



Developing an Interactive Physics Simulation System to Support the Teaching of Artificial Intelligence for Upper Secondary School Students

Nguyen Quang Linh^{1*}, Tran Kim Phuong²

¹Thai Nguyen University of Education, Thai Nguyen City, Vietnam

²Sa Pa No.1 High School, Lao Cai, Vietnam

ABSTRACT: The integration of artificial intelligence (AI) content into school science subjects remains limited, particularly in developing countries. This study aimed to develop an interactive physics simulation system supporting the teaching of solid deformation in Grade 10 and examine changes in students' AI competency after a learning process with the system. A single-group pretest-posttest design was employed over six weeks with 89 Grade 10 students at Sa Pa No. 1 High School, Vietnam, using parallel-form tests, a Likert self-assessment scale, and observation checklists, with data analyzed via paired-samples t-tests. The results demonstrated statistically significant gains in all three AI competency components-understanding AI data, applying algorithms, and evaluating models ($p < 0.001$)-indicating the feasibility of embedding AI learning within mechanics instruction. Future research should incorporate comparison groups, retention measures, and larger samples across various school contexts to validate these findings.

Corresponding Author:

Nguyen Quang Linh

Published Online:

October 26, 2026

License:

This is an open access article under the CC BY 4.0 license:

<https://creativecommons.org/licenses/by/4.0/>

KEYWORDS:

AI competency, interactive simulation, physics education, machine learning, secondary students

Cite the Article: Linh, N.Q., Phuong, T.K. (2026). Developing an Interactive Physics Simulation System to Support the Teaching of Artificial Intelligence for Upper Secondary School Students. *International Journal of Human Research and Social Science Studies*, 3(10), 951-958. <https://doi.org/10.55677/ijhrsss/09-2026-Vol03I10>

INTRODUCTION

The development of artificial intelligence (AI) is creating significant changes in the labor market, social life, and educational systems across many countries. In this context, artificial intelligence competency (AI competency) is considered one of the core competencies to be developed for high school students to help them become responsible users and co-creators of future AI systems (Miao et al., 2024). Numerous systematic reviews indicate that K-12 AI education has transitioned from the stage of debating its necessity to the phase of constructing specific competency frameworks, curricula, and learning materials (Casal-Otero et al., 2023; Su et al., 2022; Touretzky et al., 2019). In Vietnam, the 2018 General Education Curriculum identifies informatics competency, alongside problem-solving and creativity competencies, as core competencies, simultaneously encouraging the integration of new technological contents into the teaching of school subjects (Ministry of Education and Training, 2018). This orientation establishes a legal and pedagogical foundation for integrating AI content into natural science subjects, including Physics.

Regarding subject content, the Deformation of Solid Bodies section in the Grade 10 Physics curriculum contains abstract concepts and quantities such as stress, strain, elastic limit, and yield strength. Through traditional teaching methods relying on lectures and isolated demonstration experiments, students face difficulties in visualizing the quantitative relationship between applied force and deformation, as well as the transition from the elastic region to the plastic deformation region of materials. The condition of experimental equipment in many high schools, particularly in mountainous areas, often does not permit students to perform repeated measurements across various material types and force levels. Studies on computer simulations in science teaching indicate that interactive simulations possess the capacity to visualize unobservable quantities, enabling learners to modify parameters and observe immediate consequences, thereby supporting the knowledge construction process (de Jong & van Joolingen, 1998; Rutten et al., 2012; Smetana & Bell, 2012; Wieman et al., 2008).

Beyond the visualization function, simulation environments hold the potential to serve as contexts for students to experience the fundamental components of an AI system, encompassing data collection, predictive model training, and model evaluation. When students autonomously collect datasets concerning spring elongation or material bar deformation, and subsequently apply regression algorithms to predict unmeasured values, they gain the opportunity to observe how a machine learning model is constructed from data and to identify the model's limitations when extrapolating beyond the training data domain. This approach aligns with the perspective that AI should be taught through experiences embedded within subject contexts rather than merely being presented as isolated informatics content (Kasneci et al., 2023; Ng et al., 2021). From these analyses, developing an interactive physics simulation system integrated with predictive algorithms, while assessing its impact on the development of students' AI competency, constitutes a research problem of theoretical and practical significance.

