

Comparing the Relative Effectiveness of Chat GPT-generated Content and Human-generated Videos for Teaching Students Calculus

Srijan Challapalli and John Leddo

MyEdMaster, LLC

Virginia, USA

DOI: 10.46609/IJSSER.2024.v09i11.031 URL: <https://doi.org/10.46609/IJSSER.2024.v09i11.031>

Received: 10 November 2024 / Accepted: 25 November 2024 / Published: 30 November 2024

ABSTRACT

The presented research investigates and compares the relative effectiveness of Artificial Intelligence, specifically large language models (ChatGPT) and human-generated videos in teaching students the calculus topic of derivatives. 30 randomly chosen high school students were taught derivatives. 15 of them were taught through Chat GPT and the other 15 were taught through traditional educational videos. All participants of the study were given a post-test after they learned their topics. The results showed that students who learned from ChatGPT scored higher on their post-tests than did students who used human-generated videos. These results imply that using Chat GPT to learn is more effective, efficient, and overall more personalized to each individual student. Implementing AI into classrooms could enhance the learning process for students all over the world.

Introduction

Although 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

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:

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. 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 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.

Previously, we've shown that students learning from Chat GPT-generated text-based materials scored higher on a post-test than those learning from teacher-generated text-based materials (Namilae and Leddo, 2024). The present study extends this line of research to include feedback from practice problems. Specifically, as in the Namilae and Leddo (2024) study, students in the present study receive instructional material generated by humans or Chat GPT. Then, students solve practice problems and receive either human-generated feedback or Chat GPT generated feedback. After this instruction is given, they take post-tests.

Method

Participants

There were 30 Participants selected from across Loudoun County, Virginia. All Participants tested were in high school, and their ages ranged from 14-18. Participants had no prior knowledge on derivatives and have not taken any math classes past precalculus.

Materials

The math topic that was chosen for the study was derivatives. Two types of instructions were created. One set of instructions was created from well-known education websites. These websites included Khan Academy and Youtube. The links to the materials from well-known websites are shown below:

1. The derivative and tangent line equation:

<https://www.youtube.com/watch?v=fqQ6sslzyhY>

2. Derivative as slope of a tangent line:

<https://www.youtube.com/watch?v=ANyVpMS3HL4>

3. Differentiability and Continuity:

<https://www.youtube.com/watch?v=xuAiQOzIkWY>

4. Power Rule:

<https://www.youtube.com/watch?v=bRZmfc1YFsQ>

5. Product Rule:

<https://www.youtube.com/watch?v=79ngr0Bur38>

6. Quotient Rule:

<https://www.youtube.com/watch?v=WqzY3xibFL8>

The second instructional method created was using Open AI's ChatGPT 4.0 model.

ChatGPT was given the prompt, "Explain derivatives to me and then explain the power rule, the product rule, and the quotient rule." Its response is attached below:



What are Derivatives?

In calculus, a derivative represents how a function changes as its input changes. More formally, the derivative of a function $f(x)$ at a point x gives the slope of the tangent line to the function at that point. In other words, it measures the rate at which the output of the function changes with respect to its input.

For example, if you have a function that describes the position of an object over time, the derivative of that function would describe the object's velocity—how fast the position is changing over time.

The derivative of a function $f(x)$ is often written as $f'(x)$ or $\frac{df}{dx}$.

The Power Rule

The power rule is one of the most straightforward rules for finding derivatives. It applies to functions of the form:

$$f(x) = x^n$$

where n is a constant (a real number). The power rule states that:

$$\frac{d}{dx}[x^n] = n \cdot x^{n-1}$$

Example:

For $f(x) = x^3$, using the power rule:

$$f'(x) = 3 \cdot x^{3-1} = 3x^2$$

The Product Rule

The product rule is used when you have the product of two functions, $f(x)$ and $g(x)$. It states that the derivative of the product of these two functions is:

$$\frac{d}{dx}[f(x) \cdot g(x)] = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Example:

If $f(x) = x^2$ and $g(x) = \sin(x)$, then:

$$\frac{d}{dx}[x^2 \cdot \sin(x)] = (2x) \cdot \sin(x) + x^2 \cdot \cos(x)$$

The Quotient Rule

The quotient rule is used when you have one function divided by another, $f(x)/g(x)$. It states that the derivative of the quotient is:

$$\frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{[g(x)]^2}$$

Example:

If $f(x) = x^2$ and $g(x) = x + 1$, then:

$$\frac{d}{dx} \left[\frac{x^2}{x+1} \right] = \frac{(2x)(x+1) - (x^2)(1)}{(x+1)^2}$$

Simplified:

$$\frac{d}{dx} \left[\frac{x^2}{x+1} \right] = \frac{2x^2 + 2x - x^2}{(x+1)^2} = \frac{x^2 + 2x}{(x+1)^2}$$

These rules make calculating derivatives easier for many common types of functions. Would you like to go over more examples or other rules?

