

ChatGPT has Aced the Test of Understanding in College Economics: Now What?

The American Economist
2023, Vol. 68(2) 233–245
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/05694345231169654
journals.sagepub.com/home/aex



Wayne Geerling¹ , G. Dirk Mateer¹, Jadrian Wooten² and Nikhil Damodaran³

Abstract

The Test of Understanding in College Economics (TUCE) is a standardized test of economics knowledge performed in the United States which primarily targets principles-level understanding. We asked ChatGPT to complete the TUCE. ChatGPT ranked in the 91st percentile for Microeconomics and the 99th percentile for Macroeconomics when compared to students who take the TUCE exam at the end of their principles course. The results show that ChatGPT is capable of providing answers that exceed the mean responses of students across all institutions. The emergence of artificial intelligence presents a significant challenge to traditional assessment methods in higher education. An important implication of this finding is that educators will likely need to redesign their curriculum in at least one of the following three ways: reintroduce proctored, in-person assessments; augment learning with chatbots; and/or increase the prevalence of experiential learning projects that artificial intelligence struggles to replicate well.

Keywords

TUCE, ChatGPT, assessment, artificial intelligence, academic integrity

Received: March 8, 2023; Revised: March 26, 2023; Accepted: March 28, 2023

Introduction

On November 30, 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), a chatbot that quickly gained attention for—among other things—its potential to disrupt traditional assessment methods. ChatGPT allows the user to enter a prompt and receive a natural language response that is often indistinguishable from a human-generated response. The model is pre-trained

¹Department of Economics, The University of Texas at Austin, Austin, TX, USA

²Department of Economics, Virginia Polytechnic Institute and State University College of Science, Blacksburg, VA, USA

³Department of Economics, OP Jindal Global University, Sonapat, India

Corresponding Author:

Wayne Geerling, Department of Economics, The University of Texas at Austin, 2225 Speedway, Austin, TX 78705, USA.

Email: wayne.geerling@austin.utexas.edu

on large amounts of text data, allowing it to learn the patterns and structures of language. When a user enters a prompt, ChatGPT generates its response by predicting the most probable sequence of words that would follow the prompt, based on its pre-existing knowledge of language. Several studies have demonstrated ChatGPT's ability to pass standardized tests in various fields such as mathematics (Leswing, 2023), medicine (Gilson et al., 2023), law (Choi et al., 2023), and physics (West, 2023). In these studies, researchers conducted a chat session and asked the model a series of questions from common assessment tools used in their respective disciplines. In each study, ChatGPT provided responses that exceeded the median and average response of test takers. However, due to the size of these assessments, the researchers used only a subset of questions during their chat session. In our study, we followed a similar approach by asking ChatGPT-3 a series of questions from the economics discipline to determine if it could outperform the average undergraduate student in economics.

To evaluate this, we use the Test of Understanding in College Economics (TUCE), published by the Council for Economics Education (formerly known as National Council on Economic Education) or (NCEE) and in use across the United States for more than 50 years. It is one of the most widely used assessment tools for basic economic knowledge and consists of two versions: one covering microeconomic concepts and one covering macroeconomic concepts. Each version of the test has 30 multiple choice questions with four answer choices. Both versions include three questions covering international economics, but the questions are unique to each version. The TUCE is a norm-referenced measure that can be used to compare students' knowledge levels across a wide range of abilities. A score of around 50% is desirable for research purposes, as it provides appropriate levels of item discrimination and test reliability. A score of less than 50% does not necessarily indicate a failing level of knowledge in a course, as instructors may prioritize different concepts from those tested in the TUCE. When the TUCE is used as both a pre- and post-test assessment, educators can determine whether learning has occurred during the semester, while also considering the possibility that some students may have guessed correctly on the test (Smith & Wagner, 2018).

ChatGPT is not a search engine, nor does it currently have the ability to return specific information that a user may desire. ChatGPT operates using algorithms that process data, allowing it to string words together in response to a prompt. Unlike humans, ChatGPT has access to vast troves of information available on the internet and uses large language modeling to recognize patterns in the words in each prompt to mimic human writing when dispensing knowledge (McMutrie, 2023).¹ While ChatGPT is a powerful tool, its abilities are limited to the pool of information it has been trained on. ChatGPT creates responses to user prompts using a transformer-based neural network architecture based on the training data to generate contextually appropriate and coherent responses. ChatGPT doesn't actually "know" anything, but instead generates responses based on probabilities assigned to each word in the vocabulary, which are calculated through a process of iterative training on a large corpus of text. In this paper, we assess ChatGPT's performance on the microeconomics and macroeconomics versions of the TUCE and compare it to the results of college students.

In the following sections, we briefly review the literature on the role of chatbots in education and then compare ChatGPT's performance on the TUCE with the results achieved by college students after completing a semester of their principles course. We conclude by offering some practical advice on identifying alternative assessments that complement ChatGPT as a learning tool.

