

Evaluation of Automatic Multiple Choice Question Generation using Prompt Engineering

Jocelyn Chai Hui Min¹, Lu Zhiyue¹, How Chang Hong²

¹Raffles Girls' School, 2 Braddell Rise, Singapore 318871

²Defence Science and Technology Agency, 1 Depot Road, Singapore 109679

ABSTRACT

Given the myriad of advantages Large Language Models (LLMs) offer in Natural Language Processing (NLP), this project aims to evaluate the existing prompt engineering techniques and Generative Pre-Trained Transformers (GPTs) software applications to enhance the effectiveness of such GPTs in Automatic Multiple Choice Question Generation (A-MCQG) for maintenance manuals. The generated Multiple-Choice Questions (MCQs) are used to create an automated assessment system to test the learning gap of users. Ultimately, this provides a cost-effective and efficient solution for companies and individuals to train their employees or themselves on how to maintain equipment, without the need for expensive training programs or instructors.

By leveraging prompt engineering techniques like zero-shot, roleplay, flipped interaction, context manager and fact check list, we aim to extract information from maintenance manuals through text summarisation and generate questions for assessment purposes. Questions were evaluated for relevance to the given document, 15 frequently occurring item-writing flaws (IWFs) and cognitive level based on Bloom's Taxonomy. Our results show that questions generated by our proposed template using Claude achieved the highest score of 91.0% and demonstrates comparable effectiveness to manual question generation, with 51.0% of generated questions having no IWFs and 17.0% at higher cognitive levels.

INTRODUCTION

With the advent of a new technological age over the past few decades, Generative Pre-Trained Transformers (GPTs) have become a centrepiece in online learning around the world, already playing a critical role in applications such as information retrieval, essay writing and text summarisation. At its core, GPTs are Large Language Models (LLMs) built upon a large corpus of training data and intensive fine-tuning with human feedback [1]. They are well suited for a wide range of tasks in Natural Language Processing (NLP), being able to comprehend human language and generate coherent outputs that are tailored to the specific nuances of the subject matter at hand. This is attributed to a self-attention mechanism that GPTs employ to compute attention scores for each word in relation to all other words in the input sequence [2], capturing long-range dependencies and information context. As such, GPTs present a valuable asset for the creation of Automatic Multiple Choice Question Generators (A-MCQG) that can potentially enhance the efficiency and adaptability of assessments in educational settings.

In spite of the benefits of using exam-style questions as an educational tool, including the facilitation of knowledge recall and the rectification of common misconceptions, manual question construction is a challenging task that requires training, experience and resources. Several published analyses of real exam questions demonstrate their poor qualities due to the presence of item-writing flaws (IWFs) [3], which are violations of accepted item-writing guidelines that can affect student performance on MCQs by making the question either easier or more difficult to answer. This can be attributed to a lack of training in assessment development or low cognitive levels through which the questions are written [4]. This

challenge is augmented further by the need to replace assessment questions consistently to ensure they align with the ever-evolving educational landscape.

AQG techniques emerged as a solution to the challenges faced by test developers in constructing a large number of good quality questions. It offers a time-efficient means to evaluate users' comprehension and mastery of academic content that can not only expedite the assessment process, but also provide a more dynamic and interactive learning experience. By integrating GPTs with their pre-trained knowledge base into the assessment framework, educators can automate the creation of diverse and contextually relevant questions that cater to the unique complexities of various subject matters. Furthermore, the versatility of GPTs enables the customisation of question types that are calibrated to an individual's learning pace. These allow for a more objective measure of an individual's learning pace, addressing the inherent challenges faced by self-learners in quantifying and evaluating their learning progress.

Presently, limitations such as inconsistent question quality [5] and the lack of extensive research on prompt engineering are major concerns that prevent large-scale adoption of GPTs in the education system. Through our project, we aim to develop a prompt template tailored for A-MCQG by evaluating the effectiveness of existing prompt engineering techniques as well as determine the overall effectiveness of using A-MCQG technologies for assessment purposes. This can contribute to the optimisation of learning outcomes through the creation of factually accurate and targeted questions, which plays a pivotal role in assessing student knowledge and identifying learning gaps. By linking readily available machine learning technologies with educational needs, we hope to conceptualise GPT models as a tool with rich potential to streamline assessment processes and enhance educational resources.

METHODOLOGIES

To effectively utilise GPTs in A-MCQG, we investigated: 1) existing prompt engineering techniques [6], 2) readily available GPT models in their effectiveness in generating multiple choice questions (MCQs) based on a sample document provided and 3) overall effectiveness of our proposed prompting template and GPT model. Questions generated are evaluated manually based on accepted item-writing guidelines and the Bloom's Taxonomy.

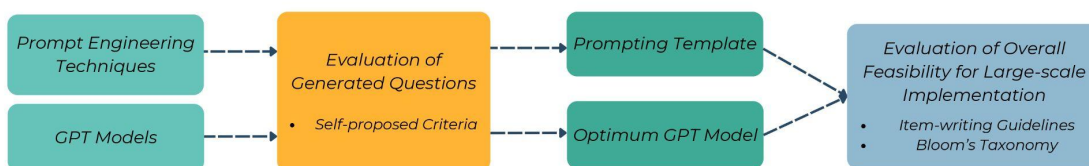


Figure 1: Conceptual Overview of Methodologies

Prompt Engineering Techniques

Prompt engineering is the practice of formulating specific prompts for LLMs to generate desired outputs. We developed a prompting template that is specially customised for the A-MCQG task by combining four different prompt engineering techniques in Anthropic's Claude.ai. Below are the four prompt engineering techniques used:

1. Fact Check List (Error Identification)

The Fact Check List is a technique that presents a comprehensive list of facts within the GPT-generated output, forming the foundational basis upon which the quiz questions are crafted. The purpose is to combat a prevalent limitation of GPTs, namely their tendency to generate convincing text that is factually inaccurate. By presenting a comprehensive

list of facts taken from the given document, this technique offers a structured means for users to conduct a more efficient review and validation of the accuracy of the output, thereby enhancing the overall credibility of the quiz.

A sample wording of the Fact Check List technique is as follows: “From now on, when you generate the quiz, create a set of facts that the quiz depends on, which can be derived from the attached document. List this set of facts at the end of the output.”

2. Context Manager (Context Control)

The Context Manager is a technique that enables users to specify and remove contextual parameters within a conversation with a GPT. The purpose is to facilitate a more targeted approach to tasks and exclude unrelated topics from consideration. This gives users greater control over the content the GPT considers or ignores when generating output, which is essential as GPTs often struggle to generate relevant responses based on prior inputs.

