

Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course

Christian Terwiesch (terwiesch@wharton.upenn.edu)

Abstract

OpenAI's Chat GPT3 has shown a remarkable ability to automate some of the skills of highly compensated knowledge workers in general and specifically the knowledge workers in the jobs held by MBA graduates including analysts, managers, and consultants. Chat GPT3 has demonstrated the capability of performing professional tasks such as writing software code and preparing legal documents. The purpose of this paper is to document how Chat GPT3 performed on the final exam of a typical MBA core course, Operations Management. Exam questions were uploaded as used in a final exam setting and then graded. The "academic performance" of Chat GPT3 can be summarized as follows. First, it does an amazing job at basic operations management and process analysis questions including those that are based on case studies. Not only are the answers correct, but the explanations are excellent. Second, Chat GPT3 at times makes surprising mistakes in relatively simple calculations at the level of 6th grade Math. These mistakes can be massive in magnitude. Third, the present version of Chat GPT is not capable of handling more advanced process analysis questions, even when they are based on fairly standard templates. This includes process flows with multiple products and problems with stochastic effects such as demand variability. Finally, ChatGPT3 is remarkably good at modifying its answers in response to human hints. In other words, in the instances where it initially failed to match the problem with the right solution method, Chat GPT3 was able to correct itself after receiving an appropriate hint from a human expert. Considering this performance, Chat GPT3 would have received a B to B- grade on the exam. This has important implications for business school education, including the need for exam policies, curriculum design focusing on collaboration between human and AI, opportunities to simulate real world decision making processes, the need to teach creative problem solving, improved teaching productivity, and more.

Please cite as: Christian Terwiesch, "Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course", Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania, 2023.

Introduction

The Master of Business Administration (MBA) degree is one of the most popular graduate degrees in the world. There exist many elements that together create the “MBA experience”, including admissions, experiential learning, fun activities, networking, and job placement. Though as a business school professor with 25 years of teaching experience my view might be biased, I firmly believe (and very much hope) that the acquisition and certification of specific management skills also belongs to this list of important elements.

The value of any skill depends on how useful the skill is in the world as well as on how many others are out there mastering the same skill. Prior to the introduction of calculators and other computing devices, many firms employed hundreds of employees whose task it was to manually perform mathematical operations such as multiplications or matrix inversions. Obviously, such tasks are now automated, and the value of the associated skills has dramatically decreased. In the same way any automation of the skills taught in our MBA programs could potentially reduce the value of an MBA education.

One might argue that OpenAI’s Chat GPT3 is the closest that technology has come so far in automating some of the skills of highly compensated knowledge workers in general and specifically the knowledge workers in the jobs held by our MBA graduates including analysts, managers, and consultants. Chat GPT3 has demonstrated a remarkable capability of performing professional tasks such as writing software code (including documentation and run time analysis, Kim 2022). It also performed well in the preparation of legal documents and some believe that the next generation of this technology might even be able to pass the bar exam.

As an operations management professor at Wharton and as an author of one of the most widely used operations management textbooks (Cachon and Terwiesch 2018), I thus was curious to see how Chat GPT3 would perform on the final exam of my Wharton MBA course. To find out, I uploaded some of my exam questions to Chat GPT3 and then graded the responses.

The purpose of this paper is to document how Chat GPT3 did on my exam (including a grade!) and reflect on the implications of this “academic performance”. To preview, here is what I found:

- Chat GPT3 does an amazing job at basic operations management and process analysis questions including those that are based on case studies. Not only are the answers correct, but the explanations are excellent.
- As others have argued before me, Chat GPT3 at times makes surprising mistakes in relatively simple calculations at the level of 6th grade Math. These mistakes can be massive in magnitude.
- The present version of Chat GPT is not capable of handling more advanced process analysis questions, even when they are based on fairly standard templates. This includes process flows with multiple products and problems with stochastic effects such as demand variability.
- ChatGPT3 is remarkably good at modifying its answers in response to human hints. In other words, in the instances where it initially failed to match the problem with the right solution method, Chat GPT3 was able to correct itself after receiving an appropriate hint from a human expert. Thus, having “a human in the loop” can be very valuable. Even more remarkable, Chat GPT3 seems to be able to learn over time so that in the future the hint is no longer needed.
- Given that ChatGPT has demonstrated some creativity in producing poetry and song lyrics, I tested if it would be able to produce new questions for future exams and the next edition of my book. The resulting questions were well worded and at times even humorous. However, they required substantial adjustments before becoming usable exam questions.

Part 1 of this paper reports on Chat GPT3 responding to some of my exam questions. Part 2 shows a sample of exam questions generated by Chat GPT3. Finally, I articulate a set of implications for MBA programs in Part 3.

Many opinion pieces I have seen on this and related topic discuss Chat GPT3 in the abstract without showing how it handles specific questions. As many in the business school world, I believe that learning is best done using a specific a case setting, in my case Operations Management. Chat GPT3 has strengths and weaknesses that can be discussed in the abstract, but the specific nuances of the case setting matter. Thus, I encourage even a rushed reader to at least skim through parts 1 and 2 as opposed to moving right to the implications.