The study aims at two objectives. First, to design an interactive simulation system supporting students in investigating the Deformation of Solid Bodies in the Grade 10 Physics curriculum, incorporating elements of artificial intelligence via a data collection module and a predictive regression model. Second, to evaluate the development of students' AI competency components through an experimental teaching process utilizing a single-group pretest-posttest design without a control group.

The research was conducted to answer the following two questions:

- Question 1: How should the simulation system for the Deformation of Solid Bodies be constructed to ensure the scientific accuracy of the physics content while concurrently integrating the essence of artificial intelligence, including the roles of data, algorithms, and the model evaluation process?
- Question 2: How does students' artificial intelligence competency change after participating in the learning process with the interactive simulation system, assessed across the components of data understanding, application of algorithms to physics problem-solving, and AI model evaluation?

Artificial intelligence competency in general education

The concept of artificial intelligence competency, frequently denoted by the terms AI literacy or AI competency, has been defined variously across research literature. Long and Magerko (2020) define AI literacy as a set of competencies enabling individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and utilize AI as a tool within online, domestic, and workplace contexts; they concurrently propose 17 component competencies organized around five conceptual AI questions. Ng et al. (2021), based on a systematic review, conceptualize the structure of AI literacy into four dimensions: knowing and understanding AI, using and applying AI, evaluating and creating with AI, and AI ethics. From a curriculum perspective, Touretzky et al. (2019) propose five big ideas for K-12 AI education, comprising perception, representation and reasoning, machine learning, natural human-AI interaction, and societal impact. Recent reviews on AI literacy in K-12 and adult education indicate a diversity in conceptual frameworks but share a common emphasis on the tripartite axes of knowledge, skills, and attitudes towards AI (Casal-Otero et al., 2023; Laupichler et al., 2022).

At the international orientation level, UNESCO's AI competency framework for students identifies 12 component competencies across four dimensions: human-centered mindset, ethics of AI, AI techniques and applications, and AI system design; each competency is described along three progression levels: understand, apply, and create (Miao et al., 2024). Within the scope of this study, grounded in the UNESCO framework and the specific content characteristics of the Deformation of Solid Bodies section, three competency components were selected for fostering and evaluation: (1) AI data understanding competency, manifested through students collecting, organizing, cleaning, and commenting on the quality of experimental datasets; (2) AI algorithm application competency, manifested through students utilizing regression models to predict spring elongation and material yield strength; (3) AI model evaluation competency, manifested through students analyzing prediction errors, comparing model results with measured data, and identifying model limitations during extrapolation. These three components correspond to the techniques and applications dimension at the understand and apply levels in the UNESCO framework, while aligning with the cognitive characteristics of Grade 10 students.

Interactive simulations and machine learning in physics education

Computer simulations have been investigated as instructional media in science education for decades. De Jong and van Joolingen (1998) analyzed the potential of scientific discovery learning with simulations, while identifying learners' difficulties in hypothesis generation, experimental design, and data interpretation, thereby highlighting the necessity of accompanying pedagogical support. Rutten et al. (2012)'s review of quantitative studies from 2001 to 2010 demonstrated that simulations positively impact learning outcomes when utilized to supplement or partially replace traditional lessons. Smetana and Bell (2012), through reviewing 61 empirical studies, concluded that computer simulations are equally or more effective than traditional instruction in developing conceptual knowledge and science process skills, provided they are integrated into structured instructional sequences. In physics, the PhET project has supplied evidence regarding the efficacy of interactive simulations for learning mechanics, electricity, and thermodynamics concepts, wherein interface design featuring parameter sliders and immediate feedback is deemed a supportive factor for learners' exploratory processes (Perkins et al., 2006; Wieman et al., 2008). Podolefsky et al. (2010) analyzed the

mechanism of implicit scaffolding in simulation design, where constraints and cues embedded in the interface allow learners relatively free exploration while maintaining focus on defined cognitive goals.