Literature Review

Chatbots are a technology application that promotes interpersonal communication and learning. They provide information and knowledge through interactive methods and easy-to-operate interfaces (Hwang & Chang, 2021). With the exponential growth in the mobile device market over the past decade, the popularity of chatbots is being driven by their ability to provide an interactive medium through which to learn, one not constrained by time and place (Zhou et al., 2020). Early computer programs used in education were mainly limited to drill-and-practice exercises and did not incorporate the sophisticated techniques of artificial intelligence. However, AI has since been identified as an applicable technology for computer-assisted learning (Stubbs & Piddock, 1985). Artificial intelligence has the potential to address challenges in learning, including improving transfer of knowledge, dispelling misconceptions, and promoting critical thinking skills among students (Mollick & Mollick, 2022), and can be utilized as an effective teaching assistant in online learning environments by helping to enhance students' understanding and engagement through personalized feedback, real-time analysis, and adaptive instruction. A Georgia Tech computer science professor made headlines in 2016 for using artificial intelligence to build a virtual teaching assistant (Goel & Polepeddi, 2018). The chatbot known as "Jill" received very positive student evaluations, and students only seemed to suspect something was amiss when their teaching assistant responded quickly at all hours of the day.

Interaction with technologies, either by natural language or speech, is possible because as technology develops, users become more used to interacting with digital entities. Chatbots are now used across a wide range of domains, including marketing, customer service, technical support, education and training (Smutny & Schreiberova, 2020). Personal digital assistants like Siri (Apple), Alexa (Amazon), Cortana (Microsoft), and Google Assistant (Google) lie at the forefront of technology in voice recognition and "artificial intelligence" and have effectively replaced much of the day-to-day tasks once performed by assistants or secretaries (Smutny & Schreiberova, 2020). The use of digital technologies is now expected by the current generation of young people who were born into an era of the internet and smartphones (Selwyn, 2021).

Despite the global proliferation in the use of chatbots, studies exploring the benefits of using chatbots in educational settings have only recently emerged (Ferrell & Ferrell, 2020). These benefits include providing users with a pleasant learning experience by allowing for real-time interaction (Kim et al., 2019), enhancing peer communication skills (Hill et al., 2015), improving the learning efficiency of learners (Wu et al., 2020), and helping instructors manage large in-class activities (Schmulian & Coetzee, 2019).

With the advent of AI-type technology, scholars are now able to apply machine learning and natural language technology to the creation of chatbots, making their application in education a new topic of academic research (Følstad & Brandtzæg, 2017). Recent empirical studies have focused on understanding the optimal role for chatbots. In a study of educational chatbots for Facebook messenger to support learning, Smutny and Schreiberova (2020) highlight the possibility for chatbots to become a smart teaching assistant in the future. Other studies have examined the use of chatbots in language learning. Based on a review of 25 empirical studies, Huang et al. (2021) find that educational chatbots can foster students' language learning via interaction activities underpinned by intended learning objectives. In a similar study, Kim et al. (2019) conclude that chatbots have a positive effect on students' communication skills by expanding the quantity of their interactions, increasing their motivation, and raising their interest in learning.

Chatbots have come a long way in the last two decades. The rise of machine learning with access to very powerful computers and processing power able to train these datasets form the backbone of these systems. Coupled with "natural language processing," this has paved the way for chatbots to be introduced into the field of education via digital transformation. Because of its

scalability and adaptability, it offers unique possibilities as a communication and information tool for digital learning (Wollny et al., 2021). While it's not exactly clear how this field will evolve in future, as these machine learning-driven systems become more advanced and capable of replicating a broader range of human-like traits, there will be a greater acceptance of its use in shaping the education landscape of the 21st century.

While not the focus of this paper, we would be remiss if we did not mention the potential mischief ChatGPT will cause in the short term. To understand the potential impact of ChatGPT on academic integrity, it is important to acknowledge that cheating is not a new issue, and ChatGPT is simply the latest tool that can be used for a variety of purposes, ethical considerations aside.

Academic Dishonesty

The emergence of ChatGPT in November 2022 has raised fears about widespread cheating on exams and assignments. These concerns are similar to the ones that were raised when students started using calculators, phones, and laptops in the classroom, with fears that they would rely too heavily on technology and forget the basics, or that the technology would distract and facilitate cheating (Surovell, 2023). However, these fears were proven unfounded as educators adapted their teaching methods to incorporate technology.

The issue of cheating has been evolving over time, with the introduction of the internet in the 1990s leading to the ability to copy and paste information from the web, a form of subconscious appropriation—cryptomnesia—of work (Sisti, 2007; Cojocariu & Mareş, 2022). More subtle forms of plagiarism, such as rearranging phrases from a source without proper citation have also become more prevalent (Das & Panjabi, 2011).

With the increasing use of online learning management systems and the digitalization of education, the market for plagiarism has become more sophisticated. Students have found new ways to cheat, such as using smartphones (Srikanth & Asmatulu, 2014) and social media (Best & Shelley, 2018) to generate answers for exams and assignments. Anti-plagiarism tools like TurnItIn and Safeassign have been developed to counter these threats and have shown some success in reducing plagiarism (Batane, 2010). However, students have resorted to contract cheating, outsourcing their academic work, to avoid detection (Lancaster & Clarke, 2007).