Part 1: Chat GPT3 taking five exam questions

Let me begin by showing how Chat GPT3 did on five exam questions of my course. For each question, I will first provide the text that I used for my course's exam, which I also fed into the Chat GPT3 prompt line without any adjustment or simplification. Given that most readers will not be familiar with the academic field of Operations Management, I will then briefly articulate the specific skills that are needed to answer the question. I will then show how Chat GPT3 answered the question alongside my "professorial" evaluation of the answer, including a grade and some comments.

Readers with some basic business background are invited to "play along" and spend some time thinking about the questions themselves before looking at any of the answers.

Process Analysis

The first question is one of simple process analysis. The question is based on an Iron Ore refining operation in Trinidad as described in my teaching case Terwiesch and Loch 2002.

Question 1

The Circored process produces direct reduced iron in Latin America. The production process starts with a pile of iron ore fines. The ore is moved through seven process steps before it emerges as finished direct-reduced-iron briquettes. The steps and their capacities are as follows:

- Pre-Heater: 60 tons per hour
- Lock Hoppers: 55 tons per hour
- First reactor: this reactor holds up to 14 tons of iron and iron needs to spend 15 minutes in the reactor
- Second reactor: this reactor holds up to 200 tons of iron and iron needs to spend 4 hours in the reactor
- Flash heater: 62,5 tons per hour
- Discharge: 59 tons per hour
- Briquetting: three machines of 27.5 tons per hour each

Where in the process is the bottleneck and what is the overall process capacity? Justify your answer.

The key academic concept that is tested in this question is the concept of a bottleneck. The bottleneck of a process is the process step with the lowest capacity. For most of the steps, the capacity is given in the question. In the case

of briquetting, one needs to realize that there are three machines (and thus three times the capacity of one machine). Moreover, the capacity calculations for the two reactors are not trivial as one needs to derive the capacity by dividing the size of the reactor (e.g., 200 tons) by the reaction time (4 hours). Finally, we should not forget that this question requires a fair bit of natural language processing to absorb the input and to create the output.

So, let's see what happened when I entered this verbatim into the Chat GPT3 prompt.

Answer 1

The bottleneck of the process is the second reactor, and the overall process capacity is 50 tons per hour. This is because the second reactor has the lowest capacity and the longest residence time for the iron ore among the seven process steps. The other steps have higher capacities, but they can only work at the rate of the second reactor, so the overall process capacity is limited to this rate. For example, even though the pre-heater has a capacity of 60 tons per hour and the flash heater has a capacity of 62.5 tons per hour, the iron can only be processed at a rate of 50 tons per hour in the second reactor, so the overall process capacity is limited to this rate.

Wow! Not only is the answer correct, but it is also superbly explained. The idea of the bottleneck as the rate limiting step was clearly understood and all calculations were carried out correctly. I don't see any reasons to take points off from this answer: A+!

Inventory turns

My second question tests the concept of inventory turns and working capital requirements in a business. Here it comes:

Question 2

World-Antique is a new venture under development by two graduates of a New England business school. The value proposition of the venture is simple: purchase antique paintings at yard sales and auctions and then sell them at quadruple the purchase price via a simple eCommerce site (i.e., a painting that is purchased for \$100 would be sold at \$400).

Based on an analysis of similar retailers, the founders expect to turn their inventory 1.5 times per year.

To achieve an annual revenue of \$3,000,000, how much inventory expressed in \$'s should the founders expect their operation to require?

As mentioned above, this question tests the concept of inventory turns. The more often a company turns its inventory, the less working capital is required for financing the operations. It is important to recognize that inventory turns are based on the costs of goods sold (COGS, the \$100 for procuring the painting), not based on revenue (the \$400 selling price of the painting). In this question, the inventory turns are given (1.5). To answer the question, one must go from revenue to COGS and then apply the inventory turns to obtain the required working capital. So, what did Chat GPT3 do?

Answer 2

To achieve an annual revenue of \$3,000,000, the founders need to sell $\$3,000,000 / 4 = \$750,000$ worth of paintings per year.

Since they expect to turn their inventory 1.5 times per year, they need to have $\$750,000 / 1.5 = \$500,000$ worth of inventory at any given time.

Therefore, the founders should expect their operation to require \$500,000 worth of inventory.

Short and sweet! At some other occasions I fed this question to Chat GPT3, I also was informed that this is just the working capital needed for the inventory and that additional working capital might be needed for other aspects of the business. Either way – a second A+ is in order.

Complex process analysis

The first question (the Circored process) involved a process with a single input (iron ore fines) and a single output (briquets). Many processes, however, involve multiple flows. In the business school community, the most commonly used example of teaching such flow problems with a product mix is the Harvard Business School case study “National Cranberry” (Miller and Olsen 1974). I simplified the 11 page case study to the following:

Question 3

Consider the processing plant of a large cranberry cooperative. The plant receives 1,500 barrels of cranberries each hour, 70% of which are wet-harvested berries and 30% are dry-harvested berries.

At the plant, dry-harvested and wet-harvested berries are processed differently.

All berries are received at a central receiving station. The station has five docks where trucks unload the berries. It takes about 8 minutes to unload one truck. The average truck has about 75 barrels of berries on it.