Parallel to simulation research, the application of AI in education has expanded into various avenues such as adaptive learning systems, automated assessment, and learning analytics (Chen et al., 2020; Zawacki-Richter et al., 2019). However, the majority of these studies approach AI as a supportive tool for the teaching process, whereas the approach of treating AI as a learning object tied to science subject content remains under-explored, especially at the upper secondary level in developing countries (Casal-Otero et al., 2023; Su et al., 2022). Integrating machine learning algorithms into physics simulations engenders a dual learning structure: students simultaneously acquire mechanics knowledge through simulation manipulation and comprehend machine learning principles through the cycle of data collection, model training, and prediction evaluation. Within this structure, experimental data generated by the students themselves becomes the input for the regression model, and the prediction accuracy becomes the object of analysis, thereby visualizing the relationship between empirical data and mathematical models. This approach inherits the guided discovery learning perspective (de Jong & van Joolingen, 1998) and expands the simulation's function from a visualization medium to an AI experiential environment. In Vietnam, research on K-12 AI education is in its nascent stages and primarily concentrates on the Informatics subject; hence, this study contributes empirical evidence regarding the feasibility of integrating AI content into Physics instruction oriented towards competency development under the 2018 General Education Curriculum (Ministry of Education and Training, 2018).

METHODS OF RESEARCH

Research design

The study employed a single-group pretest-posttest experimental design. According to Campbell and Stanley's (1963) classification, this is a form of pre-experimental design wherein changes in the dependent variable are ascertained by comparing measurement results from the same subject group at two points in time: prior to and following the pedagogical intervention. This design was selected due to the practical constraints of organizing instruction in high schools, where randomly assigning students to experimental and control groups, as well as ensuring equivalence in teachers, instructional time, and learning materials between the two groups, is challenging to execute within an academic year framework. Limitations concerning the design's internal validity, including maturation effects, testing effects, and concurrent history events, are acknowledged and addressed in the results discussion (Fraenkel et al., 2012). The pedagogical experiment was conducted over six weeks, from July 14 to August 24, 2025.

Research participants

The research participants comprised 89 students from two classes, 10A and 10B, at Sa Pa No. 1 High School, Lao Cai province. The sample was selected via convenience sampling based on intact classrooms. Prior to the experiment, the students had not studied the Deformation of Solid Bodies content within the mainstream curriculum and had not participated in any distinct artificial intelligence courses. The study obtained consent from the school administration and homeroom teachers; students and parents were informed of the research purposes, and data were anonymized and utilized strictly for scientific objectives.

Research instruments and materials

The central learning material of the study is the Solid Deformation simulation system, developed using JavaScript combined with SVG graphics, operating directly on web browsers without requiring installation. The system consists of three modules. The physics simulation module permits students to select material types, adjust applied force, cross-sectional area, and initial length of the specimen, while visually observing deformation and real-time force-elongation graphs. The data module enables students to record measurement pairs into a data table, edit, eliminate anomalous values, and export data. The AI module integrates a regression model trained on student-collected datasets, displaying the regression line alongside model evaluation metrics; it allows students to predict spring elongation at unmeasured force values and estimate the material's yield strength, while simultaneously displaying warnings when predicted values fall outside the training data domain. The simulation's physics content was cross-referenced with the achievement requirements of the Grade 10 Physics curriculum and appraised by two physics faculty members prior to use.