Rigby et al. (2015) conducted an empirical study to understand why university students in the United Kingdom cheat. They used a hypothetical situation to gather data and found that the willingness to cheat through contract cheating varied among students. About half of the students surveyed were willing to pay for an essay. The likelihood of cheating increased among students who had a higher risk tolerance or English as a second language. The authors also found that as the risk of getting caught and facing penalties increased, the perceived value of the essay decreased. Overall, the authors found that students were willing to pay up to \$445 for an assessment.

The rise of artificial intelligence in higher education, specifically natural language models like ChatGPT, presents a new challenge to universities. Unlike anti-plagiarism tools that compare a student's work with existing sources, ChatGPT can generate original content in seconds. While ChatGPT-generated papers have received good grades, they lack the depth of understanding that is expected in higher education. Additionally, it is very difficult to detect plagiarism when using ChatGPT.

Tools like ChatGPT are likely to become a common part of the writing process, just as calculators and computers have become essential tools for learning mathematics and science. The challenge of universities is to adapt their curriculum to this new reality.

Methods and Findings

The National Council on Economic Education (NCEE) created the “Test of Understanding of College Economics” (TUCE) and an accompanying examiner’s manual to allow instructors to compare their students’ results with those of post-secondary students from across the country (Walstad et al, 2006; Walstad & Rebeck, 2008). In order to make these comparisons, the authors normalized thousands of students from various institutions based on a 30-question assessment that was given at the start and end of the term. The purpose of these pre- and post-tests was for educators to measure learning over the semester, including the impacts of changing the structure of the class away from chalk-and-talk (Emerson & Taylor, 2004; Boyle & Goffe, 2018).

Additionally, the normed sample provides a baseline understanding of the level of knowledge that the average college student in the United States has at the beginning and end of their economics principles courses. On average, student performance improves over the course of a semester as students go from answering an average of 9.39 questions correctly at the start of the term to an average of 12.77 questions correctly at the end of their principles of microeconomics course. For macroeconomics, students improved from 9.80 to 14.19 questions. Despite a full semester learning economics principles, most students answer around 40–50% of questions correctly. Figure 1 and 2 show the distribution of pre- and post-test scores for both the microeconomics and macroeconomics version of the exam. Given these distributions, where would a large language model like ChatGPT place if it was administered the TUCE?

The authors conducted a new chat session on the ChatGPT (hereafter referred to as ChatGPT) on February 8, 2023, using the GPT-3 version of the language model. They provided one question from each of the two versions of the TUCE at a time, along with its answer choices. ChatGPT returned an answer, which was recorded as correct if it matched the TUCE answer key, and incorrect if it was wrong or if multiple answers were provided. The authors didn’t assign any partial credit on ChatGPT’s response since the TUCE is administered as a multiple choice test to

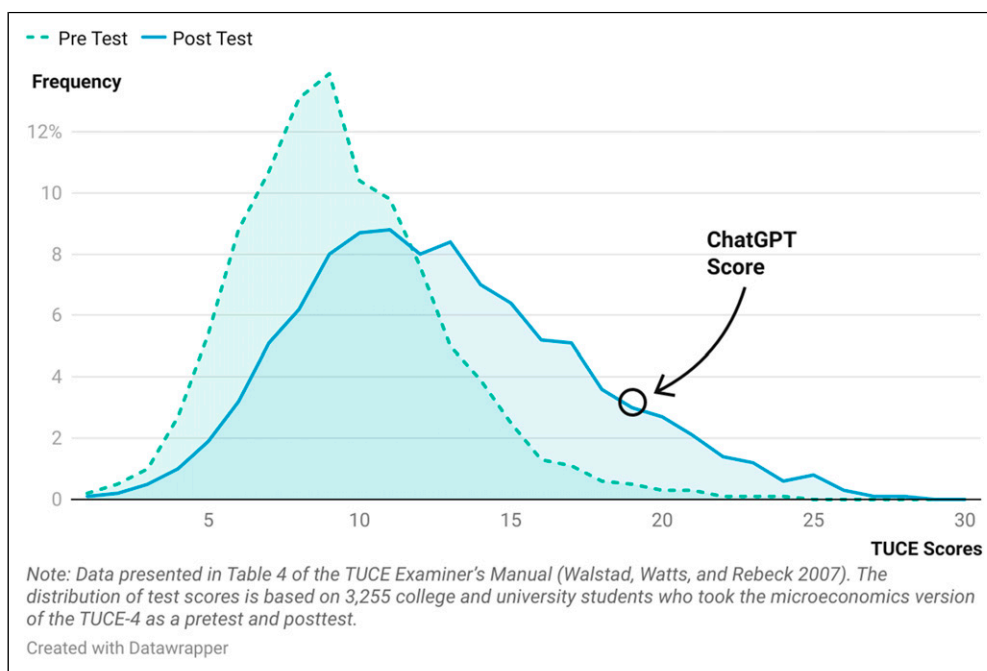


Figure 1. Distribution of pre- and post-test scores on microeconomics TUCE-4: Matched sample.

