Supplemental Table: Pros and Cons of Key Feature Formats

| Key Features to assess in these questions: * Discriminates between fruits and vegetables
* Identifies the sugar content of common fruits and vegetables
 |
| --- |
| Question Style | Purpose | Pros/Cons\* | Sample |
| Pick “N” from list(Short Menu / Multiple correct | To assess if the candidate can decide which are the correct answers among a series of possible or reasonable alternativesThis is the most common type of KFQ | Pro- good items for assessing clinical reasoning - authentic, problem-based approach- results correlate with a teaching approach that promotes clinical reasoningCon- some topics not amenable to a long list of options- some systems will not allow for scoring of multiple correct answers- requires additional consideration of scoring options | Which of following are vegetables?AppleArugulaCeleriacGrapeOrangeRadishPearPlum RosehipsTomato  |
| Fill-in-the-blank(sometimes referred to as short answer; though in the key feature context this would be a 1-2 word answer) | To assess if the candidate can recall information. This type of question can be used successfully for identifying the most likely diagnosis; It is important that subsequent questions that are related to the same scenario do not cue the candidate to the answer to this question. | Pro-uses recall- avoids cuing- avoids the need for a single answer- higher validity than MCQ- easy to write Con- takes instructor time to mark- marking may be subjective- exact answers unlikely due to misspellings, order of words, synonyms, which make computer scoring challenging | Name one vegetable: |
| Long Menu Format  | To assess if the candidate can recall information. This is similar to the “fill-in-the-blank” question, however the options are pre-determined as an extremely long, alphabetically organized list (e.g. 500 items). This is best used within a computer administered system as the computer will provide options as the candidate starts typing in their answer. The options will include both correct and incorrect items and the candidate must select the correct items.  | Pro- uses recall- avoids cuing since the candidate must start typing in their answer to see options - promotes study directed at clinical reasoning- assesses clinical decision making - is easier to score than short answer- is more discriminating than traditional MCQ or Pick N formatsCon- it is necessary that all correct options, synonyms, as well as incorrect responses that include common misconceptions, be provided in the list.- options must be clear and unambiguous- this format is not feasible for a pencil and paper exam. - response time longer than for traditional MCQ- candidates must correctly spell the beginning of the word they are looking | Patient scenario. What is the most likely diagnosis in this case? Long menu includes the entire International Classification of Diseases. |
| Match | To assess if the candidate can decide how to sort a list of items into the correct group or categoryThis type of question works well in an online administration, where the candidates can “drag” an item to the appropriate column. | Pro- easy to construct- format is understood- good for testing concepts, principles & procedures- efficient in terms of testing timeCon- using the same number of options as items results in cuing of answers.- non-homogenous options  | Organize the list into fruits and vegetables:Apple, carrot, plum, radish, tomatoFruit Vegetables |
| Short Menu format (sometimes called Extended Matching)  | To assess if the candidate can correctly determine the one best answer for a situation | Pro- easy to write- easier to score than short answer- cognitive process is understanding or application- decreased cuing- guessing is not an option when longer menus are used- more discriminating at the lower end of the score scale and higher reliability than conventional MCQ- similar to an un-cued free-response itemCon- the matching items in a set may cue each other- all options must be from a homogenous class- it is possible to oversample a small part of the curriculum/content area- test sheets do not always allow for the higher number of options; this is not an issue for electronic administration of a test | Which one of the following is a fruit? BeetBroccoli CarrotsCauliflowerCucumberKaleLettuceOnionPlumRadishSweet potato |
| Multiple Choice (One-best-answer/Traditional MCQ) | To assess if the candidate can decide which is the one correct answer among a series of possible or reasonable alternatives | Pro- when well written evokes a higher cognitive level-when using a 3-option MCQ item development time is reduced (over 4 or 5 options) and also reading time is reduced, allowing for more items on a test. - different response alternatives can provide diagnostic feedback - reliable test scores- easy to scoreCon- item-writers tend to favor lower cognitive level items as they are easier to write- require a high level of reading ability, which can be a challenge for candidates writing in a second language- particularly subject to cuing | Does the beet, the carrot, or the sweet potato, have the most sugar content by weight?BeetCarrotSweet potato |
| Multiple True/False | To assess if the candidate knows true from false answers  | Pro- preferred by students compared to Complex Multiple-Choice- avoids disadvantages of Complex Multiple-Choice- higher reliability as compared to Complex Multiple-Choice - efficient in item development and reading timeCon- tend to promote testing of recall- guessing is still influential- not possible to address degrees of truth or fallacy; answers *must* be clearly true or false- potential cuing/dependence of each option on the others reduces the ability to assess test reliability | Each item in the list is a fruit. (True or False)An apple (T/F)A carrot (T/F)A plum (T/F)A radish (T/F) |
| \* (Haladyna 2004 Chapter 5, Downing and Haladyna 2006 Chapter 12, Rotthoff, Baehring et al. 2006, Hrynchak et al. 2014, Cerutti, Blondon et al. 2016) |

# References

Cerutti, B., K. Blondon and A. Galetto (2016). "Long-menu questions in computer-based assessments: a retrospective observational study." BMC Medical Education **16**(1): 55.

Downing, S. and T. A. Haladyna (2006). Handbook of Test Development. Mahwah, NJ, Lawrence Erlbaum Assoc. Inc.

Haladyna, T. M. (2004). Developing and validating multiple-choice test items. Hillsdale, NJ, Lawrence Erlbaum Associates.

Hrynchak, P., S. Glover Takahashi and M. Nayer (2014). A growing literature on key feature questions for assessment of clinical reasoning (poster). 16th Ottawa Conference / 12th Canadian Conference on Medical Education. Ottawa, ON.

Hrynchak, P., S. Glover Takahashi and M. Nayer (2014). "Key-feature questions for assessment of clinical reasoning: a literature review." Medical Education **48**: 870-883.

Rotthoff, T., T. Baehring, H. D. Dicken, U. Fahron, B. Richter, M. R. Fischer and W. A. Scherbaum (2006). "Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial." BMC Medical Education **6**(Journal Article): 50.