Materials also included a pre-test to verify that Participants did not already know the materials and a post-test to measure how well they could solve problems after learning the material. The questions for the pre-test and post-test are shown below.

Pre Test

Power Rule

- $1/x^5$
- t^2-4/t^3

Product Rule

- $(4x^x-x)((x^3-8x^2+12)$
- $(1+\sqrt{x^3})(x^{-3}-2\sqrt[3]{x})$

Quotient Rule

- $x^2/(3x-1)$
- $2/(x^4-5)$

Post Test

Power Rule

- x^7
- $3/x^7$

- $5/2x^2$
- $3/2x^4$
- $6/(5x)^3$
- $2x^8$
- $15x^2$

Product Rule

- $(x^3+7x-1)(5x+2)$
- $x^{-2}(4+3x^{-3})$
- $x^3 \ln x$
- $y=(x^2-1)(x+5)$
- $(x^2)(x^5)$
- $(-2x^4-3)(-2x^2+1)$
- $x^2(-3x^2-2)$

Quotient Rule

- $2/(x+1)$
- $x^2/(3x-1)$
- $2/(x^4-5)$
- $2/(x^5-5)$
- $5/(4x^3+4)$
- $4x^3-3x^2/(4x^5-4)$
- $3x^4+2/(3x^3-2)$

Procedure

Participants were randomly split into the two groups. One group received the traditional instructional learning method, and the other received the AI-generated content-based learning method. Each group received their instructions on a Google doc. Both groups were asked to follow the directions on the Google doc. First, both of the groups had to complete the pre-test. After completing the pre-test, both groups were given 3 hours to prepare using only resources given to them on each of their respective documents. The group using Khan Academy used the practice questions provided on their website, and students using AI used questions generated through ChatGPT. Khan academy's feedback simply stated whether the question was right or wrong and included a video for further explanation. On the other hand, ChatGpt's explanation was far more detailed and went through each step that students did to find their error. We gave the 4.0 model a specific prompt which has shown to produce the best result for learning. Below is the exact input we asked our participants to use when solving questions.

User Input:

Problem: (Plug in the problem you wish to receive assistance from)

Solution: (Plug in the answer you got)

Work you did:

Step 1

(Type in your first step)

Step 2

(Type in your second step)

Step 3

(Type in your third step and you can add more steps if needed)

Analyze the solution, say if it is right or wrong, if incorrect provide corrective instruction

Example:

Problem: (Find the derivative of the function) $x^2 - 2x$

Solution I got $2x-2$

Work I did:

Step 1

Used power rule for the first term and

Step 2

Took the standard derivative for the second term

Step 3

Got the derivative as $2x^{-3}$

Analyze the solution, say if it's right or wrong, if incorrect provide corrective instruction

Finally, after three hours of learning and practicing, the participants were given their post-tests. They took the post-test either in front of a proctor or in a Google Meet videoconference with their cameras turned on. These measures were done to prevent cheating on the post-test.

Results

There were a total of 21 questions on the post-test. Students' responses to the questions were scored on the basis of correctness, with correct answers being given a score of 1 and incorrect answers being given a score of 0. No partial credit was given. The mean number of correctly answered questions was 15.98 for the Khan Academy video condition and 17.13 for the Chat GPT condition. While this difference was not large, due to low variability among the individual student scores, the difference was statistically significantly different, $t = 2.16$, $df, 28$, $p = .04$. This suggests that those who were taught using Chat GPT performed slightly better than those watching Khan Academy videos.

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 generated by human educators. 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 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, cf. Leddo, Li and Zhang, 2022) 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 (Ahmad & 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). We’ve also found that CSA-based assessments do not have to be delivered by teachers or software. Students can be taught to reliably self-assess their own knowledge using CSA (Cynkin and Leddo, 2024; Dandemraju, Dandemraju and Leddo, 2023) and after self-assessing, students can fill in their own knowledge gaps by reviewing material, leading to a 15-point or 1.5 letter grade improvement in performance compared to those who simply review material without self-assessing first (Ravi and Leddo, 2024). One of the drivers of personalized AI-generated content could be the (self-)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.

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.

Cynkin, C. and Leddo, J. (2023). Teaching students to self-assess using Cognitive Structure Analysis: Helping students determine what they do and do not know. *International Journal of Social Science & Economic Research*, 08(09), 3009-3020.

Dandemraju, A., Dandemraju, R., & Leddo, J. (2024). Teaching students to self-assess their knowledge of chemistry using cognitive structure analysis. *International Journal of Social Science & Economic Research*, 09(02), 541–549.

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.

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*.

Namilae, A & Leddo, J. (2024). Comparing the Effectiveness of Chat GPT and Teacher-generated Content for Teaching Students. *International Journal of Social Science and Economic Research*, 9(7), 2554-2564.

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.

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*

Ravi, D & Leddo, J. (2024). Improving Student Performance by Having Students Assess and Remediate Their Own Knowledge Deficiencies. *International Journal of Social Science and Economic Research*, 9(10), 4717-4724.

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*.