-12
10 – 19	Low Proficient	3	0	-3
0 – 9	Not Proficient	0	0	-
Total		19	19	-
Mean		24.58	33.16	+8.58
MPS		61.45	82.89	+21.44
SD		5.52	2.21	-3.31

Following the four-week intervention period, the number of learners classified as Proficient increased from 4 to 16, while 3 learners reached the Highly Proficient category. The totals demonstrate that all 19 participants were accounted for in both administrations. These results indicate a substantial shift in observed performance, while leaving open the contribution of repeated testing, regular instruction, motivation, and other concurrent learning experiences.

The mean score increased from 24.58 in the pretest to 33.16 in the posttest, an observed mean gain of 8.58 points. The MPS increased from 61.45% to 82.89%, exceeding the 75% proficiency benchmark at posttest. The 95% confidence interval for the mean difference is approximately 5.88 to 11.28 points based on the reported t statistic. These results indicate a statistically and educationally notable within-group improvement.

The standard deviation decreased from 5.52 in the pretest to 2.21 in the posttest. The reduced dispersion suggests that posttest scores were more closely clustered. This descriptive change is interpreted as evidence that the intervention reduced achievement gaps.

Taken together, the descriptive results indicate substantial within-group improvement in thermodynamics test performance. The findings support continued classroom experimentation with AI-assisted retrieval.

Difference Between Pretest and Posttest Scores After the Implementation of the AI-Assisted Retrieval Model

To determine whether the observed improvement was statistically significant, a paired-samples t-test was conducted. Results are presented in Table 5.

The paired-samples t-test showed a statistically significant mean difference between posttest and pretest scores, $t(18) = 6.67$, $p = .000002$, with a mean increase of 8.58 points. This result indicates that the average posttest score was higher than the average pretest score in this sample.

Table 5 Paired Samples t-test Results Comparing Pretest and Posttest Scores

Comparison	Mean Difference	t-value (df)	p-value	95% CI for Mean Difference
Posttest – Pretest	8.58	6.67 (18)	< .001	5.88 to 11.28

The statistical result is consistent with a meaningful increase in scores during the study period. Retrieval practice and explanatory feedback are plausible contributors, consistent with prior research, the statistical significance is therefore described as evidence of within-group change.

This finding is consistent with Yang et al. (2021), who reported that retrieval practice produces significant learning gains across disciplines. The structured retrieval activities in the AI-ARM required learners to actively engage with content, leading to improved retention and understanding.

Moreover, the integration of AI-assisted feedback enhanced the effectiveness of the intervention. Zawacki-Richter et al. (2022) highlight that AI tools provide immediate and adaptive feedback, which contributes to improved learning outcomes.

The alignment between statistical results and learner reflections further strengthens the validity of the findings, demonstrating that the intervention effectively addressed learning gaps.

Magnitude of the Effect of the AI-Assisted Retrieval Model on Learners' Thermodynamics Conceptual Mastery

To determine the practical significance of the intervention, effect size was computed using Cohen's d. The results are shown in Table 6.

Using the paired-sample formulation, Cohen's dz was estimated as the mean paired difference divided by the standard deviation of the paired differences. An equivalent calculation is $dz = t/\sqrt{n}$. Using the reported $t = 6.67$ and $n = 19$ gives $dz = 1.53$. This represents a large standardized within-group difference.

Table 6 Effect Size of the AI-Assisted Retrieval Model

Effect Size	Interpretation
Cohen's $dz = 1.53$	Large within-group standardized difference

The large standardized difference indicates that the magnitude of the observed within-group change was substantial relative to the variability of the paired differences.

Additionally, Yang et al. (2021) emphasize that retrieval practice not only improves retention but also enhances conceptual understanding and transfer, which contributes to large effect sizes in educational interventions.

Thematic Analysis of Learners' Reflections. The qualitative findings provided deeper insight into how the AI-Assisted Retrieval Model (AI-ARM) influenced learners' conceptual understanding and quantitative reasoning

in thermodynamics. Analysis of learners' reflection logs and perception responses generated five major themes: Simplification of Complex Concepts, Conceptual Clarity, Guided Error Correction, Increased Confidence, and Improved Application. These themes collectively explain the substantial improvement observed in the quantitative results and help clarify the mechanisms through which the intervention supported learning.

