



www.ijtes.net

Figure-Based Approach in Creating ChatGPT-4o-Resistant Multiple-Choice Questions for Introductory Biology Courses: An Instructional Guide

Kyeng Gea Lee 
Bronx Community College of The City University of New York, U.S.A.

Mark J Lee 
Duke University, U.S.A.

Soo Jung Lee 
Eastchester High School, U.S.A.

To cite this article:

Lee, K.G., Lee, M.J., & Lee, S.J. (2024). Figure-based approach in creating ChatGPT-4o-resistant multiple-choice questions for introductory biology courses: An instructional guide. *International Journal of Technology in Education and Science (IJTES)*, 8(4), 689-709. <https://doi.org/10.46328/ijtes.589>

The International Journal of Technology in Education and Science (IJTES) is a peer-reviewed scholarly online journal. This article may be used for research, teaching, and private study purposes. Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material. All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations regarding the submitted work.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Figure-Based Approach in Creating ChatGPT-4o-Resistant Multiple-Choice Questions for Introductory Biology Courses: an Instructional Guide

Kyeng Gea Lee, Mark J Lee, Soo Jung Lee

Article Info

Article History

Received:

13 August 2024

Accepted:

01 October 2024

Keywords

ChatGPT-4o

Figure-based

Multiple-choice

Online assessment

Biology education

Abstract

Online assessment is an essential part of online education, and if conducted properly, has been found to effectively gauge student learning. Generally, text-based questions have been the cornerstone of online assessment. Recently, however, the emergence of generative artificial intelligence has added a significant challenge to the integrity of online assessments. In particular, it has been reported that large language models, like ChatGPT-4o, show high performance on text-based questions. In comparison, ChatGPT-4o exhibited significantly reduced performance on figure-based questions in our study. In an effort to counter the recent encroachment of ChatGPT-4o into online assessment, we propose a step-by-step instructional guide for a method in creating figure-based multiple-choice questions that are resistant to ChatGPT-4o. This involves generation of a ChatGPT-4o-resistant figure, writing the question text based on it, and evaluating the final question on ChatGPT-4o. If successfully created, ChatGPT-4o response could be subject to random guessing. Our results showcase four representative examples for introductory biology courses and illustrate a systematic approach to compose questions based on qualitative analysis of ChatGPT-4o responses. In combination with other assessment methods, our method aims to serve as a tool to alleviate the current challenge that educators face for online assessments.

Introduction

Online assessment of student learning is regarded essential in online learning (Muzaffar et al., 2021) and has gained popularity in both distance and in-person contexts of education (Aristeidou et al., 2024). During the COVID-19 pandemic, adoption of online assessment became a matter of necessity rather than a possibility (Patael et al., 2022). A study conducted during this period demonstrated that individual student performance on unproctored online exams highly correlated to that of in-person exams, supporting the reliability of online assessments (Chan & Ahn, 2023). In accordance with this finding, proper design of online assessment, either formative or summative, is vital in ensuring effective online learning (Yildirim et al., 2023).

Online assessment, mainly in the form of summative examination, is no longer a matter of necessity in the post-pandemic era. However, Dawson et al. (2024) projected that it is still likely to persist since many instructors have adopted it and would continue employing the method. Their study concluded that the traditional open book versus

closed book restriction is not applicable for online examinations due to technological advances in the web that now includes generative artificial intelligence (AI). To promote the validity of online assessments, the study proposed a revision of exam restrictions and designs that regulate access to information, people, and tools, all of which can be available through the web. It must be noted, however, that such revision may not be limited to online assessments, but also need to extend across the general spectrum of educational field due to concerns on the integrity of assessments stemming from the influence of generative AI (Farazouli et al., 2024; Lee et al., 2024; and Bin-Nashwan et al., 2023).

In particular, Susnjak and McIntosh (2024) concluded that the emergence of generative AI based on large language models (LLMs) such as ChatGPT is posing a significant challenge to the integrity of online exams. Their study demonstrated that advanced reasoning capability of ChatGPT-4 can be invoked through an iterative self-reflective strategy which is based on sequential prompts. This strategy can guide ChatGPT-4 to engage in critical thinking and ultimately arrive at the correct responses on multimodal (visual and textual) exam questions of complexity. Therefore, ChatGPT would not just assist in information recall, but can also act as a surrogate thinker for higher order questions that require critical thinking. Given such capability, Lo et al. (2024) reported that students may develop a dependency on ChatGPT which can in turn hamper their productivity and intellectual growth and this aspect would act as a threat component based on the SWOT (strength, weakness, opportunities, and threats) analysis. It is evident that such a threat would become more intense with advances in the capabilities of newly released AI models. Lately, this trend is progressing with speed as documented by increasing number of studies that report high achievement of ChatGPT on some of the standard examinations of higher education. For instance, a scoping review (with a literature search cut-off date of July 20, 2023) reported that ChatGPT-4 passed most examinations of higher education that are based on multiple-choice questions (Newton & Xiromeriti, 2023). The majority of these were from medical subjects that included specialty/board examinations, medical licensing examinations, and medical school admissions examinations, with ChatGPT performing at a level similar to human subjects. Since the review was conducted prior to the release of image capable version of ChatGPT-4 (September 25, 2023; OpenAI, 2024A), the scope of this study was on text-based multiple-choice questions.

As of the submission date of our study, ChatGPT-4o (omni) is the "latest, fastest, and highest intelligence model" of OpenAI with improved capabilities over ChatGPT-4 to process text, image, and audio inputs and outputs (OpenAI, 2024A; OpenAI, 2024B). Literature search on ChatGPT-4o and multiple-choice question yielded three preprint articles that were relevant to our study. All three articles evaluated the performance ChatGPT-4o on medical licensing examinations that included figure-based questions. Miyazaki et al. (2024) investigated the performance of ChatGPT-4o on the entire content of a 2024 version of Japanese Medical Licensing Examination. The study reported correct response rates of 93.48% for image-based questions and 93.18% for text-based questions, with no significant statistical difference in the response between the two question formats.

On the other hand, Gajjar et al. (2024) evaluated the performance of ChatGPT-4o on the United States Medical Licensing Examination (USMLE) and found that the rates of correct response for image-based questions were 70.8% for Step 1, 92.9% for Step 2, and 62.5% for Step 3. For text-based questions, the rates were near or above 90% for each of the Steps. This finding is in agreement with a study by Newton et al. (2024) who reported high

performance of ChatGPT-4o on the United Kingdom Medical Licensing Examination and the USMLE, but slightly reduced performance on image-containing questions.

