TPM Vol. 32, No. 3, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

A COMPARATIVE ANALYSIS OF AI-GENERATED AND INSTRUCTOR-DESIGNED LESSON PLANS: UNDERGRADUATE ACADEMIC ACHIEVEMENT IN NUCLEAR PHYSICS

ZUBAIR YOUNAS[A] MUHAMMAD NAEEM SARWAR[B], SHAHID RAZA[C] SUMAIRA KULSOOM[C]

[A]ASSISTANT EDUCATION OFFICER, SCHOOL EDUCATION DEPARTMENT, PUNJAB, PAKISTAN,
aeozubairyounas@gmail.com
[B] DEPARTMENT OF STEM EDUCATION, DIVISION OF EDUCATION, UNIVERSITY OF EDUCATION, LAHORE, 54770, PAKISTAN,
[C]DEPARTMENT OF EDUCATIONAL LEADERSHIP AND POLICY STUDIES, DIVISION OF EDUCATION, UNIVERSITY OF EDUCATION, LAHORE, 54770, PAKISTAN,

Abstract
This study presents a comparison of the academic success of undergraduate students studying nuclear physics when they were taught using lesson plans developed by ChatGPT versus lesson plans developed by their teacher. We evaluated the effectiveness of artificial intelligence in designing lesson plans and the potential of such tools to complement traditional teaching methods. We had a qualitative and a quantitative component to our study. The former consisted of a comparison of the academic achievement of the two groups of students. The second section determined the perception of the students and the instructors regarding the corresponding lesson plans and their effectiveness in understanding nuclear physics. Both groups were given pre- and post-tests. We were interested to know how the instruction was given, the extent to which students were involved in the instruction given, and the level of satisfaction they had with the whole experience. There were substantial learning acquisitions in the experimental as well as control group. However, the lesson plans developed with ChatGPT seemed to be just as effective, if not more so in some areas, as the teacher-developed plans in facilitating students' deeper understanding of the material. This study looks at the possible ways AI can help traditional teaching methods and emphasizes a collaborative, partnership approach to lesson design. The results feed into the increasing conversation about what happens when educators integrate AI into their programs and the effect it will have on something as foundational as STEM education.
Keywords:
Academic Achievement; Nuclear Physics Education; ChatGPT-Generated Lesson Plans; Teacher-Developed Lesson Plans; Artificial Intelligence in Education

INTRODUCTION

Nuclear physics explains phenomena like nuclear binding energy, nuclear forces, and nuclear radiation that are not directly observable and may contradict everyday experience, making them hard for students to conceptualize (Atlabachew & Kriek, 2024). The subject requires advanced mathematical skills and precise manipulation, which many undergraduates find overwhelming, especially when combined with fast-paced curricula (Keblbeck et al., 2024; Royesh et al., 2021). Students may hold incorrect prior knowledge about key topics such as radiation health effects, detector operations, and nuclear reactions. These misconceptions can persist even after demonstrations or instruction if not explicitly addressed (Atlabachew et al., 2024; Sones et al., 2014). Limited Access to hands-on experiments is often restricted due to safety, cost, or equipment limitations, further distancing students from the physical reality of nuclear phenomena (Frybortova et al., 2024; Geschwind et al., 2024).
In accordance with the hitherto optimistic and rigorous Physics Education Research (PER), the orthodox teaching style can be deemed inadequate to cultivate deeper conceptual insights in physics. To tradeoff this vulnerable area, an imprescriptible experiential learning, instructor-guided teaching, and interactive discussions are opted as a typified paradigm and thought to be an effective remedy to encounter and attenuate the students' illusions and to foster their class engagement (Atlabachew et al., 2024; Sones et al., 2014; Geschwind et al., 2024). The conventional wisdom attributed to the PER has yielded critical aspects about the students' understanding and appears to be a prospective designing instruction that is capable enough to engage with both valid and invalid student conceptions, thus fostering an integrated learning (Atlabachew et al., 2024). Nevertheless, so forth, the execution and efficacy of these approaches are intermittent, paradoxical, and seems scarce most often owing to the several retaliations entailing dynamic implementation of an inflexible curriculum, insufficient pedagogical

preparation of instructors, and the minimal incorporation of modern educational tools such as simulations, information technology, pharmacology education, and non-contrastive learning (Cardoso et al., 2020; Li et. al., 2025; Cao et. al., 2025).

Physics Education Research (PER) has emerged in U.S. universities as a scholarly field alongside traditional physics domains, critically evaluating higher education teaching and learning. Kortemeyer (2020) provided example information on how to develop Discipline-Based Educational Research (DBER) in Europe. Polverini and Gregorcic (2024) present the large language models (LLMs) to the physics education community with examples of how prompt engineering can enhance the capabilities of ChatGPT on conceptual physics problems, and discuss how such models can be used in teaching and learning.

The potential of physics education research through generative AI models such as ChatGPT has been demonstrated by solving concept inventories (e.g., the Force Concept Inventory (FCI) and making synthetic data with diverse responses (Kieser et al., 2023; Liu et. al., 2025; Song et. al., 2025). Physics teacher education research has revealed the complexity of the intersection of mathematics and physics, and students often focus on calculations instead of conceptual understanding (Jacobsson et al., 2023). A case study revealed that ChatGPT, though advanced linguistically, repeatedly created contrary responses to elementary physics questions, making it not reliable as a tutor but helpful for training teachers in judging misconceptions (Gregorcic & Pendrill, 2023; Jing et. al., 2024).

A significant ratio of students remains confused about atomic and nuclear phenomena and cannot distinguish between processes at the atomic and sub-atomic scales involving the nuclei (Cardoso et al., 2020). Typical misunderstandings about the nature of nuclear properties, energy quantization, and the mechanisms of radioactive decay are key examples in this regard. Even some students believe that the energy of atoms is not quantized, or that photons can be partially absorbed in different scenarios, or misunderstand that emission of photons has nothing to do with electronic transitions (Savall-Alemany et al., 2019). Advanced methods, including collaborative learning, simulations, cyber space, and computer-assisted solutions, may lead the students to imagine the abstract concepts. These not only eliminate the misconceptions but also establish the deeper understandings of fundamental concepts (Cardoso et al., 2020; Liu et. al., 2025; Pan & Xu, 2025). The role of teachers remains primarily in identifying and diagnosing the misconceptions. Suitable training may enhance the problem-solving skills of students. The demonstrations that explain the nuclear physics for real-time applications encourage devoted participation in platforms for the emergence of positive attitudes and valuable motivation, being crucial to achieving the success endeavours (Bouchée et al., 2023).

Artificial intelligence (AI) tools, particularly LLMs, generative large language models, including ChatGPT, have been used with a fundamental role in education. LLMs can rapidly create attractive teaching lesson plans, puzzles, and activities, lessening teacher workload and supporting increased content updates (Hu et al., 2024). These models can formulate lesson plans, explanations, and descriptions on a prompt, which suggest systematized learning materials, which may align with curriculum goals. Large Language Models organize the content and objectives rationally and clearly, and present the subject matter proficiently, particularly in mathematics and allied STEM fields, even for online learning (Zhang et al., 2025; Huang et. al., 2025). Recent work in STEM education demonstrates that content generated by AI has the potential to enhance student understanding of complex content through providing alternative explanations, stepwise derivations, and interactive problem-solving opportunities (Chan et al., 2025). Artificial Intelligence tools can generate materials to different levels of learning, provide numerous explanations, and modify depending on the student feedback or performance, to serve differentiated instruction (Li, 2024). AI has exhibited its expertise in mathematics and physics, specifically, to offer analytical feedback and different scaffolding, permitting the students to participate with the material at their own pace. Large Language Models (LLMs) can create dialogue-driven quizzes, exercise questions, and personalized feedback, promoting active participation and comprehension (Krupp, Lars et al., 2025)

In line with the advantages, LLMs can provide probable yet untrue or superficial descriptions, particularly when it comes to complex, multi-stage STEM tasks. AI-generated materials sometimes miss domain-specific gradations or are unable to incorporate wider conceptual frameworks, limiting profound comprehension. The problems consist of latent bias, intellectual dishonesty, and less practical validation in real-world classrooms (Hu et al., 2024). Nuclear physics offers steep conceptual barriers and abstract content that traditional methods often struggle to address (Bernabei et al., 2023). LLMs can provide manifold explanations; by producing different analogies and step-by-step reasoning, LLMs can aid students in overcoming conceptual difficulties, a vital challenge in physics education research (PER); adaptive AI can adjust explanations and practice problems to individual student backgrounds, helping students master fundamental concepts. LLMs can rapidly generate new examples, visualizations, and context-specific explanations, addressing the diversity of student misconceptions and learning styles. Automating routine content generation allows educators to focus on higher-order teaching and individualized support (Zhang et al., 2025;Chan et al., 2025; Guo et al., 2025).