A sample wording of the Context Manager technique is as follows: “The quizzes you create will be in a multiple choice format, where each question will have 4 plausible alternatives with no "all of the above" option. Do not make up fake questions or answers and only use information given in the document provided.”

3. Flipped interaction (Interactive)

The Flipped Interaction is a technique that probes the GPT to ask questions to obtain the information it needs to perform a task, allowing the GPT to drive the conversation instead of the user. This ensures that the GPT stays focused on the task at hand by asking questions it deems relevant to achieving the specified goal. This motivates the GPT to select the format, number, and content of the interactions in a more targeted manner, ultimately enhancing the quality of the output. The prompt can also be customised to include the number of questions asked at any one time, the order of the questions and any other formatting considerations depending on the task at hand, which makes it more customisable.

A sample prompt for Flipped Interaction is as follows: “Ask the instructor clarifying questions one at a time to gather details of the quiz such as the sub-topics to be tested. Wait for the instructor to answer before you move on to the next question. Then ask the instructor if they have any other instructions regarding the quiz. Generate the quiz only after asking all the clarifying questions needed to gather information for the quiz.”

4. Role-play (Output Customisation)

The Role-play is a technique that instructs the GPT to assume a specific role or persona that aids it in selecting the types of output to generate and the details to focus on. The purpose is to enhance personalisation for specific tasks such that the GPT can adapt to the individual user’s needs and preferences. Through this technique, the GPT model adeptly applies the understanding of its persona acquired through extensive training on a large corpus of text data from diverse domains, to generate more detailed, accurate, and contextually relevant responses. The role-playing directive may also enable the model to more accurately simulate the language, tone, and content associated with that role based on its training data.

A sample prompt for Role-play is as follows: “You are a creator of highly effective diagnostic quizzes. Your goal is to help the instructors create quizzes for their lessons that will help students both retrieve information as they take the quiz and give the instructors a sense of the students’ progress.”

To assess the impact of individual prompt engineering techniques on the effectiveness of A-MCQG, we employed a zero-shot prompt as a baseline and a comprehensive template prompt integrating all 4 of the above mentioned prompt engineering techniques.

A sample prompt for Zero-shot is as follows: “Create a clearly written quiz with 5 multiple-choice questions for each sub-topic in the attached document and an answer key.”

A sample conversation using our template is as follows:

Table 1: Sample of Proposed Prompt Template

Actor	Prompts/ Replies	Techniques
User:	You are a creator of highly effective diagnostic quizzes. Your goal is to help the instructors create quizzes for their lessons that will help students retrieve information as they take the quiz and give the instructors a sense of the students’ progress. The quizzes you create will be in a multiple choice format, where each question will have 4 plausible alternatives with no "all of the above" option. Do not make up fake questions or answers and only use information given in the document provided. If you understand this, please respond with “I understand” and do not elaborate or respond with any other words.	Roleplay Context Manager
Claude:	I understand	
User:	First introduce yourself to the instructor. Then ask the instructor clarifying questions one at a time to gather details of the quiz such as the sub-topics to be tested. Wait for the instructor to answer before you move on to the next question. Then ask the instructor if they have any other instructions regarding the quiz. Generate the quiz only after asking all the clarifying questions needed to gather information for the quiz.	Flipped Interaction
Claude:	Hello, my name is Claude and I've been asked to create a highly effective diagnostic quiz for your lesson. Before I generate the quiz, I have a few clarifying questions: What specific sub-topics from the material would you like the quiz to focus on?	
User:	The sub-topics in the given document are The Nature of Matter, Isotopes, States of Matter and Catalyst.	
Claude:	Thank you for specifying the sub-topics to focus on from the material. My next question is: Would you like an equal number of questions on each sub-topic, or should certain sub-topics be weighted more heavily in the quiz?	

User:	Equal number of questions on each sub-topic	
Claude:	Got it, I'll plan to have an equal number of questions on each sub-topic. Next question: How many total questions would you like in the quiz?	
User:	5 questions for each sub topic	
Claude:	Okay, 5 questions for each of the 4 sub-topics, so there will be a total of 20 questions in the quiz. Do you have any other instructions for me regarding the structure, format or content of the quiz questions and answers? Or any other preferences I should keep in mind as I generate the quiz?	
User:	When you generate the quiz, create a set of facts that the quiz depends on and the answer key for the quiz, which can be derived from the attached document. List both the set of facts and the answer key at the end of the output.	Fact Check List
Claude:	Understood. I will generate the quiz based on facts derived from the provided document. After the quiz, I will list out the set of facts I created that the quiz depends on. Please let me know if you have any other instructions before I proceed with generating the quiz.	
User:	No, proceed to generate the quiz.	

Evaluation of Generated Questions

To evaluate the quality of the generated questions using different prompt engineering techniques and the performance of the GPT systems (Claude.ai, Bing chat and Bard), a dataset of 20 questions were generated, with 5 from each of the sub-topic given in the sample document EASA Module 2 - Physics 1: The Nature of Matter, Isotopes, States of Matter and Catalyst. All data obtained were tabulated into a table using Google Sheets, which will then be used for graph plotting and data analysis. We used 2 human evaluators, who are Junior College going students with fair enough knowledge about Physics, to evaluate the questions for IWFs. They identified number of errors in each question based on the following 5 parameters that we deemed as top 5 most important in assessing quality of the generated questions:

1. Relevance (25%): Whether the question generated is based on information from the given document
2. Factual Accuracy (30%): Whether the question generated is based on scientific evidence
3. Grammar and Readability (10%): Whether the question generated is grammatically correct and easy to comprehend
4. Distractor (10%): Whether the distractors generated are plausible and difficult to differentiate from the correct answer
5. Answerability (25%): Whether the question generated is answerable with reference to the question stem (i.e. only has one correct answer)

Evaluation of Prompt Techniques

Each prompt engineering technique was systematically removed from the prompting template. The modified prompts were independently tested using aforementioned parameters to evaluate the effects of removing each technique, enabling a nuanced understanding of the role of each technique in shaping the impact of the prompt template.

Evaluation of Overall Effectiveness of Suggested Prompt Template

From a pedagogical standpoint, it is imperative to examine the impact of generated questions based on our suggested prompt template on learning outcomes and the development of critical thinking skills. To conduct a comprehensive review of the feasibility of large-scale A-MCQG implementation in the educational landscape, we investigated the quality of the generated questions using 2 main standards.

