



Guided Inquiry Assisted by PhET for Elementary Students' Critical Thinking on Force: A Quasi-Experimental Study

Amelia Nabilla Zahra^{1*}

Universitas Muhammadiyah
Prof. Dr. Hamka, Indonesia

Zulherman²

Universitas Muhammadiyah
Prof. Dr. Hamka, Indonesia

Article Info

Article history:

Received: April 20, 2026

Revised: Aug 11, 2026

Accepted: Sept 01, 2026

Keywords:

Critical Thinking; Elementary
Science Education; Guided
Inquiry; PhET Simulation;
Force

Abstract

Purpose: This study investigates the contribution of memory skill and memory effort to university students' arithmetical literacy and examines whether their predictive patterns differ between students at different stages of study. Understanding these relationships is relevant because arithmetical literacy requires not only procedural competence but also the ability to retain, retrieve, and effectively use mathematical information when solving quantitative problems.

Method: The study employed a quantitative approach involving 300 university students, consisting of 150 third-semester students and 150 fifth-semester students from five higher education institutions. Data were collected using instruments measuring memory skill, memory effort, and arithmetical literacy. Multiple regression analysis was conducted separately for the two semester groups to determine the simultaneous and individual contributions of memory skill and memory effort to arithmetical literacy.

Findings: The results showed that the combined regression models were statistically significant for both groups. Memory skill and memory effort jointly explained 10.9% of the variance in arithmetical literacy among third-semester students and 20.2% among fifth-semester students. At the individual predictor level, memory skill significantly contributed to arithmetical literacy among third-semester students but did not show a significant contribution among fifth-semester students. In contrast, memory effort significantly contributed to arithmetical literacy in both groups. These findings indicate that the role of memory-related factors in arithmetical literacy varies according to students' stage of study.

Significance: The study extends understanding of the cognitive factors associated with arithmetical literacy by distinguishing between memory capacity-related skill and the effort students invest in remembering and retrieving mathematical information. The findings suggest that strengthening students' strategic memory effort may provide a more consistently relevant support for arithmetical literacy across different stages of university study.

To cite this article: Zahra, A. N., & Zulherman, Z. (2026). Guided Inquiry Assisted by PhET for Elementary Students' Critical Thinking on Force: A Quasi-Experimental Study. *Al Qodiri: Jurnal Pendidikan, Sosial dan Keagamaan*, 24(3), 17-28.

INTRODUCTION

Critical thinking in science education involves more than recalling factual knowledge; it requires students to interpret information, examine evidence, compare explanations, and justify conclusions. These processes are closely aligned with scientific investigation, where claims must be evaluated against observations and data (Butcher et al., 2023; Soysal, 2021). Recent meta-analytic evidence indicates that inquiry-oriented science learning can support higher-order thinking, although the magnitude of improvement varies across educational levels, instructional designs, and forms of guidance (Antonio & Prudente, 2024; Arifin et al., 2025). This variability is especially relevant in elementary classrooms because younger learners often require explicit support to coordinate observation, reasoning, and explanation. The instructional problem is therefore not simply whether

* Corresponding author:

Amelia Nabilla Zahra, Universitas Muhammadiyah Prof. Dr. Hamka, INDONESIA
amelian.zahra77@gmail.com

students are exposed to active learning, but whether classroom activities give them structured opportunities to use evidence while constructing scientific understanding.

In elementary science, teacher-directed explanation remains useful for introducing concepts, but it can limit opportunities for students to practice the reasoning processes that critical thinking requires when it dominates classroom interaction. Research on inquiry-based science education shows that productive inquiry depends on the quality of teacher support, assessment, questioning, and classroom management rather than on student activity alone (Grangeat et al., 2021; Lazonder & Harmsen, 2016; Strat et al., 2024). Studies in primary settings likewise show that teachers recognize inquiry as a route to critical thinking but encounter practical difficulty in translating it into classroom routines (Essalih et al., 2024). In the Indonesian elementary context, barriers to students' critical thinking have also been associated with learning conditions that do not consistently engage learners in higher-order reasoning (Alditia et al., 2024; Yaumie et al., 2025). These findings provide a rationale for examining forms of guided inquiry that preserve teacher support while giving elementary students a more active role in scientific investigation.

The topic of force offers a relevant setting for this problem because students must connect observable changes in motion with explanations about interactions that are not always evident from verbal description alone. Computer simulations can make such relationships visible by allowing learners to manipulate variables and observe immediate changes in a controlled environment (Alabidi et al., 2023; Banda & Nzabahimana, 2021). Reviews of simulation-based science learning have consistently reported benefits for conceptual understanding when digital tools are integrated with purposeful instructional activity rather than used as isolated demonstrations (Cromley et al., 2023; Hillmayr et al., 2020; Rutten et al., 2012; Smetana & Bell, 2012). Foundational work on PhET further emphasizes interactive representations, rapid feedback, and exploratory manipulation as design features intended to support sense-making (Finkelstein et al., 2005; Wieman et al., 2008). For elementary students learning about force, these affordances may provide concrete evidence that can be discussed, interpreted, and connected to scientific explanations.