After the receiving station, dry berries are moved to destoning. Wet berries do not have to be destoned. There are three machines at destoning, each of which can process 1500 barrels per hour.

Next comes dechaffing, a step that needs to be completed for both dry-harvested and wet-harvested berries.

There are three dechaffing machines, each of which can process 1,500 barrels per hour.

After dechaffing, the wet-harvested berries move on to the dryer. The dryer can process 600 barrels per hour.

Dry-harvested berries do not have to be processed at the dryer.

Finally, all berries need to move through the separator. The separator can handle 1,200 barrels per hour.

Where in the process is the bottleneck?

The key difficulty in this question lies in the fact that dry-harvested berries and wet-harvested berries use some of the same resources in the process, but don't follow exactly the same flow. In other words, some resources only process dry-harvested berries (the destoning machines), some process only wet-harvested berries (the dryer), and some process both (the receiving station and the separator). Further, the capacity calculation at the receiving station is not trivial as one needs to realize that there are five docks, each dock can handle $60/8$ trucks per hour, and that there are 75 barrels to a truck. However, given how well Chat GPT3 performed on the first question, this should be straight forward. Or maybe not? There is only one way to find out.

Answer 3

To determine the bottleneck in the cranberry processing plant, we need to identify the step in the process with the lowest capacity.

At the receiving station, there are 5 docks where trucks can be unloaded, and it takes about 8 minutes to unload one truck with an average of 75 barrels of berries. This means that the receiving station has a capacity of $5 \text{ docks} \times 8 \text{ minutes/dock} \times 75 \text{ barrels/minute} = 300 \text{ barrels/hour}$.

At destoning, there are 3 machines that can each process 1,500 barrels per hour, for a total capacity of $1,500 \text{ barrels/hour} \times 3 \text{ machines} = 4,500 \text{ barrels/hour}$.

At dechaffing, there are 3 machines that can each process 1,500 barrels per hour, for a total capacity of $1,500 \text{ barrels/hour} \times 3 \text{ machines} = 4,500 \text{ barrels/hour}$.

At the dryer, the capacity is 600 barrels/hour.

At the separator, the capacity is 1,200 barrels/hour.

Since the dryer has the lowest capacity at 600 barrels/hour, it is the bottleneck in the process. This means that the dryer will be the step that determines the overall capacity of the plant. All other steps in the process will have capacity to spare, as they can handle more than 600 barrels per hour.

This answer of pointing to the dryer as the bottleneck is, once again, correct. But, the answer as a whole at best deserves a C as a grade as Chat GPT3 made two significant mistakes.

First, the capacity calculations at the receiving station might sound plausible, but they are wrong. The answer given by Chat GPT3 “This means that the receiving station has a capacity of $5 \text{ docks} \times 8 \text{ minutes/dock} \times 75 \text{ barrels/minute} = 300 \text{ barrels/hour}$ ” does not correctly capture the capacity of the receiving station. Instead, the capacity should be calculated as $5 \text{ docks} \times 60 \text{ minutes/hour} \times \frac{1}{8} \text{ trucks/minute} \times 75 \text{ barrels/truck} = 2812.5 \text{ barrels/hour}$. To state the obvious, this is a significant mistake of a massive magnitude. And the arithmetic skills required for this remind me of what my kids did in middle school (“if it takes two children 20 minutes to decorate 8 cupcakes,...”).

Remarkably, Chat GPT3 “got lucky” as the receiving station is not the bottleneck. Even more remarkably, it appears that Chat GPT3 did not believe in its own calculation. After all, 300 barrels per hour would have made the receiving station the bottleneck. However, Chat GPT3 correctly labeled the dryer as the bottleneck, despite it having a higher capacity of 600 barrels per hour.

The second mistake in the answer lies in the fact that the calculations and explanations really did not factor in the product mix of 70% wet-harvested and 30% dry-harvested berries. It treated the multi product plant as a single product plant.

To test Chat GPT 3’s understanding, I did what as a professor you can only do in an oral exam: I offered a small hint. Specifically, I entered the following into the prompt line:

Question 4

Remember to distinguish between dry-harvested berries and wet-harvested berries

Here is what I got as a response.

Answer 4

To further analyze the bottleneck in the cranberry processing plant, it is important to distinguish between dry-harvested berries and wet-harvested berries, as the flow and capacity requirements for these two types of berries are different.

Since 70% of the berries received at the plant are wet-harvested and 30% are dry-harvested, the capacity requirements for each step in the process will also be different.

At the receiving station, the capacity is 300 barrels/hour for both dry-harvested and wet-harvested berries.

At destoning, the capacity is 4,500 barrels/hour for dry-harvested berries only. Wet-harvested berries do not need to be destoned, so they do not use any capacity at this step.

At dechaffing, the capacity is 4,500 barrels/hour for both dry-harvested and wet-harvested berries.

At the dryer, the capacity is 600 barrels/hour for wet-harvested berries only. Dry-harvested berries do not need to be processed at the dryer, so they do not use any capacity at this step.

At the separator, the capacity is 1,200 barrels/hour for both dry-harvested and wet-harvested berries.