A review of the current literature reveals that, while there is growing research on the use of AI-generated lesson plans and their comparison to teacher-designed plans in general science and mathematics education, no study has directly compared these two approaches specifically within the context of nuclear physics (Lee & Zhai, 2024). Existing studies have explored AI-generated lesson plans in science education broadly, including chemistry, biology, and general physics, but none have focused on nuclear physics or conducted a direct, head-to-head comparison between AI-generated and teacher-designed lesson plans in this domain (Peikos & Stavrou, 2025).

The necessity of further research on the effectiveness and subject-specific use of AI-generated lesson plans is pointed out by several studies. These studies frequently mention gaps in empirical evidence for STEM Education. Some studies have compared AI-generated and human-created lesson plans in other disciplines like mathematics or chemistry, but not nuclear physics (Goncalves Costa et al., 2024).

Overall, there is currently no published research that directly compares AI-generated and teacher-designed lesson plans in nuclear physics. This represents a critical gap, as the effectiveness of AI tools in addressing subject-specific misconceptions and supporting conceptual change in physics remains unknown. Addressing this gap is necessary to understand whether AI can be integrated responsibly and effectively into nuclear physics curricula.

This study, therefore, aims to compare the impact of ChatGPT-generated lesson plans with teacher-designed lesson plans in undergraduate nuclear physics education. Based on objectives (1) to evaluate the effectiveness of AI-generated lesson plans in improving students' academic achievement, (2) to assess the effectiveness of teacher-developed plans, (3) to compare differences in student learning outcomes across the two approaches, and (4) to explore students' perceptions of engagement, understanding, and retention, we hypothesize that;

AI-generated lesson plans may be particularly effective in supporting lower-order cognitive skills (remembering, understanding), whereas teacher-designed plans are more likely to promote higher-order thinking skills (analyzing, evaluating, creating).

This comparative analysis directly addresses a pressing research question in PER:

Can AI-generated lesson plans match or complement the pedagogical depth of traditional human-developed instruction in one of the most conceptually challenging areas of STEM?

THEORETICAL FRAMEWORK

The study design and analysis were also based on frameworks that describe how students learn physics concepts, how misconceptions are perpetuated, and how the instruction strategy can either reinforce or break these misconceptions. Instead of making the general education theories survey, this section engages with three theories that are most pertinent to nuclear physics education and the research question under study: (1) Conceptual Change Theory, (2) Knowledge in Pieces, and (3) Bloom Revised Taxonomy. Combined, these frameworks offer a consistent framework of association between theory, methodology, and analysis in this study of AI-generated and teacher-designed lesson plans.

The concept of students entering classrooms with already developed conceptions that shape their perception of new information has long been involved in research work in the field of physics education (PER). Conceptual Change Theory (Posner et al., 1982; Vosniadou, 2008) is a theory of learning that states that not only do people acquire new knowledge, but they also have to reorganize the previously established beliefs. Similarly, other approaches like adaptive multi-teacher are also useful for conceding the smart ideas. In nuclear physics, these prior conceptions are often resistant to change. For instance, students may believe that radioactive decay is a deterministic countdown that fission produces only a few particles, or that nuclear reactions are analogous to atomic processes (Duit, R. et al., 2014; Yu et al., 2023; Kortemeyer, 2020).

This framework helped to design the achievement test and interview protocols operationally based on those misconceptions that were well documented. It also made known to the hypothesis that the AI-generated lesson plans, which are inclined to focus on systematic repetition and simplified explanations, might be especially well-positioned to solve the surface-level comprehension, whereas lesson plans created by the teacher might provide deeper conceptual change based on customized explanations and interactive communication. Conceptual change theory was used in the analysis to give an appreciation of whether the misconceptions of the students remained, changed, or were replaced with scientifically sound models.

Whereas conceptual change theory is more concerned with the restructuring of coherent frameworks, the Knowledge-in-Pieces (KiP) approach by diSessa (1993) is more concerned with the fact that students' reasoning tends to be a result of fragmented and contextually specific knowledge elements, or "p-prims" (Harlow, D.B., & Bianchini, J. A., 2025). In nuclear physics, for example, students may apply intuitive primitives such as "decay = breaking apart into smaller pieces" or "radiation = dangerous substance" (Sabella, M.S., & Redish, E. F., 2007; Polverini et al., 2023). These elements may be productive in some contexts but misleading in others.

This framework was operationalized in the study when the analysis of the interview data was conducted, though the statements of the students were analyzed to provide fragmented reasoning. An example of a student's statement that "the nucleus emits radiation because it is unstable as an object in a shaking" would be an example of a p-prim analogy that assists with sense-making, however, not a scientific explanation. Lesson plans generated by AI (and lots of them based on verbal analogies) have a way of reinforcing such p-prims, whereas teacher-authored ones can take these p-prims directly on, placing the explanations in the context of physics models and empirical research. The inclusion of the framework of the KiP allowed the analysis to predict not only the learning of students of the correct answer but also the influence of their discontinuous reasoning on their learning paths.

In order to assess cognitive performance, this study used the Revised Taxonomy by Bloom (Krathwohl, D. R. 2002; Huang et. al., 2024) as a measure to distinguish the lower and higher levels of thinking. Nuclear physics education involves mastery at several levels: memorizing definitions, comprehension of decay laws, application of equations, analysis of decay chains, and evaluation of models. Previous Physics Education Research studies recommend that instructional design differentially supports these levels: teaching of algorithm may reinforce

recall, while inquiry-based methods endorse analysis and evaluation (Pendrill & Gregorcic, 2019; Werth, A., West, C. G., & Lewandowski, H. J., 2022).
Bloom's Taxonomy was also operationalized in the construction of the achievement test, in which items were placed onto cognitive levels such as remembering and understanding (lower-order) to analyzing and evaluating (higher-order). This also informed the hypotheses: AI-generated lesson plans can help support lower-order skills by offering guided repetition and working examples; conversely, teacher-created lesson plans are assumed to be useful in facilitating higher-order skills by encouraging contextualization and problem-based activities. Test results in analysis have been disaggregated according to the level of cognition, and this enabled the research to determine various effects of AI and teacher instruction across the Bloom hierarchy.

METHODOLOGY

Data were collected both quantitatively and qualitatively. For quantitative phase quasi-experimental design was applied. Pre and post tests were conducted. Experimental groups was taught through Chat Gpt generated lesson plans while control group was taught through teacher made lesson plans. Qualitative data were collected through focus group discussions. The students' responses were collected after the intervention.
The study compared the academic achievement of the two groups in undergraduate-level nuclear physics to determine the effectiveness of each instructional approach. The type of lesson plan (ChatGPT-generated vs. Teacher-developed) was the independent variable, while students' academic achievement in nuclear physics (measured by post-test scores) was the dependent variable. Prior knowledge of physics, duration of lessons, and teaching environment were kept constant to ensure accurate comparison. The study involved a total of 88 undergraduate students enrolled in a Nuclear Physics course at a public sector university in Pakistan. Participants were from the [Bachelor of Science/BS] program, majoring in Physics. The average age of participants was 20.4 years (SD = 1.2), with ages ranging from 19 to 23. All participants had previously completed introductory physics courses and voluntarily agreed to participate in the study. Students of different genders were included, reflecting the demographic distribution of the department.
The participants were undergraduate students enrolled in a nuclear physics course at a public sector university. The selection criteria for participants were that the students were enrolled in the same nuclear physics course with a common curriculum and assessment structure. Students were divided into two sections to facilitate the quasi-experimental design (one section using ChatGPT-generated lesson plans and the other using teacher-developed lesson plans). The total sample size was 88 students, randomly assigned to each instructional group.
The lesson plans covered key topics in nuclear physics and were structured to ensure consistency in content delivery. Both the AI-generated and teacher-developed lesson plans were constructed to align with an identical set of learning objectives derived from the undergraduate nuclear physics curriculum. The instructional design for each plan maintained a consistent structure, including time allocation for conceptual explanations, procedural problem-solving, interactive activities, and assessments. Although only the Alpha Decay lesson segment was presented in detail within the manuscript for illustrative purposes, comprehensive instructional lesson plans covering the entire 16-week duration of the nuclear physics course were developed to ensure full transparency and allow for independent evaluation of instructional equivalence for both the AI-generated and teacher-developed groups.
To ensure objectivity in evaluating the pedagogical quality of both lesson plan sets, an independent, double-blind review was conducted using a standardized rubric adapted from prior lesson plan quality frameworks (Baytak, 2024; van den Berg & du Plessis, 2023). The rubric assessed curriculum alignment, cognitive level distribution (based on Bloom's Taxonomy), instructional clarity, integration of technology, and learner-centered design features. Two physics education experts, blind to the study conditions, independently rated both sets of lesson plans. Inter-rater reliability was assessed using Cohen's Kappa, which indicated substantial agreement ($\kappa = 0.82$).
Classroom delivery was standardized across both groups. The same instructor facilitated all sessions using either the AI-generated or teacher-developed lesson plans according to the group assignment. Session durations, classroom settings, instructional materials (e.g., visual aids, handouts), and assessment procedures were identical across groups to minimize variability unrelated to the lesson plan source.