We also have found that figure-based exam questions reduce the accuracy of ChatGPT-4 and -4o compared to text-based questions alone. In addition, we have discovered a number of patterns on such questions that can lead to incorrect responses by ChatGPT-4o. To date, there is no study that addresses how figure-based multiple-choice questions can be systematically created that can increase the rate of incorrect responses by ChatGPT-4o.

In this study, we incorporate our findings and provide a detailed instructional guide for a method to create such questions. The key component of our guide is to use the response of ChatGPT-4o as part of the tool to create the question and this aspect will be fully illustrated. It will showcase four question type examples with qualitative analysis of responses by ChatGPT-4o in the question creation process. The significance of our finding and its applicability will be explored in the context of biology education in the discussion. Successful application of the method can contribute in preserving the reliability and value of online assessments.

Method

Our study presents both an instructional guide for the question creation method and a qualitative analysis of the results produced in the process. To reflect that and keep the methodology concise, this section is formatted to follow an instructional guide in creating ChatGPT-4o-resistant questions. The guide is an expansion of four major steps summarized as below:

- Step 1: Design and generate a question figure with label(s). This step is subdivided into four parts.
- Step 2: Evaluate the figure in ChatGPT-4o.
- Step 3: Compose question text based on the evaluation result of the figure in step 2.
- Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

Step 1: Design and generate a question figure with label(s).

Step 1a: Generate the base image.

The most challenging aspect of creating figure-based question is the availability of unlabeled base images. Textbook publishers generally provide unlabeled images as part of instructor test resources or PowerPoint (Microsoft Corporation) presentation files. Instructors can easily access those resources or assets, and we recommend accessing unlabeled images directly from publishers whenever possible. However, if a desired image in the textbook is not provided with an unlabeled version, the source figure (Figure 1A; Figure 19.9 from Betts et al., 2022) may be modified using imaging software (e.g., Photoshop or GIMP) to manually remove the label lines and other items (Figure 1B) that are merged (burnt) into the figure layer. We used GIMP with Resynthesizer plugin (<https://github.com/bootchk/resynthesizer>) to generate the unlabeled versions in this study. While modifying labeled image requires manual processing, it is viable option if there are no other alternatives.

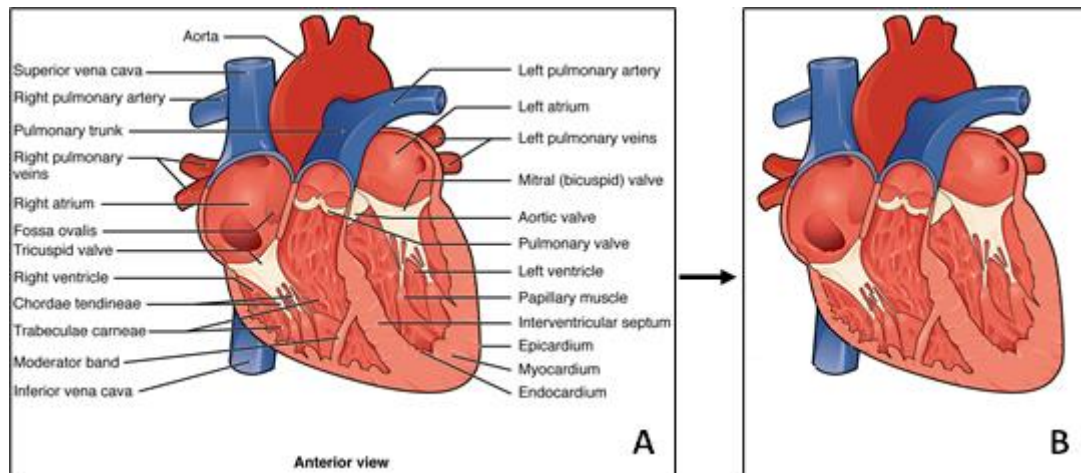


Figure 1. A) Source Image; B) Base Image with Labels Removed

[Unmodified (A) and modified (B) from Figure 19.9 in *Anatomy and Physiology 2e* by OpenStax (Betts et al., 2022); Unmodified source image: Access for free at openstax.org.]

Step 1b: Apply image enhancements.

Import (or simply copy and paste) the acquired or generated base images into a software application with image manipulation capabilities, like Microsoft PowerPoint, for general modifications such as resizing, cropping, and contrast enhancements. Images should be processed to a width and length resolution range that does not distort or blur out important details. We used a resolution range of 300 x 300 pixels. In PowerPoint, the image size is generally indicated in inches or centimeters, but it allows the option of entering the desired dimension followed by "px" for pixels in the image size boxes.

Step 1c: Add labels to the base image.

Place the desired labels to the enhanced image. The most important part of this step is to eliminate any textual cue that can lead to identification of the labeled structures. Single letter or number labels can be used, but words that have no relevance to the identification of the structure may also be a possibility. Black label lines and letters with white shadows can enhance visibility in the figures. In this study, we generated figures with a single-label (X) or four labels (W, X, Y, and Z). However, the number and style of labels to be used will depend on the scope of the question and be at the discretion of the instructor. The basic strategy to add the labels is illustrated in the Results and explained in the Discussion.

Step 1d: Merge the layers of the figure by saving it into png or jpg formats.

It is important that the layers of the labeled figure are merged into a single layer. While not apparent in PowerPoint, it is possible that the labels placed outside the base-figure area may not be visible (because of layering) to ChatGPT-4o if the figure is directly copied and pasted into its prompt box. To prevent this, the layers of the figure in PowerPoint should be flattened (merged) by simply copying and pasting the entire figure onto an imaging application such as Paint in PC or Preview in macOS. Both applications are part of the operating systems and

freely available. Using either of these, the figure can be saved into an image file (png or jpeg image formats) at the resolution that maintains the original figure dimensions. For Paint this may be the native resolution while for Preview it may be 96 pixels/inch. This figure can then be uploaded into the ChatGPT-4o prompt box for figure and question evaluations in the subsequent steps, or deployed as part of a question in learning management systems (e.g., Blackboard, D2L Brightspace, Canvas, and Google Classroom).

Step 2: Evaluate the figure in ChatGPT-4o.