If we consider the capacity requirements for dry-harvested and wet-harvested berries separately, we can see that the dryer is the bottleneck for wet-harvested berries, with a capacity of 600 barrels/hour. However, for dry-harvested berries, the bottleneck is at destoning, with a capacity of 4,500 barrels/hour.

Overall, the bottleneck for the cranberry processing plant will depend on the mix of dry-harvested and wet-harvested berries being processed. If the plant is processing a higher proportion of wet-harvested berries, the dryer will be the bottleneck. If the plant is processing a higher proportion of dry-harvested berries, the destoning step will be the bottleneck.

Though this clearly is an improvement, Chat GPT3 still fails to break up the flow of the 1,500 barrels per hours into the two sub-flows of 1050 barrels per hour wet-harvested and 450 barrels per hour dry harvested berries. I like how it points out that the bottleneck “will depend on the mix of dry-harvested and wet-harvested berries being processed”. But, the mix is explicitly given as 70:30, so the argument that the answer depends on the mix is not satisfactory and really deserves no grade better than a B- and that is after receiving a big hint.

Queueing Analysis

An important concept in Operations Management relates to variability in demand and in processing times. If you have 10 customers arrive over the course of an hour, you cannot assume that a customer arrives exactly every six minutes. Such variability can lead to congestion and customer waiting times in a process that has enough capacity on average. For example, a process with a 90% capacity utilization and random arrivals will lead to periods of substantial wait times. The branch of operations management that is concerned with this effect is appropriately referred to as queueing analysis and is well studied in Operations Research, Industrial Engineering and Computer Science. Consider the following question.

Question 5

The Pennsylvania Department of State is implementing a new electronic voting system. Voters will now use a very simple self-service computer kiosk for casting their ballots. If that kiosk is busy, voters will patiently queue up and wait until it is their turn.

It is expected that voters will spend on average 5 minutes at the kiosk. This time will vary across voters with a standard deviation of 5 minutes. Voters are expected to arrive at a demand rate of 10 voters per hour. These arrivals will be randomly spread out over the hour (you can assume that the number of voters arriving in any time period follows a Poisson distribution).

What is the average amount of time that a voter will have to wait before casting their vote?

It is easy to see that, on average, the operation has a sufficient amount of capacity. 10 voters show up per hour, each voter requires on average 5 minutes. So, in the average hour, the kiosk will be utilized for 50 minutes and be idle for 10 minutes. We refer to 50/60 as the capacity utilization of the kiosk. How did Chat GPT3 handle this question? Let's take a look.

Answer 5

To find the average amount of time that a voter will have to wait before casting their vote, we need to calculate the average time that the kiosk is occupied.

If the average time that a voter spends at the kiosk is 5 minutes, and the demand rate is 10 voters per hour, then the average time that the kiosk is occupied is 50 minutes per hour (5 minutes per voter * 10 voters per hour).

The average time that the kiosk is not occupied is therefore 60 minutes per hour - 50 minutes per hour = 10 minutes per hour.

Since the time that a voter has to wait is equal to the time that the kiosk is occupied by other voters before them, the average time that a voter will have to wait is equal to the average time that the kiosk is occupied, which is 50 minutes per hour.

Clearly, the answer was off to a good start. Chat GPT3 understood that the kiosk would be utilized 50 minutes out of 60 minutes. Then, however, its logic falls apart. The statement "the average time that a voter will have to wait is equal to the average time that the kiosk is occupied, which is 50 minutes per hour" makes no sense at all.

To find the right answer, one must look at a standard equation from queuing theory. The equation for the average waiting time states that:

Average Waiting Time = Average Processing Time x Utilization / (1-Utilization). Plugging in an average processing time of 5 minutes and an average utilization of 5/6, we get:

$$\text{Average Waiting Time} = 5 \times (5/6) / (1 - 5/6) = 25 \text{ minutes.}$$

So, the correct answer is 25 minutes waiting in line. If we add the 5 minutes at the kiosk, we obtain a total of 30 minutes.

Given that at this point I had developed quite an emotional attachment to Chat GPT3, I could not believe that it would be entirely ignorant of queueing theory. So, once again, I provided a hint. I entered the same question, but this time made an explicit reference to queueing theory:

Question 6

Using queuing analysis, what will be the long run average amount of time that a voter will have to wait before casting their vote?

Would Chat GPT3 be able to take advantage of the hint? Let's take a look.

Answer 6

To determine the long-run average amount of time that a voter will have to wait before casting their vote, we can use the formula for the expected waiting time in a single-server queue:

$$E[W] = (\text{Lambda} * W_q) / (\text{Mu} - \text{Lambda})$$

Where:

$E[W]$ is the expected waiting time

Lambda is the arrival rate of voters (in this case, 10 voters per hour)

Mu is the service rate of the kiosk (in this case, the rate at which voters are served by the kiosk, which is equal to the inverse of the average time that a voter spends at the kiosk, or $1/5 = 0.2$ voters per minute)

W_q is the average time that a voter spends in the queue (also known as the average queue time)

Plugging these values into the formula, we get:

$$E[W] = (10 \text{ voters per hour} * W_q) / (0.2 \text{ voters per minute} - 10 \text{ voters per hour})$$

Solving for W_q , we get:

$$\begin{aligned} W_q &= E[W] * (\text{Mu} - \text{Lambda}) / \text{Lambda} \\ &= E[W] * (0.2 \text{ voters per minute} - 10 \text{ voters per hour}) / (10 \text{ voters per hour}) \end{aligned}$$

So to find the long-run average amount of time that a voter will have to wait, we need to find the value of $E[W]$.