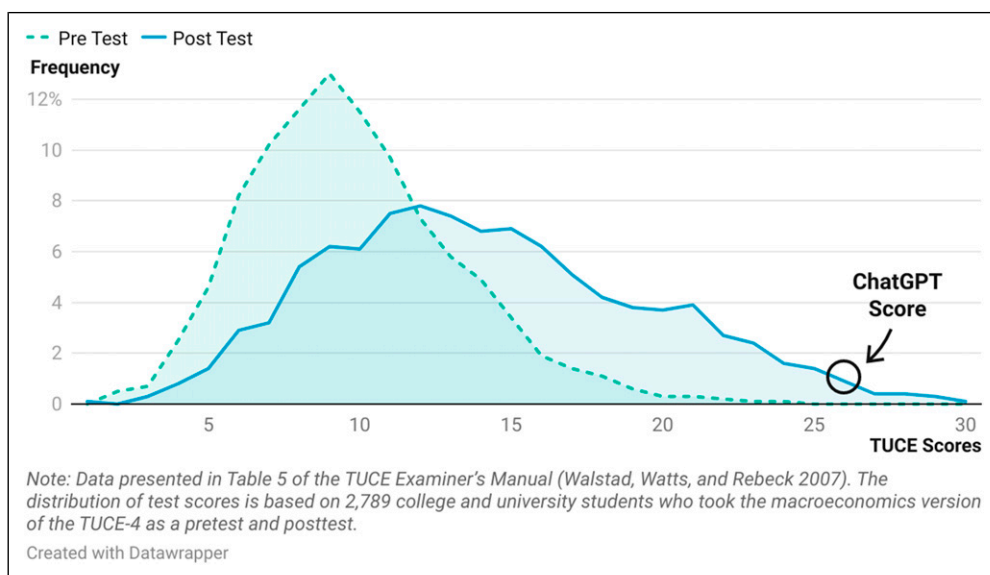


Figure 2. Distribution of pre- and post-test scores on macroeconomics TUCE-4: Matched sample.

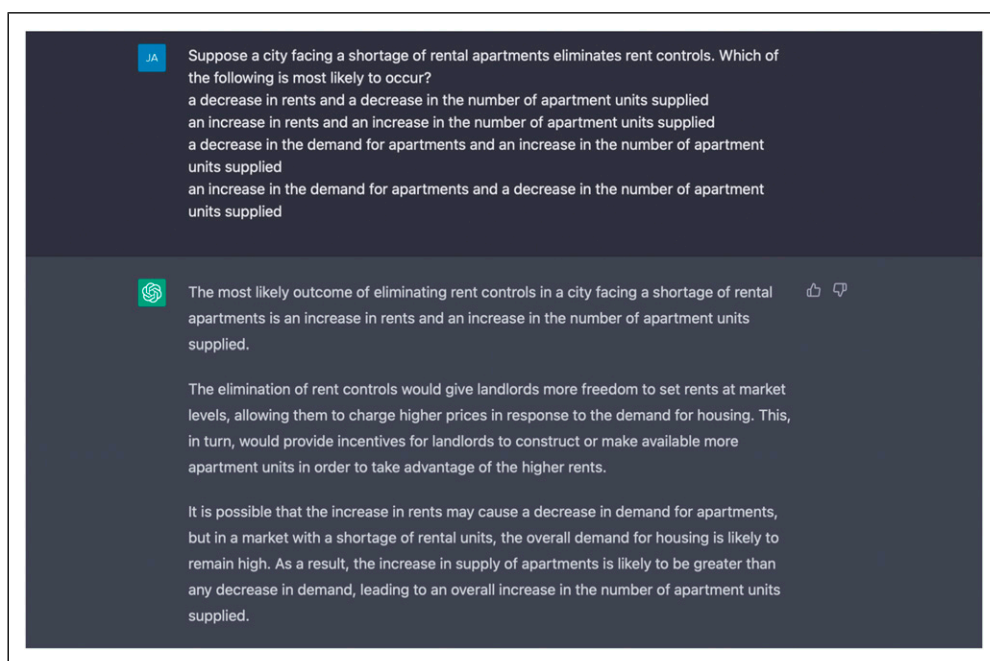


Figure 3. ChatGPT interface demonstrating question and answer methodology.

students in a proctored environment. It's also worth noting that the authors did not provide any feedback, such as thumbs up or down, to ChatGPT while it was generating responses during the chat session. Figure 3 illustrates the text input and the results for Question 2 on the microeconomics exam.

In our trial, ChatGPT answered 19 of 30 microeconomics questions correctly and 26 of 30 macroeconomics questions correctly, ranking in the 91st and 99th percentile, respectively. The incorrect responses often included odd behavior, such as when ChatGPT claimed that all answer choices were correct or provided an answer that was not among the four options. This sort of behavior isn't likely to occur among students taking a multiple choice test. It should also be noted that ChatGPT could not process images at the time of this writing, which resulted in one microeconomics question being provided with missing context. We have included a table in the appendix for both forms of the TUCE which states the concept being tested for each question and whether ChatGPT answered the question correctly or not (Walstad et al. 2006).²

To compare ChatGPT's performance with that of a typical economics student, we examined its percentile scores based on the results in Table A1 and Table A2 of the 4th Edition of the TUCE. If we consider only the pre-test scores, ChatGPT would rank in the top 1% of both microeconomics and macroeconomics exam takers. However, if we compare its scores with those of students who have completed a full semester of economics, it would still rank in the top 9% of microeconomics exam takers and continue to rank in the top 1% of macroeconomics exam takers.