Teacher-Made Developed Lesson Plan
Lesson Plan: Alpha Decay
Course Title: Nuclear Physics
Level: Bachelor's in Physics
Topic: Alpha Decay
Duration: 1.5 hours
Lesson Objectives
The lesson should enable the students to:
1. Define and explain the process of alpha decay
2. Write the equations of alpha decay
3. Understand the changes in neutron, proton, and mass numbers of the nuclei
4. Know the conditions of alpha decay

5. Understand the changes in the nuclei after the alpha decay
6. Determine the Q-values of the process
7. Should realize the importance of alpha-decay

Lesson Outline

1. Introduction (10 minutes)

- o Briefly explain the background and introduction to alpha decay
- o Explain why alpha decay occurs
- o Discuss the importance of alpha decay for nuclei to gain stability
- o Briefly explain the conditions of alpha-decay
- o Give sample examples of materials that show alpha decay
- o Mention some of the applications of alpha decay

2. Theory and Concept (30 minutes)

- **Definition and Process of Alpha Decay** (10 minutes)
  - o Start by defining the radioactivity and instability of elements with higher atomic numbers
  - o Define alpha decay with the help of some examples
  - o Discuss the changes in proton, neutron, and mass numbers of the nuclei
  - o Write some equations for alpha decay for some radioactive materials
- **Properties of Alpha Particles** (5 minutes)
  - o Tell about the mass and charge number of alpha particles
  - o Discuss the penetration power and ionizing ability of alpha particles
- **Q-Value of Alpha Decay** (10 minutes)
  - o Discuss how to determine the Q-value during the alpha decay
  - o Give some examples and calculate the Q-values
  - o Explain the kinetic energy of alpha particles
- **Conceptual Understanding through Illustrations** (5 minutes)
  - o Draw some conceptual illustrations for alpha decay (as has been given in Figure 1)
  - o Illustrate the effect of alpha decay on parent atoms

3. Experimental Evidence (10 minutes)

- Give some examples of experiments that confirm alpha decay, like Rutherford's gold foil experiment

4. Problem-Solving Session (20 minutes)

- Give some equations as examples to show the changes in proton and neutron numbers after the alpha decay
- Determine the Q-values of a few alpha decay reactions

5. Summary and Recap (10 minutes)

- Recap key points:

1. Define alpha decay
2. Why does alpha decay occur?
3. Properties of alpha particles
4. Importance of alpha decay

- A question-answer session with the students
- Ask some questions to students to check their understanding

6. Homework/Assignments (5 minutes)

- Ask the students to explore the applications of alpha decay
- Ask them to prepare the notes for their better understanding and exam work
- Ask them to find some exciting questions to increase the stock of knowledge about alpha decay

7. Materials and Resources (5 minutes)

- Share the PowerPoint slides with students
- Share the handouts on the topic
- Share some sources of visual aids, like animations/simulations of alpha decay, like PhET
- Share the References of some good books about alpha decay
- Share internet websites or YouTube channels which may contain authentic helping materials

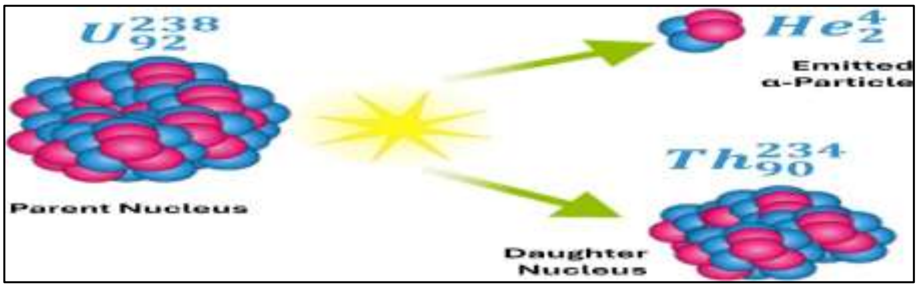


Figure 1. Conceptual schematic of alpha decay (authors' own illustration).'

ChatGPT Developed Lesson Plan

Prompting Procedure and AI Configuration

The AI-generated lesson plans were developed using OpenAI's ChatGPT, version GPT-4, accessed on March 18, 2024. The model was used with default parameters, with a temperature setting of 0.7 and top-p of 1.0 to balance creativity and coherence. The prompting process was undertaken by the corresponding author, who has extensive experience as a physics educator and instructional designer. To ensure consistency, the prompts were guided by predefined course objectives sourced from the university's nuclear physics curriculum.

The specific prompt provided to ChatGPT for lesson plan generation was as follows:

"Generate a detailed, structured lesson plan on the topic of Alpha Decay for undergraduate students studying Nuclear Physics. The lesson plan should include learning objectives, theoretical explanations, problem-solving activities, assessment methods, and integration of visual aids or simulations."

The prompting process was iterative. After generating an initial draft, the output was reviewed against three criteria: (1) alignment with course objectives, (2) conceptual accuracy based on standard nuclear physics content, and (3) instructional structure following lesson planning best practices. If deficiencies were identified, the prompt was refined by adjusting specificity or requesting revisions, and a new output was generated. A total of three iterations were undertaken before finalizing the AI-generated lesson plan used in the study.

Lesson Plan: Alpha Decay
Course Title: Nuclear Physics
Level: Bachelor's in Physics
Topic: Alpha Decay
Duration: 1.5 hours
Lesson Objectives

By the end of this lesson, students should be able to:

1. Define alpha decay and describe the basic process.
2. Understand the properties of alpha particles.
3. Derive the Q-value of alpha decay and explain its significance.
4. Explain the role of quantum tunneling in alpha decay.
5. Analyze experimental evidence supporting alpha decay.
6. Solve numerical problems related to alpha decay.
7. Relate the concept of alpha decay to its applications in science and technology.

Lesson Outline

1. Introduction (10 minutes)

- **Objective:** Set the context for the topic and engage students.
  - Begin with a thought-provoking question: "Why do some nuclei emit particles spontaneously?"
  - Discuss the importance of nuclear stability and its connection to radioactive decay.
  - Provide an overview of the three main types of radioactive decay (alpha, beta, gamma) and emphasize the focus on alpha decay.
  - Present real-world applications of alpha decay (e.g., radioactive dating, nuclear medicine, smoke detectors).
  - Show a brief animation or diagram to visualize alpha decay.

2. Theory and Concept (30 minutes)

- **Definition and Process of Alpha Decay** (10 minutes)
  - Define alpha decay as the emission of an alpha particle (2 protons, 2 neutrons) from an unstable nucleus.
  - Explain the resulting changes in atomic number and mass number with an example (e.g.,).
- **Properties of Alpha Particles** (5 minutes)
  - Discuss the physical and chemical properties of alpha particles:
    - Mass and charge.
    - Speed and kinetic energy.
    - Low penetration power but high ionizing ability.
  - Relate these properties to their practical implications.
- **Q-Value of Alpha Decay** (10 minutes)
  - Derive the Q-value formula:
  - Explain the significance of the Q-value in determining the energy released during decay.
  - Discuss examples of Q-values for common alpha-emitting isotopes.
- **Quantum Tunneling Mechanism** (5 minutes)
  - Introduce the concept of quantum tunneling and its role in alpha decay.
  - Illustrate the tunneling process using a diagram of the potential energy barrier.
  - Explain how tunneling probability influences the decay constant and half-life.

3. Experimental Evidence (10 minutes)

- Highlight key experiments that validate alpha decay:
  - Rutherford's identification of alpha particles.
  - Detection methods (e.g., cloud chamber, scintillation detectors, Geiger-Müller counters).