The assessment toolkit comprised three types. First, a knowledge and algorithmic thinking test administered at two points (pre- and post-intervention) using a parallel-form structure, consisting of multiple-choice and constructed-response questions measuring the three competency components defined in Section 2.1; the test's reliability was verified via Cronbach's alpha on pilot data. Second, an observational checklist for data thinking behaviors utilized by teachers during software-interactive lessons, recording manifestations such as how students organize measurements, handle anomalous values, and interpret model metrics. Third, an AI competency self-assessment questionnaire utilizing a 5-point Likert scale, constructed based on competency descriptions in the UNESCO framework (Miao et al., 2024) and published AI literacy scales (Laupichler et al., 2022; Ng et al., 2021), subsequently linguistically adapted for Grade 10 students.

Implementation and data collection procedures

The experimental procedure comprised three phases. Phase 1 occurred during the first week, wherein students completed the pre-test and the first self-assessment questionnaire to determine baseline AI competency levels. Phase 2 spanned four weeks, during which students studied lessons regarding Solid Deformation with the support of the simulation system following a sequence of phases: problem identification, simulation manipulation and data collection under varying material and force parameters, predictive model training on the acquired data, model utilization for prediction and empirical verification of predictions, and discussion concerning model errors and limitations. During this phase, teachers utilized the observation checklist to record students' group-based data thinking behaviors. Phase 3 took place in the final week; students completed the parallel-form post-test and the second self-assessment questionnaire. All answer sheets and questionnaires were coded by student IDs to pair pre- and post-intervention data for each individual.

Data analysis methods

Quantitative data were analyzed using JASP software. Descriptive statistics, including means and standard deviations, were calculated for total scores and individual competency indicators at both measurement points. Paired-samples t-tests were utilized to examine statistically significant differences between pre- and post-intervention results with a significance level of 0.05; the assumption of normality for the difference scores was checked via the Shapiro-Wilk test, with the Wilcoxon signed-rank test serving as an alternative if the assumption was violated. Effect sizes were calculated using Cohen's *d* for paired samples and interpreted according to reference thresholds of 0.2, 0.5, and 0.8, corresponding to small, medium, and large effects (Cohen, 1988), with reference to recommendations for calculating and reporting effect sizes in repeated-measures designs (Lakens, 2013). Qualitative data from the observation checklists and constructed-response answers were analyzed based on predefined behavioral indicators to provide supplementary evidence for interpreting the change in each competency component.

RESULTS AND DISCUSSION**Results of designing the interactive simulation system**

The design outcome of the study is the web-based simulation system for the Deformation of Solid Bodies, programmed in JavaScript combined with SVG and Canvas graphics, requiring no installation or server connection during usage. Architecturally, the system comprises three modules interconnected via a common data structure. The physics simulation module allows students to select three object types (spring, steel bar, copper wire) with adjustable parameters including applied force (0 to 500 N), cross-sectional area, and initial length; the object's deformation and the force-elongation graph are updated in real-time, incorporating an installed 2% random measurement noise to simulate real-world experimental errors. The design of parameter sliders accompanied by immediate visual feedback inherits validated design principles from PhET simulation research (Perkins et al., 2006; Wieman et al., 2008), while restricting parameter domains and sequentially unlocking functions according to lesson progression implements the implicit scaffolding mechanism (Podolefsky et al., 2010) to mitigate the risk of undirected exploration observed in simulation learning (de Jong & van Joolingen, 1998).

The data module and the artificial intelligence module are organized into a four-step cycle reflecting the workflow of a supervised machine learning system: data collection, data cleaning, model training, and model evaluation. Students record measurement pairs into a data table, flag and remove anomalous values, and then activate the linear or quadratic polynomial regression model training function on their group's dataset. The interface displays the regression curve superimposed on data points, the equation coefficients, the coefficient of determination R^2 , and the mean prediction error; when a student requests a prediction at a force value outside the training data domain, the system displays an extrapolation warning alongside an expanded uncertainty band. This display method aims to concretize two major ideas of K-12 AI education: machine learning from data and the limitations of AI systems (Touretzky et al., 2019), simultaneously corresponding to component competencies regarding understanding the role of data and critically evaluating AI outputs within current competency frameworks (Long & Magerko, 2020; Miao et al., 2024). Appraisal results from two physics faculty members and two high school teachers, using a 12-criterion evaluation rubric, indicated that criteria regarding scientific accuracy, alignment with Grade 10 Physics curriculum requirements, and interface usability achieved consensus scores of 3.5 or higher on a 4-point scale; two feedback points concerning the extrapolation warning language and graph update speed were accommodated prior to experimental implementation.

