The Effect of the Guided Inquiry Learning Model with Laboratory Experiments on Students' Higher-Order Thinking Skills

Made Devayana Krisna Pendit[1*], I Wayan Santyasa[1], & I Nengah Edi Budiarta[1]
[1]Physics Education, Universitas Pendidikan Ganesha, Indonesia
*Corresponding Author: devayana@student.undiksha.ac.id

Received: 18th May 2026; **Accepted**: 26th June 2026; **Published**: 30th June 2026
DOI: https://dx.doi.org/10.29303/jpft.v12i1.12118

Abstract *– This study aimed to examine the differences in the effects of guided inquiry learning using laboratory experiments, guided inquiry learning using virtual experiments, and conventional learning on students' higher-order thinking skills (HOTS). A quasi-experimental method with a one-way non-equivalent pretest–posttest control group design was employed. The study population consisted of five Grade 11 science classes at SMA Negeri 2 Singaraja. A sample of three classes comprising 120 students was selected using cluster random sampling. Students' HOTS were measured using a 10-item essay test. The instrument demonstrated acceptable item validity, with item-total correlation coefficients ranging from 0.31 to 0.62, and high reliability (Cronbach's α = 0.845). Students' initial HOTS scores were treated as the covariate. Data were analyzed using one-way analysis of covariance (ANCOVA), followed by the Least Significant Difference (LSD) post hoc test. Statistical significance was determined at the 5% level. The results showed significant differences among the three instructional approaches in terms of students' HOTS ($F = 68.173$, $p < 0.05$). The LSD test indicated that guided inquiry learning using laboratory experiments produced significantly higher HOTS than guided inquiry learning using virtual experiments and conventional learning. These findings suggest that integrating laboratory experiments into guided inquiry learning is more effective in enhancing students' higher-order thinking skills in physics than virtual experiments or conventional instruction.*

Keywords*: Conventional Learning; Guided Inquiry Learning; Laboratory Experiments; Virtual Experiments; HOTS*

INTRODUCTION

The national education system aims to develop the abilities and character of a respectable nation to educate the nation's civilization (Republik Indonesia, 2003). This perspective aligns with the educational philosophy of Ki Hajar Dewantara, who emphasized that instruction should be attuned to the natural development of learners and the contextual demands of their era (Irianti & Fitrotul, 2023). In the 21st century, educational practices are expected to foster four essential competencies (4Cs) that needed to face and adapt the changes in civilization demands (Sari et al., 2024). The 4Cs are critical thinking, creative thinking, collaboration, and communication. To encourage these competencies, the integration of Higher-Order Thinking Skills (HOTS) into both instructional processes and assessment frameworks is crucial (Kurniawan et al., 2021).

HOTS-based learning is already familiar among teachers Saepuzaman et al. (2022b). But the HOTS-based practical consensus has not been implemented comprehensively in the teaching and learning process (Budiarta, 2023). This is due to several factors, namely: (1) the learning process in the classroom does not effectively train higher-order thinking skills (HOTS); (2) teachers are not accustomed to developing HOTS-based test instruments; and (3) students are not familiarized with working on HOTS-based questions from teachers (Saepuzaman et al., 2022b). This is in line with research by (Kurniawan et al., 2021), which states that still few teachers master the techniques of creating HOTS-based assessments. Many teachers use assessments that emphasize the components of knowing, remembering, and memorizing, which are classified in the cognitive levels of LOTS (Lower Order Thinking Skills) and MOTS (Middle Order Thinking Skills).

The limited action in developing Indonesian students' cognitive level to a higher level (HOTS) is evidenced by the 2022 Programme for International Student Assessment (PISA) data. The data indicate that 34% of students achieved Level 2, while 31% reached Level 3. On the other hand, only 1% of students attained the higher proficiency Levels 5 and 6. The cognitive levels used in the PISA test are Bloom's Revised Taxonomy (OECD, 2023). More specifically, in the Physics subjects and heat chapter, 55% of students were in the sufficient category, while the remaining 45% were in the low and very low conceptual understanding categories (Hara et al., 2023). Students' suboptimal conceptual understanding results in suboptimal higher-level thinking abilities (Aisyah, 2025).

Classroom instruction that relies on memorization, combined with teachers' limited skill to visualize physics concepts, is the reason why students' HOTS are inadequate (Suyatna et al., 2020). Besides that, the teacher-centered approach doesn't optimally promote students' engagement in constructing their own understanding (Wahid, 2023). Theoretically, to solve this problem, there is a need to reorient learning practices toward student-centered learning by implementing various innovative learning models. But the study reported by Wakhidyah et al. (2025), show that after studying various learning models, 61.34% of students were only able to complete the C4 cognitive level (analysis), while only 12% of students completed the heat problem up to the C6 cognitive level (creation). It can be concluded that improving students' HOTS is a challenging thing.

