VOL. 11/ ISSUE (2) JUNE 2025 Lekh Raj Aryal* MS et al., 2025
Pp: 67-74

Pakistan Journal of Society, Education and Language (PJSEL)
Journal Homepage: https://pjsel.jehanf.com/index.php/journal *ISSN*
2521-8123 (Print)
2523-1227 (Online)

Artificially Intelligent? Who Benefits From the Use of Gen AI in High School Education?

Lekh Raj Aryal* MS[1], Yen Chau, MS[2], Tonie Germany-Brown, BS[3],
Binod Khanal, PhD[4]

Original Article

1. Cypress Springs High School
Email*: Lekharyal1@gmail.com
2. Cypress Springs High School
Email: Yen.chau@cfisd.net
3. Cypress Springs High School
Tonie.germanybrown@cfisd.net
4. Prairie View A&M University
Email: bikhanal@pvamu.edu

Keywords

Generative AI, High school education, Chemistry, Motivation, Effectiveness

Abstract

Generative artificial intelligence (gen AI) has already significantly impacted the education sector. The concern of who benefits more from this technology is becoming increasingly salient. A key concern lies in how students will utilize the technology within classroom settings and how it will benefit them. Against this backdrop, we used an experiment to seek the answer to how the motivation to use gen AI would impact the score. Our experiment shows that those who received the information treatment scored 3 points higher, reflecting a 20% relative increase compared to the control group. Furthermore, subsample analyses revealed that students with lower interest in the subject matter and those with less reported grades in past tests benefited from the treatment. The implications of the findings from this research and the direction of future research on the effective use of gen AIs in the classrooms are discussed.

1. Introduction

The use of advanced technology, like artificial intelligence, is new, and it is believed that it impacts education in many ways (Zawacki-Richter et al., 2019). After Open AI launched ChatGPT in November 2022, leading tech companies like Microsoft and Google introduced many other AI tools. With the advent of generative artificial intelligence (Gen AIs) like ChatGPT, Microsoft Copilot, Google Gemini, etc., the education and tech industry has changed significantly. Gen AIs can be useful for monotonous academic jobs (like language editing, coding, and checking grammar) and for generating novel ideas (Aydin and Karaarslan, 2023; Laak and Aru, 2024). Generative AIs have become increasingly common among all professionals, including students (Baidoo-Anu and Ansah, 2023).

There are mixed views on the effectiveness of generative AI in the classroom, particularly among high school students (Yu, 2023). Educators and policymakers are more worried (Mhlanga, 2023) than excited about the potential misuse of the technology. The misuse would be mostly in doing assignments and exams (take-home or open notes) (Shalevsak, 2023). However, other viewpoints posit that these technologies are here to stay and can be utilized to make teaching and learning effective (Li et al., 2024; Naik et al., 2024). There is a growing demand for a paradigm shift in teaching and assessment practices, both in and out of the classroom. Future changes should focus on integrating these generative AIs as a teaching method, as these technologies are here to stay (Yu, 2023).

Disproportionate use and access of generative AIs inside and outside of the classroom would create (or even strengthen) the divide among the students. Studies have already highlighted the digital divide among racial minorities in the United States (Jackson et al., 208). The disproportion in usage could arise from their interest or motivation in the subject matter or their access to the technology, for example, high-speed internet or laptop devices (Huang et al., 2023).

Gen AI tools like ChatGPT, Gemini, or Copilot can be used effectively in STEM topics. Many studies have highlighted the potential use of gen AIs in educating students at the school level (Iyamuremye & Ndihokubwayo, 2024; Babalola et al., 2024; Clark, 2023). However, few studies, like Clark (2023), have suggested the use of technology with caution since many of the responses are simply wrong or full of misconceptions. Although there is a quick advancement of the gen AI technology, studies on the potential use of gen AI in education (including chemistry) are pertinent. Against this backdrop, we experimented with chemistry students on using generative AI in class exercises among high school students in Texas. Specifically, we randomly treated half of the students with short information about the use of generative AI (Please see the appendix) before the test. The other half were not treated with such information, but if they had prior knowledge about generative AIs, they could do so. Thus, we formed two groups: one treated group that received information text about the use of generative AI and another control group that did not receive the text information.