Guided Inquiry and PhET simulations are pedagogically complementary because they address different parts of the learning process. Guided Inquiry organizes students' activity around problems, predictions, observations, analysis, and conclusions, while the simulation provides a manipulable environment in which those processes can be enacted. Evidence from virtual and remote laboratory research shows that the learning value of digital environments depends strongly on how guidance is aligned with phases of inquiry (Gamor, 2021; Zacharia et al., 2015). PhET simulations can also provide implicit scaffolding through their interactive design, but teacher questions and task structure remain important for directing attention toward scientifically relevant relationships (Moore et al., 2013). Technology-enhanced inquiry therefore should be understood as an interaction between instructional guidance and digital affordances, not as a simple substitution of software for conventional teaching (Cao & Hsu, 2023; Chen & Chen, 2025).

Prior research generally supports inquiry-based approaches for developing scientific reasoning and higher-order thinking, but it also shows that outcomes are sensitive to context and implementation. Meta-analyses have reported positive average effects of inquiry-based science teaching, particularly when learners receive appropriate guidance and engage with epistemic as well as procedural aspects of investigation (Furtak et al., 2012; Lazonder & Harmsen, 2016). More recent syntheses similarly indicate positive effects on critical thinking while documenting substantial heterogeneity across studies (Antonio & Prudente, 2024; Arifin et al., 2025). Classroom studies suggest that discourse, teacher facilitation, and opportunities to construct evidence-based arguments are important mechanisms through which inquiry can support critical thinking (Jeon et al., 2021; Ramnarain, 2023; Soysal, 2021). Indonesian studies have also reported improvements in critical thinking under Guided Inquiry or related inquiry models (Ilhamdi et al., 2020; Inel-Ekici & Ekici, 2022; Musliman & Kasman, 2022; Pahrudin et al., 2021; Poerwanti et al., 2022; Samadun et al., 2023), although differences in grade level, subject matter, and research design limit direct comparison.

A parallel literature supports the use of PhET and other virtual environments in science instruction, but the evidence base is concentrated more heavily in secondary and higher education than in elementary settings. Reviews and meta-analyses indicate that simulations can improve

conceptual learning and achievement, especially when embedded within active or inquiry-oriented instruction (Banda & Nzabahimana, 2021; Cao & Hsu, 2023; de Jong et al., 2013; de Medeiros et al., 2024; Salame & Makki, 2021; Santos & Prudente, 2022). Recent work has begun to document the value of PhET in elementary science, yet systematic evidence still identifies elementary education and higher-order outcomes as comparatively underrepresented areas (Diab et al., 2024; Ningrum et al., 2026). Studies combining Guided Inquiry and PhET in Indonesian science and physics contexts have reported positive results for critical thinking, conceptual learning, or science process skills (Bahtiar et al., 2024; Fitriani et al., 2024; Ismawati et al., 2023; Mardiyanti & Jatmiko, 2022; Subeki et al., 2022). However, much of this work concerns older learners or different science topics, leaving limited evidence about fourth-grade students' critical thinking while learning force through the combined approach. This gap warrants a context-specific examination rather than an assumption that findings from secondary education transfer directly to elementary classrooms.

Accordingly, this study examined critical thinking outcomes among fourth-grade students who learned the topic of force through Guided Inquiry assisted by PhET simulation and students who received conventional science instruction. The study used a quasi-experimental nonequivalent control group design with intact classes, which permits comparison of pretest and posttest performance but requires caution when the groups differ before the intervention. The study focused on critical thinking as reflected in interpretation, analysis, evaluation, inference, and explanation in an essay-based assessment. Its empirical contribution lies in extending the literature on technology-supported inquiry to an elementary science context that remains less represented in recent PhET reviews. Practically, the findings may inform how teachers combine structured inquiry processes with interactive simulation when designing learning activities on force. The conclusions are limited to the observed sample and are interpreted in light of baseline group differences rather than as definitive causal proof of instructional effectiveness.

METHOD

Research Design

This study used a quantitative quasi-experimental approach with a nonequivalent control group design. The design was selected because the research was conducted with existing fourth-grade classes and did not involve random assignment of individual students. Both groups completed a pretest before instruction and a posttest after instruction. The experimental class received Guided Inquiry assisted by PhET interactive simulation, whereas the control class received conventional science instruction. This design allowed the study to compare group performance across the two measurement occasions while retaining the natural classroom organization. Because the classes were intact, between-group differences at baseline were considered when interpreting the posttest results.

Table 1. Research Design

Group	Pretest	Learning Model	Posttest
Experimental	O ₁	Guided Inquiry assisted by PhET simulation	O ₂
Control	O ₃	Conventional instruction	O ₄

Note. O₁ = experimental group pretest; O₂ = experimental group posttest; O₃ = control group pretest; O₄ = control group posttest.

Participants and Research Context

The study was conducted during the 2025/2026 academic year with fourth-grade students from SDIT Setya Dharma and SD Pelita Harapan. The source manuscript identifies these schools as the research settings but does not report a more specific geographic location or exact intervention dates. The population comprised fourth-grade students in the two schools who were learning the topic of force in Merdeka Curriculum Phase B. Purposive sampling was used to select one intact fourth-grade class from each school, and all students enrolled in the selected classes were included. The experimental group comprised 50 students and the control group 48 students, giving a total sample of 98. The classes were selected on the basis of comparable grade level, curriculum implementation, and learning materials; however, the pretest analysis later identified a statistically significant difference between the groups, so baseline equivalence cannot be assumed.