The significant challenge of optimizing students' HOTS suggests that learning has not yet embraced the paradigm of deep and meaningful learning. Deep and meaningful learning includes a series of inquiry process that enhance critical thinking, creative thinking, problem-solving, and metacognition (Mystakidis, 2021), that was the four dimension of HOTS (Dessie et al., 2023). It implies that the guided inquiry learning models is the perfect solution to enhance HOTS as a part of inquiry-based learning (Kaçar et al., 2021) and in line with the deep and meaningful learning paradigm (Mystakidis, 2021).

Several studies have described that guided inquiry models can improve students' critical thinking skills (Maolani et al., 2025), problem-solving skills (Antonio & Prudente, 2024), and creative thinking (Adhiriyanthi et al., 2021). It giving information that the guided inquiry model has good performance at all three dimensions of HOTS separately. However, there is still not much research that reveals the success of the guided inquiry learning model in increasing students' HOTS in physics learning, especially in heat and temperature chapter.

Besides implemented learning models in classroom instruction, deep and meaningful learning can be realized when classroom instruction integrates knowledge and reasoning with real-world activities. A relevant solution to this problem is to integrate learning with laboratory activities (Alarcon et al., 2023). More specifically, integrating the learning process with real laboratories (laboratory experiment design) improves the student learning outcomes more significantly. Furthermor, from the Simbolon & Silalahi (2023) research, it implies that integrating real and virtual laboratories improves student learning outcomes.

Based on the established evidence from previous research, there is a need to improve students' higher-order thinking skills (HOTS) in physics learning. According to these needs, this research not only implemented guided inquiry as a part of innovative learning models. But the previous studies also show that the learning model should integrates with laboratory activities. So that, this research integrates guided inquiry learning with lab experiments to evaluate its effectiveness in engaging students' HOTS. More specifically, this study compares three instructional conditions: (1) guided inquiry using lab experiments (1st experimental class); (2) guided inquiry using virtual experiments (2nd experimental class); and (3) a conventional model (control class).

RESEARCH METHODS

This study is classified as quasi-experimental research using a one-way non-equivalent pretest-posttest control group design. The details about this research design are shown in Table 1.

Table 1. Research Design

O_1	X_1	O_2
O_3	X_2	O_4
O_5	X_3	O_6

(Santyasa, 2023)

with the following conditions:

O_1, O_3, O_5 = students' initial HOTS score
X_1, X_2, X_3 = various learning model
O_2, O_4, O_6 = students' final HOTS score
------- = full randomize did not conducted

The independent variable used is the learning model, which includes three categories: guided inquiry using laboratory experiments (1st experimental class), guided inquiry using virtual experiments (2nd experimental class), and the conventional model (control class). The dependent variable in this study is students' final HOTS scores. Finally, to reduce research bias, students' initial HOTS scores were used as a covariate variable.

The research population was 11th grade mathematics and natural sciences (MIPA) students at SMA Negeri 2 Singaraja, which has 200 students distributed into five classes. The three classes selected from this population as research samples using the random class sampling technique. Consequently, the research sample consisted of 120 students, with 40 students assigned to the first experimental group, 40 students to the second experimental group, and 40 students to the control group.

Students' initial and final higher-order thinking skills (HOTS) were assessed using an extended-response (essay-type) instrument containing 10 items, developed in line with HOTS indicators within the heat topic. The instrument underwent psychometric evaluation, including tests of content validity, reliability, item discrimination index, and item difficulty index, conducted using Microsoft Excel 2021 and IBM SPSS Statistics 25.0. Each question was scored on a scale of 0–5, with the guideline that shown in Table 2.

Table 2. Assessment rubric for 0-5 scheme

Scale	Classification
0	The student did not give a solution to the question.
1	The student tried to give a solution, but it was totally incorrect.
2	The student gave a solution that contained a correct idea, but was still inadequate.
3	The student gave a solution that was partially correct, and many misconceptions were found.
4	The student gave a solution that was partially correct, with a few misconceptions
5	The student gave a solution that was completely correct.

(Santyasa, 2017)

With the rubric that shown in Table 2, the students' HOTS score ranged from 0 to 50. Then, the score converted into 100-scale score assessment guideline from benchmark assessment (*Penilaian Acuan Patokan*/PAP) (Arikunto, 2009) as shown in Table 3.

Table 3. Benchmark Assessment Guidelines

Score Intervals	Classification
85-100	Very High
70 – 84	High
56 – 69	Sufficient
40 – 55	Low
0 – 39	Very Low

Furthermore, HOTS performance was analyzed across specific HOTS indicators adapted from (Saepuzaman, et al. 2022a), namely: general clarification (I1), inference (I2), using exact strategies and tactics (I3), fluency (I4), flexibility (I5), and elaboration (I6). The distribution of questions for each indicator is presented in Figure 1.