As multiple new studies explore the usefulness of gen AIs in education, this study shares insights from an experiment conducted in class and adds to the growing literature on using gen AI tools in the education sector.

2. Experiment

We chose three sections of grade 10 for the experiment. Half of the students from each section were randomly selected for the treatment group or the control group. Treatment groups and control groups were placed in two separate classrooms. The treatment group received a short description of the use of gen AI models in education. Besides the written description, the proctor (at least one of the authors) described what the Gen AIs are and how to use them. Since the schools had allowed the use of Microsoft Copilot, we specifically mentioned Copilot in the treatment message.

The treatment text reads as follows:

Do you know?

Microsoft Copilot is a helpful tool for solving tough problems in subjects like Chemistry, Math, and Computer Science. It can guide you through writing chemical equations and solving equations. Instead of just giving answers, Copilot helps you understand each step, making it easier to learn and solve complex problems confidently.

It's like having a smart study buddy by your side.

We asked 15 questions on Chemical bonding and Lewis Structure. The students were recently (aorund a month before the experiment) introduced to this concept. Test questions were pre-tested among three 12th-grade Chemistry students to confirm the quality of the questions. They were informally asked to comment on the readability and familiarity of the test questions. On the experiment set-up, students from two classes were randomly divided into two groups, informed about the test procedure, and given a set of 15 questions. One group was given access to Copilot information to help in answering, while the other group, the control group, received no such information. After completing the task, students were asked to complete a few survey questions that asked about their interest in Chemistry and their reported test grade from the previous test.

The topic we covered in the test is related to the understanding of chemical bonds. The sub-topics included are Ionic Bonding, Covalent Bonding, Octet Rule, Lewis Dot Structure, Polarity, VSEPR theory, H-bonding, Ionic size, Dipole Moment, and naming binary compounds. The questions were designed carefully by testing them in the regular Google search and Microsoft Copilot. Expectedly, the answers from the Microsoft Copilot were precise compared to the Google search.

3. Empirical Model

We modeled the test score using simple and multivariate linear regressions. Since this study involves an experiment based on randomization, an ordinary least square (OLS) design would suffice to make a causal statement on the relation between the treatment and the outcome variable.

We modeled the test score against the treatment status of the students, their interest in chemistry, their past grade in chemistry tests, and their race. The simple econometric model is given by:

$$Test_i = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 Chemistry\_interest_i + \alpha_3 Chemistry\_grade_i + \alpha_4 White_i + \varepsilon_i \quad (1)$$

Where, $Test_i$ is test score of the student *i* and range from 0 to 15. $Treatment_i$ is an indicator variable where 1 equals if the student is treated, otherwise 0. $Chemistry\_interest_i$, $Chemistry\_grade_i$, and $White_i$ indicate interest of students on chemistry, chemistry grade in the previous test, and the indicator variable for white race, respectively. ε_i is an idiosyncratic disturbance (random error term) whose mean expected value is assumed to be zero.

We ran subsample analyses besides this main model (given by equation (1)). In the first subsample analysis, we categorized samples as highly and less interested in chemistry. Further, in the second subsample analysis, we categorized samples into high and low chemistry grades (in the past tests).

4. Data

Table 1 gives the summary statistics of the dependent and independent variables used in the study. Further, it also compares the statistics for the treated and control groups. In this experiment, the average test score of the students is almost 11 out of 15. Expectedly, students in the treated group scored higher on the test than those in the control group. After the test, we asked about students' interest in Chemistry, ranging from 1 to 10, where 1 is least interested, and 10 is very interested. The logic behind asking this question is that the use of these technologies could differ by the student's interest in the subject matter. Further, we also asked about their chemistry grade from the past test. It ranged from 35 to 100, with a mean score of slightly more than 87. Finally, we asked whether the student was white or not. 15 percent of the students are white. Treated students are slightly more interested in chemistry, scored higher grades in the past, and are likelier to be white than those in control groups.