Discussion

The rise of artificial intelligence in higher education, specifically natural language models like ChatGPT, presents a new challenge to educators. Unlike anti-plagiarism tools that compare a student's work with existing sources, ChatGPT can generate original content in seconds (McMutrie, 2023). This makes it difficult to detect plagiarism when using ChatGPT. To help close that gap, OpenAI Text Classifier, DetectGPT, GPTZero, Turnitin and many others claim to be able to detect the use of ChatGPT. However, relying solely on detection tools may not be a sufficient solution, as they may not always be able to identify instances where ChatGPT has been used to generate original content that closely resembles a human-written response or when ChatGPT responses have been modified by the student. As such, educators may need to adopt new strategies and approaches to address the challenges posed by AI-generated content, such as designing assessments that are more resistant to automated answers or emphasizing critical thinking and analytical skills that cannot be easily replicated by AI.

Moreover, ChatGPT has many advantages over non-AI forms of cheating: it is free, simple to use, and generates content much quicker than earlier methods.⁴ The emergence of ChatGPT has raised fears about widespread cheating on unproctored exams and other assignments. The short-term solution for many educators involves returning to in-person, proctored assessments. The main advantage of this approach is that violations of academic integrity can usually be reduced if the assessment is run properly. There are, however, certain drawbacks, including equity issues for students in remote or online classes when assessment is scheduled on-campus as well as the logistical challenges associated with large lectures.

Beyond this back to the future approach, there are other techniques that can be utilized in an online environment. Assessments that are time-constrained reward students who know the material, while others who do not know the material as well search their notes, ask their classmates, and seek answers through any means (including ChatGPT). The time spent searching means that they cannot complete as many questions, even if they are successful in obtaining the information.

A number of educators have begun to create assessments that teach students how to use ChatGPT as a resource and also use ChatGPT as part of the assessment (Schulten, 2023). One popular recommendation among the teaching community so far has been to produce ChatGPT responses with errors and have students work in small groups to identify and correct those errors. In essence, students are asked to "fact check" the system to ensure that the responses are accurate Figure 4.

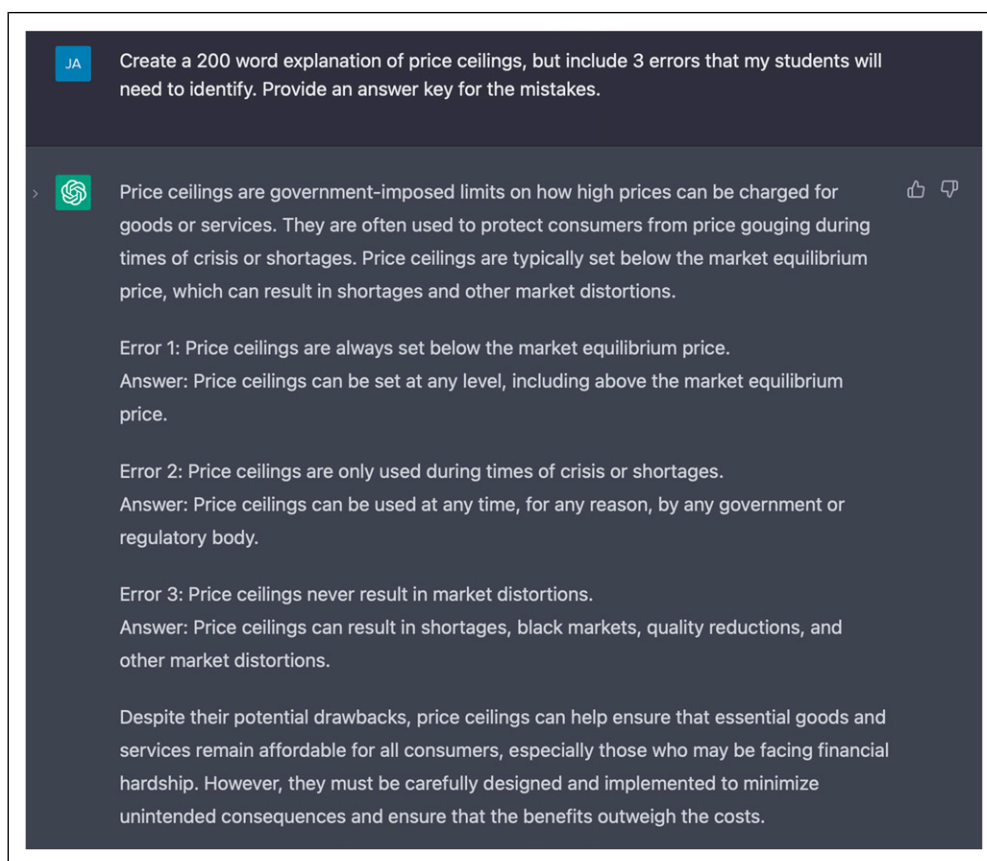


Figure 4. ChatGPT prompt and response for a hypothetical assignment in a principles of microeconomics course.