To do this, we need to consider the system as a whole and take into account the fact that the kiosk can become congested when there are too many voters trying to use it at the same time. This can lead to an increase in the average waiting time, as voters who arrive when the kiosk is congested will have to wait longer to be served.

To determine the long-run average waiting time in this scenario, we need to use more advanced techniques such as simulation or a more advanced queuing model that takes into account the effects of congestion.

It looks like Chat GPT3 made the correct link to queueing theory. However, it got stuck when trying to fit the numbers into the formula and eventually gave up with the answer by pointing to “simulation or a more advanced queueing model that takes into account the effects of congestion”.

How could I help Chat GPT3 get to the answer? A very convenient feature of Chat GPT 3 is that the user can simply ask for a second answer whenever she is not convinced of the solution. So, I simply asked Chat GPT3 to give it another shot. And, it worked, at least almost! Here is the second response.

Answer 6 (second try)

To find the long run average amount of time that a voter will have to wait before casting their vote using queueing analysis, we need to consider the arrival rate of voters at the kiosk and the service rate of the kiosk (the rate at which voters are served by the kiosk).

The arrival rate is the rate at which voters arrive at the kiosk, which is 10 voters per hour. The service rate is the rate at which the kiosk can serve voters, which is equal to the inverse of the average time that a voter spends at the kiosk, which is $1/5$ voters per minute = 12 voters per hour.

The long run average number of voters in the system (waiting and being served) is given by the formula:

$$L = \lambda / (\mu - \lambda)$$

where λ is the arrival rate and μ is the service rate.

Plugging in the values, we get:

$$L = 10 \text{ voters per hour} / (12 \text{ voters per hour} - 10 \text{ voters per hour}) = 10/2 = 5 \text{ voters}$$

The long run average time that a voter spends in the system (waiting and being served) is given by the formula:

$$W = L / \lambda = 5 \text{ voters} / 10 \text{ voters per hour} = 0.5 \text{ hours}$$

This is the long run average amount of time that a voter will have to wait before casting their vote. Converting this to minutes, we get:

$$W = 0.5 \text{ hours} * 60 \text{ minutes per hour} = 30 \text{ minutes.}$$

Note that the answer is slightly off. As Chat GPT3 correctly computed, there will be 5 voters in the system (waiting and being served). So, the correct answer for the waiting time before casting the vote should be 25 minutes. But this comes close enough to a passing grade, say a solid B. Again, this is after receiving a hint – the initial answer was more in the C range.

One more note on this question: when I returned to Chat GPT 3 the next day asking the same initial queueing question (the one without the hint referencing queueing theory), Chat GPT3 immediately answered the question

correctly – a hint was no longer necessary. It either is capable of learning from past feedback or I just got lucky (as we saw above, there seems to exist some randomness in the quality of the answers).

Toyota Production System

In addition to questions involving calculations, some of my exam questions are more conceptual in nature. Often times, such conceptual questions are in a multiple choice format. So, my final question looks at the effect of using Kanban cards to move work through a process.

Question 7

Which of the following statements is correct?

- I. Deploying the Kanban system leads to pulling work through the system instead of pushing work.
- II. Deploying the Kanban system leads to pushing work through the system instead of pulling work.
- III. The Kanban system controls the work in process inventory.
- IV. Kanban will always reduce the amount of inventory in the process
- V. Kanban requires cross-training all workers.

The Toyota Production System with its focus on flow and just-in-time delivery is an important part of any operations course or book and the Kanban method is a key concept of it.

Kanban cards are a replenishment mechanism in a process where a resource (machine or operator) only produces something when there exists an explicit demand for it. This demand could come from the customer (external demand) or from the subsequent resource in the process (internal demand). Either way, work is now pulled through the process, and it is demand that drives the flow, not machine or operator availability. As we can see, Chat GPT 3 had no problems with this question.

Answer 7

The correct statements are:

- I. Deploying the Kanban system leads to pulling work through the system instead of pushing work.
- III. The Kanban system controls the work in process inventory.
- IV. Kanban will always reduce the amount of inventory in the process (if implemented correctly).

Kanban is a method for managing the flow of work in a production process. It is based on the principles of just-in-time production, which aims to minimize inventory and waste by only producing what is needed, when it is needed. In a Kanban system, work is pulled through the process as it is needed, rather than being pushed through the process regardless of demand. The Kanban system helps to control the amount of work in process inventory by using visual signals (such as cards or electronic displays) to indicate when work should be started and when it is ready to move to the next stage in the process.

Kanban does not necessarily require cross-training all workers, although it may be beneficial to have a more flexible workforce that can handle multiple tasks.

This is a really good answer. We can and should debate the validity of statement IV. When I wrote this question for my students, I intended this to be a “trick question” and the statement to be false (if one puts too many Kanban cards into the process, strictly speaking, inventory could go up). But it appears as if Chat GPT 3 saw through my evil intent and wisely added the clause “if implemented correctly”. The explanation is once again excellent, including the final statement about the usefulness of cross-training. A+.