Research Instrument

Critical thinking was measured with an essay test on the topic of force. The instrument was developed around five indicators reported in the source manuscript: interpretation, analysis, evaluation, inference, and explanation. Before the main study, the instrument underwent expert evaluation for content validity, and the test items were also subjected to empirical validity and reliability testing. The source manuscript reports that all items met its validity criteria and that the instrument was considered reliable for measuring students' critical thinking skills. Numerical validity coefficients, reliability coefficients, the number of items, the number of validators, and the detailed scoring rubric were not available in the manuscript and are therefore not added here. These details should be reported in a final submission if the original research records are available because they are necessary for independent evaluation of measurement quality.

Procedure

The research procedure followed the sequence reported in the source manuscript. First, the researchers prepared the essay test, lesson plans, and learning materials and completed the stated instrument validation process. Both classes then completed a pretest to establish their initial critical thinking performance. During the instructional phase, the experimental group learned through Guided Inquiry assisted by PhET interactive simulation, while the control group received conventional instruction on the same science topic. After the instructional activities, both groups completed the posttest. The resulting scores were tabulated and analyzed using the statistical procedures described below. The source manuscript does not provide the number or duration of instructional sessions, so these details are not inferred.

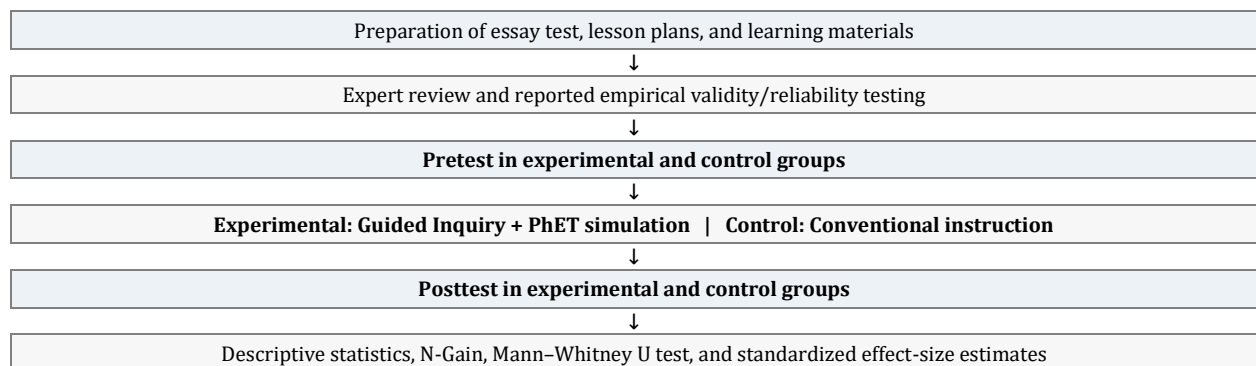


Figure 1. Research Procedure

Data Analysis

The analysis combined descriptive and inferential statistics. Pretest and posttest performance was summarized using means, standard deviations, minimum and maximum scores, and normalized gain (N-Gain) values. The original analysis stated that the data did not satisfy the assumptions required for parametric group comparison, so the Mann–Whitney U test was used to compare the experimental and control groups at each measurement occasion. Statistical significance was evaluated at $\alpha = .05$ using IBM SPSS Statistics. Cohen's d and Hedges' g were also reported as standardized between-group differences for the pretest and posttest. Because these effect-size estimates describe group separation at each measurement occasion, they are interpreted together with the pretest results rather than as isolated estimates of the intervention's causal effect.

RESULTS AND DISCUSSION

Results

The results are presented in three stages: descriptive change and N-Gain, between-group comparisons using the Mann–Whitney U test, and standardized between-group differences. The numerical values are retained from the source manuscript. Particular attention is given to the pretest comparison because the nonequivalent control group design does not guarantee that the two intact classes began from the same level.

Table 2. Descriptive Statistics and N-Gain Scores

Group	n	Pretest Mean $\pm$ SD	Pretest Range	Posttest Mean $\pm$ SD	Posttest Range	Mean N-Gain	Category
Experimental	50	16.12 $\pm$ 5.13	10–24	22.28 $\pm$ 6.26	11–36	0.2739	Low
Control	48	13.92 $\pm$ 3.70	8–26	19.65 $\pm$ 4.21	12–29	0.2126	Low

Both groups obtained higher mean scores at posttest than at pretest. The experimental group's mean increased from 16.12 $\pm$ 5.13 to 22.28 $\pm$ 6.26, while the control group's mean increased from 13.92 $\pm$ 3.70 to 19.65 $\pm$ 4.21. The experimental group also had the higher mean N-Gain (0.2739 compared with 0.2126). Nevertheless, both N-Gain values were categorized as low. The descriptive pattern therefore indicates improvement in both groups and a somewhat larger normalized gain in the experimental group, but it does not by itself establish that the intervention caused the difference.