To assess if ChatGPT-4o is able to identify the label(s) correctly, upload the labeled figure prepared as above to ChatGPT-4o with the following prompt: "*What do you see in this figure?*". ChatGPT-4o will generally respond by describing the figure and identifying the given label(s). While ChatGPT-4o may describe the figure fairly accurately, the key factor is the identification of the labels. In the event that ChatGPT-4o correctly identifies the label in a single-label figure, change the label and repeat the figure evaluation. If it misidentifies the label, then the figure meets the prerequisite for step 3. For figures with multiple labels, it is not necessary that ChatGPT-4o misidentifies all the labels. There are ways to use partially misidentified labels, which will be addressed in Step 3, below, and in the Result section. It is possible that ChatGPT-4o correctly identifies the labels of certain figures regardless of label modification. A recommended rule would be to abandon figures that fail the figure evaluation more than three times even after label modifications.

Step 3: Compose question text based on the evaluation result of the figure in step 2.

The general strategy is to minimize any contextual cue in the stem of the question text, and to add distractors (incorrect choices) that conform to the figure label misidentified by ChatGPT-4o. For figures with a single-label that is misidentified by ChatGPT-4o, one of the incorrect choices should be written to accommodate this misidentification to lead ChatGPT-4o to select that choice. For figures with multiple labels that are completely misidentified by ChatGPT-4o, assuming a multiple-choice question with four choices, three of the choices should be written to accommodate the misidentification such that ChatGPT-4o will incorrectly select it, while one of the choices, the actual correct answer, needs to disagree with the misidentification so ChatGPT-4o does not select. For figures with multiple labels that are partially misidentified, the general rule would be to create as many incorrect choices as possible that accommodate the misidentification. While it is a bit more challenging to create ChatGPT-4o-resistant questions with partially misidentified labels, by crafting strong question text and answer choices, it can be overcome. We outline some strategies through several examples in the Result section, and provide relevant rationales in the Discussion section.

Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

To check whether ChatGPT-4o could answer the question correct, evaluate the final question by uploading the labeled figure followed by the question text. ChatGPT-4o generally responds with some explanations on the subject matter and selects its answer. In other instances, however, it may choose multiple answers. In such cases, a second prompt is required to force it to commit to one answer: "This is a multiple-choice question. There should

be one correct answer." The response can be noted as the final answer. The evaluation should be repeated according to the number of choices to assess reproducibility. For instance, all multiple-choice questions in this study consist of four choices and, therefore, the evaluation consisted of four trials. When evaluating the questions, it is important to include the result of all the trials. If the number of correct responses is more than one in the four trials (i.e., greater than 25%), the question should be revised by altering the figure labels, question texts, or both (generally to a limit of two times), and be reevaluated. The rationale for evaluating with repeating trials will be addressed in the discussion.

Creation of Multiple-Choice Questions without Following the Instructional Guide

In order to ascertain the effectiveness of the instructional guide for the method presented above, three multiple-choice questions were generated using the same base images as the questions that were created following the instructional guide. However, labels were added without following the key strategies of Step 1 for label placement that are illustrated in the results and explained in the discussion for ChatGPT-4o-resistance. The most critical deviation in all cases was the omission of figure evaluation in Step 2. Since question text composition in Step 3 is dependent on Step 2, the strategies of Step 3 could not be used. For consistency, however, the question text was derived from the questions created following the guide with minimal modification, if needed, to make the new questions valid with one correct answer.

Note on the Copyright of the Images in This Study

The images included in this study are modifications (adaptations) of Figures 23.2, 13.12, and 19.9 in *Anatomy and Physiology 2e* by OpenStax (Betts et al., 2022; <https://openstax.org/details/books/anatomy-and-physiology-2e>). This book is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0/>) which permits to share (copy and redistribute the material in any medium or format for any purpose, even commercially) and to adapt (remix, transform, and build upon the material for any purpose, even commercially). Unmodified source images: Access for free at openstax.org.

ChatGPT-4o Settings

In evaluating the figures and the final questions, each trial was carried out as separate conversation with the "memory" setting off, a feature that allows ChatGPT-4o refer to other conversations. This was to prevent ChatGPT-4o from cross-referencing previous figure evaluations that may influence its response. Also, the setting to "Improve the model for everyone" was kept off so that any entry made as part of question composition would remain as a query alone (i.e., to prevent incorporation of the question content into the language model).

Data Recording and Supplementary Material

The responses of ChatGPT-4o in both figure evaluation and in final question evaluation were recorded as saved webpages in PDF format. It must be noted that the embedded figures appear cropped in the webpages by current

settings of OpenAI and they have been saved into PDF as such. However, ChatGPT-4o was queried with the full-size figure and had access to all parts of the figure in responding to the given prompts. Full-size figure can be retrieved by clicking onto the figure in the original webpage of the conversation. The PDF files of all recorded ChatGPT-4o responses in this study have been merged into a single PDF and referenced as figures in the supplementary material. The supplementary material can be accessed via:

https://1drv.ms/b/s!Amlc0GNnQ0y_c9XZtsS85GXSado

Results

Following the prescribed 4-step instructional guide, four ChatGPT-4o-resistant figure-based questions were generated, each depicting a specific approach. These are based on some of the essential topics in a typical introductory biology course and examples of exam questions that students could encounter. The questions were generated to showcase the steps and strategies involved in this instructional guide.

Composition and Analysis of a Single-Label Question: The Duodenum Question

Step 1: Design and generate a question figure with label.

The question figure was generated following the steps of the guide for a single-label figure. For this example, original labels were removed using GIMP with Resynthesizer plugin from the figure source (Figure 23.2; Betts et al., 2022), resized, and a single label was placed using PowerPoint. Label X in the figure is pointing at the duodenum within the selected region of the digestive system (Figure 2).

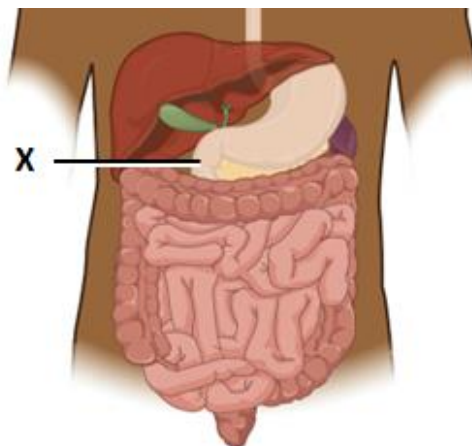


Figure 2. The Duodenum Figure

[Modified from Figure 23.2 in *Anatomy and Physiology 2e* by OpenStax (Betts et al., 2022); Unmodified source image: Access for free at openstax.org.]

Step 2: Evaluate the figure in ChatGPT-4o.

Once the figure was produced, it was queried to evaluate whether ChatGPT-4o is able to recognize the label. The

result of the figure evaluation is summarized in table 1.