• Discuss modern techniques for measuring alpha decay.

4. Problem-Solving Session (20 minutes)

• Guide students through numerical problems:
1. Calculate the Q-value for a given alpha decay reaction.
2. Estimate the decay constant using the tunneling probability.
3. Relate the energy released to the kinetic energy of the emitted alpha particle.

• Encourage collaborative problem-solving by pairing students for practice.
• Provide immediate feedback and address misconceptions.

5. Summary and Recap (10 minutes)

• Recap the key points covered:
1. Definition and process of alpha decay.
2. Properties of alpha particles.
3. Q-value and its significance.
4. Quantum tunneling as a mechanism for alpha decay.

• Clarify any lingering questions.
• Conduct a short quiz (3-5 multiple-choice or conceptual questions) to reinforce learning.

6. Homework/Assignments

• Read about applications of alpha decay in nuclear medicine, energy production, and radiocarbon dating.
• Solve additional numerical problems from the course textbook or the provided worksheet.
• Write a short essay (200-300 words) on the significance of quantum tunneling in modern physics.

Materials and Resources

• PowerPoint slides with diagrams, equations, and key points.
• Handouts summarizing the derivation of the Q-value and the tunneling mechanism.
• Visual aids: animations/simulations of alpha decay and tunneling.
• Demonstration equipment (e.g., cloud chamber, Geiger counter, or alpha source).
• Recommended reading: Chapter on Alpha Decay from the Nuclear Physics textbook.

Assessment Methods

• Monitor in-class participation and engagement.
• Evaluate quiz performance at the end of the session.
• Assess the quality and accuracy of homework assignments.
• Provide additional support to students struggling with numerical problems.

Note for Instructor:

• Use visual aids and real-world analogies to explain abstract concepts like quantum tunneling.
• Encourage interactive questioning to promote critical thinking.
• Highlight the connection between theoretical concepts and their practical applications in science and technology.

Comparison of ChatGPT-generated and Teacher-generated Lesson Plans

The ChatGPT-generated lesson plan looks very impressive at a glance. "The contents are worthy of consideration for integration into undergraduate physics instruction, given their alignment with curriculum standards and potential to enhance conceptual understanding." However, the comparison of ChatGPT and teacher-generated lesson plans and their deep analysis brings several important aspects to evaluate their merits. "In both the teacher-generated lesson plans and the ChatGPT-developed lesson plans, the learning objectives were presented, relevant, and aligned with the curriculum."

Moreover, after iterative refinement of prompts provided to ChatGPT, the objectives became more specific and measurable. The content coverage was almost comparable except for a small portion about quantum tunneling in the ChatGPT-generated lesson plan, which should not be a part of this lesson plan, as per the level of undergraduate students. The accuracy of scientific content in both lesson plans was accurate and up-to-date. However, the teaching strategies suggested by ChatGPT were superior, particularly those that involve some exercises to enhance the knowledge of students more effectively. It offered diverse learning needs and methodologies and provided a better understanding of the fundamental concepts. Very importantly, the AI-generated lesson plan contains instructional materials, including visual aids, diagrams, animations, and suggestions for simulations of alpha decay. It also included supporting resources and additional materials to enhance understanding. It is interesting to mention that the teacher may recommend the supporting materials through ChatGPT, which could be accessible to all the students.

The assessment techniques given in AI-generated lesson plans included quizzes, discussions, and hands-on activities as per the learning objectives of the course. These look more efficient and productive than described in the teacher-generated lesson plan. An important feature of the AI-generated lesson plans is that they offer efficient feedback mechanisms to include opportunities for immediate and constructive exposure. The objectives were obtained through engagement and interactivity, like group discussions, experiments, or Q&A sessions. The AI-generated lesson plan particularly focuses on developing the critical thinking of the students through activities for students to analyze and synthesize information. This aspect has not been focused on significantly in the traditional teacher-generated lesson plans. In addition, the AI-generated lesson plans considered learning time management, which was not included in the traditional teacher-developed lesson plans. It is important to note that the time allocated for each part of the lesson must be appropriate to offer productive learning.

The analysis of the two lesson plans showed that the AI-generated lesson plans care more about the lesson pace to ensure all content is covered without overwhelming students. The contents of the AI-generated lesson plans were more worthy regarding the involvement of digital tools, including interactive simulations like PhET, to explain alpha decay. It was found that the AI-generated lesson plans may offer innovations that are more achievable and practical, and incorporate modern tools to enhance engagement and understanding. The traditional teacher-developed lesson plan, although it covers such content, could be random or incompatible and could have less real-world relevance, such as nuclear medicine and radiometric dating. The AI-generated lesson plans demonstrated the instructor's role as more influential and the instructor leveraging expertise to enrich the lesson plans.

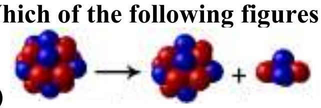
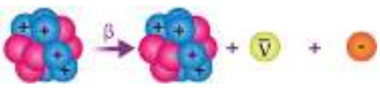
Instrumentations

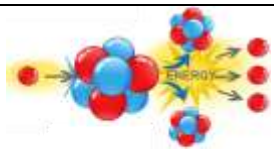
a) Physics Achievement Test

A Physics achievement test was developed by the researcher. This test was conducted before and after the intervention. After completing the instructional period, they took a post-test to measure their academic achievement. The assessments were made up of multiple-choice questions, tasks requiring problem-solving, and conceptual questions that all fit Bloom's taxonomy of learning. After the intervention, the students responded to focus group discussions about the lesson plans. The focus group discussions went into the three areas of engagement, clarity, and effectiveness of the lesson plans. Observations of the classroom were made to see how the students were interacting with the lesson. Before the intervention, both groups completed the same test to assess their existing knowledge in nuclear physics.

This test was the same for both groups to ensure fairness. The experimental group was taught using lesson plans generated by ChatGPT for sixteen weeks and covering particular topics in the nuclear physics curriculum. The control group was taught using teacher-developed lesson plans for the same period, covering the same topics. The teaching duration and frequency of lessons were identical for both groups to maintain consistency. After the sixteenth-week instructional period, both groups took the post-test, which assessed their learning outcomes and academic achievement in nuclear physics. Students completed the perception questionnaire to provide feedback on their learning experience with the respective lesson plans. Observations were conducted during the lessons to record student engagement and teaching effectiveness. This provided qualitative data on the classroom dynamics for each instructional method. The Nuclear Physics achievement test covers the topics "Definition and Process of Alpha Decay," "Properties of Alpha Particles," and Q-Value of Alpha Decay, in the undergraduate nuclear physics curriculum. The researchers prepared test questions according to the learning objectives. The Nuclear Physics achievement test was developed to measure students' performance in the subject of Nuclear Physics. The achievement test consisted of 25 items. The scientific procedure was used to develop the test, which employed a proper table of Specifications to maintain the difficulty level of the test (Mitchell & Jolley, 2010; Yin & Wang, 2025). The Table of Specification helped to include the first three levels of the cognitive domain proportionately. Moreover, all basic rules of test construction were observed while designing the test. To have a balanced test, 25% easy, 50% moderate, and 25% difficult questions were included in the test. However, the recommended difficulty level of the test was 0.6. Rudner and Schafer (2002) was preferred after undergoing item analysis through pilot testing of the sample.