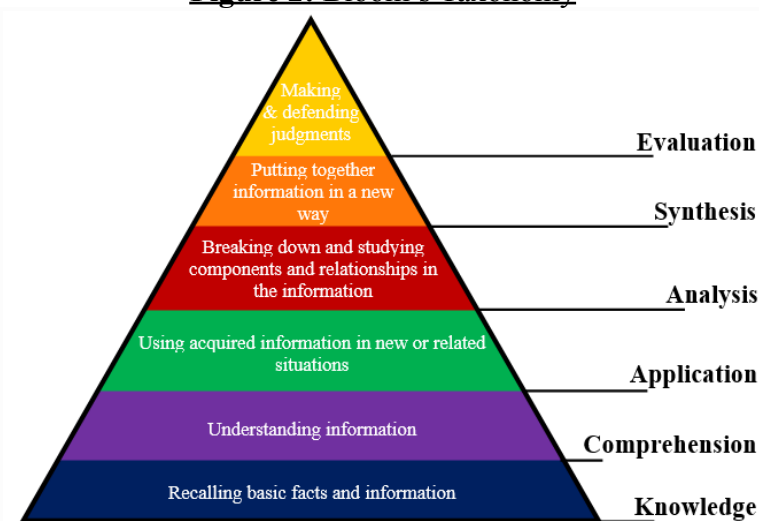
1. Item-writing Flaws

We reviewed existing literature to determine the most cited sources for MCQ construction, from which we identified 32 commonly occurring item-writing guidelines. Using these guidelines, we conducted a preliminary review of a random sample of 25 MCQs to determine the most commonly occurring IWFs. In total, 15 of the cited IWFs were found in the sample of 20 MCQs, which were subsequently used to evaluate the quality of the 100 MCQs generated from the provided document (refer to Appendix 2).

2. Bloom's Taxonomy

Bloom's taxonomy outlines 6 domains which assess progressively higher levels of cognitive function: knowledge, comprehension, application, analysis, synthesis and evaluation. We focused on the first four domains in our analysis, which have greater relevance to the context of MCQ generation. Knowledge and comprehension are regarded as lower cognitive domains that encourage superficial learning while application and analysis are regarded as higher cognitive domains that encourage critical thinking. To facilitate the categorisation of MCQs and maintain inter-rater reliability, we simplified Bloom's taxonomy into two distinct levels of cognition: B1 and B2. A B1 question assessed recall of facts or basic comprehension whereas a B2 item assessed application or analysis of facts.

Figure 2: Bloom's Taxonomy

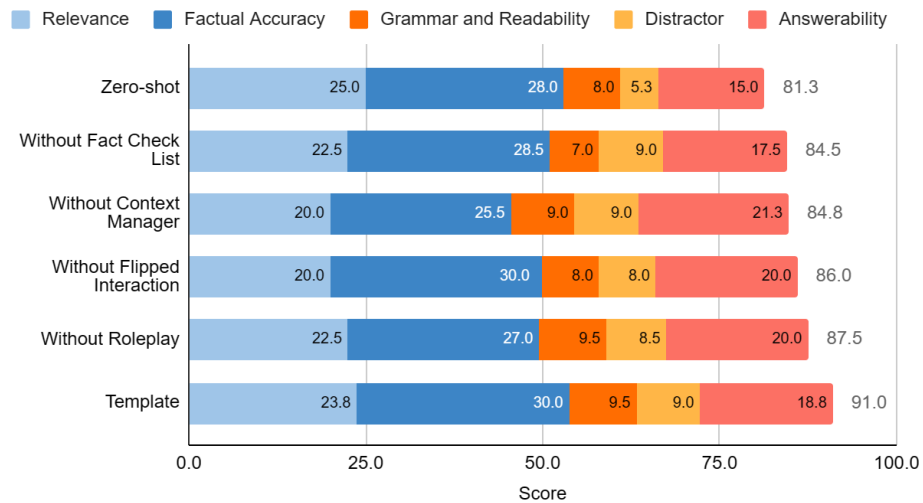


RESULTS AND DISCUSSION

Performance of Prompt Engineering Techniques

As can be seen from Figure 3, our prompting template showed the highest score of 91.0%.

Figure 3: Performance using different prompt engineering techniques



In terms of “Relevance”, questions generated by the prompting template achieved a high score of 23.8%. The score decreases notably to 20.0% when Context Manager or Flipped Interaction are removed. This is likely because Context manager specifies context parameters while Flipped interaction facilitates a more targeted approach to the task, both enabling the template prompt to generate more contextually relevant questions.

In terms of “Factual Accuracy”, the template-generated questions achieved a full score of 30.0%. The score decreases to 27.0% when Context Manager is removed. This is likely because Context Manager restricts the GPT to obtaining information solely from the document, thus preventing the generation of unfactual information.

In terms of 'Grammar and Readability,' questions generated by the prompting template achieved a high score of 9.5%. The removal of the Fact Check List resulted in a notable decrease to 7.0% due to the presence of language use that affected “Readability”. However, this may be attributed to subjectivity in the evaluation process influenced by individual language understanding, leading to variability in results.

In terms of “Distractors”, questions generated by the prompting template achieved a higher score of 9.0% than 5.3% of zero-shot. This score decreases when Flipped interaction and Roleplay are removed from the prompting template. Flipped interaction facilitates active engagement and increases the GPT’s contextual understanding while Roleplay enables the GPT to simulate a specific perspective, both could have enabled the model to better anticipate and incorporate common misconceptions into the distractor choices.

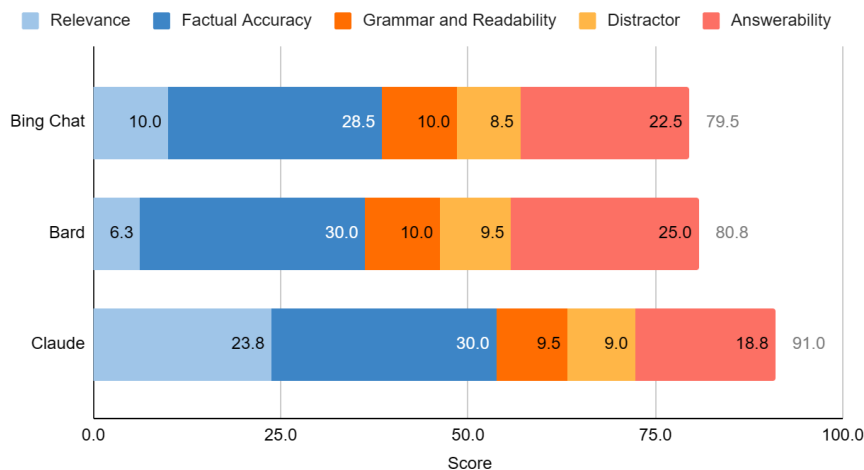
In terms of "Answerability," questions generated by the prompting template exhibit a lower score of 18.8% compared to 21.3% when Context Manager is removed. This disparity could be attributed to the restrictive nature of Context Manager, which limits the GPT to acquiring information solely from the document. This could hinder the GPT’s ability to identify synonyms and alternative phrases that are similar to but distinct enough to be distinguished from the correct option, introducing ambiguity into the answers of the generated questions.