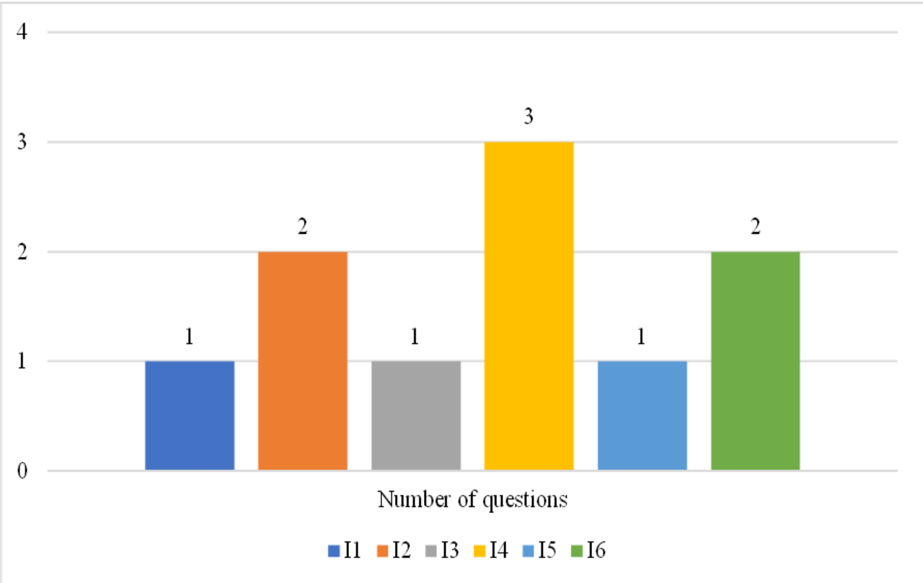

Figure 1. Distribution of questions for each HOTS indicator

Students' initial and final HOTS scores were analyzed using descriptive analysis and one-way analysis of covariance (ANCOVA). Descriptive analysis was used to calculate the average HOTS score and standard deviation. Then, ANCOVA is used to determine the difference in average score for each data variant, using a significance level of 0.05. If the p-value < 0.05, the null hypothesis (H_0) is rejected and alternative hypothesis (H_a) is accepted.

Inference statistics analysis using ANCOVA, followed by a post-hoc analysis using the Least Significant Difference (LSD) test. The LSD test aims to test the significance of HOTS differences. As a comparison, the *p*-value of the LSD compares with the significance level alpha. If the *p*-value of the LSD test between two test classes is greater than the 0.05 significance level, then there is a significant difference in HOTS. Then, to test how much of the total sample is affected by a given learning model, the partial eta square value in the "Univariate Test" is observed. This statistic describes the proportion of total variability caused by a factor, through the guidelines that are shown in Table 4.

Table 4. Criteria for Partial Eta Squared

Score Intervals	Criteria
0.0099 – 0.0588	Small Effect
0.0589 – 0.1379	Medium Effect
> 0.1379	Large Effect

(Richardson, 2011)

RESULTS AND DISCUSSION

Results

As the first step of this research, a pretest was given to the students using a previously validated test instrument. Pretest is given before the models are implemented in each class. The summary of the pretest results is shown in Table 5.

Table 5. Summary of the Pretest Result

Statistics	1st Experimental Class	2nd Experimental Class	Control Class
Class population	40	40	40
Minimum score	6	6	8
Maximum score	44	38	48
Range	38	32	40
Mean	25.90	20.90	32.95
Deviation Standard	8.509	6.468	9.543
Classification	Very Low	Very Low	Very Low

Furthermore, based on these summaries, students' initial higher-order thinking skills (HOTS) were analyzed at a more specific level according to each HOTS indicator. A detailed of the average initial HOTS score for each HOTS indicator is provided in Table 6.

From Table 6, an overview about average score of initial HOTS for each indicator, is presented in Figure 2.

Table 6. Average Initial Score Each HOTS Indicator

Class	Average Score Each HOTS Indicator					
	I1	I2	I3	I4	I5	I6
1st Experimental	35.5	35.0	27.5	18.8	19.5	24.8
2nd Experimental	38.5	26.0	19.0	14.2	18.0	19.5
Control Class	38.5	34.8	39.0	29.7	31.0	31.5

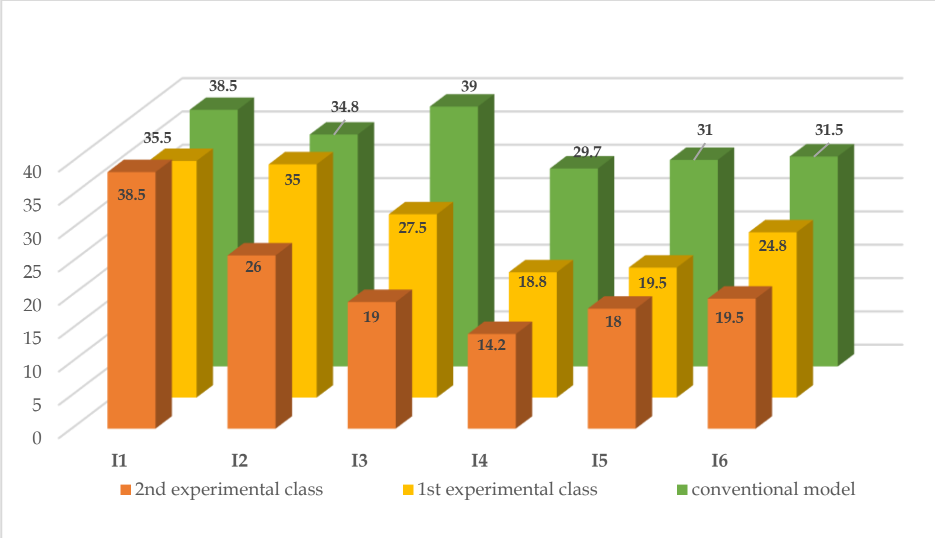

Figure 2. Average Score for Each Initial HOTS Indicator

After the pretest, treatment was administered to each class. The 1st experimental class learned with guided inquiry model using laboratory experiments, 2nd experimental learned with guided inquiry model using virtual experiments, and the control class learned with conventional model. After the treatment, a posttest was administered to each class. The summary of the posttest results is shown in Table 7.

