

Gen AI and Business School assessments – stress testing to identify opportunities and challenges

Cloda Jenkins (Associate Dean Education Quality)
Sean O’Grady (Lead Learning Designer, IDEA Lab)
Peter Atkinson (Senior Learning Designer, IDEA Lab)

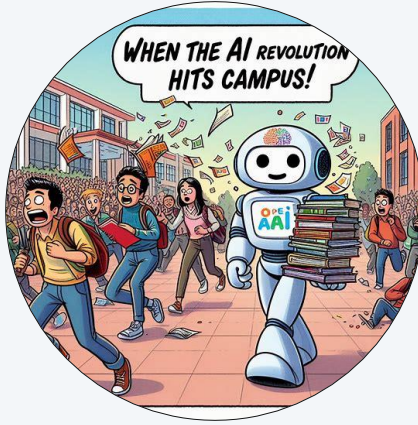
Imperial Festival of Teaching and Learning
Wednesday 1st May 2024

Plan for today

1. What is the Business School's approach to GenAI?
2. How did we stress test our assessments?
3. What did we learn about running an exercise like this?
4. What did we learn about our assessments and GenAI?
5. How to keep the momentum going?



November 30 2022



Panic instinct?



Innovation

Entrepreneurs

Duty to our students

Note: Images were generated by Microsoft Co-Pilot, as well as reminders of key date

Embrace value and manage risks

Learn by doing



Academic integrity guidance for students

- Written
- Video

Guidance for teaching staff

- Written
- Presentations
- Teams Site

Faculty using

- Teaching
- Assessments
- Research

Assessment Stress Test

- Education Quality
- IDEA Lab
- Departments

Work with our AI experts, Imperial Working Group, library, colleagues other Departments, discipline networks, Business School networks, Student Shapers

Established a
process and
buy-in to it

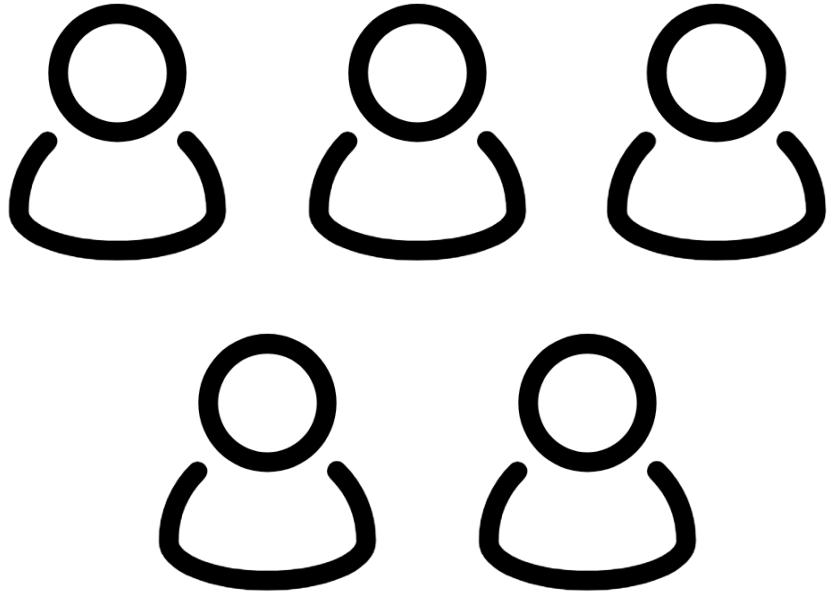
Resourced by School

Stress Test Plan

1. IDEA Lab and Education Quality teams jointly designed a clear process of how Stress Test would work
2. Got buy-in from Senior Leadership in Business School
3. Faculty were informed of plan and given option to opt-out from Stress Test (only one did)
4. Documented all modules, who taught them, what assessments were and links to their location on VLE
5. Established team of Stress Testers to execute the plan, from August to Jan bulk of work

How we did it

Who did the testing?



- Five 'naïve' testers

What tools
did we use?



ChatGPT



Claude



Bard



Gamma



Scite.ai

How did we work individually?



- Each tester assigned a number of modules / assignments to work on
- Recorded results on a form later to be shared with faculty

The process



1. Record the 'conversations' in full
2. Summarise our findings (approach, difficulties reflections)
3. Rank susceptibility to AI according to six criteria

The template

☰ ≡ +

Response analysis: ChatGPT-4

Assessment task	
Prompt	
Response	
Prompt	
Response	

Summary: Assessment 2

How did you approach the task / interact with the AI to find answers?	
Difficulties / errors encountered	
Any other reflections / comments	

Scoring (naive)

1 (low susceptibility to AI) to 10 (high susceptibility to AI)

Accuracy
(did the AI provide information that is factually correct and up to date?)