Table 3. Mann–Whitney U Test Results

Measurement	Mann–Whitney U	Z	p-value	Interpretation
Pretest	919.500	-2.004	.045	Significant
Posttest	919.500	-1.998	.046	Significant

Note. Statistical significance was determined at $\alpha = .05$.

The Mann–Whitney U test showed a statistically significant difference between the groups at pretest ($U = 919.500$, $Z = -2.004$, $p = .045$) and at posttest ($U = 919.500$, $Z = -1.998$, $p = .046$). The posttest result confirms that the two classes differed after instruction. However, the significant pretest result shows that the classes also differed before the instructional treatment. Consequently, the posttest comparison cannot be interpreted as an unambiguous treatment effect without accounting for this initial nonequivalence. Within the available analysis, the most defensible conclusion is that the experimental group ended with higher descriptive performance while baseline differences remain a competing explanation for part of the observed separation.

Table 4. Standardized Between-Group Differences

Measurement	Effect Size	Estimate	95% CI	Magnitude
Pretest	Cohen's d	0.627	0.220–1.031	Medium
Pretest	Hedges' g	0.622	0.218–1.023	Medium
Posttest	Cohen's d	0.618	0.211–1.022	Medium
Posttest	Hedges' g	0.613	0.209–1.014	Medium

Note. Magnitude labels follow the conventional interpretation used in the source manuscript. The estimates are treated as between-group standardized differences at each measurement occasion.

The standardized differences were of medium magnitude at both measurement occasions. Cohen's d was 0.627 at pretest and 0.618 at posttest, while Hedges' g was 0.622 and 0.613, respectively. The close similarity of the pretest and posttest estimates is important because a medium group difference was already present before the intervention. These values therefore provide information about the magnitude of group separation, but they should not be described as evidence that the treatment itself produced a medium effect. A stronger estimate of intervention-related change would require an analysis that explicitly adjusts for baseline performance or directly compares change between groups using an appropriate model.

Discussion

The central pattern of the study is that the experimental class finished with a higher posttest mean and a slightly higher descriptive N-Gain than the control class, but the groups were not equivalent at baseline. This distinction changes how the result should be interpreted. The posttest difference is consistent with the possibility that Guided Inquiry assisted by PhET supported students' critical thinking, yet the significant pretest difference means that some of the posttest separation may reflect initial class characteristics rather than instructional change alone. Quasi-experimental research can provide useful classroom evidence when random assignment is not feasible, but causal claims depend on how well pre-existing differences are addressed. The findings are therefore better positioned as context-specific evidence that the combined approach was associated with favorable

outcomes in the experimental class, not as definitive proof that the intervention alone produced the observed difference.

The descriptive improvement in the experimental group is pedagogically plausible because Guided Inquiry repeatedly places students in situations that require them to interpret a problem, propose or examine explanations, analyze observations, and justify conclusions. Those actions overlap directly with the critical thinking indicators measured in this study. Meta-analytic research has found positive average effects of inquiry-based science instruction on higher-order thinking and critical thinking, while also showing that the strength of outcomes varies across studies (Antonio & Prudente, 2024; Arifin et al., 2025). Earlier synthesis likewise indicates that inquiry is most productive when students receive sufficient guidance and engage with epistemic reasoning rather than merely completing procedural steps (Furtak et al., 2012; Lazonder & Harmsen, 2016). This mechanism is consistent with Indonesian findings that Guided Inquiry can support critical thinking by increasing students' participation in scientific reasoning tasks (Ilhamdi et al., 2020; Poerwanti et al., 2022; Samadun et al., 2023). In the present study, the potential value of the model lies in the structured reasoning opportunities it created, although the available data do not isolate which inquiry phase contributed most strongly to student performance.

PhET simulation may have complemented the inquiry process by providing immediate and manipulable representations of force-related phenomena. Interactive simulations allow learners to change conditions, observe consequences, and revisit a phenomenon more efficiently than is possible through verbal explanation alone, which can make evidence available for analysis and discussion (Banda & Nzabahimana, 2021; Rutten et al., 2012). Research on Newtonian concepts has shown that simulations embedded within inquiry environments can support conceptual and procedural understanding (Alabidi et al., 2023). PhET-specific research also describes how interactive design can guide exploration through visual feedback and constraints, creating a form of implicit support for inquiry (Moore et al., 2013; Wieman et al., 2008). More recent reviews report broadly positive learning outcomes for PhET, while emphasizing that instructional design and teacher facilitation influence what students gain from the technology (de Medeiros et al., 2024; Ningrum et al., 2026). This suggests that the simulation in the present study should not be treated as the intervention by itself; its educational function depended on how it was embedded within Guided Inquiry.