Test Questions

Which of the following equations shows alpha decay? **A)** $\mathbf{X_Z^A \rightarrow Y_{Z-1}^{A-1} + H_1^1}$ **B)** $\mathbf{X_Z^A \rightarrow Y_Z^{A-1} + n_0^1}$ **C)** $\mathbf{X_Z^A \rightarrow Y_{Z-2}^{A-4} + He_2^4}$ **D)** $\mathbf{X_Z^A \rightarrow Y_Z^A + \gamma}$ **Correct Answer: C**	**Which of the following nuclei is most likely to undergo alpha decay?** A) Light elements with Z < 20 B) Medium-mass elements with stable neutron-proton ratios C) Isotopes with equal numbers of protons and neutrons D) Heavy elements with Z > 82 **Correct Answer:** D
In the context of alpha decay, what is the significance of the Q-value? A) It determines the speed of the emitted alpha particle. B) It represents the energy released during the decay process. C) It is the probability of quantum tunneling. D) It measures the half-life of the parent nucleus. **Correct Answer:** B	**Which of the following figures shows alpha decay?** A)  B)  C) 

	 D) **Correct Answer: A**

Quantitative Data Analysis

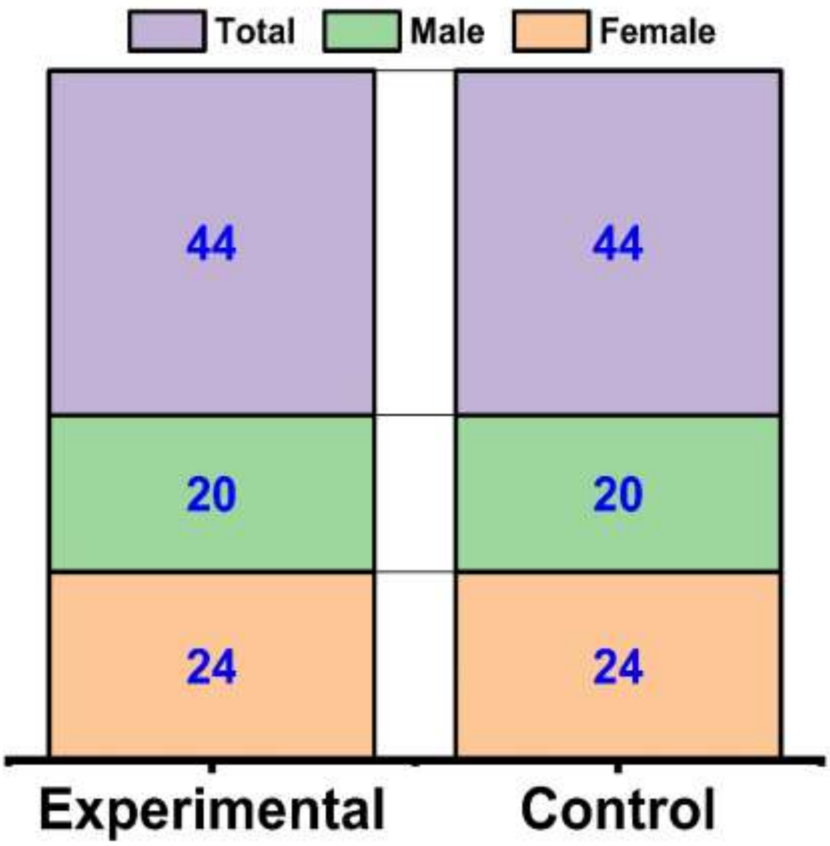


Figure 2 Gender Distribution of the Study Group

Table 1 Summary table for Kolmogorov-Simirnov tests

Group	KS Statistic	p-value	Normality Assumption
Experimental (Pre-test)	0.123	0.935	Normal
Control (Pre-test)	0.106	0.380	Normal

Table 2 Summary table for Kolmogorov-Simirnov tests

Group	KS Statistic	p-value	Normality Assumption
Experimental (Post-test)	0.087	0.107	Normal
Control (Post-test)	0.100	0.825	Normal

Tables 1 and 2 showed that Kolmogorov–Smirnov tests indicated no violations of normality (all $p > .05$).
Tables 1 and 2 show the Kolmogorov-Smirnov test results. This test was conducted to assess the normality of the data distribution for nuclear physics achievement scores in both the experimental and control groups before and after the intervention. For the pre-test, the experimental group exhibited a KS statistic of 0.123 ($p = 0.935$), while the control group showed a KS statistic of 0.106 ($p = 0.380$). Since the p-values for both groups are greater than 0.05, we fail to reject the null hypothesis, indicating that the data in both groups followed a normal distribution before the intervention. Similarly, for the post-test, the experimental group had a KS statistic of 0.087 ($p = 0.107$), and the control group had a KS statistic of 0.100 ($p = 0.825$). Again, both p-values exceed the conventional 0.05 threshold, suggesting that the post-test scores in both groups also followed a normal distribution. Taken as a whole, these results show that the normality assumption holds for both groups at both time points. This confirms

that we can use parametric statistics to analyze the differences in nuclear physics achievement between our experimental and control groups.

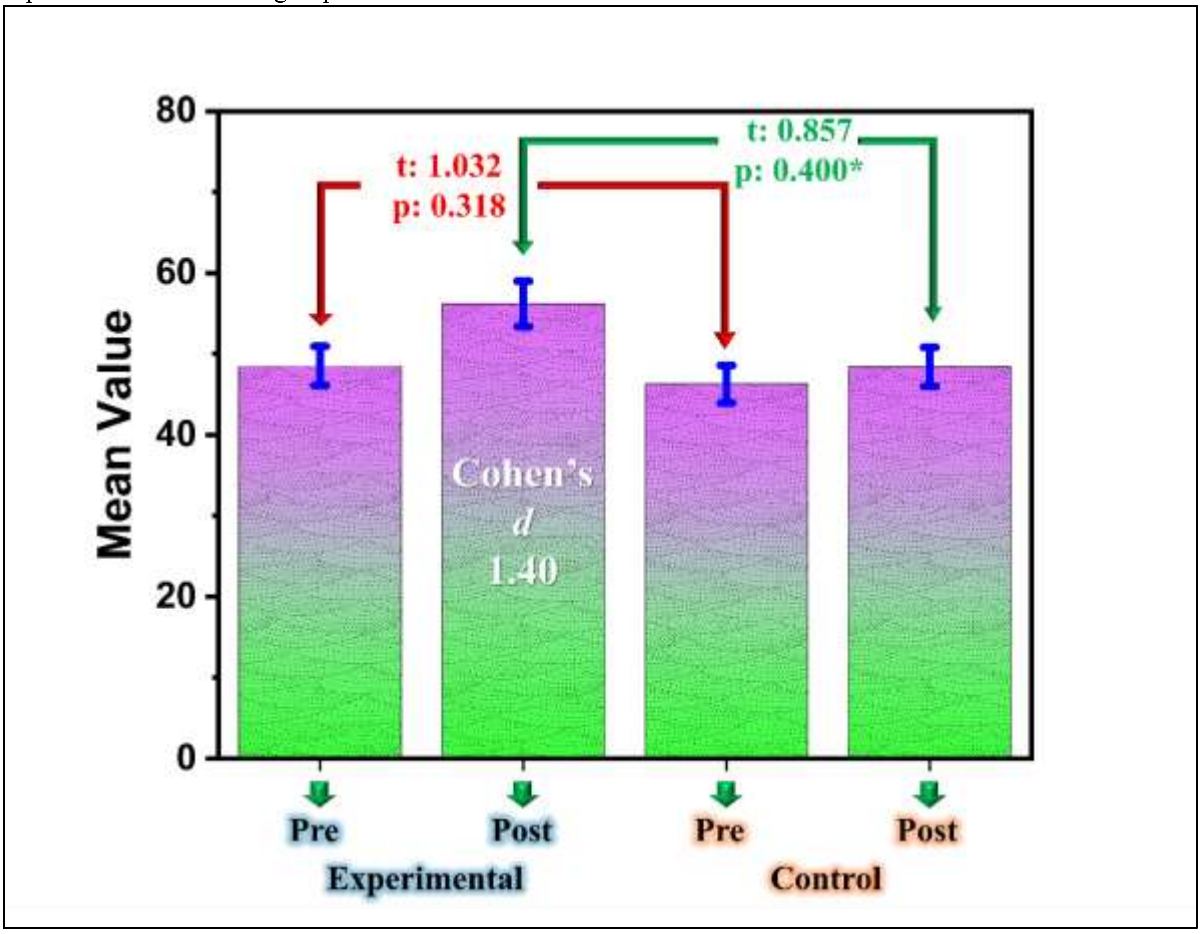


Figure 4.
Comparison of post-test scores between the experimental (AI) and control (teacher-developed) groups. No significant difference was observed (t = 0.9, p = 0.370; Cohen's d = 0.19, small effect size).

Before the intervention, both groups completed the same test to assess their existing knowledge in nuclear physics. An independent-samples t-test confirmed no significant difference in pre-test scores between the experimental group (M = 29.68, SD = 9.52) and the control group (M = 28.12, SD = 9.21), t(86) = 0.75, p = .457, confirming initial equivalence.

Following the intervention, both groups demonstrated significant improvement in academic performance. A paired-samples t-test revealed that the experimental group, which used AI-generated lesson plans, showed a significant increase from pre-test (M = 29.68, SD = 9.52) to post-test scores (M = 47.41, SD = 18.93), t(43) = 7.03, p < .001. Similarly, the control group, taught using teacher-developed lesson plans, exhibited significant gains from pre-test (M = 28.12, SD = 9.21) to post-test scores (M = 55.84, SD = 25.54), t(43) = 5.87, p < .001.

