International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

Comparing the Effectiveness of Chat GPT and Teacher-generated Content for Teaching Students

Ananth Namilae and John Leddo

MyEdMaster, LLC

DOI: 10.46609/IJSSER.2024.v09i07.031 URL: https://doi.org/10.46609/IJSSER.2024.v09i07.031

Received: 17 July 2024 / Accepted: 30 July 2024 / Published: 10 August 2024

ABSTRACT

This research explores the relative teaching effectiveness of educational content created by large language models (LLMs), specifically ChatGPT, and traditional educational content created by teachers. 20 middle and high school students were taught about dying stars, a topic for which they had no significant prior knowledge. Half were given content created by Chat GPT and the other half were given content created by human teachers. Following the instructional period, all students were given a post-test to measure how much they learned. Results showed that students who learned using Chat GPT-generated material scored 33% higher on the posttest than those who learned using teacher generated materials. Results suggest that LLMs offer not only the opportunity to increase speed and save money in content generation, but may improve the learning process as well. Future research can explore whether LLMs can enhance learning even further by producing content that is customized to each student's learning needs.

Introduction

While formal educational institutions have been around for hundreds of years, the methods used have changed very little. By and large, students sit in classrooms, are taught the same instruction by teachers, use the same textbooks and other instructional materials, do the same homework and take the same tests. In other words, education still employs the traditional one-size-fits-all approach, rather than being customized to the needs of each student. This would not be a problem if it were not for the fact that, just as educational methods have not changed much over the years, neither have educational outcomes. In the United States, the majority of students still perform below grade level in core subjects. According to the National Assessment of Educational Progress, of 4th graders, 67% perform below grade level in reading, 73% perform below grade level in math and 63% perform below grade level in science. Of 8th graders, 68% perform below grade level in reading, 73% perform below grade level in math and 67% perform

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

below grade level in science. Of 12 graders, 63% perform below grade level in reading, 76% perform below grade level in math and 78% perform below grade level in science.

One of the areas in which education has changed is through the use of computers at home and in the classroom. Computers and supporting technology (such as the Internet) open the world of education to unlimited resources. Of course, simple exposure to unlimited information is not the same as improving education since information is useless until it is learned. One application of computer technology that has garnered considerable attention in education is artificial intelligence (AI). AI presents opportunities to better enhance the learning experience, customize educational content to individual needs, and optimize lesson delivery. The integration of AI into educational lessons represents more than just a technological upgrade; it signifies a shift that promises to make learning more personalized, efficient, and accessible.

AI in education is gaining traction, with numerous studies highlighting its benefits. According to Holmes et al. (2019), AI can tailor educational content to meet individual student needs, thus fostering a more personalized learning environment. Furthermore, AI can assist teachers by automating administrative tasks, thereby allowing them more time to focus on direct instructional activities (Luckin et al., 2016).

One of the most exciting advancements in AI/machine learning is large language models (LLMs), such as Chat GPT. While LLMs can perform a variety of functions, one of the most common is content generation. Research indicates several advantages of using LLMs for educational content generation:

1. **Efficiency and Scalability**: LLMs can produce vast amounts of content quickly, addressing the needs of diverse student populations. This is particularly beneficial in large-scale educational settings such as MOOCs (Massive Open Online Courses) (Moore et al., 2022).

2. **Personalization**: LLMs can tailor educational materials to individual student needs, providing customized learning experiences that adapt to the learner's pace and understanding (Ni et al., 2022).

3. **Enhanced Engagement**: Studies suggest that LLM-generated content can be more engaging for students due to its ability to incorporate contemporary language and context, making learning more relatable and interesting (Sarsa et al., 2022).

Given the potential applications for LLMs, the next question to examine is LLM effectiveness. The effectiveness of LLM-generated educational content has been evaluated in several studies:

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

1. **Quality of Content**: Research by Moore and colleagues (2022) assessed the quality of student-generated questions using GPT-3. The findings revealed that the questions were comparable in quality to those created by human educators, demonstrating the model's potential to produce high-quality educational content.

2. **Learning Outcomes**: An empirical study by Singh et al. (2021) investigated the impact of LLM on enhancing student-generated content and found that LLMs could help students improve their own content.

3. **Instructor Support**: Studies have also examined the role of LLMs in aiding instructors. For example, the work by Wang et al. (2022) demonstrated that LLMs could assist teachers in creating diverse and challenging questions, thereby enhancing the overall learning experience without significantly increasing the teachers' workload.

Furthermore, AI's capabilities greatly enhance self-directed learning, enabling students to take charge of their educational journeys. As stated by the International Society for Technology in Education (ISTE), AI can assist students in developing their critical thinking skills, and promote creativity, and aim for independent learning. AI can enhance self directed learning by offering personalized learning experience, facilitating active engagement, and providing immediate feedback. AI tools, such as ChatGPT, enable students to enhance their learning paths by selecting topics, finding resources, and receiving proper guidance, which enhances both independence, as well as a deeper understanding in the concepts (Halaweh, M. et al.,2023).