The combination of teacher guidance and interactive simulation is particularly relevant for elementary learners because inquiry can impose substantial demands on planning, evidence coordination, and explanation. Research on computer-supported inquiry has repeatedly shown that the timing and type of guidance matter for successful learning (Gamor, 2021; Zacharia et al., 2015). Technology-enhanced collaborative inquiry likewise benefits from activity structures that help students participate in specific inquiry stages rather than leaving interaction with digital tools unstructured (Chen & Chen, 2025). Studies of primary teachers also indicate that inquiry-based critical thinking is shaped by classroom routines, teacher questioning, and practical implementation conditions (Essalih et al., 2024). The Guided Inquiry framework used in this study is therefore conceptually appropriate because it retains teacher facilitation while giving students a role in investigation. The present data, however, contain no observation measure of implementation fidelity, so the study cannot determine how consistently these forms of guidance were enacted across lessons.

The findings align in direction with earlier studies that combined Guided Inquiry or inquiry-oriented instruction with PhET. Fitriani et al. (2024), Ismawati et al. (2023), and Mardiyanti and Jatmiko (2022) reported favorable critical thinking outcomes in science or physics contexts, while Subeki et al. (2022) found benefits for science process skills. Bahtiar et al. (2024) also reported improved understanding of elementary science concepts under PhET-supported learning. International evidence similarly suggests that virtual laboratories and digital simulations can support learning when they are tied to active scientific tasks (Cao & Hsu, 2023; de Jong et al., 2013; Santos & Prudente, 2022). The present study extends this line of work to fourth-grade learning about force and uses an essay measure focused explicitly on interpretation, analysis, evaluation, inference, and explanation. Its contribution is therefore contextual rather than a claim that the approach has been shown to outperform conventional instruction under all conditions.

One result requiring particular attention is that both groups remained in the low N-Gain category. The experimental class had the larger mean N-Gain, but 0.2739 still indicates limited normalized improvement under the categorization used in the study. This outcome tempers an

interpretation based only on the statistically significant posttest comparison. Reviews of inquiry and simulation research show that positive average effects coexist with considerable variation related to instructional support, implementation, learning duration, outcome measurement, and educational level (Arifin et al., 2025; Dah et al., 2024; Hillmayr et al., 2020; Ningrum et al., 2026). Contextual research also shows that inquiry can produce different patterns across outcomes and school conditions, which cautions against assuming that active learning automatically yields large achievement gains (Gómez & Suárez, 2020; Ramnarain, 2023). Because the present manuscript does not report session duration, implementation fidelity, or an inferential comparison of gain scores, it cannot identify the source of the low gains, and explanations for this pattern should remain provisional.

The baseline difference is also consistent with the effect-size results. Cohen's *d* and Hedges' *g* were already in the medium range at pretest and remained in approximately the same range at posttest. If the posttest effect size were interpreted in isolation, it could be mistaken for a medium treatment effect; viewed longitudinally, however, the similar pretest estimate shows that substantial separation preceded the intervention. This is why nonequivalent control group studies require analytical strategies that distinguish pre-existing differences from change associated with instruction. The current results still have descriptive value because the experimental group improved and retained higher performance, but the evidence does not quantify a baseline-adjusted treatment effect. Future analyses using the original student-level data could consider an appropriate baseline-adjusted model if its assumptions are met, or another change-focused analysis suited to the distribution and design.

Placed within the broader literature, the study addresses an area that remains less developed than research on secondary and higher education. PhET reviews have historically drawn heavily on older learners, while recent elementary studies are only beginning to clarify how simulations function with younger students (Banda & Nzabahimana, 2021; Diab et al., 2024; Ningrum et al., 2026). Inquiry studies with children also show that age-appropriate scaffolding and task design are important for sustaining meaningful investigation (Liang et al., 2021). The present study contributes evidence from fourth-grade science by examining critical thinking rather than only conceptual achievement, and it connects interactive simulation with a structured inquiry model. That contribution is useful precisely because the result is not uniformly strong: the experimental group showed a favorable descriptive pattern, but both *N-Gains* were low and baseline nonequivalence limits causal interpretation. Taken together, the findings support continued investigation of Guided Inquiry with PhET in elementary science using designs that measure implementation quality, report instrument properties in detail, and more directly estimate learning change.

Implications

For classroom practice, the findings support the use of PhET as an integral part of a structured inquiry sequence rather than merely as a stand-alone visual aid. Teachers can design simulation-based activities by guiding students through a sequence of questioning, making predictions, manipulating variables, observing changes, comparing results, and drawing evidence-based conclusions. Through this process, students are not only exposed to visual representations of scientific concepts but are also encouraged to actively interpret information and evaluate relationships among variables. Such instructional organization can help ensure that students' interaction with the digital environment is explicitly connected to the development of critical thinking skills. This recommendation is consistent with previous research indicating that the effectiveness of inquiry-based and simulation-supported learning depends substantially on the quality of instructional guidance and the way digital tools are integrated into classroom activities (Chen & Chen, 2025; Gamor, 2021; Moore et al., 2013).

For research practice, the study also highlights the importance of carefully examining baseline equivalence before interpreting differences in posttest performance, particularly in studies using intact-class designs. Differences that already exist between groups before the intervention may affect the interpretation of subsequent outcomes and should therefore be clearly considered in the analytical process. Future studies would also benefit from providing more detailed information about scoring procedures, instrument reliability coefficients, intervention duration, and the consistency with which the instructional treatment is implemented across groups. Clear documentation of these

methodological aspects would enable readers to evaluate the robustness of the findings more accurately. In addition, such reporting would strengthen the interpretability, transparency, and replicability of future research examining the use of PhET simulations and structured inquiry in educational settings.