An independent-samples t-test comparing post-test scores between the two groups revealed no statistically significant difference, t(86) = 0.9, p = .370, indicating comparable performance. The effect size, calculated using Cohen's d, was 0.19, reflecting a small effect in favor of the control group. These results suggest that both AI-generated and teacher-developed lesson plans effectively improved students' understanding of nuclear physics, with no significant advantage observed for either approach.

In addition to basic group comparisons, repeated measures ANOVA was conducted to analyze the longitudinal effects of the intervention, examining both main effects of time and instructional group, as well as interaction effects. Assumptions of normality and sphericity were checked before analysis, and effect sizes (Cohen's d) were calculated to interpret the practical significance of the findings.

Table 3 Pre- and Post-Test Comparison with Interpretations

Group	Pre-Test Mean (SD)	Post-Test Mean (SD)	Paired t-test (p-value)
Experimental	29.68 (9.52)	47.41 (18.93)	t = 7.03, p = 0.000
Control	28.12 (9.21)	55.84 (25.54)	t = 5.87, p = 0.000
Between Groups Post-Test	—	—	t = 1.76, p = 0.082

The results of the paired-samples t-tests indicated a statistically significant improvement in academic performance within both groups from pre-test to post-test. For the Experimental group, students taught using AI-generated lesson plans demonstrated a significant increase in scores from pre-test (M = 29.68, SD = 9.52) to post-test (M =

47.41, SD = 18.93), t(43) = 7.03, p < .001. Similarly, the Control group, taught using teacher-developed lesson plans, showed a significant improvement from pre-test (M = 28.12, SD = 9.21) to post-test (M = 55.84, SD = 25.54), t(43) = 5.87, p < .001.
An independent-samples t-test comparing post-test scores between the two groups revealed no statistically significant difference, t(86) = 1.76, p = .082. This suggests that both instructional approaches were effective in enhancing students' understanding of nuclear physics, with neither group demonstrating a clear advantage over the other in terms of academic achievement following the intervention.

Table 4 Repeated Measures ANOVA Results by Group

Group	Source	F	Num DF	Den DF	p-Value
Experimental	Time	49.45	1.0	43.0	< .001
Control	Time	34.41	1.0	43.0	< .001

A within-group repeated measures ANOVA was conducted to examine the effect of time (pre-test vs. post-test) on student achievement scores separately for each group. . Both experimental and control, F(1, 43) = 49.45, p < .001, F(1, 43) = 34.41, p < .001 respectively indicating a significant improvement in academic performance over time. These findings showed that both experimental and control groups showed improvement during intervention.
Qualitative Data Analysis:
A thematic analysis approach (Braun & Clarke, 2006) was applied in this study for the analysis of qualitative data. Both lesson plans and students' responses were analyzed thematically. Thematic analysis is a reliable technique for qualitative data analysis, especially to explore complex phenomena (Nowell et al., 2017). A structured content analysis rubric was adapted for the analysis of both ChatGPT and teacher-made lesson plans (Baytak, 2024; van den Berg and du Plessis, 2023). In this rubric, the indicators were stated in alignment with the standards of the curriculum, the complexity of the instructions in accordance with the Bloom Taxonomy (Spanos, 2024), the quality of the instructional design according to TPACK principles (Niess, 2016), and the successful use of technological instruments to facilitate learning.
In the focus group interviews, audio-tapes were transcribed word-to-word and analyzed in an inductive thematic coding process (Lo, 2023). Open coding was done by two independent researchers who sought key themes and patterns in the data with regards to student engagement, understanding of nuclear physics concepts, and lesson plan effectiveness as perceived by them. Constant comparison was used to refine the codes and pattern recognition, and the emergent themes were discussed among the research team to promote conformability and reduce bias by the researcher (Nowell et al., 2017; Braun & Clarke, 2006; Zhao et. al., 2024). To improve credibility, several methods were used in the study such as Investigator triangulation, member checking on a small number of participants to confirm the interpretations and an audit trail of the coding decisions made. The inter-coder agreement as measured with the help of the Cohens Kappa was past 0.85, and this implies that the coding reliability is high (McHugh, 2012). Based on the discussions of the focus group, we came to understand that the students are considering their experience in undergoing the AI-based instructional intervention in a number of ways.
The students were very interested and engaged, which accentuated that the lessons created by AI were interesting. They explained how the interactive nature of the lessons made them all listen and appear interested in the content of the lessons. For instance, one student stated, "I felt more a part of the learning process because the visuals kept my attention." Another put forth, "It wasn't boring like regular lectures; I actually wanted to participate." The feedback, such as this, shows that the use of AI tools could be contributing to the creation of an environment where students are more willing to engage and participate.
As far as understanding goes, the online simulations and video materials were obvious contributions to developing the ideas of our students. "I finally got nuclear fission because I could see it in action," was one comment. "The 3D visuals made the kind of stuff that's usually impossible to follow much easier to understand," was another.
In the case of technology acceptance, a number of students expressed favorable views towards the application of AI in education. One remarked, "I didn't expect to like it this much—it was easy to follow," while another said, "It made me curious to learn more about AI itself." These responses reflect a general comfort with and interest in integrating technology into their learning routines.
When comparing the intervention to traditional teaching, participants described AI-based lessons as "less monotonous," "more personalized," and "way more fun." This suggests that students perceived the AI-supported instruction to be more adaptable to their pace and preferences.
Despite the benefits, a few challenges were noted. Initial confusion, minor technical issues, and the need for teacher support were mentioned. For instance, one student shared, "It was a bit confusing at first, and I needed help using the tools," while another pointed out, "The app froze once or twice, which interrupted the session."
In general, the responses indicate that the students have a positive experience with AI-based lessons, in reference to engagement, comprehension, and novelty, although with a few usability challenges. The thematic visualization above is a summary of the major themes and sub-themes that were found during the analysis.
In the course of the qualitative analysis of AI-based and teacher-created lesson plans, there were cases of slight conceptual errors in both. In the AI-written lesson plans, it was found that some hallucinations sometimes occurred, including the definition of the concepts of nuclear physics being simplified too much, or the fact that

scientifically inaccurate examples were included. Nevertheless, the number of such errors was quite low, and the team of researchers fixed such errors before their implementation. On the same note, lesson plans created by teachers had slight weaknesses and these were mainly associated with unclear words or discrepancy in the taxonomy levels of Bloom. No severe conceptual errors were found that could significantly impact the validity of the instructional content.

DISCUSSION

This research was carried out to show the impact of ChatGPT lesson plans for a Nuclear Physics class on the academic achievement of undergraduates in a BS Physics program at the university. There were 40 students in the experimental group while 40 in control group. The groups were divided at random, like flipping a coin, and then both groups were taught for a total of 16 weeks. At the end of the treatment, both groups took an achievement test. The two groups were compared on the basis of their test scores.

The findings indicate that both AI-generated and teacher-developed lesson plans significantly improved students' academic performance in nuclear physics. However, no statistically significant difference was observed between the two groups in post-test scores ($t(86) = 0.9$, $p = .370$), suggesting comparable effectiveness. The calculated Cohen's d effect size of 0.19 reflects a small practical advantage in favor of the control group, but this difference is minimal. Furthermore, Repeated Measures ANOVA confirmed significant within-group learning gains for both groups over time ($p < .001$), reinforcing the conclusion that both instructional approaches positively contributed to student understanding. These results align with emerging literature suggesting that while AI tools like ChatGPT can support lesson planning and instructional delivery, they currently perform comparably to traditional, teacher-developed approaches rather than exceeding them (Kasneci et al., 2023; Lo, 2023; Wardat et al., 2023).

This research is also of paramount necessity for teachers and implementers of education who wish to develop their own lesson plans incorporating ChatGPT and similar artificial intelligence models to enhance the learning process. This study suggests that these models can serve as valuable and beneficial tools for teachers and students in the learning process, enabling both parties to utilize them with great efficiency and effectiveness.

Whalen and Mouza (2023) argue that some of the possible positive impacts of ChatGPT on educators can include its capability to assist with lesson planning, give immediate feedback, and customize learning experiences of students. According to Kilinc (2023), lesson plans created with the help of ChatGPT could be used to introduce unique teaching approaches and implement technology in the curriculum to guide teachers in meeting the instructional objectives, leading to providing platforms for smart education (Bhaskar and Gupta, 2024; Liu et. al., 2024).

It is also highlighted by Cooper (2024) that outputs created by ChatGPT have to be assessed by subject-matter experts in the framework of the course. In the case of such professional assessment, the results demonstrate that the lesson plans created with the help of ChatGPT have a great prospect of increasing the academic performance of students. These findings are in line with those of past studies in the literature.