The current emphasis of “teaching with ChatGPT” has focused on humanities courses, but will likely evolve to the social sciences in due course (Cowen & Tabarrok, 2023). The current outlook among economics educators is to use ChatGPT as a source of knowledge, which is dangerous in its current stage since the program is merely predicting responses (Gecker, 2023). It’s important to emphasize to students that just because ChatGPT provides a response that looks reasonable doesn’t mean that the response is accurate.

ChatGPT presents some challenges, but they can be overcome by designing a learning environment that fosters knowledge acquisition. Artificial intelligence can enhance students’ learning experience and help them achieve more in less time, but there are ways to engage students in meaningful learning experiences that can’t be replicated by a program like ChatGPT. Research suggests that economic education can be effectively taught through hands-on experiences like classroom demonstrations, experiments, service learning, undergraduate research, case studies, and cooperative learning (Dorestani, 2005; Boyle & Goffe, 2018). This type of experiential learning goes beyond simple memorization and fosters a deeper understanding of the subject. Students can be asked to write brief essays that apply economic principles to solve interesting questions they personally observe (Geerling, 2013), form student groups to synthesize music with economics (Geerling, 2019), or work on art-inspired projects that require students to apply economic concepts (Al-Bahrani et al., 2016).

Assessments that evaluate higher-level thinking skills like analysis, evaluation, and creation can help engage students in meaningful learning experiences while making it more difficult for

ChatGPT to circumvent the process. The Economic Instructor's Toolkit (Picault, 2019, 2021) is a valuable resource that provides information on a growing list of class activities and student projects that foster higher-level learning. Whether teaching in-person or online, incorporating hands-on experiences into the curriculum can make a big impact on students' learning outcomes.

Conclusion

The purpose of this study was to evaluate the performance of ChatGPT in principles of economics tests, as assessed by the TUCE. The results found that ChatGPT ranks at the 99th percentile in macroeconomics and the 91st percentile in microeconomics, when compared to students who take the exams at the end of a semester-long principles course. It is hardly surprising that ChatGPT outperforms the average college student in a standardized test of economics comprehension delivered in multiple choice format with textbook answers, but the extent of this performance gap is quite revealing. ChatGPT was trained on a vast amount of text for its predictive algorithm, which gives it a significant advantage over its human counterparts.

Our findings have significant implications for assessment strategies in the ChatGPT-era. It is crucial to rethink assessment strategies to include both traditional methods, such as proctored exams, in-class writing assignments, or experiential learning opportunities, and to find ways to utilize chatbots as a teaching aide or as part of assessments in the future. It is important to note that ChatGPT is not the only disruptive technology in education. The advent of artificial intelligence in education is a reality that cannot be ignored, and it is time to embrace the new era with innovative and effective assessment strategies.

Appendix

Table A1. ChatGPT Performance on Microeconomics Version of TUCE

Question	Concept	Correct
1	Supply and demand	Yes
2	Price ceilings	Yes
3	Supply and demand	No
4	Perfect competition	Yes
5	Factors of production	No
6	Externalities	Yes
7	Income distribution	Yes
8	Opportunity cost	Yes
9	Supply and demand	No
10	Utility	No
11	Perfect competition	Yes
12	Monopoly	Yes
13	Diminishing marginal returns	Yes
14	Profit maximization	N/A
15	Externalities	Yes
16	Taxation	Yes
17	Monopoly	Yes
18	Elasticity	No
19	Demand	Yes
20	Profit maximization	DNA

(continued)

Table A1. (continued)

Question	Concept	Correct
21	Market structure	No
22	Duopoly	Yes
23	Economic rent	Yes
24	Profit maximization	Yes
25	Public choice	Yes
26	Externalities	No
27	Public goods	Yes
28	Comparative advantage	No
29	Trade barriers	Yes
30	Exchange rates	No

Table A2. ChatGPT Performance on Macroeconomics Version of TUCE.

Question	Concept	Correct
1	Components of GDP	No
2	Inflation	Yes
3	Aggregate demand	Yes
4	Potential GDP	Yes
5	Money supply	Yes
6	Tools of monetary policy	No
7	Tools of monetary policy	Yes
8	Automatic fiscal policy	Yes
9	Crowding out	Yes
10	Inflation expectations	Yes
11	Unemployment rate	Yes
12	Real interest rate	Yes
13	Supply shocks	Yes
14	Aggregate demand	Yes
15	Aggregate demand	Yes
16	Tools of monetary policy	No
17	Fiscal policy	Yes
18	Tools of monetary policy	Yes
19	Real GDP	Yes
20	Multiplier effect	Yes
21	Economic growth	Yes
22	Money creation	Yes
23	Tools of fiscal policy	Yes
24	Monetary vs. Fiscal policy	Yes
25	Tools of monetary policy	Yes
26	Policy lags and limitations	Yes
27	Automatic fiscal policy	Yes
28	Exchange rates	No
29	Open-economy macroeconomics	Yes
30	Trade balance	Yes