Limitations

The study has several limitations that constrain the scope of its conclusions. First, participants were drawn from only two intact fourth-grade classes in two elementary schools and the study focused on a single science topic, so the findings should not be generalized automatically to other grade levels, topics, or educational settings. Second, the quasi-experimental design did not use random assignment, and the groups differed significantly at pretest, which limits causal attribution of the posttest difference. Third, the source manuscript does not provide the number of test items, detailed scoring rubric, number of validators, or numerical validity and reliability coefficients, preventing a fuller evaluation of measurement quality. Fourth, the number and duration of instructional sessions and any measure of implementation fidelity were not reported. Finally, N-Gain was presented descriptively, while the reported inferential tests compared the groups separately at pretest and posttest rather than directly estimating baseline-adjusted change.

Suggestions for Future Research

Future studies should examine the effectiveness of the approach using larger and more diverse elementary school samples and across a broader range of science topics. Expanding the sample characteristics and instructional contexts would provide a stronger basis for assessing the consistency and generalizability of the findings. Future research should also consider designs that improve baseline comparability between groups before the intervention is implemented. Where feasible, researchers may employ random assignment or cluster assignment to reduce systematic differences between comparison groups. When randomization is not possible, appropriate baseline-adjusted analyses can be used to account for initial differences and provide stronger evidence regarding intervention-related changes.

Researchers should also provide more complete information concerning instrument development, including item structure, scoring procedures, validity evidence, and reliability coefficients, so that the quality of the measurements can be evaluated more clearly. In addition, documenting the duration of the intervention and the fidelity of implementation would help distinguish the effects of the instructional model from variations in how the intervention is enacted in different classrooms. Such methodological transparency would strengthen the interpretation and reproducibility of future studies. Further research may also extend the scope of the outcomes examined by investigating whether Guided Inquiry assisted by PhET can support other higher-order learning outcomes. These may include problem-solving skills, scientific literacy, and scientific argumentation, which are also important components of students' science learning and reasoning.

CONCLUSION

This quasi-experimental study examined critical thinking in fourth-grade students learning about force through Guided Inquiry assisted by PhET simulation and through conventional instruction. The experimental group obtained a higher posttest mean and a slightly higher mean N-Gain than the control group, but the N-Gain for both groups remained in the low category. Mann-Whitney U tests showed significant differences at both pretest and posttest, and medium standardized between-group differences were also present at both measurement occasions. These results indicate that the experimental class ended the study with favorable critical thinking performance, while also showing that group nonequivalence existed before the intervention. Accordingly, Guided Inquiry assisted by PhET can be regarded as a promising approach for structuring elementary science investigation and interactive visualization, but the present data do not provide definitive causal evidence that the intervention alone produced the observed posttest difference. Stronger baseline control, fuller measurement reporting, and documentation of implementation are needed before broader effectiveness claims are made.

ACKNOWLEDGMENT

The authors would like to thank Universitas Muhammadiyah Prof. Dr. Hamka, particularly the Faculty of Teacher Training and Education, for academic support during this research. The authors also thank the principals, teachers, and fourth-grade students of SDIT Setya Dharma and SD Pelita Harapan for their cooperation and participation in the research process. The first author also acknowledges the research supervisor for guidance and constructive input during completion of the study.

AUTHOR CONTRIBUTIONS STATEMENT

ANZ: Conceptualization, methodology, investigation, data collection, formal analysis, data interpretation, writing the original draft, reviewing, editing, and finalizing the manuscript.

REFERENCES

- Alabidi, S., Alarabi, K., Tairab, H., Alamassi, S., & Alsalhi, N. R. (2023). The effect of computer simulations on students' conceptual and procedural understanding of Newton's second law of motion. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(5), em2259. <https://doi.org/10.29333/ejmste/13140>
- Alditia, L. M., Fadillah, N., & Hanis, M. (2024). Analysis of barriers to elementary school students' critical thinking skills in science subjects. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, 16(1), 97–111. <https://doi.org/10.37640/jip.v16i1.2060>
- 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>
- Arifin, Z., Sukarmin, Saputro, S., & Kamari, A. (2025). The effect of inquiry-based learning on students' critical thinking skills in science education: A systematic review and meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(3), em2592. <https://doi.org/10.29333/ejmste/15988>
- Bahtiar, B., Ibrahim, I., & Maimun, M. (2024). The effect of PhET simulation-based learning on the ability to understand elementary science concepts in work and energy material. *Jurnal Pendidikan Ilmu Pengetahuan Alam (JP-IPA)*, 5(1), 28–35. <https://doi.org/10.56842/jp-ipa.v5i1.286>
- Banda, H. J., & Nzabahimana, J. (2021). Effect of integrating physics education technology simulations on students' conceptual understanding in physics: A review of literature. *Physical Review*