Table 7 Thematic Analysis of Learners' Reflections

Responses	Theme
"AI helped organize and simplify the explanation."	Simplification of Complex Concepts
"Making the complex topics into simpler ideas."	Conceptual Clarity
"It explains step-by-step so I understand where I went wrong."	Guided Error Correction
"I became more confident solving problems."	Increased Confidence
"I can now connect the lesson to real-life situations."	Improved Application

The first theme, Simplification of Complex Concepts, reflected learners' reports that explanations made difficult material more manageable. For example, one learner wrote, "AI helped organize and simplify the explanation." Another noted, "Making the complex topics into simpler ideas." These statements suggest that learners perceived the feedback as useful for reducing complexity and organizing information. They are best interpreted as learner perceptions rather than direct evidence of learning causation.

Conceptual Clarity described learners' perception that repeated explanation and retrieval helped them distinguish related ideas and understand underlying relationships. This theme is compatible with retrieval-practice research emphasizing active reconstruction of knowledge and provides qualitative context for the observed improvement in conceptual-domain scores.

Guided Error Correction was reflected in comments such as, "It explains step-by-step so I understand where I went wrong." Learners appeared to value feedback that identified not only an incorrect answer but also the reasoning that needed revision. This finding is consistent with literature emphasizing explanatory feedback as a support for error correction and self-regulation.

Increased Confidence emerged in statements such as, "I became more confident solving problems." The responses suggest that repeated opportunities to attempt problems, receive explanations, and revise answers may have made learners more willing to engage with difficult tasks. Because confidence was based on self-report rather than a validated self-efficacy scale, this finding is interpreted as a perceived change in confidence.

Improved Application was represented by the statement, "I can now connect the lesson to real-life situations." This theme is relevant to the application domain, which showed a descriptive increase in performance. The convergence between the score pattern and learner reflections strengthens the interpretation that participants perceived improved ability to connect thermodynamics with practical contexts.

Taken together, the five themes provide a coherent account of learners' reported experiences: explanations were perceived as clearer, errors were easier to identify, confidence increased, and concepts were more readily connected with practical situations. These findings complement the quantitative results by suggesting possible learning processes associated with the intervention. They do not independently establish causal effectiveness.

Synthesis and Reflection

The study documented a statistically significant and educationally notable increase in thermodynamics test performance among the 19 participating learners. The findings are consistent with literature on retrieval practice, explanatory feedback, and AI-supported learning.

The qualitative themes further suggest that learners experienced the AI-assisted retrieval activities as helpful for simplifying explanations, correcting errors, increasing confidence, and connecting concepts with real-world situations. These perceptions provide plausible explanations for the observed gains but cannot rule out repeated testing, regular classroom learning, maturation, motivation, or other concurrent exposures to thermodynamics.

The study therefore supports continued, carefully supervised exploration of AI-assisted retrieval as a supplementary remediation strategy. Future action-research cycles should strengthen the design by including a comparison group where feasible, increasing the sample size, using participant-level data for complete statistical reporting, documenting AI prompts and platform/model details, monitoring AI accuracy, and applying stronger qualitative credibility procedures.

Action Plan

The results will be disseminated through a multi-level cascading approach, with the emphasis placed on sharing a promising classroom practice and its limitations rather than presenting AI-ARM as a proven causal intervention.

At the school level, findings may be shared with Physics teachers, Master Teachers, and SHS Coordinators through LAC sessions, INSET, and research colloquia. Presentations should distinguish clearly between statistically significant within-group improvement and causal effectiveness, and should include the study's limitations and responsible-AI safeguards.

To support practical application, STEM teachers may be engaged through demonstration teaching and capability-building workshops. Training should include retrieval-practice design, standardized prompting, teacher verification of AI responses, learner privacy, responsible use by minors, and procedures for correcting inaccurate AI-generated explanations.

At the school leadership level, findings may inform discussions about supplementary remediation during homeroom periods. Any wider adoption should be treated as an implementation pilot accompanied by monitoring and, where feasible, comparison of learner outcomes across instructional conditions.

At the division level, the study may be presented as preliminary classroom evidence and a basis for further research should incorporate larger samples, comparison conditions, and standardized documentation of AI use.

Continuous improvement should include learner and teacher feedback, periodic review of AI-generated explanations for accuracy, monitoring of learner performance, and refinement of retrieval tasks. Subsequent cycles should strengthen the research design and preserve complete statistical and qualitative documentation.