Table 1. Summary of the Duodenum Figure (Figure 2) Evaluation

Prompt	What do you see in this figure?
ChatGPT-4o Response (abbreviated)	The figure shows a diagram of the human digestive system. The structures visible in the diagram include: ... The label "X" is pointing to the pancreas.
Figure Evaluation	While it described the details of the figure fairly accurately, it failed to identify structure X correctly. Since it misidentified the label, the figure was regarded satisfactory for the next step. (Evaluation record: Table S1 and Figure S1 of the supplementary material)

Step 3: Compose question text based on the evaluation result of the figure in step 2.

Using the ChatGPT-4o response in figure evaluation from Step 2, the following question text was composed:

The duodenum question:

Which statement is correct about structure X in the figure?

It produces bile to facilitate lipid digestion.

It secretes hydrochloric acid.

It secretes pancreatic juice for chemical digestion.

It's part of the alimentary canal.

Since ChatGPT-4o misidentified X as the pancreas, a reinforcing/gratifying distractor (third choice) was added that described an essential function of the pancreas (by actually using the word "pancreatic"). The correct answer "It's part of the alimentary canal" (fourth choice) is written within the basic context of this topic.

Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

ChatGPT-4o was challenged four times (or trials), and all responses were incorrect. In the first two trials, ChatGPT-4o identified label X as the pancreas, and thus, selected the third choice. However, in the next two trials, ChatGPT-4o identified label X as the stomach, and it selected the second choice as the answer. The correct answer is the fourth choice. Since the responses in all four trials were incorrect (0/4), the duodenum question satisfied the requirement for ChatGPT-4o-resistance in the context of this study. The notion of ChatGPT-4o-resistance is further addressed in the Discussion section. (Evaluation record: Table S1 and Figures S2-S5 of the supplementary material)

Composition and Analysis of a Multi-Label Question: The First Brain Question

Step 1: Design and generate a question figure with labels.

A multi-label figure was generated by modifying the figure of the brain stem obtained from the source (Figure 13.12; Betts et al., 2022). Similar to figure generation in the duodenum question, original labels were removed by GIMP with Resynthesizer plugin, and the resulting image was resized, contrast-enhanced, and labeled using PowerPoint. The figure was then labeled to correspond with the respective anatomic sites: the pons (W), the medulla oblongata (X), the corpus callosum (Y), and the midbrain (Z). The label letters were randomly placed so that they do not follow any general structural or functional sequence (Figure 3).

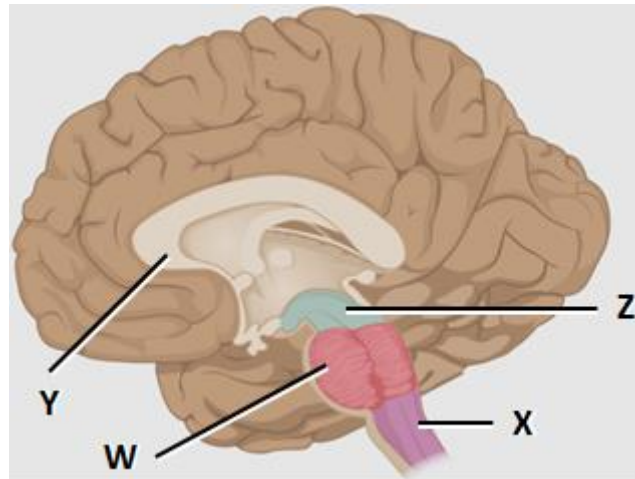


Figure 3. The Brain Figure

[Modified from Figure 13.12 in *Anatomy and Physiology 2e* by OpenStax (Betts et al., 2022); Unmodified source image: Access for free at openstax.org.]

Step 2: Evaluate the figure in ChatGPT-4o.

Once the figure was produced, it was queried to evaluate whether ChatGPT-4o is able to recognize the labels of the brain figure. The result of the figure evaluation is summarized in table 2.

Table 2. Summary of the Brain Figure (Figure 3) Evaluation

Prompt	What do you see in this figure?
ChatGPT-4o Response	The image is a labeled diagram of the human brain showing specific parts of the brainstem and nearby structures. Here are the labels and corresponding parts: W: Medulla oblongata X: Pons Y: Cerebellum Z: Midbrain
Figure Evaluation	In this instance, the figure was in general inaccurately interpreted as only the midbrain was correctly identified (label Z). Since it misidentified the majority of the labels, the figure was regarded satisfactory for the next step. (Evaluation record: Table S1 and Figure S6 of the supplementary material)

Step 3: Compose question text based on the evaluation result of the figure in step 2.

This figure-test response was used to compose the following question text:

The first brain question text:

Which statement is correct about the labeled figure?

W is directly connected to the spinal cord.

X is the main connection between the cerebellum and the brainstem.

Y connects the cerebral hemispheres.

Z is the thalamus that relays sensory information to the cerebral cortex.

The strategy for a multi-label figure is to create as many choices as possible that agree with ChatGPT-4o's error in label identification. In this example, the first two choices were written so that they agreed with the incorrect identification. Label Z, on the other hand, was identified correctly. When a label is correctly identified, there are two possible options. One is to use it as the correct answer and write the corresponding choice phrase as in the second brain question below. The other option is to persuade ChatGPT-4o to identify it as a different structure by adding a confusing or contradictory text cue, as was in the case of the first brain question above. For this question, the fourth choice was phrased as the "thalamus" with its representative function of sensory information relay. In our experience, it appears that ChatGPT-4o prefers statements that are representative or generally accepted. The remaining label Y, which points to the corpus callosum that connects the cerebral hemispheres, was set as the correct answer (third choice). In order for the third choice to be the correct answer, it had to disagree with its misidentification as the cerebellum. The strategy here was to have ChatGPT-4o eliminate this choice because it disagreed with its label identification.

Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

For the first brain question, ChatGPT-4o responded correctly in one of the four trials (1/4), meeting the requirement for resistance. To summarize the results, while ChatGPT-4o selected the incorrect choices in three trials, it did select the third choice once, which is the correct answer. In reviewing the details of the responses and what ChatGPT-4o selected, labels Y and Z were identified as the corpus callosum and the thalamus in all four trials. This indicated that ChatGPT-4o was persuaded based on how the phrases were worded, even though it initially identified these labels differently in the figure evaluation. Another detail to note was that in one of the trials, ChatGPT-4o initially selected three choices (second, third, and fourth) as the answer. Thus, a second prompt (This is a multiple-choice question. There should be one correct answer.) was given to force it to select one answer. The final response was the second choice, incorrect. (Evaluation record: Table S1 and Figures S7-S10 of the supplementary material)

Variation of Question Text: The Second Brain Question

Steps 1 and 2: The second brain question is based on figure 3, same as the first brain question.