Given that LLMs have been repeatedly shown to generate high quality content, the question remains whether students taught using this LLM-generated content can learn as well as those taught using human-generated content. If this were to be the case, then the implications are revolutionary. Not only could the speed of content creation be increased but the cost of doing so be greatly reduced. Moreover, this opens the door to the possibility of providing students with customized content based on their individual learning needs with no material increase in cost or disruption to the operations of the educational system.

The present study addresses the research question as to whether LLM-generated content can teach as effectively as human-generated content. We defer the issue of customization of content to individual learner needs, so that the general research question can be addressed first. This study aims to evaluate the effectiveness of LLM vs teacher-generated content in middle and high school science education.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

Method

Participants

There were a total of 20 participants who were recruited from Fairfax County, Virginia and also from Florida and California. All participants ranged from 6th graders to 9th graders. They had never been enrolled in any astronomy classes, nor had any previous knowledge of the concept of dying stars.

Materials

The topic chosen for the present study was dying stars. Two versions of the instructional material were created. One version was taken from traditional educational sources contained in school databases. These databases are listed below:

- Go Gale
- Go Gale (Science)
- Go Gale (one file news)
- PebbleGo
- Britannica
- World Book Encyclopedia

The second version of the instructional materials was created by Open AI's ChatGPT. ChatGPT was given the instruction "explain dying stars." Its response is shown below:

"Dying stars refer to stars that have reached the later stages of their life cycle and are undergoing significant changes in their structure and behavior. The life cycle of a star is determined primarily by its mass, with more massive stars experiencing a more rapid and dramatic evolution. There are several ways in which stars can die, depending on their initial mass and other factors:

Low-Mass Stars (Less than about 1.5 times the mass of the Sun):

- Red Giant Phase: As a low-mass star exhausts its core hydrogen fuel, it expands and becomes a red giant. During this phase, the star's outer layers expand, cool down, and become less dense.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

- Helium Core Fusion: In the core of the red giant, helium fusion begins to occur, forming carbon and oxygen through the triple-alpha process. This release of energy counteracts the force of gravity, preventing the star from collapsing under its own weight.

- Helium Shell Burning: Surrounding the inert helium core, there are still hydrogen-rich layers that undergo fusion in a shell around the core.

- Planetary Nebula: After the helium is depleted in the core, the outer layers of the star are ejected into space, forming a beautiful glowing shell of gas and dust called a planetary nebula.

- White Dwarf: The remaining core, mostly composed of carbon and oxygen, collapses under its own gravity and becomes a dense and hot stellar remnant known as a white dwarf. It no longer undergoes fusion and gradually cools over billions of years.

Intermediate-Mass Stars (Between about 1.5 and 8 times the mass of the Sun):

- These stars follow a similar path as low-mass stars during the red giant phase and helium core fusion.

- However, intermediate-mass stars have enough mass to undergo further nuclear fusion processes.

- Carbon Fusion: Once helium is depleted in the core, the temperature and pressure increase, enabling carbon fusion to occur, forming heavier elements like neon, magnesium, and oxygen.

- Supernova (Type II): When the core can no longer support its own weight due to iron accumulation, it collapses catastrophically. This collapse creates a supernova explosion, during which the star releases an enormous amount of energy, outshining entire galaxies for a brief period. The remaining core can become either a neutron star or a black hole, depending on its mass.

High-Mass Stars (More than about 8 times the mass of the Sun):

- High-mass stars follow a similar path to intermediate-mass stars, but their higher mass allows for the fusion of progressively heavier elements, such as silicon, sulfur, and calcium.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

- Core Collapse Supernova (Type II): The core of a high-mass star eventually undergoes a supernova explosion in a more energetic manner than the one in intermediate-mass stars. The remnants can become neutron stars or black holes.

In summary, the fate of a dying star depends on its mass. Low-mass stars end their lives as white dwarfs, intermediate-mass stars go through supernova explosions, and high-mass stars end up as either neutron stars or black holes. These stellar deaths play a crucial role in enriching the universe with heavier elements and shaping the dynamics of galaxies."

In order to address the study's research question regarding the relative instructional effectiveness of human-generated and AI-generated content, a post-test was created. The questions on the posttest were constructed so as to include topics that were included in both the human-generated study materials and the AI-generated study materials so as not to bias the test in favor of either instructional material source. The questions given on the test were as follows

1. Input Name (question not marked)
2. How does a supernova type II or supernova explosion form?
3. What are neutron stars?
4. Name the outcome(s) of when a high mass star dies..
5. Explain in simple sentences about the Red Giant Phase ?
6. How are white dwarfs formed ?
7. How do the neutron stars form?
8. Why do stars die ?
9. Name a singular scenario in which a dying star ends its life.
10. Name one phase in which a star dies.
11. What gas is expended in creating a red giant or supergiant?
12. What are supernovae?
13. How are planetary nebulae formed ?
14. What do low mass stars eventually become?
15. What are red giants?