During gathering of the data, the results from Zero-shot were inconsistent and random between different runs, both in the number and quality of questions generated. The data presented for Zero-shot used the best-score attempt. This phenomenon could be attributed to the fact that given no specific guidance or context, the GPT relies solely on its pre-trained knowledge, resulting in unpredictability and a lack of control over the generated output. This leads to questions that may vary widely in terms of quality, relevance, and coherence. On the other hand, our template prompt offers a higher level of control, leading to the consistent generation of higher quality questions.

Performance of GPT models

As can be seen from Figure 4, questions generated using Claude showed the highest score of 91.0%. In terms of “Relevance”, the questions generated by Claude showed a significantly higher score of 23.8% as opposed to 10.0% and 6.3% for Bing Chat and Bard respectively. This suggests that Claude is adept in tasks related to NLP, including understanding and extracting of information, causing it to be the most suited for A-MCQG tasks that require relevant and accurate information retrieval from textual documents. In terms of “Answerability”, the questions generated by Claude showed a considerably lower score of 18.8% as opposed to 22.5% and 25.0% for Bing Chat and Bard respectively. This suggests that while Claude demonstrates proficiency in generating relevant and factually accurate questions, it faces challenges in producing plausible distractors. However, with minor adjustments, the generated distractors can be refined to ensure that each question has a single best answer. In terms of “Factual Accuracy”, “Grammar and Readability” and “Distractor”, all three models showed relatively similar data. This could be due to the fact that these GPTs are general language models, which are trained to provide coherence in language but not fine-tuned in a specific topic, and are thus able to perform simple NLP tasks like A-MCQG.

Figure 4: Performance of different GPT models using prompting template



Performance of Overall Effectiveness of Proposed Prompt Template (100 GPT-generated MCQs)

1. Item-writing Flaws

In total, 100 MCQs were reviewed within this study and a total of 79 IWFs were detected. Of these 100 questions, 49 of them contained at least one IWF and 22 of them contained more than one IWF. Albeit high, the frequency of IWFs found in this study is strikingly similar to Costello’s evaluation of IWFs in Computer Science, Humanities, Medicine and Health, Psychology and Mathematics Massive Open Online Courses (MOOC) in which

47.4% of MCQs contained item-writing violations [7]. The proportion of flawed questions in this study is also substantially less than the 54.9% found in computing, social science and health sciences MOOCs (Kirwan, 2018) [8]. This suggests that the performance of our prompt template is comparable to traditional methods of generating questions. Furthermore, the majority of the generated questions (27.0%) contained only one item-writing violation, suggesting that the quality of these MCQs could be substantially improved with minor edits.

The most frequently occurring flaw observed in our dataset was implausible distractors (25.3%), followed by questions with more than one or no correct answer, which was identified in 14 instances (17.7%). This could be due to limitations of the GPT in coming up with three equally plausible distractors in a four-option question without compromising the validity of the correct answer. A number of violations that help students answer questions based on cues given in the stem or options were also present. Specifically, IWFs such as logical cues in stem (10.1%) and word repeats in stem and correct answer (7.6%) were the third and fourth most frequently occurring flaws. Other less common flaws that could favour test-wise students and potentially affect the validity of the test results include longest option is correct (2.5%) and grammatical cues in sentence completion (2.5%).

Table 2: Total number of IWFs in 100 GPT-generated MCQs

Number of IWFs	Number detected
None	51
One	27
Two	15
Three	7

Table 3: Frequency of IWFs in 100 GPT-generated MCQs

IWFs	Number detected	Percentage of total
Implausible distractors	20	25.6%
More than one or no correct answer	13	16.7%
Logical cues in stem	8	10.3%
Word repeats in stem and correct answer	6	7.7%
Unfocused stem	6	7.7%
Fill-in-the-blank	6	7.7%
Ambiguous or unclear information	5	6.4%
Use of none of the above	3	3.8%
Gratuitous information in stem	2	2.6%
Longest option is correct	2	2.6%
True/ false question	2	2.6%
Grammatical cues in sentence completion	2	2.6%
Negative worded stem	1	1.3%

Use of all of the above	1	1.3%
Convergence cues	1	1.3%

2. Bloom's Taxonomy

The overwhelming majority of the questions (83.0%) were generated at the lower cognitive level (B1) while 17.0% of questions were generated at the higher cognitive level (B2). The amount of higher cognitive questions are higher when compared to Mehmood's study pegged at the secondary school level (9.5%) but lower when compared to Momsen's study pegged at the undergraduate level (37.0%). This can be attributed to the fact that the nature of content in the given document is more informative than application-based. It could also indicate that GPTs are limited in their ability to generate questions of higher cognitive domains, which require more creative thinking, leading to challenges in having a standard template across different topics. If other question formats such as short-answer questions or essay questions are generated simultaneously to test higher cognitive domains, this could offset the low cognitive level of the MCQ component of the overall assessment and allow for greater variability in question difficulty.

CONCLUSION

According to the analysis we conducted, we can conclude that the most optimal approach for A-MCQG using prompt engineering tasks involving information extraction from a given document is a combination of roleplay, context manager technique, fact check list and flipped interaction using Claude. It is also reasonable to conclude that A-MCQG using GPT models is a feasible alternative to manual generation of MCQs and has potential to be implemented into the educational landscape as our study suggests that the quality of the GPT-generated questions are comparable to human-generated ones. However, slight modifications to the proposed template as well as fine-tuning the GPT model to specific domains is necessary to generate more content-relevant and better quality MCQs that can effectively assess student learning outcomes.

Limitation of Study

Human evaluation, combined with our limited number of evaluators, could have led to inter-rater subjectivity in identifying and classifying the errors present. This gave rise to random errors caused by differing viewpoints, resulting in variability in our result.

Future Work

Future work can explore other forms of question generation, such as short-answer and open-ended questions, as well as open-source GPT models like Mistral and Phi-2 since they can be hosted on premise and will allow generation of questions on classified manuals.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to our mentor, Mr How Chang Hong, for his invaluable support and guidance throughout the entire duration of this project. We would also like to thank Ms Esther Chew Mei Ping for her assistance in the administrative part of this project as well as DSTA for providing us with the opportunity to participate in Research@YDSP 2023.