Table 1: Summary Statistics

Variables	Overall mean (N=103)	Treated (n=51)	Control (n=52)	Min	Max
Test score	10.961 (3.592)	12.627 (2.374)	9.327 (3.844)	0	15
Chemistry interest	4.835 (2.529)	4.902 (2.655)	4.769 (2.422)	1	10
Past Chemistry Grade	87.825 (10.890)	90.217 (6.512)	85.481 (13.574)	35	100
White	0.146 (0.354)	0.216 (0.415)	0.077 (0.269)	0	1

5. Results

Table 2 shows the effect of the treatment on the test scores from two different specifications and robust standard error (Robust SE). Specification 1 gives estimates from a simple linear regression model, while Specification 2 gives estimates from the multivariate regression model. From Specification 1, the estimates show that students treated with the Gen AI information scored 3 points more than those from the control group. This result is similar to the findings from Shalevsak (2023. Consistently, the multivariate linear regression estimates also suggest similar results. Although interest in the subject, past grades in Chemistry, and race are positively associated with the score, these relationships are statistically insignificant. This result is expected yet interesting. The students can use the Gen AI (like Microsoft Copilot) and score better on tests. The 'constant' variables in result tables are intercept value of the model, which indicates the baseline value of the test scores. The R-squared value ranges from 21% -25%, which indicates the low explanatory power of the control variables in our model. Low R-squared value limits the prediction power of the model. Adding more relevant variables would have increased the explanatory power of the models, however, this is a limitation of our study that we could not collect more variables.

Table 2: Treatment Effect on Test Scores (N=103)

	Specification 1		Specification 2	
	Coefficient	Robust SE	Coefficient	Robust SE
Treatment	3.301***	0.628	2.907***	0.645
Chemistry interest			0.120	0.118
Past Grade in Chemistry			0.046	0.030
White			1.158	0.758
Constant	9.327***	0.533	4.754*	2.718
R-squared	0.213		0.253	

Note: * and *** indicate statistical significance levels of <10% and 1%, respectively. Specification 1 has only one "Treatment" variable, while Specification 2 includes other control variables.

Table 3 reports the subsample analysis based on the interest of students in chemistry. We categorized students into two groups: those who reported an interest in chemistry less than or equal to 5 were defined as less interested in chemistry, and those with 5-10 ratings were defined as interested in chemistry. The treatment is found to benefit those who have less interest in chemistry. Those highly interested in chemistry and are treated score 3.5 points, while less interested treated groups only increased the score by 2 points.

Table 3: Subsample Analysis by Interest in Chemistry

	Chemistry interest=1 (n=42)		Chemistry interest=0 (n=61)	
	Coefficient	Robust SE	Coefficient	Robust SE
Treatment	2.000**	0.908	3.503***	0.894
Past Chemistry grade	0.026	0.033	0.048	0.046
White	2.257***	0.518	0.637	1.078
Constant	8.088***	2.722	4.477***	4.053

Note: ** and *** indicate statistical significance levels of <5% and 1%, respectively.

In Table 4, we conducted a subsample analysis of the quality of the students based on their reported grades in previous Chemistry grades. We have two categories of students here: more than 90 (out of 100) and less than or equal to 90. The estimates from the subsample analysis are also similar to that of Table 3. The impact of treatment on the low-grade students (non-As) is higher than that of high-grade students.

Table 4: Subsample Analysis by Past Chemistry Grade

	Chemistry grade>90 (n=43)		Chemistry grade<=90 (n=60)	
	Coefficient	Robust SE	Coefficient	Robust SE
Treatment	2.122**	0.944	3.534***	0.869
Chemistry interest	0.046	0.175	0.155	0.185
White	2.082***	0.585	1.045	1.272
Constant	10.017***	1.432	7.972***	1.262

Note: ** and *** indicate statistical significance levels of <5% and 1%, respectively.

6. Discussion

The use of new advanced technologies, like gen AIs, in education should be done carefully to ensure it will help the student's learning process. This study has two significant outcomes. The first is that the students encouraged to use gen AIs were likelier to score higher on the test. Unlike Clark (2023), our results show that gen AI tools could be helpful for students in chemistry tests. The second is that the treatment had a larger impact on those not highly interested in the subject matter and those whose past grades were below the median score. The second result highlights how technology could help those less interested in the subject matter and who are relatively weak students. While the smarter and more motivated students could be those more inclined to use the technologies, the easy-to-use technology, like gen AIs, could be helpful for relatively weaker students to do better in tests and in understanding complex topics like chemistry. This result aligns with the findings from Huang et al. (2023), which found that students with moderate motivation improve learning outcomes and engagement because of the use of artificial intelligence tools.