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

16. How does the red giant phase form ?
17. Name an element typically found in Red Giants.
18. At a rough estimate, how much more massive is a low mass star to the sun? (just write the number in decimal format)
19. How much more massive is a medium mass star compared to the sun? (write an individual number or decimal)
20. How many more times or solar masses are high mass stars to the sun?
21. Name the fate of a high mass star in as much detail as possible.
22. How do black holes form ?

Procedure

Participants were randomly assigned to either the school database or the ChatGPT- generated materials condition. There were 10 assigned to each condition. Each group received its instructional materials via Google Forms. Participants were then allowed to study the materials they were given. After this period, access to the study guide was revoked, and participants were given a test on the subject matter. The test was given via Google Forms as well and was carried out through Google Meet. During this time, the experimenter was present in the Google Meet. Each Participant was asked to share his or her screen during the test, so that the experimenter could ensure that there was no cheating on the test.

Results

Participants' answers to test questions were scored for correctness. Answers that were deemed to be correct were given a score of 1. Answers that were deemed to be incorrect were given a score of 0. Since the questions were in free-text format, partial credit of .5 was awarded if an answer was partially correct. Given that there were 22 questions on the tests, a maximum possible score was 22. Of the 20 participants, four (two per condition) were eliminated from the analysis because their tests were incomplete or contained answers that were predominantly meaningless (e.g., "IDK" throughout the test). Of the remaining 16 Participants, the mean test score for the ChatGPT condition was 18.25, and the mean test scores for the school database group was 13.75. This difference was statistically significant, $t(14) = 2.56$, $p = .02$. These results suggest that Participants using the ChatGPT materials outperformed those using the school database materials.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

Discussion

The results of the present study showed that students learning from materials created by ChatGPT, a form of generative AI, performed better on a post-test than those learning from materials taken from school databases. Further research is needed to determine if this is a robust finding, i.e., that content created by generative AI is as effective as that created by humans in teaching students concepts. Such research should investigate a variety of subject areas and age groups. Moreover, the instructions given to ChatGPT were relatively simple. More complex instructions could prove to result in even more effective content as we shall discuss shortly.

If AI-generated content can be as effective instructionally as human-generated content, this could be educationally revolutionary. Traditional content creation is a slow and expensive proposition. Considerable time and resources are spent researching, organizing material, outlining, writing, editing and publishing material. While information evolves rapidly and exponentially, educational materials do not. Imagine the cost savings and improved contemporariness of instructional content if AI-generated content were as effective or more so than human-generated content. Schools could provide students with instructional materials far more cheaply and timely than they do now.

Moreover, the present study used very simple instructions to Chat GPT for generating content. In this respect, the present study may not have exploited the full potential of AI-generated content to improve learning. For example, Mahajan et al. (2021) found that beginning students performed equally well learning computer-programming concepts when taught using a conceptual or procedural format, but advanced students learned better when taught using a procedural one. Wang et al. (2021) showed that explaining the rationales behind procedures being taught produced three times better learning in math than simply teaching procedures. Teaching students metacognitive strategies has improved performance in subjects of reading, math and grammar (Leddo et al., 2019; Leddo et al., 2020; Leddo, Sangela and Bekkary, 2021). Variables such as these were not included in the present study but could be tested in future studies involving AI-generated content.

Another application of AI-generated content is customization. Currently, our educational system employs a "one-size fits all" paradigm. Every student gets the same lecture, textbook, homework and test. This makes sense in a human-centered teaching world. It is difficult for a human teacher to customize lectures for each student in a class. It would probably be impossible for textbook publishers to write textbooks that are personalized to each student's learning needs. Giving personalized homework and tests would greatly increase teacher workload. However, there is evidence that customized materials make a difference. For example, our research has shown that standard homework helps average students improve greatly in math, but does little

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

for gifted students (Bhandarkar et al., 2016), while gifted students need more challenging homework to improve (Bahl et al., 2018).

While customized content may not be practical when humans are the creators, it may become a trivial process when AI is the creator. Perhaps the most powerful application of using AI to create personalized content comes during the primary teaching and remediation of content. By assessing students' learning needs, learning styles and speed of learning, AI could generate, for each student, a personalized program of instruction. One tool for supporting this is the Cognitive Structure Analysis (CSA) assessment technique we have created. Instead of simply assessing whether students can provide answers to problems, CSA assesses the concept knowledge students have of a topic area by asking questions regarding facts (e.g., "What is a variable?"), strategies (e.g., "How do you solve an equation with variables on both sides of the equal sign?"), procedures (e.g., "What do you do when there's a coefficient in front of a variable?"), and rationales (e.g., "Why do you perform the same operation to both sides of an equation?")