Table 7. Summary of the Posttest Result

Statistics	1st Experimental Class	2nd Experimental Class	Control Class
Class population	40	40	40
Minimum score	60	62	56
Maximum score	94	90	78
Range	34	28	22
Mean	76.15	73.05	69.75
Deviation Standard	8.094	6.341	5.803
Classification	High	High	Sufficient

Generally, based on Table 5 and Table 7, an overview of students' HOTS scores, before and after treatment, can be visualized in Figure 3.

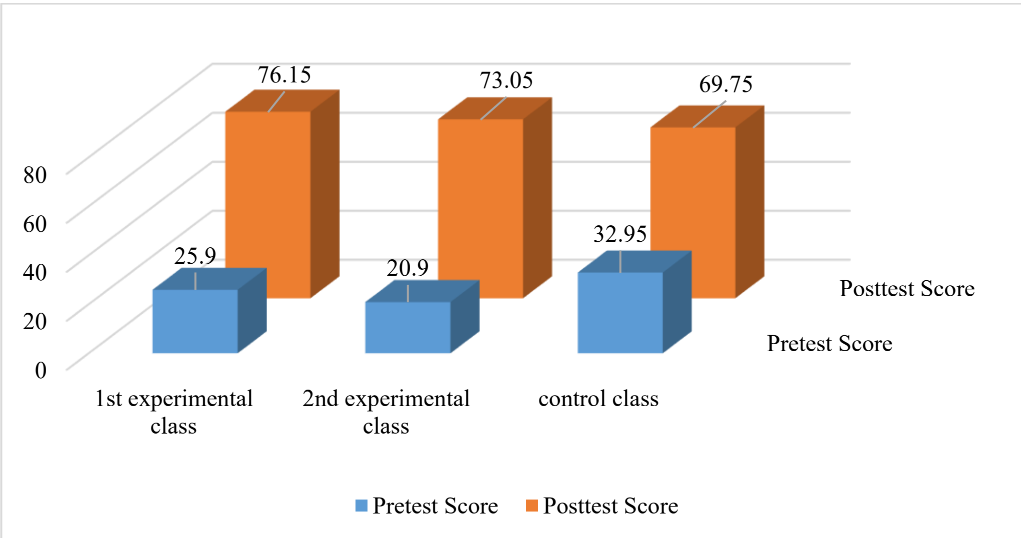

Figure 3. Comparation between Students' Initial and Final HOTS

Based on these summaries, students' final HOTS were analyzed at a more specific level according to each HOTS indicator. A detailed of the average final score each HOTS indicator is provided in Table 8.

Table 8. Average Final Score Each HOTS Indicator

Class	Average Score Each HOTS Indicator					
	I1	I2	I3	I4	I5	I6
1st Experimental	80.0	79.5	87.0	70.8	70.5	76.3
2nd Experimental	80.5	76.0	82.0	66.3	67.5	74.8
Control Class	76.5	71.5	76.0	64.8	63.0	72.5

From the data about the average final HOTS score for each indicator as shown in Table 8, it can be visualized in Figure 4.

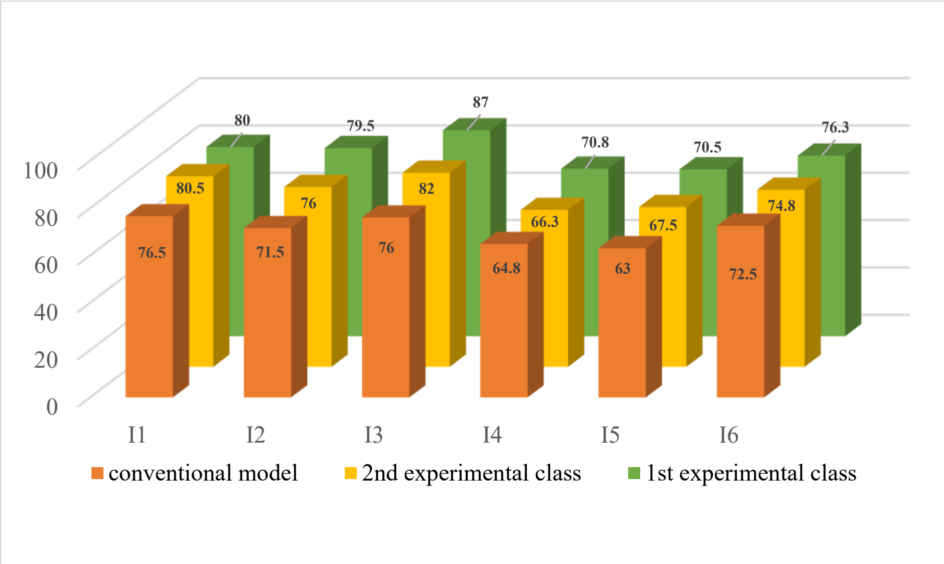

Figure 4. Average Score for Each HOTS Indicator

After the students' final HOTS data were analyzed descriptively, the analysis continued with a one-way analysis of covariance (ANCOVA). Three assumption tests were required normality, homogeneity, and linearity. Firstly, Table 9 summarizes the results of the normality test.