REFERENCES

- [1] Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. *Procedia CIRP*, 119, 84–90. <https://doi.org/10.1016/j.procir.2023.05.002>
- [2] Vaswani, A. (2017, June 12). Attention is all you need. arXiv.org. <https://arxiv.org/abs/1706.03762>
- [3] Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2019). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30(1), 121–204. <https://doi.org/10.1007/s40593-019-00186-y>
- [4] Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. H. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, 6(6), 354–363. <https://doi.org/10.1016/j.nepr.2006.07.002>
- [5] Lee, U., Jung, H. C., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12249-8>
- [6] White, J. (2023, February 21). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. arXiv.org. <https://arxiv.org/abs/2302.11382>
- [7] Costello, E., Brown, M., & Holland, J. C. (2016). What Questions are MOOCs Asking? – An Evidence-Based Investigation. *ResearchGate*. https://www.researchgate.net/publication/289957387_What_Questions_are_MOOCs_Asking_-_An_Evidence-Based_Investigation
- [8] Costello, E., Holland, J., & Kirwan, C. (2018). The future of online testing and assessment: question quality in MOOCs. *International Journal of Educational Technology in Higher Education*, 15(1). <https://doi.org/10.1186/s41239-018-0124-z>
- [9] Ahmad, M. (2019). Analysis of question papers of Physics at secondary level in Pakistan in the light of revised Bloom's Taxonomy. *www.academia.edu*. https://www.academia.edu/40789168/Analysis_of_Question_Papers_of_Physics_at_Secondary_Level_in_Pakistan_in_the_Light_of_Revised_Blooms_Taxonomy
- [10] Momsen, J. L., Offerdahl, E. G., Kryjevskaja, M., Montplaisir, L., Anderson, E. M., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE- Life Sciences Education*, 12(2), 239–249. <https://doi.org/10.1187/cbe.12-08-0130>

APPENDIX 1- Generated questions

Module 1 using Claude:

Here is the 20 question multiple choice quiz on The Nature of Matter, Isotopes, States of Matter and Catalyst:

1. What are the three subatomic particles found in an atom? a) Proton, neutron, electron
b) Nucleus, electron shell, valence electron c) Isotope, molecule, compound d)
Element, mixture, catalyst
2. What charge does a proton have? a) Positive b) Negative
c) Neutral d) Varies
3. What charge does an electron have? a) Positive b) Negative c) Neutral d) Varies
4. What is the name of the chart that arranges elements according to their properties? a)
The quantum table b) The elemental table
c) The periodic table d) The atomic roster
5. True or False: An atom's mass is unrelated to the number of protons, neutrons, and
electrons it contains. a) True b) False
6. What are isotopes? a) Atoms of different elements with different masses b) Atoms of
the same element with different masses
c) Charged particles orbiting the nucleus of an atom d) A state of matter
7. How do isotopes of the same element differ from each other? a) By their electric
charge b) By the number of neutrons c) By the number of electrons d) By the number
of protons
8. How are isotopes of the same element similar? a) They have identical masses b) They
have the same atomic number c) They have a different atomic number d) They have a
different number of protons
9. Which isotope of hydrogen has 1 proton and 2 neutrons? a) Deuterium b) Protium c)
Tritium d) Hydrogen-3
10. What makes tritium an unstable isotope? a) It has too many neutrons b) It has too
many protons c) It is radioactive d) It is a gas at room temperature
11. Which state of matter has particles with the highest degree of motion? a) Solid b)
Liquid c) Gas d) Plasma
12. What are the three primary states of matter for aviation maintenance? a) Solid, liquid,
gas b) Solid, liquid, plasma c) Gas, liquid, plasma
d) Solid, gas, plasma
13. What causes matter to change from one state to another? a) Changes in pressure b)
Addition or subtraction of energy c) Neutron emission d) Nuclear fusion
14. Which state of matter freely conducts electricity? a) Solid b) Liquid
c) Gas d) Plasma
15. Which state of matter has the least motion between particles? a) Gas b) Liquid c)
Plasma d) Solid
16. What does a catalyst do? a) Slows down a chemical reaction b) Stops a chemical
reaction c) Causes or accelerates a chemical reaction d) Absorbs extra heat from a
chemical reaction
17. How is a catalyst affected by the reaction it influences? a) It is consumed over time b)
It exerts a balanced force c) It is not affected d) It decays rapidly
18. Which of the following best describes the role of a catalyst? a) It is the main reactant
b) It is the product of the reaction c) It changes the activation energy d) It causes side
reactions
19. What is the opposite of a catalyst?
a) Base b) Inhibitor c) Reactant d) Product

20. Can a catalyst alter the chemical equilibrium of a reaction? a) Yes, it can favor the forwards or reverse reaction b) No, it accelerates both the forward and reverse reactions equally c) It only accelerates decomposition reactions d) It has no effect on equilibrium

Facts created that the quiz depends on:

1. Atoms contain subatomic particles called protons, neutrons, and electrons
2. Protons have a positive charge, electrons have a negative charge
3. The periodic table arranges elements by their properties
4. An atom's mass relates to the particles it contains
5. Isotopes are atoms of the same element with different masses
6. Isotopes differ by number of neutrons
7. Isotopes of one element share the same atomic number
8. Hydrogen has isotopes like protium, deuterium, and tritium
9. Plasma, gas, and solid are key states of matter in aviation
10. Adding/subtracting energy causes the matter's state to change
11. Plasma freely conducts electricity
12. Catalysts accelerate chemical reactions without being consumed
13. Inhibitors slow down reactions
14. Catalysts reduce activation energy but don't shift equilibrium

Here is the answer key for the quiz:

1. a
2. a
3. b
4. c
5. b
6. b
7. b
8. b
9. c
10. c
11. d
12. a
13. b
14. d
15. d
16. c
17. c
18. c
19. b
20. b

Module 2 using Claude:

Here is the 10 question quiz I generated on statics and kinetics along with the set of facts I used to create the questions and their answer key:

1. Given a force of 200 N at 60° relative to the X-axis, what is the X component of this force rounded to the nearest whole number? A) 173 N B) 100 N C) 115 N D) 150 N
2. A seesaw with 2 people on it is an example of which type of lever? A) First class B) Second class C) Third class D) None of these