In his research, Rundle et al. (2024) described that by using ChatGPT, instructors can create unique learning experiences based on the needs of students. Wardat et al. (2023) explained that ChatGPT has the potential to alter widely used pedagogical techniques across the science education field and provoke radical change in the same field. According to Adiguzel et al. (2023), educators and researchers should include ChatGPT in education in a manner that is ethical and responsible. When these stakeholders properly incorporate ChatGPT in their conversations about the future of education, then a more interesting and technologically progressive education system can be created that can address the changing demands of learners (Barrot, 2023).

Analysis of ChatGPT-generated lesson plans reveals that they consider a range of important factors, including differentiation, learner needs, and instructional goals. These factors involve highlighting the importance of learning objectives at the beginning of the lesson, connecting them to real-life contexts, promoting inclusive and cooperative learning, and providing timely feedback. In terms of pedagogical sequence, the plans follow a path from simple to more complex tasks, utilize discussion groups, and involve peer learning. ChatGPT's plans also link new learnings to previous knowledge, employ creative methods, and provide feedback. Finally, they inform students of the learning objectives for the next lesson (Abdullah et al. 2022).

Moreover, ChatGPT's plans are a good use of time and have a flexible structure. Such a comprehensive and effective plan was expected to have an impact that would be statistically significantly higher than the impact of the plans prepared by teachers. The instructional design necessitates also factoring in the student's preparedness level, the features of the learning environment, the learning materials on hand, and a host of other student characteristics. If ChatGPT learns these data, a more effective plan will unfold. This suggestion may be considered in the next research. On the contrary, the plans implemented in the control group (which were prepared by teachers) seem to have a similar effect. These plans are (mostly) not as flexible and adaptable as ChatGPT's plans and are prepared in line with (mostly) the course curriculum and the textbook framework stipulated by the University. However, factors such as the prior knowledge of the students, the particular characteristics of the learning environment, and many other (numerous) student traits mentioned above could have made the plans prepared by the control group teachers more effective than ChatGPT's.According to Keiper et al. (2023), ChatGPT does not yield solutions that can be put to use right away, and it still necessitates that human touch, which makes a given reply seem all the more personal. For this reason, we can consider ChatGPT and our plans for it to be a kind of intellectual division of labor that benefits teachers. Plan work can be done collaboratively with ChatGPT. Integrating both lesson planning approaches to complement one another can effectively mitigate the limitations associated with each. However, when interacting with ChatGPT, educators should avoid using overly superficial or excessively brief prompts. Instead, instructors are encouraged to engage in deeper inquiry by employing a series of continuous, contextually rich conversational prompts. When crafting these prompts, it is beneficial for teachers to incorporate detailed descriptions of the learning environment, typical student characteristics, and specific considerations for students with disabilities. Importantly, educators must also uphold critical thinking and apply human creativity to ensure pedagogical soundness and relevance when leveraging ChatGPT as an instructional tool (Barrot, 2023). AI should be use as a tool rather than a goal (Kim & Adlof, 2024). Furthermore, considering that ChatGPT is trained on a vast and globally sourced dataset without a built-in mechanism for assessing the quality or credibility of the information it utilizes, some of its recommendations may lack scientific rigor or reflect subjective perspectives. Additionally, ChatGPT may overlook important cultural nuances, leading to outputs that are not contextually appropriate for all educational settings (Barrot, 2023). Educators must consider all these concerns when integrating ChatGPT into their instructional practices

CONCLUSIONS

This study concludes that AI-generated lesson plans can be as effective as those developed by teachers when it comes to enabling students to grasp the concepts of nuclear physics, provided that the plans are designed with certain qualities in mind. Both methods were effective in enhancing students' academic performance. Yet, when the statistics and effect sizes were compared, there was no clear advantage seen for either method. The results indicate that AI tools can assist in lesson planning; however, they should not be regarded as replacements for teachers' lesson designs.

Future research needs to examine how effective lesson plans made by artificial intelligence are over a longer period of time and across a variety of different subjects, not merely nuclear physics. Improving how we instruct AI and enhancing AI models can help ensure that errors are reduced and that the lesson plans provided to the AI models are more closely aligned with established frameworks like Bloom's Taxonomy. Investigators ought to conduct research with a greater number of students drawn from diverse backgrounds and over an extended period of time to ascertain the long-term effects of AI-driven pedagogies on higher-order thinking skills.

Ethical Considerations

All participants were informed about the purpose and procedure of the study. Written consent was obtained from each student before their participation. Participants' data were anonymized to protect their privacy. The data collected were used solely for research purposes. Students were made aware that their participation was voluntary and that they can withdraw from the study at any time without any consequences.

Limitations

The findings from this study may not be generalizable to other subjects or educational levels beyond undergraduate physics students. The relatively small sample size may limit the ability to detect smaller effects, though it should be sufficient for identifying larger, meaningful differences between the groups. The ChatGPT-generated lesson plans were dependent on the current level of AI development, and any limitations in the AI's understanding of pedagogical content could influence the outcomes.

REFERENCES

1. Abdullah, M., Madain, A., & Jararweh, Y. (2022. ChatGPT: Fundamentals, applications and social impacts. 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS).
2. Adıgüzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. Contemporary Educational Technology.
3. Atlabachew, T. K., & Kriek, J. (2024). Using students' critical aspects to deepen understanding of nuclear physics concepts. European Journal of Physics, 45(6), 65701, https://doi.org/10.1088/1361-6404/ad6064
4. Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 10074
5. Baytak, A. (2024). The content analysis of the lesson plans created by ChatGPT and Google Gemini. Research in Social Sciences and Technology, 9(1), 329–350.
6. Bhaskar, P., & Gupta, P. K. K. (2024). Delving into educators' perspectives on ChatGPT in management education: A qualitative exploration. Interactive Technology and Smart Education.
7. Bouchée, T., Thurlings, M., de Putter-Smits, L., & Pepin, B. (2023). Investigating teachers' and students' experiences of quantum physics lessons: Opportunities and challenges. Research in Science & Technological Education, 41(2), 777–799. https://doi.org/10.1080/02635143.2021.1948826
8. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