Results of evaluating students' artificial intelligence competency development

Pre-test and post-test results. The Cronbach's alpha coefficient of the test reached 0.82 in the pre-measurement and 0.85 in the post-measurement, falling within an acceptable range for measurement instruments in educational research (Fraenkel et al., 2012). The Shapiro-Wilk test on the score differences between the two measurements yielded $W = 0.983$, $p = 0.291$; therefore, the normality assumption was satisfied, and the paired-samples t-test was employed. Table 1 presents the comparison of mean scores by each competency component and the total score converted to a 10-point scale.

Table 1. Paired-samples t-test results for pre-test and post-test (N = 89)

Competency component	Pre-impact M (SD)	Post-impact M (SD)	t(88)	p	Cohen's d
AI data understanding	4.82 (1.53)	7.16 (1.38)	14.27	< 0.001	1.51
Applying AI algorithms	3.95 (1.61)	6.48 (1.52)	13.02	< 0.001	1.38
Evaluating AI models	3.41 (1.47)	5.87 (1.59)	11.60	< 0.001	1.23
Total score (10-point scale)	4.06 (1.14)	6.50 (1.18)	15.84	< 0.001	1.68

Note: Each component is graded on a 10-point scale; d values are calculated for paired samples.

The data in Table 1 demonstrate that the mean scores of all three competency components increased with statistical significance following the intervention ($p < 0.001$). The effect size ranged from 1.23 to 1.51 across components and reached 1.68 for the total score, exceeding the 0.8 threshold deemed as a large effect by Cohen's (1988) conventions. The highest increase pertained to the AI data understanding component, while the lowest increase corresponded to the AI model evaluation component; the post-intervention mean score of the model evaluation component (5.87) was also lower than the other two components, indicating that this is a more challenging component to develop within a 6-week timeframe.

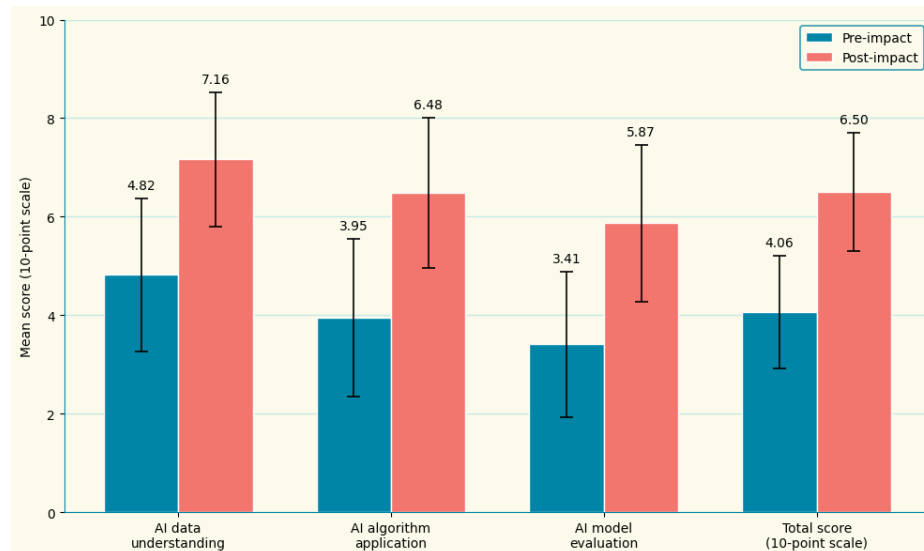


Figure 1 visually represents the changes in mean scores and standard deviations of the components.