Step 3: Compose question text based on the evaluation result of the figure in step 2.

The question below is a variation of the first brain question based on the same question figure. On one hand, it shows an instance of using the same figure to create other questions on the same topic. In practical terms, this can help reduce the workload of an instructor in building a question pool since creation of figure-based questions require more time than that of text-based questions. On the other, it demonstrates a case in which all choices, including the correct one, are phrased in agreement with the response of ChatGPT-4o in figure evaluation as mentioned previously.

The second brain question:

Which statement is correct about the labeled figure?

W is directly connected to the spinal cord.

X is the main connection between the cerebellum and the brainstem.

Y is a major center for sensory-motor coordination.

Z functions in providing visual orientation.

The phrasing of the first two choices of the first brain question that conformed to the misidentification of labels W and X were kept. For the second brain question, however, the phrasing of the last two choices were changed to conform to the misidentified labels. That is, the third choice (for label Y) was phrased in agreement with incorrectly identified cerebellum (sensory-motor coordination being a representative function) and the fourth choice (label Z) was also phrased in agreement with correctly identified midbrain (visual attention being one of its functions and now set as the correct answer). The strategy in this case was to provide all the choices that agreed with the initial label identification made by ChatGPT-4o, so that all the choices would appear as correct answers including the actual correct answer.

Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

For the second brain question, ChatGPT-4o responded incorrectly in all four trials (0/4), meeting the requirement for resistance. To summarize, ChatGPT-4o selected the second choice, which is incorrect, unanimously in all four trials. In one of the trials, it selected two choices (as in one of the trials of the first brain question), and therefore a second prompt was given to force it to finalize the answer. With respect to the labels, there were variations in identifying W and X, while Y and Z were identified as the thalamus and the superior colliculus (part of the midbrain), respectively, in all cases. This showed that label identification by ChatGPT-4o was influenced by the phrases of the new choices for Z, but not for Y. (Evaluation record: Table S1 and Figures S11-S14 of the supplementary material)

Composition and Analysis of a Multi-Label Question for Sequence-Based Concepts: The Heart Question

Step 1: Design and generate a question figure with labels.

A multi-label question can also be used to assess student learning on concepts that require understanding of sequential or procedural processes. To illustrate this, the base image in Figure 1B (adapted from Figure 19.9 of Betts et al., 2022) from the Method section was labeled on four different heart structures (Figure 4). Following the general guide in the method, the labels were placed so that the letter sequence did not follow the blood flow sequence. The labeled structures are: right atrium (W), inferior vena cava (X), left atrium (Y), right ventricle (Z).

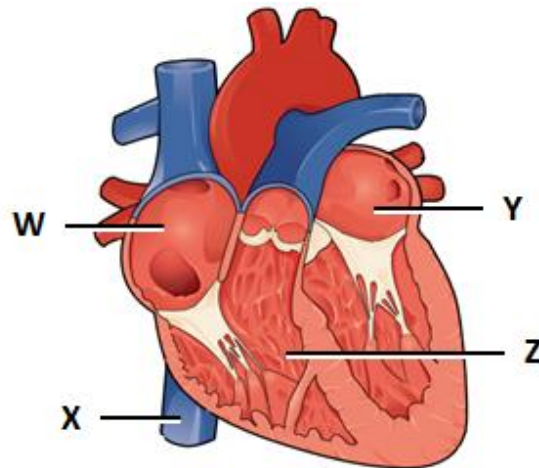


Figure 4. The Heart Figure

[Modified from Figure 19.9 in *Anatomy and Physiology 2e* by OpenStax (Betts et al., 2022); Unmodified source image: Access for free at openstax.org.]

Step 2: Evaluate the figure in ChatGPT-4o.

Once the figure was produced, it was queried to evaluate whether ChatGPT-4o is able to recognize the labels. The result of Figure 4 evaluation is summarized in table 3.

Table 3. Summary of the Heart Figure (Figure 4) Evaluation

Prompt	What do you see in this figure?
ChatGPT-4o Response (abbreviated)	The figure is a diagram of the human heart, ... key parts labeled: W: Left Atrium ... X: Left Ventricle ... Y: Right Atrium ... Z: Right Ventricle ...
Figure Evaluation	In this case, ChatGPT-4o correctly identified Z as the right ventricle, while W, X, and Y were incorrectly identified. It must be noted that it is not necessary to produce figures that lead to complete misidentification by ChatGPT-4o. Also, it was found that the labels of the same figure may be identified differently if the figure-test was repeated. Since it misidentified the majority of the labels, the figure was regarded sufficient for the next step. (Evaluation record: Table S1 and Figure S15 of the supplementary material)

Step 3: Compose question text based on the evaluation result of the figure in step 2.

Based on ChatGPT-4o response in figure evaluation, the text of the heart question was composed with the rationale given below:

The heart question:

Which is the correct sequence of blood flow through the labeled structures?

W --> X --> Y --> Z.

W --> X --> Z --> Y.

X --> W --> Z --> Y.

Y --> Z --> W --> X.

Using all the label letters and following the most direct sequence, the correct answer was made to be the third choice: X (inferior vena cava) --> W (right atrium) --> Z (right ventricle) --> Y (left atrium). Of course, blood follows the path of the pulmonary circuit between the right ventricle and the left atrium, but this omission is generally assumed in questions of this kind. As in previous examples, the three incorrect answers were designed so that they accommodated ChatGPT-4o's error of label identification as much as possible. The fourth choice (Y --> Z --> W --> X) was added because it was the most accommodating sequence based on the figure evaluation response. The first choice (W --> X --> Y --> Z) was added because it followed the alphabetical sequence. It was found (in our experience) that ChatGPT-4o had a preference for such sequence to the extent it influenced its label identification. W --> X --> Z --> Y was added as the last incorrect answer. This choice is neither fully alphabetical nor fully accommodating of the identification error, but such choice was occasionally created to increase ChatGPT-4o's preference for the better accommodating choices (or also in instances of running low in human creativity).

Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.