The second finding raises additional questions that merit further investigation. For instance, our results are based on the encouragement to use Gen AI tools while taking the test. However, this could bring us to the topic of this paper: Do generative AIs make some groups of students artificially intelligent? This needs to be answered by more detailed and comprehensive studies. Further, this could raise the issue of difficulty in grading the examinations (primarily online) and homework. This issue could demand a different approach to examining students' grades when technology use is pervasive. This study was carried out among a few students over a short period in one school. Further, we could not run subsample analysis based on the students' race due to the lack of statistical power. We recommend experimenting with gen AI technologies in the classroom for a more extended period, like an entire semester, among a larger number of students. Including more than one school in the study would ensure the robustness of the study. Finally, a more comprehensive study should be conducted by collecting more variables on students' understanding of the subject matter.

7. Conclusion

This study highlights the transformative potential of generative artificial intelligence (gen AI) in education, demonstrating that its targeted use can significantly enhance student performance, particularly among those who traditionally struggle or exhibit lower motivation. The experimental findings reveal a notable improvement in scores among students exposed to motivation treatment, suggesting how students can use gen AI in tests and exams, particularly in assessments that are open-book, open-note, or conducted remotely.

Subsample analyses further emphasize that gen AI can be an equalizing tool, offering additional support to students with lower prior achievement and interest levels. These results suggest that integrating gen AI into classrooms, mainly through structured interventions, could help address disparities in student outcomes while fostering engagement and curiosity in subject matter areas.

However, the findings also prompt a broader discussion about how to ensure the responsible and equitable use of gen AI technologies. As some school administrations and countries have even banned the use of some gen AI tools in school (Ogugua, 2023), now the discussion around the use of gen AI tools in schools is reshaping in a way that it should be well integrated with the conventional teaching method to enhance students' learning (Yu, 2023).

Future research should explore scalable implementation strategies, long-term impacts on learning, and methods to balance AI's role with developing critical thinking and problem-solving skills. Given the deep digital divide among students (by race, rurality, countries' development status, etc.), future studies should explore the implications of access to the latest technology in learning outcomes. By continuing to examine the effects of gen AI on diverse student populations, educators and policymakers can better harness its potential to enhance teaching and learning outcomes effectively.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2023.2253858

Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. In Ö. Aydın (Ed.), *Emerging computer technologies* (pp. 1-2).

Babalola, V. T., Ahmad, S. S., & Tafida, H. S. (2024). ChatGPT in organic chemistry classrooms: Analyzing the impacts of social environment on students' interest, critical thinking and academic achievement. *International Journal of Education and Teaching Zone*, 3(1), 60-72. https://doi.org/10.57092/ijetz.v3i1.155

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62.

Clark, T. M. (2023). Investigating the use of an artificial intelligence chatbot with general chemistry exam questions. *Journal of Chemical Education*, 100(5), 1905-1916. https://doi.org/10.1021/acs.jchemed.3c00027

Iyamuremye, A., & Ndihokubwayo, K. (2024). Exploring secondary school students' interest and mastery of atomic structure and chemical bonding through ChatGPT. *Educational Journal of Artificial Intelligence and Machine Learning*, 1(1), 1-13.

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684.

Jackson, L. A., Zhao, Y., Kolenic III, A., Fitzgerald, H. E., Harold, R., & Von Eye, A. (2008). Race, gender, and information technology use: The new digital divide. *CyberPsychology & Behavior*, *11*(4), 437-442.

Laak, K. J., & Aru, J. (2024). Generative AI in K-12: Opportunities for learning and utility for teachers. In *International conference on artificial intelligence in education* (pp. 502-509). Cham: Springer Nature Switzerland.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. https://doi.org/10.3390/educsci13040410

Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. In *FinTech and artificial intelligence for sustainable development: The role of smart technologies in achieving development goals* (pp. 387-409). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37776-1_17

Naik, A., Yin, J. R., Kamath, A., Ma, Q., Wu, S. T., Murray, R. C., Bogart, C., Sakr, M. & Rose, C. P. (2025). Providing tailored reflection instructions in collaborative learning using large language models. *British Journal of Educational Technology*, *56*(2), 531-550.