Self-assessment questionnaire results. The self-assessment scale comprised 18 indicators divided into three groups corresponding to the three competency components; the Cronbach's alpha coefficient reached 0.88 and 0.90 across the two measurements. The comparison results are presented in Table 2.

Table 2. Students' self-assessment results on artificial intelligence competency using a 5-point Likert scale (N = 89)

Self-assessment indicator group	Pre-impact M (SD)	Post-impact M (SD)	t(88)	p	Cohen's d
AI data understanding	2.74 (0.61)	3.86 (0.55)	10.57	< 0.001	1.12
Applying AI algorithms	2.51 (0.66)	3.65 (0.58)	9.91	< 0.001	1.05
Evaluating AI models	2.38 (0.59)	3.42 (0.63)	9.25	< 0.001	0.98
Entire scale	2.54 (0.55)	3.64 (0.52)	11.23	< 0.001	1.19

Note: The scale ranges from 1 (completely unable to do) to 5 (able to do independently).

The mean self-assessment score of the entire scale increased from 2.54 to 3.64, corresponding to a shift from the level of able to do with substantial support to able to do with limited support. The order of increase among the three indicator groups was consistent with the test results, wherein the AI model evaluation group possessed the lowest post-intervention score (3.42). The consistent trend between self-assessment data and test data contributes to solidifying the convergent validity of the measurements, although self-assessment scores may be susceptible to subjective factors and social desirability bias.

Data thinking behavior observation results. The observation checklist was implemented on 18 study groups (4 to 5 students per group) during software-interactive lessons in the first and fourth weeks of Phase 2. Table 3 presents the percentage of groups exhibiting the recorded behavioral manifestations.

Table 3. Percentage of student groups exhibiting data thinking behaviors in week 1 and week 4 (n = 18 groups)

Data thinking behavioral manifestations	Week 1 (% groups)	Week 4 (% groups)
Recording measurement data in a tabular structure, unifying units of measurement	38.9	88.9
Detecting and processing anomalous measurement values before model training	22.2	77.8
Commenting on the quantity and distribution of data points	16.7	72.2
Comparing the model's predicted values with verification measurements	27.8	83.3
Stating the model's limitations when extrapolating beyond the training data domain	11.1	61.1

The observational data indicate an increase across all five manifestations, wherein the manifestation of stating the model's extrapolation limitations held the lowest initial rate (11.1%) and reached 61.1% in week 4. The analysis of constructed-response answers in the post-test supplied supplementary qualitative evidence. Regarding the data understanding component, 71 out of 89 students (79.8%) could describe at least two data cleaning operations, such as removing anomalous values or conducting repeated measurements to reduce errors. Concerning the algorithm application component, 64 students (71.9%) correctly utilized the regression equation provided by the software to predict elongation at unmeasured force values and cross-referenced it with Hooke's law within the elastic region. For the model evaluation component, 49 students (55.1%) demonstrated that the linear regression model yielded substantial errors when the material exceeded its elastic limit and explained the cause based on the fundamentally altered force-elongation relationship; the remaining students primarily stopped at remarking on the magnitude of the error without connecting it to the domain of the training data.

DISCUSSION

The research findings provide initial evidence regarding the feasibility of integrating artificial intelligence content into Grade 10 mechanics instruction via an interactive simulation environment. The concurrent improvement trends across three data sources (tests, self-assessment questionnaires, observation checklists) align with assertions in recent reviews that K-12 AI education yields more favorable outcomes when AI content is anchored within subject contexts and student practical activities rather than being presented in isolation (Casal-Otero et al., 2023; Su et al., 2022). Allowing students to generate their own datasets through simulation manipulation, and subsequently observe a model trained on that very data, facilitates the connection between the concept of training data in AI and the concept of measurement error in physics—an interdisciplinary linkage recommended in UNESCO's competency framework within the techniques and applications dimension (Miao et al., 2024; Ng et al., 2021). Results pertaining to the efficacy of simulation-based learning also maintain consistency with prior reviews on simulations in science education (Rutten et al., 2012; Smetana & Bell, 2012).