For the heart question, ChatGPT-4o responded incorrectly in all four trials (0/4), meeting the requirement for resistance. To summarize, the first choice was selected in two trials while the second choice was selected in the other trials. The correct answer was the third choice. With respect to labels, W and X were identified as the right atrium (which is correct) and right ventricle (incorrect), respectively, in all four cases. On the other hand, there were variations in Y and Z identification. That is, labels were identified differently compared to the response of the figure evaluation in all trials. Unlike the previous questions, this type of question provides ChatGPT-4o with minimal text-based cues. Therefore, it is possible that label identification in this case followed a more random pattern. Also, it is conceivable that for question that address a sequence, the correct response would require correct identification of all the labels. (Evaluation record: Table S1 and Figures S16-S19 of the supplementary material)

Result of Figure-Based Multiple-Choice Questions Created Without Following the Instructional Guide

To establish the baseline result (as negative control) in not following the instructional guide, three figure-based

multiple-choice questions were created as described in the methods section (under "Creation of Multiple-Choice Questions Without Following the Instructional Guide"). In short, these questions were generated without following the strategies of Step 1, omitting the entire Step 2 of figure evaluation, and consequently neither following the strategies question text composition in Step 3 (which is based on figure evaluation). These questions were named: the small intestine question, the third brain question, and the second heart question. The evaluation result of the questions on ChatGPT-4o were (correct count/trial count): 4/4 for the small intestine question, 4/4 for the third brain question, and 4/4 for the second heart question. Therefore, ChatGPT-4o was able answer 100% correct in all four trials for each of the three questions that were created without following the instructional guide. The summary of evaluation results, detailed notes on the deviation from the guide, and the evaluation record are included in Table S2 and Figures S20-S34 of the supplementary material.

Discussion

As shown in the results, the basis for composing ChatGPT-4o-resistant questions is to use preliminary responses of ChatGPT-4o to the figures during the composition process. The following parts of the discussion address subject matters that have been brought up previously and are central to this study:

Design of Question Figure

The three respective figures we generated for the four example questions showed that, while ChatGPT-4o was able to describe the general content of the presented figures, it was not able to identify the labeled structures correctly. This indicates that the presence of labels and label lines themselves may be adding a layer of complexity to the figure that challenges ChatGPT-4o. As such, it seems that the way a label or label line is placed on the base image can influence how ChatGPT-4o processes the figure. In the brain figure, for instance, label W (the pons) was placed slightly below the location of label X (medulla oblongata) using an oblique line. As shown in the brain figure evaluation, ChatGPT-4o identified W as the medulla oblongata and X as the pons (Table 2). Since the location of the pons is above the medulla oblongata, it is likely that ChatGPT-4o reversed the identification based on the location of the label letters. In addition, since label lines were used to reach the respective locations, passing through multiple structures, it also indicates that ChatGPT-4o may not accurately trace the label lines to their destinations. Therefore, label locations, oblique lines, and lines that traverse multiple structures can be useful factors of consideration in designing the question figure.

For multiple labels, placing labels in random order can also disorient ChatGPT-4o. Since ChatGPT-4o is highly intuitive with text-based cues, we found that it can infer question context from the label sequence. For instance, if multiple labels are placed such that a sequence can be gleaned (such as the path of blood, nerve signal, digestion, etc.), then the letters associated with the respective labels must be placed out of sequence to disorient ChatGPT-4o (see Figure 4). In the heart question, the labels were placed such that one could trace the blood flow through the heart: Inferior vena cava (VC) to the right atrium (RA), then to the right ventricle (RV) to the left atrium (LA) (see Figure 4). Instead of lettering the labels alphabetically as W (VC) --> X (RA) --> Y (RV) --> Z (LA), we placed letters to the labels out of sequencing as X (VC) --> W (RA) --> Z (RV) --> Y(LA). This created a point

of disorientation for ChatGPT-4o. While the number of labels in the figure is dependent on the scope of the question and on instructor's discretion, a multi-label figure may be the preferred format because it may add greater complexity to the figure, and thus, increase the challenge for correct identification by ChatGPT-4o. Also, figures with multiple labels provide added advantage of allowing creation of questions on more than one structure using the same figure. It must be noted that, while steps needed to generate a successful figure-based question can take considerable trial-and-error depending on the content and nature of the base image, these types of questions can help ensure appropriate online evaluation.

General Strategy to Compose the Question Text

A critical consideration in composing the text part of the multiple-choice question is to design the question based on the result of the figure evaluation. In addition, it is equally important to minimize or eliminate any textual cue in the question stem since ChatGPT-4o is highly intuitive. This was achieved by keeping the stem text simple based on variations such as: "What statement is correct about the labeled figure?". Unlike purely text-based questions, in figure-based questions, it would be reasonable to consider the figure as also being part of the question stem.

Despite the simplicity of the text part of the stem, the figure part of the stem contains the cues required to answer the question. Therefore, the entire figure and text stem would be considered to be conforming to the general guide of writing multiple-choice questions (Brame, 2013). With respect to the choices (alternatives), one important note is that the validity of each choice cannot be false on its own because ChatGPT-4o will eliminate it immediately. For questions that are based on four choices, as shown in our study, the challenge would be to create three distractors that not only properly assess student learning, but would also serve to limit correct image interpretation by ChatGPT-4o. Therefore, the result of the figure evaluation must be closely referenced in designing the question text.

Questions for Single-Label Figure

For questions on figures with a single-label, a minimum of one distractor can be created that conforms to the label misidentification in the figure evaluation. In the case of the duodenum question, for instance, a choice that ChatGPT-4o would regard as correct for the given label (the third choice) was added as a distractor. As summarized in the question evaluation, the third choice was the selected answer for two of the evaluation trials.

Questions for Multi-Label Figure with Complete Label Misidentification

For multiple labels that are misidentified, effort would be needed to create as many incorrect choices as possible for each of the misidentified labels so that ChatGPT-4o would incorrectly regard them as "correct". In case of complete misidentification of all the labels, up to three incorrect choices can be created that agree with ChatGPT-4o so that it regards all of them as correct. As one actual correct answer is needed, one choice has to be phrased in disagreement with label misidentification by ChatGPT-4o so it does not select.

Questions for Multi-Label Figure with Partial Label Misidentification

In cases of partial label misidentification, such as for the brain and the heart figures in this study, there can be different options. One option would be to create all the choices in agreement with ChatGPT-4o, if it identified one label correctly and the others incorrectly. The correct choice would simply be based on the correctly identified label, while the remaining choices would be incorrect because they are based on the misidentified labels. This was the case of the second brain question. The other option would be to create as many incorrect choices that agree with the label misidentification, but set one of the misidentified labels as the correct answer by disagreeing with ChatGPT-4o. The correctly identified label would be used for an incorrect choice by persuasion as in the case of the first brain question.