- Physics Education Research*, 17(2), 023108.
<https://doi.org/10.1103/PhysRevPhysEducRes.17.023108>
- Butcher, K. R., Hudson, M., Lane, M., & Larson, M. (2023). Critical thinking during science investigations: What do practicing teachers value and observe? *Teachers and Teaching*, 29(6), 594–614. <https://doi.org/10.1080/13540602.2023.2191186>
- Cao, X., & Hsu, Y. (2023). Systematic review and meta-analysis of the impact of virtual experiments on students' learning effectiveness. *Interactive Learning Environments*, 31(10), 7433–7454. <https://doi.org/10.1080/10494820.2022.2072898>
- Chen, F., & Chen, G. (2025). Technology-enhanced collaborative inquiry in K–12 classrooms: A systematic review of empirical studies. *Science & Education*, 34, 1731–1773. <https://doi.org/10.1007/s11191-024-00538-8>
- Cromley, J. G., Chen, R., & Lawrence, L. E. M. (2023). Meta-analysis of STEM learning using virtual reality: Benefits across the board. *Journal of Science Education and Technology*, 32(3), 355–364. <https://doi.org/10.1007/s10956-023-10032-5>
- Dah, N. M., Mat Noor, M. S. A., Kamarudin, M. Z., & Syed Abdul Azziz, S. S. (2024). The impacts of open inquiry on students' learning in science: A systematic literature review. *Educational Research Review*, 43, 100601. <https://doi.org/10.1016/j.edurev.2024.100601>
- de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340(6130), 305–308. <https://doi.org/10.1126/science.1230579>
- de Medeiros, R. N., Jr., Naia, M. D., & Lopes, J. B. (2024). PhET interactive simulations in physics teaching practices: A meta-analysis. *Revista Brasileira de Ensino de Física*, 46, e20240186. <https://doi.org/10.1590/1806-9126-RBEF-2024-0186>
- Diab, H., Daher, W., Rayan, B., Issa, N., & Rayan, A. (2024). Transforming science education in elementary schools: The power of PhET simulations in enhancing student learning. *Multimodal Technologies and Interaction*, 8(11), 105. <https://doi.org/10.3390/mti8110105>
- Essalih, S., Ourahay, M., & Khzami, S.-E. (2024). Perceptions and practices of science teachers for the development of critical thinking through inquiry: A qualitative study of Moroccan primary schools. *Education 3-13*, 52(5), 662–677. <https://doi.org/10.1080/03004279.2022.2116942>
- Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics - Physics Education Research*, 1(1), 010103. <https://doi.org/10.1103/PhysRevSTPER.1.010103>
- Fitriani, S. R., Doyan, A., Taufik, M., & Wahyudi, W. (2024). The effect of the guided inquiry model assisted by PhET media on students' critical thinking skills. *Jurnal Penelitian Pendidikan IPA*, 10(6), 3527–3532. <https://doi.org/10.29303/jppipa.v10i6.4852>
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82(3), 300–329. <https://doi.org/10.3102/0034654312457206>
- Gamor, K. I. (2021). Insights on identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science. *Educational Technology Research and Development*, 69(1), 239–242. <https://doi.org/10.1007/s11423-020-09928-5>
- Gómez, R. L., & Suárez, A. M. (2020). Do inquiry-based teaching and school climate influence science achievement and critical thinking? Evidence from PISA 2015. *International Journal of STEM Education*, 7, 43. <https://doi.org/10.1186/s40594-020-00240-5>
- Grangeat, M., Harrison, C., & Dolin, J. (2021). Exploring assessment in STEM inquiry learning classrooms. *International Journal of Science Education*, 43(3), 345–361. <https://doi.org/10.1080/09500693.2021.1903617>
- Hillmayr, D., Zierwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Ilhamdi, M. L., Novita, D., & Rosyidah, A. N. K. (2020). Pengaruh model pembelajaran inkuiri terbimbing terhadap kemampuan berpikir kritis IPA SD. *Jurnal Ilmiah KONTEKSTUAL*, 1(2), 49–57. <https://doi.org/10.46772/kontekstual.v1i02.162>