Table 9. Summary of Normality Test

Data	Class	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig	Statistic	df	Sig.
Pretest	1st experimental	.145	40	.033	.956	40	.123
	2nd experimental	.095	40	.200	.983	40	.799
	control	.135	40	.063	.946	40	.056
Posttest	1st experimental	.135	40	.066	.973	40	.460
	2nd experimental	.140	40	.045	.954	40	.103
	control	.151	40	.022	.948	40	.064

Based on the normality test, the Shapiro–Wilk statistic showed a significance level of 0.123, 0.799, and 0.056 for the pretest, respectively. This indicates that the students' initial HOTS are from a normally distributed sample (sig. > 0.05). Referring to the posttest data, the Shapiro–Wilk statistic showed significance levels of

0.460, 0.103, and 0.064, respectively. It means that the posttest data is also from a normally distributed sample.

The second assumption that must be passed is that the data variance should be homogeneous. Table 10 summarizes the results of the homogeneity test.

Table 10. Summary of Homogeneity Test

Data	Criteria	Levene Statistics	df1	df2	Sig.
Pretest	Based on mean	2.559	2	117	.082
	Based on median	2.200	2	117	.115
	Based on trimmed mean	2.200	2	117	.093
Posttest	Based on mean	1.410	2	117	.248
	Based on median	1.349	2	117	.264
	Based on trimmed mean	1.419	2	117	.246

From the Table 10, it showed that the significance based on means of 0.082 and 0.248 for the pretest and posttests, respectively. As the Levene statistical significance value is greater than 0.05 it can be concluded that the pretest and posttest data are from the homogeneous variances.

The last assumption is that the covariate variable must be correlated with the dependent variable. To analyze it, we used a linearity test. A summary of the linearity test is presented in Table 11.

Table 11. Summary of Linearity Test

Data		Criteria	Sum of Squares	df	Sig.
Pretest* Posttest	Between Groups	(Combined)	2,402.12	21	.000
		Linearity	1,630.39	1	.000
		Deviation from Linearity	771.724	20	.491

Based on Table 11, it can be inferred that: 1) from the linearity row, it is found that the statistical significance is 0.001, which is lower than 0.05. This means that there is a linear correlation between the covariate variable (initial HOTS) and the dependent variable (final HOTS). 2) From the deviation from linearity row, it was found that the statistical significance = 0.491, greater than 0.05. It means the initial and the final HOTS scores do not deviate significantly from the regression line.

Referring to the assumption test, the results showed that (1) the data were normally distributed, (2) the data variance was homogeneous, (3) there was a linear relationship between the covariate and the dependent variables, and (4) there is no significant deviation of data from the regression line. After the assumptions test was fulfilled, a one-way analysis of covariance (ANCOVA) was conducted. The summary of the ANCOVA results is shown in Table 12.

Table 12. Summary of One-Way ANCOVA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4,129.653	3	1,376.55	75.097	.000
Intercept	32,341.279	1	32,341.28	1,764.363	.000
Covariate Variable	3,310.186	1	3,310.19	180.586	.000

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Learning Models	2,499.260	2	1,249.63	68.173	.000
Error	2,126.314	116	18.33		
Total	645,444.000	120			
Corrected Total	6,255.967	119			
R Squared = .660 (Adjusted R Squared = .651)					

Based on the Table 12, it can be explained that: 1) from the effect source of covariate variable, it is founded that statistic value F = 180.586 with the sig. = 0.001 lower than 0.05. This mean that ANCOVA prover to utilize for control the covariate effect. 2) From the effect source of learning models, it founded that statistic value F = 68.173 with sig. = 0.001 lower than 0.05. This mean that there is a significant different of HOTS among students learn by guided inquiry using laboratory experiments, guided inquiry using virtual experiments, and conventional learning.

After the students' posttest results were analyzed using ANCOVA, a post-hoc LSD test was conducted. Table 13 summarizes the LSD test results.

Table 13. Summary of LSD Test

(I) Class	(J) Class	Mean Difference (μ_i-μ_j)	Std. Error	Sig.[b]
1st experimental class	2nd experimental class	3.100*	1.524	.044
	control class	6.400*	1.524	.000
2nd experimental class	1st experimental class	-3,100*	1,524	.044
	control class	3,300*	1,524	.032

Referring to Table 13, it can be explained that: (1) the difference in the average HOTS score of 1st experimental class to the control class is 6.4 with *p*-value significance level 0.001; (2) the difference in the average HOTS score of 2nd experimental class to control class is 3.3 with *p*-value significance 0.032; and (3) the difference in the average HOTS score of 1st experimental class to 2nd experimental class is 3.1 with *p*-value significance 0.044.