The discrepancy among competency components necessitates further analysis. The AI model evaluation component exhibited a lower post-intervention score and behavioral manifestation rate than the other two components. This outcome is consistent with Laupichler et al.'s (2022) observation that higher-order competencies related to the critical evaluation of AI systems typically demand lengthier developmental periods compared to recognition and utilization competencies. In the context of this study, model evaluation required students to simultaneously coordinate physics knowledge of elastic limits, statistical knowledge of errors, and

an understanding of extrapolation mechanisms; consequently, the 6-week timeframe might have been insufficient for the majority of students to attain the level of independent application. Furthermore, several potential influencing factors include: the software's novelty might have elevated student interest and effort during the experimental phase; participating teachers received direct training from the research team; and conducting the experiment during the summer might present conditions distinct from mainstream teaching concerning time pressure and group activity class sizes.

The utilization of a single-group pretest-posttest design imposes constraints on interpreting the results. Without a control group, the score increments cannot be exclusively attributed to the simulation system; sources threatening internal validity, such as student maturation, testing effects, and concurrent events, remain uncontrolled (Campbell & Stanley, 1963; Fraenkel et al., 2012). Effect sizes calculated on paired samples in repeated-measures designs are frequently higher than those derived from independent groups due to their dependence on the correlation between the two measurements; hence, the *d* values in this study must be compared cautiously with studies employing control groups (Lakens, 2013; Morris & DeShon, 2002). Furthermore, when juxtaposed with the average reference levels for educational interventions (Hattie, 2009), the magnitude of the obtained *d* values might partially reflect the effect of experimental conditions that were better supported than standard, widespread implementations. Nonetheless, in school contexts where random assignment is impermissible, a single-group design combined with multiple data sources still yields valuable information for the initial development and testing phases of learning materials; adding control groups and expanding samples across different localities represent necessary subsequent steps to fortify the solution's empirical validity.

CONCLUSIONS

This study aimed to design an interactive physics simulation system supporting the teaching of solid deformation and to evaluate its impact on the artificial intelligence competency of Grade 10 students. The results demonstrated a statistically significant improvement across all three competency components—understanding AI data, applying AI algorithms, and evaluating AI models—with the data understanding component exhibiting the highest gains, thereby providing initial evidence for the feasibility of embedding AI learning experiences within mechanics instruction. However, the study is limited by the use of a single-group design without a control group, a short intervention duration of six weeks, and a specific sample context, which restrict causal interpretations and broad generalizability. Future research should employ quasi-experimental designs with control groups, incorporate retention measurements, and expand the simulation's application to other physics topics to further validate and generalize these findings.

REFERENCES

1. Bộ Giáo dục và Đào tạo. (2018). Chương trình giáo dục phổ thông: Chương trình tổng thể (Ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26/12/2018). <https://moet.gov.vn>
2. Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.
3. Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10, 29. <https://doi.org/10.1186/s40594-023-00418-7>
4. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
6. de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201. <https://doi.org/10.3102/00346543068002179>
7. Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.
8. Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
9. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
10. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
11. Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>

12. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. <https://doi.org/10.1145/3313831.3376727>
13. Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO. <https://doi.org/10.54675/JKJB9835>
14. Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989X.7.1.105>
15. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
16. Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., & LeMaster, R. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44(1), 18–23. <https://doi.org/10.1119/1.2150754>
17. Podolefsky, N. S., Perkins, K. K., & Adams, W. K. (2010). Factors promoting engaged exploration with computer simulations. *Physical Review Special Topics - Physics Education Research*, 6(2), 020117. <https://doi.org/10.1103/PhysRevSTPER.6.020117>
18. 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>
19. 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>
20. Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on the academic use of artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence*, 3, 100065. <https://doi.org/10.1016/j.caeai.2022.100065>
21. Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 9795–9799. <https://doi.org/10.1609/aaai.v33i01.33019795>
22. Wieman, 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>
23. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>