[1-10]

Clarity
(Was the AI's response coherent, logically developed and free from unnecessary repetition?)

[1-10]

Relevance
(Did the AI address the question effectively? i.e., did it directly address the central question or topic?)

[1-10]

Compliance
(Did the AI follow the instructions given in the task?)

[1-10]

Referencing
(Was the AI able to provide accurate and appropriate references when required in the response?)

[1-10]

Ease-of-use

[1-10]

(Was the AI tool straightforward and efficient to use in completing the assessment?).

Average score

Examples

Prompt	<div>I am about to ask your four questions related to an economics assignment. I will give you the background info and then each question at a time., starting with question a. Please provide answers for each question in turn.</div> <div>Question 1. You manage Live Nation Entertainment, the firm that holds the monopoly of ticket sales for concerts. You determine that the total demand for</div>									
	<div>concerts in the US is reasonably described by two segments, Seniors (> 70 years of age) and the rest of the population (General Public). There are 100 identical locations, and, in each location, demand is as follows:</div> <table><thead><tr><th></th><th>General Public</th><th>Seniors</th></tr></thead><tbody><tr><td>Off-peak (Mon-Fri)</td><td>1,400 – 10P</td><td>1,200 – 20P</td></tr><tr><td>Peak (Sat-Sun)</td><td>3,400 – 10P</td><td>1,000 – 20P</td></tr></tbody></table> <div>P denotes the price in \$ of the ticket. The marginal cost of an additional person in the audience is zero, but your 100 identical venues only have a capacity of 1,400 each.</div> <div>(a) Can you charge a different price to different segments? If yes, how? [5 marks]</div>		General Public	Seniors	Off-peak (Mon-Fri)	1,400 – 10P	1,200 – 20P	Peak (Sat-Sun)	3,400 – 10P	1,000 – 20P
	General Public	Seniors								
Off-peak (Mon-Fri)	1,400 – 10P	1,200 – 20P								
Peak (Sat-Sun)	3,400 – 10P	1,000 – 20P								
Response	<div>Yes, you can charge different prices to different segments in this scenario, a strategy known as price discrimination. Price discrimination is the practice of charging different prices to different groups of consumers for the same product when the price differences are not due to differences in cost. Here's how you can implement it:</div> <div><div>1. Identify the Segments: In your case, the market is already segmented into Seniors and the General Public. This segmentation is based on age, a clear and identifiable criterion.</div><div>2. Ensure Segment Separation: To successfully charge different prices, it's crucial that the segments are separated and that one segment cannot access the price offered to the other. This can be done by requiring proof of age for seniors to access their</div></div>									

Summary: Assessment 1	
How did you approach the task / interact with the AI to find answers?	<p>I first set the context for ChatGPT: that it should imagine itself as a student at a business school studying on this module. I explained that it's been given a group assignment, and I attached screenshots of all the assignment instructions and asked it whether it understood. It explained in detail what it had to do, and the summary was accurate, so I then gave it some more useful info for completing the task: the marking criteria and also the module reading list in case that helped orient its discussion or was useful for providing references in the report.</p> <p>I then asked for an outline of the executive report, with breakdown of each section and word counts for each. It got the word count <u>total</u> wrong, so I had to ask it to tweak, before eventually moving onto writing each section (or half of each section for the longer ones) at a time.</p> <p>Once the report was complete I asked it for the second task: the executive summary in PPT form.</p>
Difficulties / errors encountered	<p>There were quite a few issues encountered, or limitations in what ChatGPT provided for this assignment :</p> <ul style="list-style-type: none"> - It didn't provide citations/references at all, despite me mentioning this near the start of the conversation. It did occasionally bring in theories/frameworks that were included in the module reading list (e.g. Blue Ocean Strategy). I could have reminded it of this while it was producing the sections, but given ChatGPT's tendency to hallucinate sources anyway, it would probably make more sense for a student to incorporate these later, if they were using it in a clever way rather than just copying and pasting. - A lot of the writing it produced was quite vague, and the tendency to split everything up into very small chunks in its answers leads to a sense of it covering a lot of ground but never in much serious detail. - It struggled to produce the right amount of words for individual sections (often too few words near the second half of the report's sections). - Overall I think it struggled to produce something of this length without starting to forget instructions I had provided it earlier on in the conversation. I think it lost a sense of what the overall scenario was at some points and I'm not sure the report all held together as a coherent whole. - The content it suggested for the PPT slides for the exec. <u>summary</u> task was quite vague and limited. <p>For all these reasons, I've scored it fairly low below.</p>

Prompt	Before moving on to step 3, can you present the strategy canvas in visual form?
--------	---

Response

Strategy Canvas for Coca-Cola

Key Competitive Factors	Current Position	Shifted Position
Product Quality	5	5
Brand Image	5	5
Price	3	3
Distribution Network	5	5
Sustainability Efforts	2	4

Here's the visual representation of the Strategy Canvas for Coca-Cola:

- The solid line represents Coca-Cola's current position across key competitive factors like Product Quality, Brand Image, Price, Distribution Network, and Sustainability Efforts.
- The dashed line illustrates the envisioned shifted position after implementing the new strategy, particularly highlighting an improvement in Sustainability Efforts while maintaining strengths in other areas.