Summary

The previous pages should convince even a skeptical reader that Chat GPT 3 has remarkable skills in handling problems as used extensively in the training and testing of our MBA students. Combining the results of the questions, I would grade this performance as a B to B-. The following reference point might help put this performance into perspective. Until Wharton allowed students more flexibility in which courses they take, this Operations Management course was a required course that every student had to take. However, we did allow students to waive this course if they could demonstrate content mastery on a waiver exam. The performance of Chat GPT3 reported above would have been sufficient to pass the waiver exam, though by a very small margin.

Part 2: Chat GPT3 Generating Exam Questions

After exploring to what extent my students might be better off with this technology, I also wanted to find out if I as a professor and author could take advantage of it. I know that only few readers (most likely fellow instructors) will pity me for this, but I face the challenge of coming up with new test questions every year and for every new edition of my book. By now, I have written 1000’s of questions and, at times, I feel I have exhausted my imagination for new problems.

Can I turn to Chat GPT3 to come up with new exam questions? To find out, I entered the first process analysis question (the Circored iron ore process) into the Chat GPT3 prompt alongside with the following instruction

Question 8 Create similar questions in the context of a coal mine.

Here is the response, i.e., the question generated by Chat GPT3.

Answer 8 In a coal mining operation, the production process starts with extracting coal from underground seams. The coal is then moved through six process steps before it emerges as finished product. The steps and their capacities are as follows: Crusher: 50 tons per hour Conveyor belt: 45 tons per hour Wash Plant: this plant holds up to 20 tons of coal and coal needs to spend 1 hour in the plant Dryer: holds up to 100 tons of coal and coal needs to spend 3 hours in the dryer Screening: 62.5 tons per hour Loading: 59 tons per hour Where in the process is the bottleneck and what is the overall process capacity?

This actually is a pretty good question. Chat GPT3 picked up the format and the logic of the iron process, but was able to transfer the question into the coal domain. I am by no means an expert on coal mining and processing, but these steps exist in coal plants and their sequence is at least plausible.

To replicate this for another question type and to add some humor to this paper, I also requested a waiting time question (see the voting operation discussed above), one that involves a very specific context.

Question 9

Generate three queuing theory questions about average waiting times in the context of Homer Simpson's day to day life.

Here is the response by Chat GPT3.

Answer 9

"At Moe's Tavern, Homer Simpson and his friends often wait in line for a drink. On average, customers spend 2 minutes at the bar and arrive at a rate of 20 customers per hour. The bartender takes an average of 3 minutes to serve each customer. What is the average waiting time for Homer and his friends before they are served a drink?"

"Homer Simpson spends a lot of time waiting in line at the Springfield DMV. On average, customers spend 5 minutes at the counter and arrive at a rate of 30 customers per hour. The DMV employees take an average of 10 minutes to serve each customer. What is the average waiting time for Homer before he is served at the DMV?"

"Homer Simpson likes to buy donuts at Lard Lad Donuts. On average, customers spend 1 minute at the counter and arrive at a rate of 15 customers per hour. The cashier takes an average of 2 minutes to serve each customer. What is the average waiting time for Homer before he is able to buy donuts at Lard Lad Donuts?"

Clearly, these are plausible questions and likely are more humorous than anything I could have produced myself (though, given that I am a German Operations Management professor, one might argue that this is not a high bar to clear). It is certainly good enough that I will take advantage of this technology in the future. Nevertheless, two shortcomings in these questions need to be pointed out.

First, the question really should provide only two numbers, the arrival rate (demand, 20 customers per hour in the first question) and the time to serve one customer (3 minutes in the first question). I have no idea why the additional number ("customers spend 2 minutes at the bar") is provided and how it should be interpreted in the context of the question.

Second, for queuing theory being the right tool for the job, the demand for the process (the rate of customer arrivals) should be somewhat close to the capacity of the process (the rate of service). Consider the second question. 30 customers arrive per hour, each of them requiring 10 minutes of work. Clearly, one DMV employee would not be able to deal with this load and multiple employees would be needed. Though the question generated by Chat GPT3 speaks of "DMV employees" (plural), it does not state how many there are, making it impossible to answer the question.

Part 3: Implications for MBA Programs and Faculty

I am not the first to speculate about the impact of Chat GPT3 on education. However, I propose that the impact of Chat GPT3 on business school education in general and Operations Management in particular goes beyond what its impact will be on teaching mathematics, history, biology, or literature. “Operations” has its roots in the Latin word “opus”, which stands for “work”. The purpose of my Operations Management class thus is about helping students analyze and improve the way people work, now and in the future. The science of biology has and will not change because of Chat GPT3. How people work, in contrast, is constantly changing as technology advances. Based on the B to B- performance of Chat GPT3 in my course and its ability to generate creative (though imperfect) questions for my future exams, I see the following implications for us as business school faculty.

Implication 1: Be mindful of what Chat GPT3 can and cannot do

The moment I saw the answer to my first question, I fell in love with Chat GPT3. I had used other natural language processing and AI software before, but this simple user experience and the great answer put me in a state of awe, and I am sure it has impressed many users before me.