3. Which of Newton's Laws states: "When a force acts upon a body, the momentum of that body is changed. The rate of change of momentum is proportional to the applied force." ? A) First B) Second C) Third D) Fourth
4. An object in uniform motion has what defining characteristic?
A) Acceleration changes B) Its velocity changes C) The distance it travels remains the same over a given period of time
D) The force acting on it changes
5. The formula for centripetal force is: A) Force = Mass x Acceleration / Velocity B) Force = Mass x Acceleration x Velocity C) Force = Mass x Acceleration ÷ Radius D) Force = Mass x Velocity² ÷ Radius
6. The speed of an object tells you what about its motion? A) Direction B) Acceleration C) How fast it is moving over a period of time
D) None of these
7. If a pulley allows you to lift a mass 5 times greater than the effort force exerted, how much distance will the effort force move compared to the distance the weight is lifted?
A) The same distance B) 1/5 the distance
C) 5 times the distance D) 1/25 the distance
8. A wheelbarrow is an example of which class of lever? A) First
B) Second C) Third D) Pulley lever
9. What is indicated when an object is said to experience stress? A) A force is applied B) It gains momentum C) It changes velocity D) Internal resistance counteracts an external force
10. Which simple machine works on the principle of redirecting a force rather than gaining a mechanical advantage? A) Inclined plane B) Wedge C) Wheel and axle D) Single fixed pulley
11. What is the formula used to calculate kinetic energy? a) Potential Energy = Mass x Velocity b) Kinetic Energy = 0.5 x Mass x Velocity² c) Kinetic Energy = Mass x Acceleration d) Kinetic Energy = Force x Distance
12. Which type of friction occurs when an object is already moving across a surface? a) Static friction b) Sliding friction
c) Rolling friction d) Starting friction
13. Pascal's law states that pressure applied to a fluid acts: a) Equally in all directions b) Downward due to gravity c) In the direction opposing fluid flow d) Only in perpendicular directions
14. Which variable is directly proportional to static pressure in Bernoulli's principle? a) Velocity b) Density c) Temperature d) Volume
15. What term refers to the internal resistance to flow in a fluid? a) Static pressure b) Specific gravity c) Viscosity d) Density
16. Power is equal to which of the following equations? a) Work x Time b) Force x Velocity c) Force x Distance / Time d) Pressure x Area
17. What unit of measurement is torque given in? a) Foot-pounds b) Inch-pounds c) Pound-inches d) Pounds per square inch
18. What is the primary difference between liquids and gases? a) Density b) Compressibility c) Viscosity
d) Weight
19. True or false: No machine can convert 100% of the input energy into output energy. a) True b) False
20. What type of energy does a compressed spring possess? a) Potential energy
b) Kinetic energy c) Heat energy d) Electrical energy

Set of Facts Used:

1. 200 N force at 60 degree angle
2. Seesaw is a first class lever
3. Newton's second law describes relationship between force, mass and acceleration
4. Uniform motion means constant speed
5. Centripetal force formula
6. Speed defines rate of motion over time
7. Pulleys sacrifice distance for force gained
8. Wheelbarrow is a second class lever
9. Stress is internal resistance to external force
10. Single fixed pulley redirects force
11. Formula for kinetic energy: $\text{Kinetic Energy} = 0.5 \times \text{Mass} \times \text{Velocity}^2$
12. Sliding friction acts on objects already in motion
13. Pascal's law: Pressure applied to a fluid acts equally in all directions
14. Velocity is inversely proportional to static pressure in Bernoulli's principle
15. Viscosity refers to a fluid's internal resistance to flow
16. $\text{Power} = \text{Force} \times \text{Distance} / \text{Time}$
17. Unit of torque: Pound-inches
18. Primary difference between liquids and gases is compressibility
19. No machine is 100% efficient due to losses from friction, heat, etc
20. A compressed spring possesses potential energy

Answer Key:

1. C
2. A
3. B
4. C
5. D
6. C
7. B
8. B
9. D
10. D
11. B
12. B
13. A
14. A
15. C
16. C
17. C
18. B
19. A
20. A

Module 3 using Claude:

Here is a 20 question thermodynamics diagnostic quiz based on the provided textbook excerpt:

1. What method of heat transfer requires physical contact between objects?
a. Conduction b. Convection c. Radiation d. Combustion
2. As temperature increases, the viscosity of a liquid: a. Increases b. Decreases c. Stays the same d. None of the above

3. The amount of heat required to raise the temperature of 1 g of water by 1°C is defined as: a. Specific heat capacity b. Latent heat c. Sensible heat d. Heat of vaporization
4. According to Charles' law, if temperature is held constant, the volume of a gas is: a. Directly proportional to pressure b. Inversely proportional to pressure c. Unrelated to pressure d. Doubled when pressure is doubled
5. Which gas law states that the pressure of a mixture of gases is equal to the sum of the pressures that each gas would exert separately? a. Boyle's law b. Charles' law c. Ideal gas law d. Dalton's Law
6. In a vapor-cycle air conditioning system, what process occurs when compressed refrigerant vapor gives off heat to the outside air and condenses to a liquid? a. Evaporation b. Condensing c. Superheating d. Sublimation
7. Engine efficiency can be increased by: a. Increasing compression ratio b. Reducing friction c. Improving combustion d. All of the above
8. The main difference between conduction and convection is that: a. Conduction transfers heat through direct contact, convection through fluid motion b. Conduction relies on radiation, convection does not c. Convection only works for gases, not liquids d. Conduction works better for gases, convection for liquids
9. Which scale of temperature has the boiling point of water at 212 degrees? a. Fahrenheit b. Celsius c. Kelvin d. Rankine
10. In a piston engine, the thermal efficiency depends largely upon the: a. Engine RPM b. Fuel used c. Compression ratio d. Number of cylinders
11. What is transferred when friction between two surfaces generates heat? a. Nuclear energy b. Electrical energy c. Mechanical energy d. Chemical energy
12. In a turbine engine, what thermodynamic cycle takes place? a. Otto cycle b. Diesel cycle c. Rankine cycle d. Brayton cycle
13. The amount of heat needed to turn 1 kg of a liquid at its boiling point into a vapor with no change in temperature is known as: a. Sensible heat b. Latent heat c. Specific heat d. Heat of fusion
14. Which gas law states that the volume of gas is inversely proportional to pressure when temperature is held constant? a. Charles' law b. Boyle's law c. Ideal gas law d. Dalton's law of partial pressures
15. What is transferred when an electric current flows and encounters resistance, as in an electric stove heating element? a. Chemical energy b. Electrical energy c. Nuclear energy d. Mechanical energy
16. In conduction heat transfer, which metal is the best conductor? a. Lead b. Silver c. Brass d. Steel
17. What term refers to the energy transferred across space in the form of electromagnetic waves? a. Conduction b. Convection c. Radiation d. Combustion
18. In a piston engine, the power stroke begins when: a. The intake valve closes b. Fuel ignition occurs c. The piston passes top dead center d. The exhaust valve opens
19. What is transferred when an exothermic chemical reaction takes place, as in combustion? a. Electrical energy b. Chemical energy c. Nuclear energy d. Mechanical energy
20. Compared to liquids, gases are generally: a. Better heat conductors b. Similar heat conductors c. Poorer heat conductors d. Perfect heat conductors