- Inel-Ekici, D., & Ekici, M. (2022). Mobile inquiry and inquiry-based science learning in higher education: Advantages, challenges, and attitudes. *Asia Pacific Education Review*, 23(3), 427–444. <https://doi.org/10.1007/s12564-021-09729-2>
- Ismawati, Y., Yasa, P., & Rachmawati, D. O. (2023). Pengaruh model pembelajaran inkuiri terbimbing berbantuan PhET terhadap kemampuan berpikir kritis peserta didik. *Jurnal Pendidikan dan Pembelajaran IPA Indonesia*, 13(3), 94–100. <https://doi.org/10.23887/jppii.v13i3.74038>
- Jeon, A.-J., Kellogg, D., Khan, M. A., & Tucker-Kellogg, G. (2021). Developing critical thinking in STEM education through inquiry-based writing in the laboratory classroom. *Biochemistry and Molecular Biology Education*, 49(1), 140–150. <https://doi.org/10.1002/bmb.21414>
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718. <https://doi.org/10.3102/0034654315627366>
- Liang, H.-Y., Hsu, T.-Y., & Hwang, G.-J. (2021). Promoting children's inquiry performances in alternate reality games: A mobile concept mapping-based questioning approach. *British Journal of Educational Technology*, 52(5), 2000–2019. <https://doi.org/10.1111/bjet.13095>
- Mardiyanti, N. E. A., & Jatmiko, B. (2022). Keefektifan pembelajaran fisika dengan model inkuiri terbimbing berbantuan PhET Interactive Simulations untuk meningkatkan kemampuan berpikir kritis siswa SMA. *Jurnal Ilmiah Pendidikan Fisika*, 6(2), 327–336. <https://doi.org/10.20527/jipf.v6i2.5281>
- Moore, E. B., Herzog, T. A., & Perkins, K. K. (2013). Interactive simulations as implicit support for guided-inquiry. *Chemistry Education Research and Practice*, 14(3), 257–268. <https://doi.org/10.1039/C3RP20157K>
- Musliman, A., & Kasman, U. (2022). Efektivitas model inkuiri terbimbing untuk melatih kemampuan berpikir kritis siswa pada konsep fisika yang bersifat abstrak. *Jurnal Jendela Pendidikan*, 2(1), 48–53. <https://doi.org/10.57008/jjp.v2i01.116>
- Ningrum, S. A., Rohman, I., Gumilar, G. G., Mudzakir, A., Hana, M. N., & Nais, M. K. (2026). Effectiveness of PhET simulations on learning outcomes in science and chemistry education: A systematic review. *Multimodal Technologies and Interaction*, 10(7), 69. <https://doi.org/10.3390/mti10070069>
- Pahrudin, A., Misbah, Alisia, G., Saregar, A., Asyhari, A., Anugrah, A., & Susilowati, N. E. (2021). The effectiveness of science, technology, engineering, and mathematics-inquiry learning for 15–16 years old students based on K-13 Indonesian curriculum: The impact on the critical thinking skills. *European Journal of Educational Research*, 10(2), 681–692. <https://doi.org/10.12973/eu-jer.10.2.681>
- Poerwanti, J. I. S., Marmoah, S., & Syawaludin, A. (2022). The effectiveness of guided inquiry model and problem-based learning on critical thinking skills of elementary school students. *Jurnal Pendidikan dan Pengajaran*, 55(3), 666–678. <https://doi.org/10.23887/jpp.v55i3.50940>
- Ramnarain, U. (2023). South African science teachers' experiences of inquiry-based teaching at disadvantaged schools. *Research in Science Education*, 53(4), 727–740. <https://doi.org/10.1007/s11165-022-10095-1>
- Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153. <https://doi.org/10.1016/j.compedu.2011.07.017>
- Salame, I. I., & Makki, J. (2021). Examining the use of PhET simulations on students' attitudes and learning in general chemistry II. *Interdisciplinary Journal of Environmental and Science Education*, 17(4), e2247. <https://doi.org/10.21601/ijese/10966>
- Samadun, S., Setiani, R., Dwikoranto, D., & Marsini, M. (2023). Effectiveness of inquiry learning models to improve students' critical thinking ability. *IJORER: International Journal of Recent Educational Research*, 4(2), 203–212. <https://doi.org/10.46245/ijorer.v4i2.277>
- Santos, M. L., & Prudente, M. (2022). Effectiveness of virtual laboratories in science education: A meta-analysis. *International Journal of Information and Education Technology*, 12(2), 150–156. <https://doi.org/10.18178/ijiet.2022.12.2.1598>
- Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, 34(9), 1337–1370. <https://doi.org/10.1080/09500693.2011.605182>

- Soysal, Y. (2021). Talking science: Argument-based inquiry, teachers' talk moves, and students' critical thinking in the classroom. *Science & Education*, 30(1), 33–65. <https://doi.org/10.1007/s11191-020-00163-1>
- Strat, T. T. S., Henriksen, E. K., & Jegstad, K. M. (2024). Inquiry-based science education in science teacher education: A systematic review. *Studies in Science Education*, 60(2), 191–249. <https://doi.org/10.1080/03057267.2023.2207148>
- Subeki, R. S., Astriani, D., & Qosyim, A. (2022). Media simulasi PhET berbasis inkuiri terbimbing materi getaran dan gelombang terhadap peningkatan keterampilan proses sains peserta didik. *PENSA E-Jurnal: Pendidikan Sains*, 10(1), 75–80. <https://doi.org/10.26740/pensa.v10i1.41459>
- Wiemann, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science*, 322(5902), 682–683. <https://doi.org/10.1126/science.1161948>
- Yaumie, S. N., Jatmiko, B., Satriawan, M., Supardi, Z. A. I., & Arivi, M. A. (2025). The role of guided inquiry to improve critical thinking skills in physics education: A systematic review of Indonesian studies. *Jurnal Pendidikan MIPA*, 26(3), 1893–1908. <https://doi.org/10.23960/jpmipa.v26i3.pp1893-1908>
- Zacharia, Z. C., Manoli, C. C., Xenofontos, N. A., de Jong, T., Pedaste, M., van Riesen, S. A. N., Kamp, E. T., Mäeots, M., Siiman, L. A., & Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational Technology Research and Development*, 63(2), 257–302. <https://doi.org/10.1007/s11423-015-9370-0>