Overall, the cascading process ensures that the research moves beyond dissemination toward capacity building, institutional adoption, and sustained instructional improvement, thereby maximizing its impact on teaching and learning.

Research Topic	Objectives	Target Audience	Channels	Key Messages	Time-line	Evaluation Process
AI-Assisted Retrieval Model for Thermo-dynamics Mastery	To share the effectiveness of AI-ARM in improving conceptual understanding and test performance	Physics Teachers, Master Teachers, SHS Coordinators	LAC Session, INSET, Research Colloquium	Preliminary evidence of significant within-group gains; causal attribution requires stronger design	May–June 2026	Pre-post teacher feedback survey; adoption rate of strategy

Implementa-tion of Retrieval Practice with AI Feedback	To train teachers on integrating retrieval strategies with AI tools	STEM Teachers	Demonstra-tion Teaching, Workshop	Structured retrieval + verified AI feedback may support conceptual clarity and retention	June 2026	Lesson plan outputs; classroom observation (COT indicators)
Learner Support through AI-Based Remediation	To institutional-ize AI-assisted remediation during homeroom	School Heads, Depart-ment Heads	School-based policy proposal, meetings	AI-ARM may be piloted as supplementary remediation with monitoring and safeguards	July 2026	Monitoring of learner mastery rates ($\geq 75\%$)
Scaling Across STEM Subjects	To expand intervention to other science and math areas	Division Super-visors, Curricu-lum Planners	Division Research Conference, Policy Brief	Model warrants further evaluation in other domains using comparison conditions	Aug.20 26	Comparative performance data across subjects
Continuous Improvement of AI Integration	To refine and improve AI-assisted strategies based on feedback	Teachers and Learners	Reflection logs, feedback forms	Continuous review improves instructional quality, AI accuracy, privacy, and responsible use	On-going	Thematic analysis of feedback; improvement in future cycles

ACKNOWLEDGEMENT

The researcher expresses sincere gratitude to all individuals and institutions who contributed to the successful completion of this action research.

First and foremost, heartfelt appreciation is extended to the school head and administrators of Antipolo City Senior High School for their support in the implementation of this study, particularly in allowing the integration of the intervention during homeroom sessions.

Deep gratitude is also given to the fellow Physics and STEM teachers for their valuable insights during the validation of the research instrument and for their encouragement throughout the research process.

The researcher is especially thankful to the participating Grade 11 learners whose active engagement, honesty in reflection, and commitment to the intervention made this study meaningful and possible.

Recognition is likewise given to the Schools Division Office of Antipolo, particularly the Research Unit, for promoting a culture of action research and continuous professional growth among educators.

Finally, the researcher extends personal gratitude to family and colleagues for their unwavering support, motivation, and understanding throughout the completion of this research endeavor.

Above all, glory and gratitude are offered to God for the guidance, wisdom, and strength in accomplishing this work.

REFERENCES

1. Kelley, T. R., & Knowles, J. G. (2021). A conceptual framework for integrated STEM education and sustainability thinking. *Journal of STEM Education Research*, 4(2), 123–139. <https://doi.org/10.1007/s41979-020-00041-4>
2. Morales, M. P. E., & Santos, J. M. (2022). Classroom action research as a driver of instructional innovation in secondary science education. *International Journal of Educational Research Open*, 3, 100143. <https://doi.org/10.1016/j.ijedro.2022.100143>
3. Singh, C., & Marshman, E. (2021). Student difficulties in thermodynamics and implications for instruction. *Physical Review Physics Education Research*, 17(1), 010110. <https://doi.org/10.1103/PhysRevPhysEducRes.17.010110>
4. Wisniewski, B., Zierer, K., & Hattie, J. (2021). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 12, 678541. <https://doi.org/10.3389/fpsyg.2021.678541>
5. Yang, C., Luo, L., Vadiello, M. A., Yu, R., & Shanks, D. R. (2021). Testing enhances learning: A meta-analysis of the retrieval practice effect. *Psychological Bulletin*, 147(4), 399–435. <https://doi.org/10.1037/bul0000306>
6. Zawacki-Richter, O., Bond, M., Marin, V. I., & Gouverneur, F. (2022). Systematic review of research on artificial intelligence applications in higher education and implications for secondary education. *Computers and Education: Artificial Intelligence*, 3, 100054. <https://doi.org/10.1016/j.caeai.2022.100054>