But we should not forget that it made major mistakes in some fairly simple situations. Being off by a factor of 10x in the receiving station of Question 3 is below the academic performance of a middle school student. The average grade of Chat GPT3 was a B to B- in a domain that is well documented in thousands of pieces of knowledge that are accessible online.

We have many reasons to believe that the technology is getting better over time. But, we are still far from an A+ for complex problems and we still need a human in the loop.

Implication 2: Continue to teach the foundations

I am sure that there will be many calling for a change in course content making an argument of the type “if a computer can do it at zero marginal cost, a student should not need to spend time and money on mastering this skill” or “if a bot can pass the waiver exam of a course, clearly these skills should be removed from the curriculum!”. I have some sympathy for this argument. 35 years ago, as an undergraduate student in Germany, I learnt how to manually invert a matrix and how to solve simple optimization problems with nothing but pen and paper. After successfully displaying these skills on my final exam, I have never used them again.

But I would not go as far as making the claim that these skills were a complete waste of my time. In my view of education, an elementary school student still needs to learn that $7 \times 7 = 49$ and that the capital of Pennsylvania is Harrisburg, even though calculators have been widely used for over 50 years and students can use Google or Wikipedia to find answers for most factual questions. It is the nature of foundational skills that they are required to comprehend more advanced topics.

How does the increasing market volatility post Covid impact the suitability of a just-in-time supply chain? Should US manufacturers with most of their suppliers located in China embrace a dual sourcing strategy? To competently answer these questions one needs a solid understanding of the foundations of operations management. You need to be able to walk before you can run! So, business school faculty or elementary school teachers – as educators we still must teach the foundations.

Implication 3: Deal with the cheating when testing foundational knowledge

Many educators are interested in the Chat GPT3 discussion out of a concern that their students might be cheating on homework assignments and final exams. They should be. Though in the past I have had an open book, open notes policy for my exams and I allowed students to use computing devices, I will now join thousands of professors and teachers and explicitly ban the usage of Chat GPT3 and other technologies of this type for the purpose of homework assignments and the final exam in my foundations courses.

I realize that regulating exciting new technology is oftentimes perceived as a desire to hang on to the status quo. Nevertheless, allowing a student to access Chat GPT3 during an exam that tests facts and foundational concepts is like allowing the student to call a friend with an average academic competence and take the exam for her. Reliable tests play an important role in teaching and skill certification and this should not be compromised because of a new technology.

Implication 4: Mimic the workplace by teaching how to evaluate a proposed plan of action

MBA programs are professional degrees that prepare students for careers in the business community. Technologies like Chat GPT3 are already used in the workplace and that usage is only going to increase. Their ultimate goal is to improve managerial decision making. Managerial decision making involves creating a set of alternatives and then critically reflecting which alternative is the best for the situation at hand.

As our MBA graduates advance in their careers, they will make more and more of these decisions in groups where alternatives will be generated by consultants, co-workers, and direct reports. The skill of looking at a suggested alternative that is well presented and looks totally plausible and then being able to critically evaluate if the suggested alternative is fundamentally flawed or absolutely brilliant is thus among the most important skills in a management career. With Chat GPT3 playing the role of that smart consultant (who always has an elegant answer, but oftentimes is wrong) we thus have a perfect training ground for developing that skill. Just think back to Answer 3 (the Cranberry process). It was well presented and the numbers looked coherent and plausible – but, it was wrong nevertheless. By letting students use Chat GPT3 during case discussions, I can thus emulate the decision making process a senior executive would face in the work place.

Implication 5: Let students use Chat GPT3, but simultaneously raise the bar for assignments

The scenario that most K12 teachers are concerned about is that a student who in the past would go to the library and spend four hours on a homework assignment contrasting the views of Albert Camus and Jean-Paul Sartre on existentialism is now “getting away” with a five minute interaction with Chat GPT3. As a result of this “shortcut” the student learns less than before. If I would be a K12 teacher, I would be concerned about my students taking such shortcuts as well.

But, I am not a K12 teacher. I have the privilege of teaching highly motivated students, most of whom are making a substantial personal and financial sacrifice to be in my classroom. With or without Chat GPT3, I can get a certain amount of time from them each week (say four hours). It is now up to me to come up with assignments that are challenging enough so that they warrant that time investment. To the extent that we believe that Chat GPT3 gives the students a head start on their homework, it is my job to hold them accountable to a higher standard.

In many ways, this is similar to how we have been using group assignments for many years. Just as we would expect a better deliverable for an assignment that was given to a group of five students, we should expect more from a student that we encourage to collaborate with a technology such as Chat GPT3.

Implication 6: Ask students to imagine the new rather than tweaking the old

In his book “Zero to One”, Peter Thiel challenged entrepreneurs to move beyond the existing paths of innovation. Thiel distinguishes between “going from 0 to 1”, which corresponds to creating something fundamentally new (Bill Gates coming up with a PC operating system) and improving upon what exists by going from “1 to n” (moving from Windows 10 to Windows 11).

The knowledge of Chat GPT3 is inherently built on synthesis. Even when we ask it to engage in creative problem solving (recall Question 7), Chat GPT3 will always stay in a solution space that is defined by what it has seen in the past.