9. Cao, Z., Huang, L., Wang, T., Wang, Y., Shi, J., Zhu, A.,... Snoussi, H. (2025). Understanding the Dimensional Need of Noncontrastive Learning. IEEE Transactions on Cybernetics, 55(9), 4089-4102. doi: 10.1109/TCYB.2025.3577745
10. Cardoso, P. S. S., Nunes, M. C. S., Silva, G. P. S., Braghittoni, L. S., & Trindade, N. M. (2020). Conceptions of high school students on atomic models, radiation and radioactivity. Physics Education, 55(3), 35030. https://doi.org/10.1088/1361- 6552/ab7fc6
11. Chan, K. W., Ali, F., Park, J., Sham, K. S. B., Tan, E. Y. T., Chong, F. W. C., Qian, K., & Sze, G. K. (2025). Automatic item generation in various STEM subjects using large language model prompting. Computers and Education: Artificial Intelligence, 8, 100344. https://doi.org/10.1016/j.caeai.2024.100344
12. Cooper, P. K. (2024). Music teachers' labeling accuracy and quality ratings of lesson plans by artificial intelligence (AI) and humans. International Journal of Music Education.
13. diSessa, A. A. (1993). Ontologies in pieces: Response to Chi and Slotta. Cognition and Instruction, 10(2–3), 272–280.
14. Duit, R., Schecker, H., Höttecke, D., & Niedderer, H. (2014). Teaching physics. In N. G. Lederman & S. K. Abell (Eds.), Handbook of research on science education, Volume II (pp. 434–456). Routledge.
15. Gonçalves Costa, G., Nascimento Júnior, W. J. D., Mombelli, M. N., & Girotto Júnior, G. (2024). Revisiting a teaching sequence on the topic of electrolysis: A comparative study with the use of artificial intelligence. Journal of Chemical Education, 101(8), 3255–3263. https://doi.org/10.1021/acs.jchemed.4c00247
16. Gregorcic, B., & Pendrill, A. M. (2023). ChatGPT and the frustrated Socrates. Physics Education, 58(3), 035021.
17. Guo, Q., Zhen, J., Wu, F., He, Y., & Qiao, C. (2025). Can students make STEM progress with large language models (LLMs)? An empirical study of LLMs integration within middle school science and engineering practice. Journal of Educational Computing Research, 63(2), 372–405.
18. https://doi.org/10.1177/07356331241312365
19. Harlow, D. B., & Bianchini, J. A. (2025). Knowledge-in-pieces—Andrea A. diSessa, David Hammer. In J. A. Bianchini (Ed.), Science education in theory and practice: An introductory guide to learning theory (pp. 389–403). Springer.
20. Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. IEEE Transactions on Learning Technologies, 17, 1445–1459. https://doi.org/10.1109/TLT.2024.3384765
21. Huang, C., Huang, Q., Huang, X., Wang, H., Li, M., Lin, K.,... Chang, Y. (2024). XKT: Toward Explainable Knowledge Tracing Model With Cognitive Learning Theories for Questions of Multiple Knowledge Concepts. IEEE Transactions on Knowledge and Data Engineering, 36(11), 7308-7325. doi: 10.1109/TKDE.2024.3418098
22. Huang, C., Tu, Y., Wang, Q., Li, M., He, T.,... Zhang, D. (2025). How does social support detected automatically in discussion forums relate to online learning burnout? The moderating role of students' self-regulated learning. Computers & Education, 227, 105213. doi: https://doi.org/10.1016/j.compedu.2024.105213
23. Jacobsson, E., Wallenlind, S., & Pendrill, A. M. (2024). Educational research in physics teacher education: The role of mathematics in physics teaching. Physics Education, 59(6), 065021.
24. Jing, L., Fan, X., Feng, D., Lu, C., & Jiang, S. (2024). A patent text-based product conceptual design decision-making approach considering the fusion of incomplete evaluation semantic and scheme beliefs. Applied Soft Computing, 157, 111492. doi: https://doi.org/10.1016/j.asoc.2024.111492
25. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Hüllermeier,E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.
26. Keiper, M. C., Fried, G., Lupinek, J., & Nordstrom, H. (2023). Artificial intelligence in sport management education: Playing the AI game with ChatGPT. Journal of Hospitality, Leisure, Sport & Tourism Education, 33, 100456.
27. Kieser, F., Wulff, P., Kuhn, J., & Küchemann, S. (2023). Educational data augmentation in physics education research using ChatGPT. Physical Review Physics Education Research, 19(2), 020150.
28. Kılınç, S. (2023). Embracing the future of distance science education: Opportunities and challenges of ChatGPT integration.
29. Kim, M., & Adlof, L. (2024). Adapting to the future: ChatGPT as a means for supporting constructivist learning environments. TechTrends, 68(1), 37–46.
30. Kortemeyer, G. (2020). Physics education research: A replicable model for discipline-based educational research at European universities? ETH Learning and Teaching Journal, 2(1), 137–146.
31. Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory Into Practice, 41(4), 212–218.
32. Krupp, L., Bley, J., Gobbi, I., Geng, A., Müller, S., Suh, S., Moghiseh, A., Castaneda Medina, A., Bartsch, V., Widera, A., Ott, H., Lukowicz, P., Karolus, J., & Kiefer- Emmanouilidis, M. (2025). LLM-generated tips rival expert-created tips in helping students answer quantum-computing questions. EPJ Quantum Technology, 12(1), 33. https://doi.org/10.1140/epjqt/s40507-025-00334-5

33. Lee, G.-G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. IEEE Transactions on Learning Technologies, 17, 1643– 1660. https://doi.org/10.1109/TLT.2024.3401457
34. Li, M. Z. (2024). Using prompt engineering to enhance STEM education. 2024 IEEE Integrated STEM Education Conference (ISEC), 1–2. https://doi.org/10.1109/ISEC61299.2024.10664775
35. Liu, Y., Chen, W., Bai, Y., Liang, X., Li, G., Gao, W.,... Lin, L. (2025). Aligning Cyber Space With Physical World: A Comprehensive Survey on Embodied AI. IEEE/ASME Transactions on Mechatronics, 1-22. doi: 10.1109/TMECH.2025.3574943
36. Li, Y., Xiao, C., Ren, H., Li, W., Guo, Z.,... Luo, J. (2025). Unraveling the effectiveness of new media teaching strategies in pharmacology education under different educational backgrounds: Insights from 6447 students. European Journal of Pharmacology, 989, 177255. doi: https://doi.org/10.1016/j.ejphar.2025.177255
37. Liu, S., Li, C., Qiu, J., Zhang, X., Huang, F., Zhang, L.,... Yu, P. S. (2025). The Scales of Justitia: A Comprehensive Survey on Safety Evaluation of LLMs. IEEE Transactions on Knowledge and Data Engineering. doi: 10.48550/arXiv.2506.11094
38. Liu, Y., Cao, S., & Chen, G. (2024). Research on the Long-term Mechanism of Using Public Service Platforms in National Smart Education—Based on the Double Reduction Policy. Sage Open, 14(1). doi: https://doi.org/10.1177/21582440241239471
39. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410.
40. McHugh, M. L. (2012). Interrater reliability: The kappa statistic. Biochemia Medica, 22(3), 276–282.
41. Mitchell, M. L., & Jolley, J. M. (2010). Research design explained: Instructor's edition. Cengage.
42. Niess, J., & Diefenbach, S. (2016). Communication styles of interactive tools for self- improvement. Psychology of Well-Being, 6, 1–15.
43. Niess, M. L. (2016). Technological pedagogical content knowledge (TPACK) framework for K–12 teacher preparation: Emerging research and opportunities. IGI Global.
44. Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1609406917733847.
45. Pan, Y., & Xu, J. (2025). Human-machine plan conflict and conflict resolution in a visual search task. International Journal of Human-Computer Studies, 193, 103377. doi: https://doi.org/10.1016/j.ijhcs.2024.103377
46. Peikos, G., & Stavrou, D. (2025). ChatGPT for science lesson planning: An exploratory study basedon pedagogical content knowledge. Education Sciences, 15(3). https://doi.org/10.3390/educsci15030338
47. Polverini, G., & Gregorcic, B. (2024). Performance of ChatGPT on the test of understanding graphsi n kinematics. Physical Review Physics Education Research, 20(1), 010109.
48. Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, 66(2), 211–227.
49. Rudner, L. M., & Schafer, W. D. (2002). What teachers need to know about assessment. National Education Association.
50. Rundle, C. W., Szeto, M. D., Presley, C. L., Shahwan, K. T., & Carr, D. R. (2024). Analysis of ChatGPT generated differential diagnoses in response to physical exam findings for benign and malignant cutaneous neoplasms. Journal of the American Academy of Dermatology, 90(3), 615–616.
51. Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. American Journal of Physics, 75(11), 1017–1029.
52. Song, J., Ma, C., & Ran, M. (2025). AirGPT: pioneering the convergence of conversational AI with atmospheric science. npj Climate and Atmospheric Science, 8(1), 179. doi: 10.1038/s41612-025-01070-4
53. van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. Education Sciences, 13(10), 998.
54. Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 3–34). Routledge.
55. Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. Eurasia Journal of Mathematics, Science and Technology Education, 19(7), em2286.
56. Werth, A., West, C. G., & Lewandowski, H. J. (2022). Impacts on student learning, confidence, and affect in a remote, large-enrollment, course-based undergraduate research experience in physics. Physical Review Physics Education Research, 18(1), 010129.
57. Whalen, J., & Mouza, C. (2023). ChatGPT: Challenges, opportunities, and implications for teacher education. Contemporary Issues in Technology and Teacher Education, 23(1), 1–23.
58. Yin, Z., & Wang, S. (2025). Enhancing scientific table understanding with type-guided chain-of-thought. Information Processing & Management, 62(4), 104159. doi: https://doi.org/10.1016/j.ipm.2025.104159
59. Yu, L., Li, Y., Weng, S., Tian, H., & Liu, J. (2023). Adaptive multi-teacher softened relational knowledge distillation framework for payload mismatch in image steganalysis. Journal of Visual Communication and Image Representation, 95, 103900. doi: https://doi.org/10.1016/j.jvcir.2023.103900

60. Zhang, X., Zhang, C., Sun, J., Xiao, J., Yang, Y., & Luo, Y. (2025). EduPlanner: LLM-based multiagent systems for customized and intelligent instructional design. IEEE Transactions on Learning Technologies, 18, 416–427. https://doi.org/10.1109/TLT.2025.3561332

61. Zhao, H., Ji, T., Rosin, P. L., Lai, Y., Meng, W.,... Wang, Y. (2024). Cross-lingual font style transfer with full-domain convolutional attention. Pattern Recognition, 155, 110709. doi: https://doi.org/10.1016/j.patcog.2024.110709