Then, to test how much of the total sample is affected by a given learning model, the partial eta square value as effect size coefficient. The result of "Univariate Test" is shown in Table 14.

Referring to Table 14, it can be informed from the partial eta-squared that shown coefficient 0.540, larger than 0.1379. It means, the learning models that implemented in each classes giving large effect to the students HOTS.

Table 14. Summary of Univariate Test

Criteria	*Sum of Squares*	df	*Mean Square*	F	Partial eta-squared
Contrast	2,499.260	2	1,249.63	68.173	.540
Error	2,126.314	116	18.330		

Discussion

The first step was giving a pretest to determine students' initial HOTS. In general, the results of the HOTS pretest refer to Table 5. Table 5 shown that the average initial HOTS value from each class, sequently: (1) 1st experimental class has an average value of 25.90 and a standard deviation of 8.509; (2) 2nd experimental class has an average value of 20.90 and a standard deviation of

6.468; (3) control class has an average value of 32.95 and a standard deviation of 9.543. Generally, from the pretest, the students' initial HOTS in each class were categorized as very low.

The initial HOTS score of the average student can be expressed for each indicator. As shown in Table 6, the initial HOTS for each indicator are still in the very low category. The highest score, 39.0, is achieved by the control class, from the 'using exact strategies and tactics' indicator (I3). The lowest score, 14.2, is from the fluency category (I4). While some HOTS indicators have higher scores than others, the overall HOTS score remains in the very low category, requiring intensive attention.

Referring to the pretest results, students' initial HOTS scores were categorized as very low. That was caused by the students' prior knowledge, which was insufficient for the heat subject in the senior high school. The insufficiency of prior knowledge indicates that students' conceptual understanding of the 'heat' subject is still low (Hara et al., 2023). Students' suboptimal conceptual understanding results in suboptimal higher-level thinking abilities (Aisyah, 2025). These pretest results also align with Saepuzaman et al. (2022b) prediction that one factor contributing to students' low HOTS is that the classroom learning process does not effectively train the students' HOTS. After that, a different learning model was given to each class as treatment. The 1st experimental class was treated with a guided inquiry model using laboratory experiments and the 2nd experimental class was treated with the guided inquiry model using virtual experiments. Lastly, the conventional model was implemented in the control class.

The implementation of the learning model in each class was based on the learning aims in the heat subject. The treatment was given for 12 learning hours. After treatment, a posttest was conducted, and the general result was presented in Table 7. From Table 7, there was an increase in students' HOTS. The average final HOTS score in the first experimental class was 76.15, which was classified as good. Then, the average final HOTS score of students in the second experimental class was 73.05, which was classified as good. Lastly, the average HOTS score in the control class was 69.75, which is still in the sufficient category.

Improvements were also evident in the average scores for each HOTS indicator. The greatest improvement occurred in the using exact strategies and tactics indicator (I3), with each indicator improving from a “very low” to “high” category in 2nd experimental class and control class. In 1st experimental class, the I3 score was even in the “very high” category. Furthermore, the general clarification indicator (I1), the inference indicator (I2), and the elaboration indicator (I6) showed significant differences across the three classes, from “very low” to “very high”. Finally, the fluency indicator (I4) and flexibility indicator (I5) saw a sharp increase only in 1st experimental class, changing from “very low” to “high”, while 2nd experimental class, and the control class were in the “sufficient” category.

After descriptively identifying students’ initial and final HOTS conditions, the hypothesis was tested. The differences in HOTS between students learning with guided inquiry models using laboratory experiments, virtual experiments, and conventional models were tested using analysis of covariance (ANCOVA) and LSD test as the post-hoc analysis. In the ANCOVA test, initial HOTS was treated as a covariate that was retained to avoid confounding with students’ final HOTS. Before ANCOVA was used to examine the hypothesis, there were assumption tests that had to be met, namely the normality test, the homogeneity test, and the linearity test.

Based on the SPSS output presented in Table 9, Table 10, and Table 11, all assumptions for the ANCOVA test were met. It is because (1) the data (initial HOTS and final HOTS) in each class were normally distributed; (2) the data variances are homogeneous; (3) there was a linear relationship between the covariate (initial HOTS) and the dependent variables (final HOTS), and (4) there is no significant deviation of data from the regression line.

After the ANCOVA test was conducted, two sources of information were obtained. There are the effects of the covariate variable and the effect of the learning models. The significance level of the covariate variable was 0.001, which was lower than 0.05. That means, ANCOVA proved to be useful for controlling the students' initial HOTS as a covariate variable statistically. So, the learning model is the only source that affected the students' final HOTS.