CSA has been tested with different age groups and with different subjects and has been shown to be remarkably predictive of problem solving performance (Ahmand & Leddo, 2023; Leddo et al., 2022). Moreover, if the faulty knowledge that is assessed by CSA is remediated, students perform 10 points or a full-letter grade higher than if a teacher simply assesses and remediates a students' "show all work" problem solving process (Leddo & Ahmad, 2024). One of the drivers of personalized AI-generated content could be the assessed learning needs of each student, something we are currently working to develop now.

We recognize that as AI becomes more and more powerful, there is a fear that it can systematically replace humans in their professions. While the goal of the present research is not to produce ammunition for those calling for an AI revolution, it is noteworthy that educational content generation is an area for which enormous practical benefits can be realized if AI-generated content really does equal or exceed human-generated content in teaching effectiveness.

References

Ahmad, M. & Leddo, J. (2023). The effectiveness of Cognitive Structure Analysis in assessing students' knowledge of the scientific method. International Journal of Social Science and Economic Research, 8(8), 2397-2410.

Bahl, K., Bahl, K & Leddo, J. (2018). Optimizing homework for high aptitude students. International Journal of Advanced Educational Research, 3(5), 27-30.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

Bhandarkar, S., Leddo, J. & Banerjee, M. (2016). Comparing the relative effects of homework on high aptitude vs. average aptitude students. International Journal of Humanities and Social Science Research, 2(9), 59-62.

Britannica. www.britannica.com/. Accessed 1 July 2024.

gale one file news. www.gale.com/c/onefile-news. Accessed 1 July 2024.

Gale Science. www.gale.com/c/in-context-science. Accessed 1 July 2024.

Go Gale. go.gale.com/ps/start.do?p=LitRC&userGroupName=anon%7E25537f9f. Accessed 1 July 2024.

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, *15*(2), ep421.

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.

Leddo, J. & Ahmad, M. (2024). Do you want to raise student achievement? Then, assess and remediate student knowledge, not student performance. International Journal of Social Science and Economic Research, 9(1), 249-261.

Leddo, J., Hong, Q., Shyamala, N. & Xue, A. (2019). Improving SAT reading scores by using metacognitive reading strategies. International Journal of Advanced Educational Research, 4(4), 91-93.

Leddo, J., Li, S.and Zhang, Y. (2022) Cognitive Structure Analysis: A Technique For Assessing What Students Know, Not Just How They Perform. International Journal of Social Science and Economic Research, 7(11), 3716-3726.

Leddo, J., Sangela, S. & Bekkary, A. (2021). Improving SAT math scores by using metacognitive problem-solving strategies. International Journal of Social Science and Economic Research, 6(3), 1044-1053.

Leddo, J., Sengar, A., Liang, I. & Chilumula, R. (2020). Improving SAT writing and language scores by using metacognitive strategies. International Journal of Advanced Educational Research, 5(1), 24-26.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. Pearson.

ISSN: 2455-8834

Volume:09, Issue:07 "July 2024"

Mahajan, L., Chadeva, A. & Leddo, J. (2021). Evaluating procedural and conceptual teaching formats for beginning and advanced programming students. International Journal of Social Science and Economic Research, 6(4), 1356-1365.

Moore, S., Nguyen, H. A., Bier, N., Domadia, T., Stamper, J. (2022). Assessing the quality of student-generated short answer questions using GPT-3. In Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022.

National Assessment of Educational Progress. Retrieved from www.nationsreportcard.gov.

Ni, L., et al. (2022). Deepqr: Neural-based quality ratings for learner sourced multiple-choice questions. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12826–12834.

PebbleGO. pgoplayer.pebblego.com/. Accessed 1 July 2024.

World Book Encyclopedia. www.worldbook.com/encyclopedias.aspx. Accessed 1 July 2024.

Sarsa, S., Denny, P., Hellas, A., Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1.

Singh, A., Brooks, C., Doroudi, S. (2021). Learnersourcing in theory and practice: Synthesizing the literature and charting the future. In Proceedings of the Eighth ACM Conference on Learning@ Scale.

Wang, E., Ailneni, A. & Leddo, J. (2021).Improving mathematics learning by adding conceptual to procedural instruction. International Journal of Social Science and Economic Research, 6(9), 3491-3498.

Wang, Z., Valdez, J., Basu Mallick, D., Baraniuk, R. G. (2022). Towards human-like educational question generation with large language models. In Artificial Intelligence in Education: 23rd International Conference, AIED 2022.

