Appendix 4. Clinical Scenario Evaluation Framework

Dimension	Score Range	Evaluation Criteria
Quality	0-5	0: Did not respond to prompt as instructed 1: Minimal adherence to instructions 2: Partial adherence with significant omissions 3: Adequate adherence with minor omissions 4: Strong adherence with complete response 5: Exemplary adherence exceeding expectations
Complexity	0-5	0: Completely inappropriate for medical students 1: Significantly below appropriate complexity 2: Somewhat below appropriate complexity 3: Appropriate complexity for year 1 medical students 4: Challenging but appropriate complexity 5: Optimal complexity with graduated difficulty
Relevance	0-5	0: No relation to Histology Urinary system course objectives 1: Minimal relevance to course objectives 2: Partial relevance with significant gaps 3: Adequate relevance to core objectives 4: Strong relevance to multiple objectives 5: Comprehensive relevance to course curriculum
Correctness	0-5	0: Contains critical factual errors 1: Contains major factual inaccuracies 2: Contains several minor factual errors 3: Generally accurate with isolated minor errors 4: Highly accurate with negligible errors 5: Completely accurate with current medical knowledge
Variety	0-5	0: Complete repetition of previous responses 1: Minimal variation from previous responses 2: Limited variation in approach or content 3: Moderate variation from previous responses 4: Substantial variation in approach and content 5: Complete originality across all responses

1. If the Question Is "Detached" from the Scenario

Situation

- The scenario (e.g., about a 20-year-old with elbow pain) is followed by an MCQ that **references a completely different** patient or condition (unrelated to the original scenario).
- Essentially, the question **does not flow** from the scenario, or it addresses an entirely different clinical context.

Rubric Implications

1. **Quality of Response to the Prompt**: The scenario/Q set is not internally consistent; it fails to follow instructions requiring it to be a cohesive set. This likely lowers the score.
2. **Complexity**: Might still be moderate or adequate, but the disjointedness suggests poor structure.

3. **Relevance**: The question is not relevant to the scenario's content.
4. **Correctness**: The question could still be factually correct, but it's off-topic.
5. **Variety**: If the mismatch is frequent or repeated, variety may also be low.

Likely Score Range

- **1–2** if the question is severely off-topic or the scenario is no longer coherent.
- If there are *some* relevant details but an obvious mismatch, possibly a **3** at best.
- **0** only if it's completely nonsensical or no meaningful scenario is provided at all.

---

2. If the Scenario "Gives Away" the Answers

Situation

- The scenario is so detailed or **explicit** that the correct answer choice becomes **obvious** (e.g., "The patient's troponin levels are elevated, indicating an MI," and then an MCQ "What is the most likely cause of chest pain?" with "Myocardial infarction" as an option).
- This can mean the item lacks challenge/complexity.

Rubric Implications

1. **Quality of Response to Prompt**: The scenario addresses the topic, but might not meet typical exam standards because it's leading students too strongly.
2. **Complexity**: If the question is trivialized by giving away the key detail, complexity is low—there's not much reasoning required.
3. **Relevance**: Could still be relevant, but it's **too easy** or telegraphed.
4. **Correctness**: The scenario might still be correct, but it's not testing knowledge thoroughly.
5. **Variety**: If all questions follow the same pattern of giving away the answer, it reduces variety.

Likely Score Range

- Possibly **3** (adequate) if the scenario is at least correct and on-topic, but **some** points deducted for poor test utility.
- Could be **4** if everything else is strong (the text is relevant, well-structured, correct) but the question is too transparent.
- Rarely a **5**, because an ideal question typically requires some reasoning.

---

3. If the Answer Choices (1–2) Are Not Correct or Not the Best

Situation

- The item's MCQ or scenario has flawed answer choices:
  - Maybe multiple answers appear "correct," or the provided "correct answer" is **actually incorrect** or suboptimal.
  - Or the question is poorly constructed, so none of the offered choices truly fits.

Rubric Implications

1. **Quality of Response to the Prompt**: The question itself fails to meet typical standards for clarity.
2. **Complexity**: Could still be complex or trivial. The main problem is accuracy.
3. **Relevance**: Possibly still relevant to the scenario, but overshadowed by correctness problems.
4. **Correctness**: This is **majorly** impacted. If the correct answer choice is not truly correct, it's a factual error.
5. **Variety**: The scenario might be repeated or might not; the bigger issue is correctness.

Likely Score Range

- **1–2** if the entire question or answer set is flawed, but you can see some attempt to be on-topic.
- **0** if it is entirely unusable—completely missing the mark in both the scenario and question.
- **3–4** if only **minor** inaccuracies are there (e.g., slight wording confusion). But generally, major correctness issues push the score lower.

Overall Guidance for the 0–5 Rating

- **0**: Almost no redeeming qualities (completely off-topic, incorrect, or nonsensical).
- **1–2**: Very poor or weak. Severe mismatch, major factual errors, or scenario-question disconnect.
- **3**: Adequate. Generally coherent and correct but with noticeable flaws like partial irrelevance or lacking complexity.
- **4**: Good. Mostly correct, relevant, structured. Possibly too easy or has small mistakes but otherwise strong.
- **5**: Excellent. Meets high-quality standards in all rubrics—on-topic, challenging but correct, and well-varied from prior attempts.