Set of Facts:

1. Conduction requires direct contact for heat transfer
2. viscosity decreases as temperature increases

3. Specific heat capacity defines the heat needed to raise 1 g of a substance 1 °C
4. Charles' Law: at constant temperature, gas volume is inversely proportional to pressure
5. Dalton's Law: total pressure of mixed gases equals the sum of partial pressures
6. Condensing is when hot compressed refrigerant vapor gives off heat and condenses
7. Engine efficiency can be increased by: higher compression ratio, reducing friction, improving combustion
8. Conduction is direct contact, convection uses fluid motion to transfer heat
9. Fahrenheit scale has water boiling point at 212 degrees
10. Compression ratio largely determines thermal efficiency in piston engines
11. Mechanical energy is transferred when friction generates heat
12. Brayton cycle occurs in turbine engines
13. Latent heat is energy needed to change a liquid at the boiling point to vapor
14. Boyle's Law: Gas volume is inversely proportional to pressure at constant temperature
15. Electrical energy transferred when current meets resistance
16. Silver is the best heat conducting metal
17. Radiation is energy transferred by electromagnetic waves
18. Power stroke in piston engine begins when piston passes top dead center
19. Chemical reaction releases chemical energy
20. Gases are poorer heat conductors than liquids

Answer Key:

1. A
2. B
3. A
4. B
5. D
6. B
7. D
8. A
9. A
10. C
11. C
12. D
13. B
14. B
15. B
16. B
17. C
18. C
19. B
20. C

Module 4 using Claude:

Here is the 20 question multiple choice quiz on optics:

1. What is the speed of light in a vacuum? a) 3×10^8 meters per second b) 186,000 miles per second c) 299,792,458 meters per second d) Faster than anything else
2. What does refractive index compare? a) The frequency of light in two materials b) The wavelength of light in two materials c) The speed of light in a vacuum vs in a material d) The intensity of light in two materials
3. When light bounces directly back from a surface, this is known as: a) Refraction b)

- Diffraction c) Reflection d) Scattering
4. According to Snell's law, what determines how much light refracts between two media? a) The incident angle b) The light's wavelength c) The two refractive indices d) The attenuation
 5. Concave mirrors are used in telescopes to: a) Absorb light b) Refract light c) Diffuse light d) Magnify images
 6. Cladding in an optical fiber: a) Absorbs colors b) Conducts electricity c) Reflects light d) Emits light
 7. Why does fusion splicing produce lower loss connections than mechanical splicing? a) It physically contacts the fiber cores b) It uses a finer polish c) It permeates the cladding better d) It melts the fiber ends together
 8. What are fiber optic connectors used for? a) Absorbing signals b) Focusing light c) Joining cables and equipment d) Amplifying signals
 9. Which causes intensity reduction as a light signal travels through an optical fiber? a) Reflection b) Attenuation c) Refraction d) Diffraction
 10. What does a fiber optic transmitter convert into optical signals? a) Radio waves b) Electrical signals c) Infrared radiation d) X-rays
 11. When light passes from a fast medium to a slow medium, what happens? a) The frequency decreases b) The light accelerates c) The wavelength decreases d) The light refracts toward the normal
 12. What shape lens is thicker in the middle than at the edges? a) Concave b) Convex c) Planoconvex d) Biconvex
 13. What main factor limits fiber optic transmission distances? a) Number of modes b) Connector loss c) Attenuation d) Chromatic dispersion
 14. Where is light lost in a fiber optic splice? a) The buffer coating b) The core c) The cladding d) The jacket
 15. What emits a spark to fuse optical fibers together in fusion splicing? a) Arc generator b) Flash lamp c) Optical fuse d) Plasma welder
 16. Single-mode fiber optic cables: a) Have larger diameter cores b) Propagate multiple light modes c) Are used with infrared light d) Support multiple signals
 17. What technology uses non-spherical lenses? a) Monocentric lenses b) Fresnel lenses c) Aspheric lenses d) Achromatic lenses
 18. Angle of incidence equals what other angle in reflection? a) Angle of refraction b) Angle of diffraction c) Angle of tuning d) Angle of reflection
 19. What type of reflection occurs from a rough surface? a) Regular b) Paraxial c) Diffuse d) First surface
 20. Electromagnetic spectrum radiation with wavelengths between 700-400 nm is: a) X-rays b) Microwaves c) Visible light d) Gamma rays

Set of Facts:

1. The speed of light in a vacuum is 299,792,458 meters per second.
2. Refractive index is the speed of light in a vacuum divided by the speed of light in a material.
3. Reflection is when light changes direction by bouncing off a surface.
4. Refraction is when light bends while passing between materials.
5. Lenses and mirrors can focus light based on their shape and materials.
6. Fiber optics use glass fibers to transmit light signals.
7. Fusion splicing joins fiber optic cables with lower optical loss than mechanical splicing.
8. Connectors link fiber optic cables to transmitters, receivers and equipment.

9. Attenuation in fiber optics reduces signal strength over long distances.
10. Fiber optic data links convert electrical signals to optical signals.
11. When light enters a slower medium at an angle, the wavelength shortens and it refracts towards the normal line.
12. A lens thicker in the middle than at the edges is a convex lens.
13. Attenuation/transmission loss is the main factor limiting fiber optic transmission over long distances.
14. Light is lost in the cladding during fiber optic splicing rather than the core.
15. An arc generator emits a spark to fuse fibers in fusion splicing.
16. Single-mode fiber optic cables propagate only one signal at a time and are used with infrared light.
17. Aspheric lenses means non-spherical lenses.
18. Angle of incidence equals angle of reflection in light reflection.
19. Diffuse reflection occurs when light reflects irregularly from rough surfaces.
20. Visible light wavelengths are approximately 700-400 nanometers.

Answer Key:

1. c
2. c
3. c
4. c
5. d
6. c
7. d
8. c
9. b
10. b
11. d
12. b
13. c
14. c
15. a
16. c
17. c
18. d
19. c
20. c

Please let me know if you need any modifications or have additional feedback on these 10 questions! Once you are satisfied, I can compile the full 20 question quiz with answer key into an organised document for the optics lesson.