Ogugua, D., Yoon, S. N., & Lee, D. (2023). Academic integrity in a digital era: Should the use of ChatGPT be banned in schools?. *Global Business & Finance Review (GBFR)*, *28*(7), 1-10.

Shalevska, E. (2023). AI language models, standardized tests, and academic integrity: A chat (GPT). *International journal of Education Teacher*, *26*, 17-25.

Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity*, 51, 101440 https://doi.org/10.1016/j.tsc.2023.101440

Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, 1181712.

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(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Supplementary Material:

Questions:

1. Which set of elements typically gains electrons to form anions?
   A. Group 1
   B. Group 2
   C. Group 15, 16, 17
   D. Group 17, 18

2. How many electron dots does a chloride ion have in its electron dot structure?
   A. 6
   B. 7
   C. 8
   D. 5

3. What happens to ionic compounds when they dissolve in water?
   A. They form a solid precipitate.
   B. They break into ions surrounded by water molecules.
   C. They remain as a solid crystal.
   D. They evaporate into gas.

4. 1) Which of the following statements is FALSE?
   A) A covalent bond is formed through the sharing of electrons between atoms.
   B) A pair of electrons not shared is referred to as a "lone pair."
   C) It is not possible for two atoms to share more than two electrons.
   D) Single bonds are longer than double bonds.

5.) Which of the following statements is TRUE?
   A) An ionic bond is much stronger than most covalent bonds.
   B) An ionic bond is formed through the sharing of electrons.
   C) Ionic compounds at room temperature typically conduct electricity.
   D) Once dissolved in water, ionic compounds rarely conduct electricity.

6. Which of the following is an example of a polar covalent compound?
   A. Sodium chloride (NaCl)
   B. Calcium oxide (CaO)
   C. Hydrogen (H2)
   D. Ammonia (NH3)

7. Which of the following is a key concept of VSEPR theory?
   A. Ionization energy
   B. Molecular Shape prediction
   C. Molecular Mass
   D. Electron Affinity

8. Which molecule can form hydrogen bonds with itself?
   A. CH3OH
   B. CH4
   C. Xe
   D. C2H6

9. Which of the following statements about hydrogen bonds is correct?
   A. They are the same as ionic bonds.
   B. They involve a hydrogen atom bonded to a highly electropositive atom.
   C. They are the same as covalent bonds.
   D. They involve a hydrogen atom bonded to a highly electronegative atom.

10. The larger the difference in electronegativity between two atoms, the more ______ the bond is.
   A. Nonpolar
   B. Polar
   C. Metallic
   D. Covalent

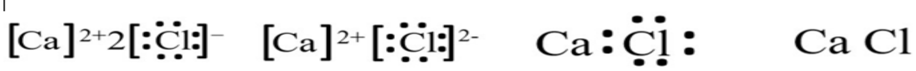
11. A molecule must have ______ and ______ to have a dipole moment.
   A. nonpolar bonds, symmetrical shape
   B. polar bonds, asymmetrical shape
   C. ionic bonds, symmetrical shape
   D. metallic bonds, asymmetrical shape

12. Which of the following ions is larger in size: Na^{1+}, Mg^{2+}, F^{1-}, or O^{2-}?
   A. Na^{1+}
   B. Mg^{2+}
   C. F^{1-}
   D. O^{2-}

13. Which molecule has a linear geometry?
A. IF_3
B. XeF_2
C. SF_6
D. ICl_5

14. Which of the following is NOT a rule for naming binary covalent compounds?
   A. The first element in the formula is named first, using the full element name.
   B. The second element is named as if it were an anion.
   C. Prefixes are used to denote the numbers of atoms present.
   D. The prefix mono- is always used for naming the first element.

15. What is the Lewis structure for the ionic compound formed between calcium and bromine?



I II III IV

A) I
B) II
C) III
D) IV

License Pakistan Journal of Society, Education and Language (PJSEL). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) 4.0 International.