Step 4: Rebranding Coca-Cola for the Target Market

Rebranding Approach

- **Eco-Friendly Visual Identity:** Introduce a new visual branding that emphasizes Coca-Cola's commitment to sustainability. The new design integrates green accents and symbols like leaves to signify eco-friendliness.
- **Sustainable Packaging Design:** Use recycled materials for packaging, showcasing this feature on labels and advertising materials. The bottle design,

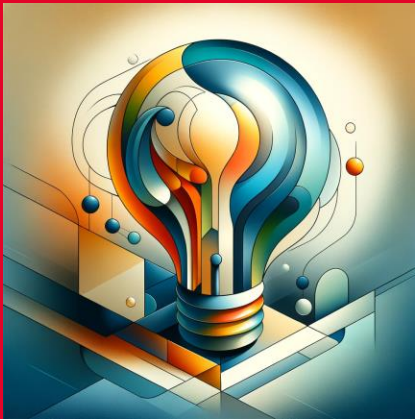
made from recycled glass or biodegradable materials, reflects this sustainable approach.

How did we work as a group?



- MS Teams group
- Regular catch-ups
- Shared Google docs
- Team members with different backgrounds
- Spread the load

Lessons learned on how to run a stress test exercise



- It takes time to do it effectively
- Working in a team of testers brings many benefits
- Assign all modules from one faculty member to one tester
- Try to standardise approach to scoring
- Be aware of costs of using multiple tools
- Ensure testers know how to get the most out of the tools, e.g.:
 - Use image/file upload functionalities to save time
 - Use multiple tools for different elements
 - Provide the tools with as much info as possible
 - Structure the tools' response

What did we learn?

Faculty response - reports

“Thank you very much for sending us the stress test results of our course. I had a quick look through. The test results are very interesting.”

“This is a great initiative – well done”

“Can you do the same for my summer module?”

In general, faculty were:

- Appreciative we had taken time to test their assessment
- Engaged with the results
- Occasionally surprised by the AI performance



Faculty response - workshops

Workshops with faculty from related subject areas told us that:

- Stress testing and reports were a catalyst for discussion
- Benefit from sharing ideas with peers
- Faculty keen to embrace AI
- Some already incorporating AI into assessment e.g. use GenAI for certain aspects
- Concerns about consistency / expectations
- Concerns about equity / access



Implications for assessment design

General findings

1. Current approach to assessment must change
2. Need to focus on what AI will do in near future, not what it can do now
3. Lots of potential for AI to support assessment process in legitimate ways
4. Need to develop 'AI literacy' – staff and students

Implications for assessment design



- Open book quiz (especially MCQ)
- Short/medium-sized writing task e.g. open question/prompt
- Basic calculations/coding



- Written tasks with multiple parts e.g. reports
- Reflective tasks / peer-reviews
- Maths/coding with explanation or specific approach required



- In-class presentations with Q&A/Video presentation
- Tasks requiring reference to specific module content e.g. lectures, discussions...
- Large/complex writing tasks requiring multiple related parts e.g. a dissertation

Approaches to assessment redesign

- Change weighting
 - Reduce weightings for more at-risk tasks (quizzes, short writing tasks)
- Change criteria
 - More emphasis on what AI cannot do e.g. Q&A performance vs slide content
- Change process
 - Incorporate some use of AI
 - See drafts/development of ideas
 - Require reference to specific module materials
- Change task
 - Written task > Presentation
 - Multiple, unrelated tasks > More complex, related tasks

Much still to learn and do

Keeping up

1. Stress test more assessments if Faculty want
2. Work with Faculty who are keen to embed in teaching
3. Work with Faculty who are keen to adapt assessments
4. Consider how to embed at programme level
5. Support students to use effectively, ethically and with integrity
6. Training for all staff – education + wider uses
7. Pilot AI tools for Student Support and Feedback and ?
8. Work out how best to keep sharing ideas and experiences
9. Review Guidance and Academic Integrity Processes
10. Keep talking widely (and reading all those blogs)

Work with our AI experts, Imperial Working Group, GenAI Futurists, library, colleagues in other Departments, discipline networks, Business School networks, Student Shapers

Any questions?

**Imperial College
Business School**

Imperial means
Intelligent Business