Module 5 using Claude:

Here is the quiz on sound and wave motion:

1. What are the three elements required for the transmission and reception of sound? a) Speaker, microphone, amplifier b) Observer, particles, waves c) Source, detector, medium d) Frequency, pitch, loudness
2. Approximately how fast does sound travel in air at 0° C? a) 768 ft/sec b) 100 ft/sec c) 1,087 ft/sec d) 2,000 ft/sec
3. What term describes the distance between two wave crests? a) Frequency b) Wavelength c) Rarefaction d) Doppler effect
4. Which type of wave involves up and down motion perpendicular to the direction the

- wave travels? a) Compressional b) Longitudinal c) Latitudinal d) Transverse
5. The pitch of a sound depends on what characteristic of the sound wave? a) Frequency b) Amplitude c) Wavelength d) Velocity
 6. In general, sound travels _____ in gases compared to liquids and solids a) Louder b) Slower c) Faster d) Softer
 7. What scale is used to measure sound intensity in decibels? a) Fahrenheit b) Celsius c) Kelvin d) Mach
 8. On the decibel scale, an increase of 10 decibels corresponds to _____ as much intensity. a) Half b) Double c) Triple d) Quadruple
 9. The Doppler effect describes changes in sound frequency due to _____. a) Humidity levels b) Source velocity c) Background noise d) Number of observers
 10. Standing waves may occur when sound waves are _____. a) Absorbed b) Reflected c) Refracted d) Dissipated
 11. What is the term for how high or low a sound is perceived? a) Timbre b) Intensity c) Pitch d) Resonance
 12. When two sound waves combine in phase, what phenomenon occurs? a) Diffraction b) Reflection c) Refraction d) Interference
 13. In a medium, particles vibrate _____ as sound waves pass through. a) Randomly b) Vertically c) Horizontally d) Longitudinally
 14. The Mach number represents the ratio of an object's speed to what other velocity? a) Speed of light b) Speed of sound c) Speed of the medium d) Speed of the observer
 15. Sound waves passing through a medium cause _____ in pressure. a) Changes b) Increases c) Decreases d) Equalization
 16. The quality or timbre of a sound depends primarily on the sound wave's _____. a) Wavelength b) Amplitude c) Frequency d) Velocity
 17. Noise canceling headphones use _____ waves to reduce apparent loudness. a) Transverse b) Compression c) Rarefaction d) Interference
 18. In general, the _____ the density of a substance, the faster sound will travel through it. a) Lower b) Higher c) Variable d) Equal
 19. The loudness of a sound depends primarily on the sound wave's _____. a) Rate of change b) Wavelength c) Amplitude d) Frequency
 20. Ultrasonic sounds have _____ that are beyond human hearing range. a) Velocities b) Frequencies c) Wavelengths d) Amplitudes

Facts Set:

1. Sound requires a source, medium and detector
2. Speed of sound in air at 0°C is 1087 fps
3. Wavelength is distance between wave crests
4. Transverse waves involve perpendicular motion
5. Pitch depends on frequency
6. Sound is slower in gases than liquids/solids
7. Sound intensity is measured in decibels
8. 10 dB increase is twice the intensity
9. Doppler effect involves source velocity changes
10. Standing waves from wave reflection
11. Pitch describes how high/low sound is perceived
12. In phase waves cause interference
13. Particles vibrate longitudinally
14. Mach number is ratio to speed of sound

15. Sound waves cause pressure changes
16. Timbre depends on frequency
17. Noise cancellation uses interference
18. Higher density increases sound speed
19. Loudness depends on amplitude
20. Ultrasonic frequencies are beyond hearing

Answer Key:

1. c
2. c
3. b
4. d
5. a
6. b
7. b
8. b
9. b
10. b
11. c
12. d
13. d
14. b
15. a
16. c
17. d
18. b
19. c
20. b

Please let me know if you need any clarification or have additional feedback on the format. I'm happy to refine my quiz generation approach further.

APPENDIX 2 - Item-writing guidelines

1. All options should be grammatically consistent with the stem and should be parallel in style and form. Non-grammatically correct options provide cues to the students who easily eliminate distractors that do not flow grammatically with the stem.
2. Each MCQ should have a clear and focused question. Teachers should avoid using MCQs with unfocused stems which do not ask a clear question or state a clear problem in the sentence completion format.
3. Each MCQ should have the problem in the stem of the question, not in the options. The options should not be a series of true/false statements.
4. The basic format for MCQs is the single best answer. Therefore, ensure that questions have one, and only one, best answer.
5. Avoid gratuitous or unnecessary information in the stem or the options. If a vignette is provided with the MCQ, it should be required to answer the question.

6. Questions and all options should be written in clear, unambiguous language. Poorly worded or ambiguous questions can confuse even knowledgeable students and cause them to answer incorrectly.
7. Make all distractors plausible as plausible distractors are vital to high quality MCQs. Students who do not know the material increase their chances of guessing the correct option by eliminating implausible distractors.
8. Avoid repeating words in the stem and the correct option. Similar wording allows students to identify the correct option without knowing the material.
9. Avoid providing logical cues in the stem and the correct option that can help the student to identify the correct option without knowing the material. An example of a logical cue is asking students to select the most appropriate pharmaceutical intervention for a problem and only having one or two options which are actually pharmaceutical interventions.
10. Avoid convergence cues in options where there are different combinations of multiple components to the answer. Question writers tend to use the correct answers more frequently across all options and students will identify as correct the answer in which all components appear most frequently.
11. All options should be similar in length and amount of detail provided in the option. If one option is longer, includes more detailed information, or it contains more complex language, students can usually correctly assume that this is the correct answer.
12. Avoid the use of negatives (e.g., not, except, incorrect) in the stem as they poorly assess students actual knowledge. If teachers wish to assess contraindications, the questions should be worded clearly to indicate that this is what is being assessed.
13. Avoid the use of “all of the above” as the last option. Students can easily identify if this is the correct answer by simply knowing that at least two of the options are correct. Similarly, they can eliminate it by knowing if only one of the options is incorrect.
14. Avoid the use of “none of the above” as the last option as it only measures students’ ability to detect incorrect answers. Furthermore, if “none of the above” is the correct option, the teacher must be certain that there are no exceptions to any of the options that the student may detect.
15. Avoid fill-in-the-blank format whereby a word is omitted in the middle of a sentence and the student must guess the correct word. All options should be placed at the end of the stem.