Acknowledgments

The authors would like to thank Charity-Joy Acchiardo, Bob Gazelle, Bill Goffe, Simon Halliday, Kris Nagy, Brian O’Roark, and Julien Picault for their helpful comments and feedback on an earlier draft. We would also like to thank the feedback from two anonymous reviewers and the editor.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Wayne Geerling  <https://orcid.org/0009-0005-1389-0844>

Notes

1. ChatGPT-3 was modeled on data from the internet and other sources, up until the end of 2021.
2. Due to copyright restrictions on TUCE material, we are unable to include the full-text of the questions. We are happy to send our findings to anyone upon request. The question order presented in the appendix is the same question order presented in the 4th edition of the TUCE available online through the Council for Economic Educators.

References

- Al-Bahrani, A., Holder, K., Patel, D., & Wooten, J. (2016). Art of econ: Incorporating the arts through active learning assignments in principles courses. *Journal of Economics and Finance Education*, 15(2), 1–16.
- Batane, T. (2010). Turning to Turnitin to fight plagiarism among university students. *Journal of Educational Technology and Society*, 13(2), 1–12.
- Best, L. M., & Shelley, D. J. (2018). Academic dishonesty: Does social media allow for increased and more sophisticated levels of student cheating? *International Journal of Information and Communication Technology Education*, 14(3), 1–14. <https://doi.org/10.4018/ijicte.2018070101>
- Boyle, A., & Goffe, W. L. (2018). Beyond the flipped class: The impact of research-based teaching methods in a macroeconomics principles class. *AEA Papers and Proceedings*, 108, 297–301. <https://doi.org/10.1257/pandp.20181052>
- Choi, J., Hickman, K., Monahan, A., & Schwarcz, D. (2023). *ChatGPT goes to law school*. Minnesota Legal Studies Research Paper No. 23-03, Available at SSRN: <https://ssrn.com/abstract=4335905>
- Cojocariu, V., & Mareş, G. (2022). Academic integrity in the technology-driven education era. In *Ethical Use of information technology in higher education* (pp. 1–16).
- Cowen, T., & Tabarrok, A. T. (2023). *How to learn and teach economics with large language models, including Gpt*. Working paper available at SSRN: <https://ssrn.com/abstract=4391863>
- Das, N., & Panjabi, M. (2011). Plagiarism: Why is it such a big issue for medical writers? *Perspectives in Clinical Research*, 2(2), 67–71. <https://doi.org/10.4103/2229-3485.80370>
- Dorestani, A. (2005). Is interactive/active learning superior to traditional lecturing in economics courses? *Humanomics*, 21(1), 1–20. <https://doi.org/10.1108/eb018897>
- Emerson, T. L. N., & Taylor, B. A. (2004). Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course. *Southern Economic Journal*, 70(3), 672–693. <https://doi.org/10.2307/4135338>

- Ferrell, O. C., & Ferrell, L. (2020). Technology challenges and opportunities facing marketing education. *Marketing Education Review*, 30(1), 3–14. <https://doi.org/10.1080/10528008.2020.1718510>
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
- Gecker, J. (2023). *Some educators embrace ChatGPT as a new teaching tool*. Austin PBS. <https://www.pbs.org/newshour/education/some-educators-embrace-chatgpt-as-a-new-teaching-tool>
- Geerling, W. (2013). An exploration of robert frank's 'the economic naturalist' in the classroom. *International Review of Economics Education*, 12, 48–59. <https://doi.org/10.1016/j.iree.2013.04.008>
- Geerling, W., Mateer, G. D., & O'Roark, B. (2019). Music then and now: Using technology to build a lyric animation module. *The American Economist*, 65(2), 264–276. <https://doi.org/10.1177/0569434519889063>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.2196/45312>
- Goel, A. K., & Polepeddi, L. (2018). Jill Watson: A virtual teaching assistant for online education. In *Learning engineering for online education* (pp. 120–143). Routledge.
- Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Huang, W., Hew, K. F., & Fryer, L. K. (2021). Chatbots for language learning – are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Hwang, G. J., & Chang, C.-Y. (2021). *A review of opportunities and challenges of chatbots in education*. <https://doi.org/10.1080/10494820.2021.1952615>
- Kim, N. Y., Cha, Y., & Kim, H. S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, 22(3), 32–53.
- Lancaster, T., & Clarke, R. (2007). Assessing contract cheating through auction sites: A computing perspective. *Higher Education Academy for Information and Computer Sciences*, 1.
- Leswing, K. (2023). March 26). *OpenAI announces GPT-4, claims it can beat 90% of humans on the SAT*. CNBC. <https://www.cnbc.com/2023/03/14/openai-announces-gpt-4-says-beats-90percent-of-humans-on-sat.html>
- McMutrie, B. (2023). February 8). *AI and the future of undergraduate writing*. The Chronicle of Higher Education. <https://www.chronicle.com/article/ai-and-the-future-of-undergraduate-writing>
- Mollick, E. R., & Mollick, L. (2022). New Modes of learning enabled by AI chatbots: Three methods and assignments. Working paper available at SSRN: <https://ssrn.com/abstract=4300783>
- Picault, J. (2019). The economics instructor's toolbox. *International Review of Economics Education*, 30, 100154. <https://doi.org/10.1016/j.iree.2019.01.001>
- Picault, J. (2021). Looking for innovative pedagogy? An online economics instructor's toolbox. *The Journal of Economic Education*, 52(2), 174. <https://doi.org/10.1080/00220485.2021.1887024>
- Rigby, D., Burton, M., Balcombe, K., Bateman, I., & Mulatu, A. (2015). Contract cheating and the market in essays. *Journal of Economic Behavior and Organization*, 111, 23–37. <https://doi.org/10.1016/j.jebo.2014.12.019>
- Schmulian, A., & Coetzee, S. A. (2019). The development of messenger bots for teaching and learning and accounting students' experience of the use thereof. *British Journal of Educational Technology*, 50(5), 2751–2777. <https://doi.org/10.1111/bjet.12723>
- Schulten, K. (2023). February 8). *Lesson plan: Teaching and learning in the era of ChatGPT*. *The New York times*. <https://www.nytimes.com/2023/01/24/learning/lesson-plans/lesson-plan-teaching-and-learning-in-the-era-of-chatgpt.html>
- Selwyn, N. (2021). *Education and technology: Key issues and debates* (3rd ed.). Bloomsbury.
- Sisti, D. A. (2007). How do high school students justify internet plagiarism? *Ethics and Behavior*, 17(3), 215–231. <https://doi.org/10.1080/10508420701519163>