Outstanding business ideas, in contrast, oftentimes go beyond optimizing what is and move to imagining what could be. For example, Chat GPT3 and its successors will likely excel in finding the optimal delivery path for a truck that has to make a given number of shipments to a given set of addresses. But will it be able to question the problem? Will it challenge the mode of delivery? I appreciate the need to teach our students how to find the shortest path through a network. But wouldn't the A+ student come up with an idea such as a strategically positioned locker that customers walk to in order to pick up their shipment or use a drone for delivery? It is up to us as educators to reward, if not require, such out of the box thinking.

Implication 7: Don't be shy using Chat GPT3 to improve the productivity of the teaching process

Nine years ago, I was among the first faculty to make my MBA course content available to the general public by creating a MOOC. Video based teaching allowed for enormous increases in teaching productivity (Terwiesch and Ulrich 2014), something that at first, many of my students and faculty colleagues were skeptical about. The strategic decision, of course, was who should benefit from that higher productivity. I now use these videos to help students who have missed class or are struggling for other reasons. This allows me to customize learning in a way I could not do before. Using this video content has resulted in improved learning and has reduced student anxiety. It has also increased test scores.

Recall my experiment with having Chat GPT3 create exam questions for me. It normally takes me about 20 hours of work to create an exam and another 10 hours for TA's to test the exam and write solutions to it. Now, I think we could get my exam writing time down to 10 hours and the TA time down to 5. In other words, we have a 100% productivity increase in the “exam writing operation”. It is easy to imagine similar magnitudes of improvement in grading processes, tutoring, and office hours.

It is now up to us to determine what to do with this increased productivity. In my view, we should return it to the students in the form of extra meetings outside class, personal attention, joint social activities, or the design of new course materials.

Afterthought

Given my new relationship with Chat GPT3, I can't help but turn to the computer one last time and ask my new digital friend what it thinks its impact on MBA education will be the in the future.

Question 10

What will be the impact of the Chat GPT3 technology on how MBA students are taught. Specifically, how will Chat GPT3 impact MBA education differently from K12 education?

Look at the answer and draw whatever conclusion you like...

Answer 10

Chat GPT-3 technology has the potential to greatly enhance MBA education by providing students with personalized, interactive learning experiences. It can be used to generate personalized study materials, provide real-time feedback, and facilitate online discussions and group projects. Additionally, Chat GPT-3 can be used to simulate real-world business scenarios, allowing students to practice decision-making and problem-solving in a safe, virtual environment.

The impact of Chat GPT-3 on MBA education may be different from its impact on K-12 education because MBA students are typically older and more experienced, and are often working professionals. They may have different learning styles and needs than K-12 students, and may require more advanced, specialized content. Additionally, MBA programs often involve a greater emphasis on practical, real-world applications, which Chat GPT-3 may be particularly well-suited to support.

References

Terwiesch, Christian, Karl T. Ulrich, "Will Video Kill the Classroom Star? The Threat and Opportunity of MOOCs for Full-Time MBA Programs", White Paper, Mack Institute 2014

Miller, Jeffrey G. "National Cranberry Cooperative." Harvard Business School Case 675-014, August 1974. (Revised November 1983.)

Terwiesch, Christian, Christoph H. Loch, "Pumping Iron at Cliffs and Associates: The Circored Iron Ore Reduction Plant in Trinidad", Case of the Wharton-INSEAD Alliance, 2004

Cachon, Gerard, Christian Terwiesch, *Matching Supply with Demand: An Introduction to Operations Management*, 4th edition, McGraw Hill, 2018

About the Author

Christian Terwiesch is the Andrew M. Heller Professor at the Wharton School of the University of Pennsylvania. He is a Professor in and the chair of Wharton's Operations, Information, and Decisions department, co-director of Penn's Mack Institute for Innovation Management, and also holds a faculty appointment in Penn's Perelman School of Medicine. His research on Operations Management and on Innovation Management appears in many of the leading academic journals ranging from Management Science to The New England Journal of Medicine. He is an award winning teacher with extensive experience in MBA teaching and executive education.

Professor Terwiesch is the co-author of *Matching Supply with Demand*, a widely used text-book in Operations Management that is now in its third edition. Based on this book, Professor Terwiesch has launched the first

Massive Open Online Course (MOOC) in business on Coursera. By now, well over half a million students enrolled in the course.

His first management book, *Innovation Tournaments*, was published by Harvard Business School Press. The novel, process-based approach to innovation outlined in the book was featured by *BusinessWeek*, the *Financial Times*, and the *Sloan Management Review* and has led to innovation tournaments in organizations around the world. His latest book, *Connected Strategies*, combines his expertise in the fields of operations, innovation, and strategy to help companies take advantage of digital technology leading to new business models. The book has been featured as the cover story of the *Harvard Business Review* and has been featured by *Bloomberg / BusinessWeek* as one of the best books in 2020.

Professor Terwiesch has researched with and consulted for various organizations. From small start-ups to Fortune 500 companies, he has helped companies become more innovative, often by implementing innovation tournament events and by helping to restructure their innovation portfolio. He holds a doctoral degree from INSEAD and a Diploma from the University of Mannheim.