The second information from ANCOVA test results, is from the effect source of learning models. The significance level of the covariate variable was 0.001, which was lower than 0.05. Thus, H_0 (null hypothesis) is rejected, and H_a (alternative hypothesis) is accepted. This means there are significant differences in HOTS among students taught using the guided inquiry model using laboratory experiments, virtual experiments, and the conventional model. The contribution of learning models as an independent variable toward students' HOTS as dependent variable is 66.0%, as shown by the *R*-squared value.

The LSD test, as a post-hoc analysis, compared three learning models into three categories. Referring to Table 13, all of the comparisons indicate that there are significant differences between the two learning models in each configuration. From the difference of average final HOTS, the guided inquiry learning model using laboratory experiments has a better performance than the students who learn with the guided inquiry learning model using virtual experiments and the conventional model.

Lastly, to test how much of the total sample is affected by a given learning model, the partial eta square value as effect size coefficient. Table 13 shows a partial eta squared value of 0.540. This means that 54% of the total variability (research sample) is influenced by the learning model that is implemented in classroom instruction. In the other words, the guided inquiry learning model using laboratory experiments, not only shows significant differences but also has a significant impact on students' HOTS.

That means, the empirical study of the advantage of the guided inquiry learning model in increasing students' HOTS is in line with several previous relevant studies. Previous studies have demonstrated that guided inquiry models can enhance students' critical thinking skills (Maolani et al., 2025; Islamyah et al., 2018), problem-solving skills (Antonio & Prudente, 2024), creative thinking (Adhiriyanthi et al., 2021), and scientific attitude (Parwati et al., 2023). Furthermore, this study had a similar perception to Tyas et al. (2023), which states that integrating learning activities with real-world experiments improves inference, using the exact strategy, and decision-making skills more effectively.

CONCLUSION

Based on the results and discussion part, it can be concluded that there is a difference in HOTS among the students who learn with guided inquiry models using laboratory experiments, virtual experiments, and conventional models at SMA Negeri 2 Singaraja. This is based on the results of the ANCOVA test. The ANCOVA results showed that F = 68,173 and *p*-value < 0.05. Then the ANCOVA was also followed by the LSD test as a post-hoc analysis. The result of LSD test indicated that the effect of the guided inquiry model using laboratory experiments higher than the virtual experiments effect, and as well as the conventional learning effect, it all on the students' HOTS.

The guided inquiry model using laboratory experiments has been proven to be effective in increasing students' HOTS. Therefore, this research implicated that, to enhance students' HOTS in physics learning, it better if students learn utilize the guided inquiry model with laboratory experiments. The limitation about this research is the students' HOTS described as six indicators, namely: general clarification (I1), inference (I2), using exact strategies and tactics (I3), fluency (I4), flexibility (I5), and elaboration (I6).

To support these learning models, it is recommended to set the class in several groups, to increase student interaction and collaboration in learning activity.

ACKNOWLEDGMENT

The researcher would like to express thanks to the Principal of SMA Negeri 2 Singaraja, the physics teacher, and the head of the laboratory for their excellent collaboration. The researcher would also like to thank the 11th grade students of SMAN 2 Singaraja for their enthusiasm in participating in the research. The researcher would also like to thank all those who assisted in completing this research, including those who cannot be mentioned individually.

REFERENCES

Adhiriyanthi, S., Solihin, H., & Arifin, M. (2021). Improving students' creative thinking skills through guided inquiry practicums learning with STEM approach. *Journal of Physics: Conference Series*, *1806*(1). IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/1806/1/012180

Aisyah, N. (2025). Analisis kemampuan pemahaman konsep matematis siswa sma dalam menyelesaikan soal bertipe higher order thinking skills (hots). *OMEGA: Jurnal Keilmuan Pendidikan Matematika*, *4*(2), 72–82. Retrieved from https://ejurnal.univamedan.ac.id/index.php/jkpm/index

Antonio, R. P., & Prudente, M. S. (2024). Effects of inquiry-based approaches on students' higher-order thinking skills in science: A meta-analysis. *International Journal of Education in Mathematics, Science and Technology*, *12*(1), 251–281. https://doi.org/10.46328/ijemst.3216

Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.

Budiarta, I. N. E. (2023). Kajian literatur sistematis: Konseptualisasi dan pengukuran higher-order thinking skills dalam pembelajaran fisika. *13*, *13*(2), 286–295.

Dessie, E., Gebeyehu, D., & Eshetu, F. (2023). Enhancing critical thinking, metacognition, and conceptual understanding in introductory physics: The impact of direct and experiential instructional models. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7). https://doi.org/10.29333/ejmste/13273

Hara, A. K., Astiti, K. A., & Lantik, V. (2023). Analisis penguasaan konsep fisika pada materi suhu dan kalor pasca pembelajaran online di kelas xi sma negeri 12 kota kupang. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *14*(2), 118–126. https://doi.org/10.37640/jip.v14i2.1548

Irianti, R. I., & Fitrotul, M. (2023). Penerapan kurikulum merdeka dalam pengimplementasian pendidikan yang sesuai dengan kodrat alam dan zaman. *Pubmedia Jurnal Penelitian Tindakan Kelas Indonesia*, *1*(2), 10. https://doi.org/10.47134/ptk.v1i2.56

Islamyah, D. G., Yasa, P., & Rachmawati, D. O. (2018). Penerapan model pembelajaran inkuiri terbimbing berbasis stem guna meningkatkan kemampuann berpikir kritis siswa kelas x mipa 4 sman tahun ajaran 2018/2019. *Jurnal Pendidikan Fisika Undiksha*, *8*(2).