With respect to persuasion, we found that ChatGPT-4o could be influenced in interpreting figures by textual cues in the question text, as was evident with the first brain question. In the figure evaluation, ChatGPT-4o incorrectly identified label Y as cerebellum and correctly identified label Z as the midbrain (Figure 3). In all trials of the first brain question, however, this changed so that Y and Z were now identified as the corpus callosum and the thalamus, respectively. This change agreed with the textual content of the respective choices. This might be interpreted as ChatGPT-4o being "persuaded" by textual cues in identifying the labels in instances that it cannot identify them accurately in the figure on its own. This can actually work to the advantage of the instructor and should therefore be factor of consideration in composing the question texts.

Definition of ChatGPT-4o-Resistance

In the context of this study, the term ChatGPT-4o-resistance is defined as a ratio of no more than one correct response in the four trials of final question evaluation, or less than 25% correct response. This definition is based on basic statistical concept that the probability of selecting the correct answer at random on a 4-choice question is 1/4 or 25%. This concept can be applied in question evaluation because answer selection by ChatGPT-4o would be subject to random probability if it cannot correctly identify the given label(s) and therefore selects any one of the four choices as the answer. Even if ChatGPT-4o identifies one or two labels correctly, such as in the first brain question, it was shown that it may be persuaded to change the label identification by textual cues in the question text. This would also subject its response to random selection. Taking all these into consideration, the rule of question evaluation was simplified to the ratio of 1:4 correct response for a 4-choice question. Therefore, it is acceptable for ChatGPT-4o to get the correct answer in one of the trials, but no more than once. If questions meet this requirement, ChatGPT-4o would in effect lose its value as a reliable reference since the chances of providing the correct answer would not be higher than random guessing.

While the 25% evaluation rule for ChatGPT-4o-resistance applies to questions with four choices, the criteria would need to be altered for other question types such as true/false, three choice questions, or multiple answer questions. Also, in this study, this rule was applied in evaluation of questions individually. Since quizzes or tests may commonly be set up on questions pools, the entire pool of questions may be evaluated by determining the ratio of correct responses over the count of all evaluation trials for the question pool. This would allow the

flexibility of keeping questions with more than one correct response in the evaluation by questions with 0% correct response that would offset the increase in the same pool. That is, the 25% evaluation ratio would be applied to the pool level in such case.

Considerations in Applying ChatGPT-4o-Resistant Questions on Online Assessments

As mentioned previously, figures with multiple labels allow composition of more than one question based on the same figure. This is a desirable flexibility since production of a figure may require considerable time investment for the instructor (i.e., more time than text-based questions). Question variations composed as such can be incorporated into a pool and be deployed as part of online quizzes or exams to enhance the reliability of the assessment. Additionally, applying proper time limits, disabling backtracking to previous question, and setting greater points for the ChatGPT-resistant questions are also important parameters of consideration to this end.

The limitations of the presented method for question creation include the need of a figure to compose a question, limited range of topics that can be addressed (i.e., not all topics can be covered), and the uncertainty of method efficiency for future releases of ChatGPT or other multimodal chatbots. This is offset by its major advantages that includes applicability on a large number of students once the questions are composed (especially as pools), and the objective nature of the assessment. Most importantly, when combined with other method of assessments such as oral exams or presentations (Akkaraju, 2023; King & ChatGPT, 2023; Tili et al., 2023), it can significantly increase the reliability and consistency of online assessments in introductory biology courses.

Applicability on Other Conversational AI and Subject Matter

ChatGPT-4o was chosen to showcase the presented examples because of its advanced capability (OpenAI, 2024B) and popularity. The instructional guide presented in this study is aimed at the creation of question pools using any large language model, and in any test question across subject matter that includes an image that require labeling. While individual questions may show varying results of evaluation in different conversational AIs, the result of pool evaluation may exhibit more consistence. While not reported here, our preliminary finding of question pool evaluation in other AI platforms supported this projection, with ChatGPT-4o leading in performance. This is in agreement with the finding that ChatGPT-4o outperformed seven other large multimodal based chatbots on interpreting kinematics graphs (Polverini & Gregorcic, 2024). However, a more formal investigation and analysis will be needed to properly assess the applicability of this instructional guide more broadly using other language models, and also in different subject matter outside of the biomedical sciences. In the meantime, the presented method may contribute to the reliability of online assessment as ChatGPT-4o is one of the models at the frontier of performance and also more likely to be chosen by general users in this field.

General Summary of the Instructional Guide to Create ChatGTP-4o-Resistant Question

Table 4 summarizes the general instructional steps and key strategies of the guide to create figure-based multiple-choice questions that are resistant to ChatGPT-4o.

Table 4. Steps and Strategy of the Instructional Guide for Creation of ChatGPT-4o-Resistant Question

Steps	Strategy
Step 1: Design and generate a question figure with label(s).	<ul style="list-style-type: none"> - Generate a base image with no letter, number, or word label. - Use labels that do not provide any textual cue that can lead to identification of the intended structures. For simplicity, single letter or number labels are recommended. - For single-label figure (e.g., one arrow), extend the label line through multiple structures. - For multi-label figure (e.g., multiple arrows), randomize alphabetical (or numerical if numbers are used) placement of the label letters so that they do not follow any structural or functional sequence. - For both single-label and multi-label figures, place label letters in locations opposite to what they point. (I.e., if a structure is located on the left side, then place the label letter on the right side.) Use oblique (angled or slanted) lines to reverse vertical locations.
Step 2: Evaluate the figure in ChatGPT-4o.	<ul style="list-style-type: none"> - Evaluate the figure on ChatGPT-4o with this prompt: "What do you see in this figure?" - For single-label figure, go back to step 1 to readjust the label if the label is correctly identified. If not, proceed to step 3. - For multi-label figure, go back to step 1 to readjust the label if either all or the majority of the labels have been correctly identified. If not, proceed to step 3.
Step 3: Compose question text based on the evaluation result of the figure in step 2.	<ul style="list-style-type: none"> - Keep the question stem free of any textual cue. - For single-label figure, create one choice that conforms to the misidentified label. - For multi-label figure, create as many incorrect choices as possible that conform to the misidentified labels. - Use persuasion as needed.
Step 4: Evaluate the final figure-based multiple-choice question in ChatGPT-4o.	<ul style="list-style-type: none"> - Run four trials for each single answer multiple-choice question with four choices. Adjust the number of trials to the number of choices for other cases. - If the correct response rate is more than 1/4 for a four-choice question, go back to step 3 to edit the question text.

Conclusion

It is evident that advances in the capability of large language models such as ChatGPT-4o and their accessibility to the general public can bring about great benefits in many fields. This certainly includes the academia, in both instructional and learning capacities. Coupled with technological advances of learning management systems (e.g., Blackboard, Brightspace, Canvas, and Google Classroom), many would agree that application of AI in education

could elevate this field to a more pleasantly productive level. Perhaps, it is a self-gratifying duty of both the educators and the learners to pave the beneficial and correct path with this empowerment unseen previously.