- Smith, B. O., & Wagner, J. (2018). Adjusting for guessing and applying a statistical test to the disaggregation of value-added learning scores. *The Journal of Economic Education*, 49(4), 307–323. <https://doi.org/10.1080/00220485.2018.1500959>
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers and Education*, 151, 103862. <https://doi.org/10.1016/j.compedu.2020.103862>
- Srikanth, M., & Asmatulu, R. (2014). Modern cheating techniques, their adverse effects on engineering education and preventions. *International Journal of Mechanical Engineering Education*, 42(2), 129–140. <https://doi.org/10.7227/ijmee.0005>
- Stubbs, M., & Piddock, P. (1985). Artificial intelligence in teaching and learning: An introduction. *PLET: Programmed Learning and Educational Technology*, 22(2), 150–157. <https://doi.org/10.1080/1355800850220207>
- Surovell, E. (2023). February 8). *ChatGPT has everyone freaking out about cheating. It's not the first time.* The Chronicle of Higher Education. <https://www.chronicle.com/article/chatgpt-has-everyone-freaking-out-about-cheating-its-not-the-first-time>
- Walstad, W., Watts, M., & Rebeck, K. (2006). Test of understanding of college economics. *Examiner's manual* (4th edn.). National Council on Economic Education.
- Walstad, W. B., & Rebeck, K. (2008). The test of understanding of college economics. *American Economic Review*, 98(2), 547–551. <https://doi.org/10.1257/aer.98.2.547>
- West, C. G. (2023). *AI and the FCI: Can ChatGPT project an understanding of introductory physics?* arXiv preprint arXiv:2303.01067.
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? – a systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 1–18. <https://doi.org/10.3389/frai.2021.654924>
- Wu, E. H. K., Lin, C. H., Ou, Y. Y., Liu, C. Z., Wang, W. K., & Chao, C. Y. (2020). Advantages and constraints of a hybrid model K-12 e-learning assistant chatbot. *IEEE Access*, 8, 77788–77801. <https://doi.org/10.1109/access.2020.2988252>
- Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53–93. https://doi.org/10.1162/coli_a_00368

Author Biographies

Wayne Geerling is a Professor of Instruction at the University of Texas at Austin. He is the author of *Quantifying Resistance: Political Crime and the People's Court in Nazi Germany* (2018) and *Socialism with a Human Face: Using Behavioural Economics to Understand East German Economic History* (2022) and has published in leading peer-reviewed journals in the fields of economics education and interdisciplinary history.

G. Dirk Mateer is a Professor of Instruction at the University of Texas at Austin. He is the author of *Economics in the Movies*, *Essentials of Economics*, and *Principles of Economics*. Dirk has received over 40 teaching awards, including the Kenneth Elzinga Distinguished Teaching Award from the Southern Economic Association in 2021.

Jadrian Wooten is a Collegiate Associate Professor of Economics at Virginia Tech. He is the author of *Parks and Recreation and Economics* and writes a weekly newsletter known as the Monday Morning Economist. His academic research focuses on teaching pedagogy as well as sports and other labor-related issues. Jadrian is committed to developing teaching resources for university and high school economics instructors and is most well known for his work on the integration of media into the economics curriculum.

Nikhil Damodaran is an Assistant Professor at the OP Jindal Global's School of Government and Public Policy. His research focuses on applying open economy macroeconomic models to understand regional economic interactions and their implications for aggregate fluctuations in a macroeconomy.