Kaçar, T., Terzi, R., Arıkan, İ., & Kırıkçı, A. C. (2021). The effect of inquiry-based learning on academic success: A meta-analysis study. *International Journal of Education and Literacy Studies*, *9*(2), 15. https://doi.org/10.7575/aiac.ijels.v.9n.2p.15

Kurniawan, T. T., Santoso, & Sri, U. (2021). Development of 4c-based hots assessment to improve critical thinking skills in grade 6 students of kudus elementary school. *JURNAL PAJAR (Pendidikan Dan Pengajaran)*, *5*(3), 675–683. https://doi.org/10.33578/pjr.v5i3.8319

Maolani, A. G., Resti, W., & Asep, I. I. (2025). Application of guided inquiry model to improve the ability and critical thinking disposition of senior high school students in physics learning. *Jurnal Pendidikan Fisika Dan Terapan*, *11*(2), 192–201. Retrieved from https://jurnal.ar-raniry.ac.id/index.php/jurnalphi/index

Mystakidis, S. (2021). Deep meaningful learning. *Encyclopedia*, *1*(3), 988–997. https://doi.org/10.3390/encyclopedia1030075

OECD. (2023). Pisa 2022 results factsheets indonesia. In *https://www.oecd.org/en/publications/pisa-results-2022-volume-iii-factsheets_041a90f1-en/indonesia_a7090b49-en.html*. Retrieved from https://oecdch.art/a40de1dbaf/C108.

Parwati, G. A. P. U., Sugiarta, I. M., & Rapi, N. K. (2023). Pengaruh model pembelajaran inkuiri terbimbing terhadap keterampilan berpikir kritis dan sikap ilmiah siswa. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *14*(1), 21–33.

Republik Indonesia. *Republik Indonesia*., Pub. L. No. Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional (2003).

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, Vol. 6, pp. 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Saepuzaman, D., Istiyono, E., & Haryanto. (2022). Characteristics of fundamental physics higher-order thinking skills test using item response theory analysis. *Pegem Egitim ve Ogretim Dergisi*, *12*(4), 269–279. https://doi.org/10.47750/pegegog.12.04.28

Saepuzaman, D., Retnawati, H., Istiyono, E., & Haryanto, H. (2022). Developing hots instruments: Is it difficult for physics teachers? *Jurnal Pendidikan MIPA*, *23*(4), 1802–1826. https://doi.org/10.23960/jpmipa/v23i4.pp1802-1826

Santyasa I, W. (2017). *Asesmen dan evaluasi pembelajaran fisika*. Singaraja: Graha Ilmu.

Santyasa, I. W. (2023). *Metodologi penelitian pendidikan*. Singaraja: Undiksha Press. Retrieved from www.penerbit.undiksha.ac.id

Sari, L. S., Kaepah, K., Fitriani, P., Tresnawati, N., & Khodari, R. (2024). Stem approach to 4c skills in elementary school students: A systematic literature review. *Scientiae*

Educatia, *13*(1), 10. https://doi.org/10.24235/sc.educatia.v13i1.18062

Simbolon, D. holden, & Silalahi, E. K. (2023). Physics learning using guided inquiry models based on virtual laboratories and real laboratories to improve learning. *Journal for Lesson and Learning Studies*, *6*(1), 55–62. https://doi.org/10.23887/jlls.v6i1.61000

Suyatna, A., Viyanti, V., & Sari, S. (2020). Dynamic fluid e-module with stem approach to stimulate hots of high school students in distance learning. *Jurnal Pendidikan MIPA*, *21*(2), 132–145. https://doi.org/10.23960/jpmipa/v21i2.pp132-145

Tyas, D. N., Isdaryanti, B., & Feylosofia, P. A. (2023). Efektivitas virtual laboratory dibandingkan real laboratory pada perkuliahan hybrid terhadap kemampuan berpikir kritis mahasiswa. *Prolife: Jurnal Pendidikan Biologi, Biologi, dan Ilmu Serumpun*, *10*(1), 707–720.

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*, Vol. 8. Frontiers Media S.A. https://doi.org/10.3389/feduc.2023.1170487

Wahid, S. (2023). Pengaruh model pembelajaran inkuiri terbimbing terhadap hasil belajar fisika ditinjau dari minat belajar peserta didik kelas xi sman 9 makassar. *Jurnal Ilmiah Ecosystem*, *23*(3), 774–785. https://doi.org/10.35965/eco.v23i3.3905

Wakhidyah, S. H., Mudakir, I., & Wahyuni, S. (2025). Analysis of high order thinking skills of high school students in solving heat and thermodynamics problems. *Jurnal Pendidikan Fisika Dan Teknologi*, *11*(1), 102–107. https://doi.org/10.29303/jpft.v11i1.8030