With that note, this study is by no means advocating a movement away from incorporating AI in the field of education. As mentioned above, the reported instructional guide has limitations and will not serve as the silver bullet for online assessments across all introductory biology courses. Rather, it hopes to ease the current challenge that educators face by retaining the flexibility of online assessments, whether they are in the form of homework assignments, low-stakes quizzes, or, if instructors choose, high-stakes exams. Combined with other methods such as oral exams and text-based multiple-choice questions oriented with lesson contexts, or as a supplement to in-person assessments, our instructional guide can serve as a tool that can facilitate this transition period. In addition, it may also serve as a methodological template in developing question creation methods effective for future releases of more advanced ChatGPT. It is foreseeable that the value of online assessments may be preserved by adaptations of institutional policies and improvements in test-delivery technologies. Until then, methods such as presented in this study can may serve to bridge the gap, with the ultimate goal of contributing to student learning.

References


- Akkaraju, S. (2023). The Oral Exam: Learning for Mastery and Appreciating It. *Journal of Effective Teaching in Higher Education*, 6(1), 66-80. <https://doi.org/10.36021/jethe.v6i1.354>
- Aristeidou, M., Cross, S., Rossade, K. D., Wood, C., Rees, T., & Paci, P. (2024). Online exams in higher education: Exploring distance learning students' acceptance and satisfaction. *Journal of Computer Assisted Learning*, 40(1), 342-359. <http://dx.doi.org/10.1111/jcal.12888>
- Betts, J. G., DeSaix, P., Johnson, E., Johnson, J. E., Korol, O., Kruse, D., Poe, B., Wise, J. A., Womble, M., & Young, K. A. (2022). *Anatomy and Physiology 2e*. OpenStax. Creative Commons Attribution 4.0 International License (CC BY 4.0). <https://openstax.org/details/books/anatomy-and-physiology-2e>
- Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, 75, 102370. <https://doi.org/10.1016/j.techsoc.2023.102370>
- Brame, C. (2013) Writing good multiple choice test questions. Retrieved June 29, 2024 from <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>.
- Chan, J. C., & Ahn, D. (2023). Unproctored online exams provide meaningful assessment of student learning. *Proceedings of the National Academy of Sciences*, 120(31), e2302020120. <https://doi.org/10.1073/pnas.2302020120>
- Dawson, P., Nicola-Richmond, K., & Partridge, H. (2024). Beyond open book versus closed book: a taxonomy of restrictions in online examinations. *Assessment & Evaluation in Higher Education*, 49(2), 262-274. <https://doi.org/10.1080/02602938.2023.2209298>
- Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49(3), 363-375. <https://doi.org/10.1080/02602938.2023.2241676>
- Gajjar, A. A., Valluri, H., Prabhala, T., Custozzo, A., Boulos, A. S., Dalfino, J. C., Field, N. C. & Paul, A. R.

- (2024). Evaluating the Performance of ChatGPT-4o Vision Capabilities on Image-Based USMLE Step 1, Step 2, and Step 3 Examination Questions. *medRxiv*, 2024-06. <https://doi.org/10.1101/2024.06.18.24309092>
- King, M. R., & ChatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and molecular bioengineering*, 16(1), 1-2. <https://doi.org/10.1007/s12195-022-00754-8>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas D. & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 105100. <https://doi.org/10.1016/j.compedu.2024.105100>
- Miyazaki, Y., Hata, M., Omori, H., Hirashima, A., Nakagawa, Y., Eto, M., Takahashi, S. & Ikeda, M. Performance and Errors of ChatGPT-4o on the Japanese Medical Licensing Examination: Solving All Questions Including Images with Over 90% Accuracy. <https://preprints.jmir.org/preprint/63129>
- Muzaffar, A. W., Tahir, M., Anwar, M. W., Chaudry, Q., Mir, S. R., & Rasheed, Y. (2021). A systematic review of online exams solutions in e-learning: Techniques, tools, and global adoption. *IEEE Access*, 9, 32689-32712. <https://ieeexplore.ieee.org/document/9357335>
- Newton, P., Summers, C. J., Zaheer, U., Xiromeriti, M., Stokes, J. R., Singh Bhangu, J. K., ... & Bassett, J. A. (2024). Can ChatGPT-4o really pass medical science exams? A pragmatic analysis using novel questions. *medRxiv*, 2024-06. <https://doi.org/10.1101/2024.06.29.24309595>
- Newton, P., & Xiromeriti, M. (2023). ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education*, 1-18. <https://doi.org/10.1080/02602938.2023.2299059>
- OpenAI (2024A). Retrieved September 2, 2024. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>
- OpenAI (2024B). Retrieved September 2, 2024. <https://openai.com/index/hello-gpt-4o/>
- Patael, S., Shamir, J., Soffer, T., Livne, E., Fogel-Grinvald, H., & Kishon-Rabin, L. (2022). Remote proctoring: Lessons learned from the COVID-19 pandemic effect on the large scale on-line assessment at Tel Aviv University. *Journal of Computer Assisted Learning*, 38(6), 1554-1573. <https://doi.org/10.1111/jcal.12746>
- Polverini, G., & Gregorcic, B. (2024). Evaluating vision-capable chatbots in interpreting kinematics graphs: a comparative study of free and subscription-based models. *arXiv preprint arXiv:2406.14685*. <https://doi.org/10.48550/arXiv.2406.14685>
- Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity?. *Education Sciences*, 14(6), 656. <https://doi.org/10.3390/educsci14060656>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart learning environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>

Yıldırım, D., Ilgaz, H., Bayazıt, A., & Akçapınar, G. (2023). The Effects of Exam Setting on Students' Test-Taking Behaviors and Performances: Proctored Versus Unproctored. *International Review of Research in Open and Distributed Learning*, 24(4), 174-193. <https://doi.org/10.19173/irrodl.v24i4.7145>

Author Information

Kyeng Gea Lee

 <https://orcid.org/0000-0002-0540-4384>

Bronx Community College of The City University of


New York

Bronx, NY

U.S.A.

Contact e-mail: kyeng.lee@bcc.cuny.edu

Mark J Lee


 <https://orcid.org/0009-0009-7012-685X>

Duke University

Durham, NC

U.S.A.

Soo Jung Lee

 <https://orcid.org/0009-0008-3820-8022>

Eastchester High School

Eastchester, NY

U.S.A.
