GENERAL PRINCIPLES

Open-ended, essay or short answer questions?

- Open-ended questions differ from multiple-choice questions, because the student is not provided with options to choose from but must generate concepts and information.
- The literature makes further distinctions between essay questions (answer is several paragraphs) and short answer questions (a sentence or two)
- In the context of the Bridges curriculum, most exam questions will require answers of a few sentences to a couple of paragraphs long. To minimize confusion, we will refer to the questions as “open-ended”.

Why Open-ended Questions?

- Well-designed open-ended questions promote deep learning, and allow assessment of analytical and critical thinking.
- Written questions offer students the opportunity to demonstrate their own judgment and vocabularies, and thereby reflect their “real-life” problem solving skills and application of knowledge more accurately than multiple-choice questions.
- Open-ended questions may promote different study habits: students are more likely to focus on broad issues, general concepts, and interrelationships.

Potential limitations:

- Open-ended questions permit only a limited sampling of content learning due to the time required for students to respond – so pick the content carefully
- They can favor students who possess good writing skills – but they also promote good writing skills and allow students to practice clear writing
- Students can go off on tangents or misunderstand the main point of the question – so ensure question is clear

II. GUIDELINES FOR WRITING OPEN-ENDED QUESTIONS AND EXAMS

The goal is to write questions that are worth students’ time and reflect important content, so that students will recognize and understand the value.

In the Bridges curriculum we use open-ended questions for

- Weekly required checkpoints for formative assessment (required weekly questions, 2 vignettes (cases or experiences), students choose one vignette, each vignette has 2-4 associated questions)
- Summative assessment (1-2 times per block, 5-6 vignettes with 2-5 questions each); total 14-16 questions on the exam.

Weekly formative checkpoint questions should incorporate all best practices and prepare students for the summative exam. Blocks should create a bank of questions that includes back-up questions for remediation and make-up exams, and ensure variation over the years to promote exam security. For the same reason, at least 2-3 new questions should be rotated in on an annual basis. New questions also help represent new and evolving content.

General best practices for open-ended questions:

- Questions should reflect the language that is used in the course materials. Test questions should be stated in simple, clear language.
- Proofread exam questions carefully and have another person proofread as well.
UCSF Bridges Curriculum Faculty Development - Writing Open-ended Questions

- Weekly formative checkpoint exam questions should be of equal difficulty to those in the summative exam
- Do not use optional questions on summative exams – this affects reliability of grading and students may not study all the course material if they know they will have a choice.
- Be careful about having too many separate questions that depend upon answers or skills required in previous questions. A student’s initial mistake will be perpetuated over succeeding questions or tasks, penalizing the student repeatedly for one error.
- Keep questions free of nonfunctional material and extraneous clues.
- Test questions should also consider diversity and be free of race, ethnic, sex and other bias. They can be written to address issues such as race, ethnicity, and sexual orientation that affect health care and were taught in the course.

To create a high-quality exam, the questions in an exam should:
- Assess achievement of instructional learning objectives
- Measure important aspects of the subject (concepts and conceptual relations) in the context of authentic scenarios (patient cases, research scenarios, etc.)
- Accurately reflect the emphasis placed on important aspects of instruction
- Emphasize deep learning and application rather than recall
- Test on multiple disciplines in an integrated manner
- Have multiple sub-parts that test different discipline-based elements of the case

Bloom’s Taxonomy and Open-ended Questions:
Bloom’s taxonomy is a framework to categorize educational objectives and assessments according to complexity, with recall of facts at the bottom and more complex tasks such as analyzing and evaluating at the top.

![Bloom's Taxonomy](image)
### Table 1: Bloom’s Taxonomy with verbs and sample question frames

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
<th>Verbs</th>
<th>Sample question frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remember*</td>
<td>Recognizing and recalling information, including terms, definitions; facts, principles, theories; methods and procedures</td>
<td>Define, List, State, Label, Name, Describe</td>
<td>List three reasons for ...? How do we ...? Which principle explains ...? What is meant by ...?</td>
</tr>
<tr>
<td>Understand*</td>
<td>Understanding the meaning of information, including restating (in own words); translating from one form to another; or interpreting, explaining, and summarizing.</td>
<td>Restate, Paraphrase, Explain, Summarize, Interpret, Illustrate</td>
<td>Restate in your own words ...? Explain the principle of ...? Summarize the main features of ...?</td>
</tr>
<tr>
<td>Apply</td>
<td>Applying general rules, methods, or principles to a new situation, including classifying something as a specific example of a general principle or using a formula to solve a problem.</td>
<td>Apply, Demonstrate, Use, Compute, Solve, Predict</td>
<td>How is ... an example of ...? How is ... related to ...? Why is ... significant?</td>
</tr>
<tr>
<td>Analyze</td>
<td>Identifying the organization and patterns within a system by identifying its component parts and the relationships among the components.</td>
<td>Compare, Contrast, Categorize, Distinguish</td>
<td>Compare .. to ...? Classify ... according to ... Outline/diagram ...</td>
</tr>
<tr>
<td>Evaluate</td>
<td>Using evidence and reasoned argument to judge how well a proposal would accomplish a particular purpose; resolving controversies or differences of opinion.</td>
<td>Judge, Appraise, Recommend, Justify, Defend, Criticize, Evaluate</td>
<td>How would you argue for or against ...? How would you decide about ...? What priority would you give ... and why?</td>
</tr>
<tr>
<td>Create</td>
<td>Discovering/creating new connections, generalizations, patterns, or perspectives; combining ideas to form a new whole.</td>
<td>Develop, Create, Propose, Formulate, Design, Invent</td>
<td>What would you infer from ...? What ideas can you add to ...? How would you create a ...?</td>
</tr>
</tbody>
</table>

*shaded levels provided as example only – open-ended questions should be at the higher levels of Bloom’s taxonomy*
Constructing Open-ended Questions:

The goal is to write open-ended questions that measure higher cognitive processes – apply, analyze, evaluate or create/synthesize.

1. Questions will typically be based on a (patient) vignette, which should be concise and contain all relevant information for the learner to be able to answer the questions, without extraneous or redundant information.
2. In writing questions, carefully choose verbs that elicit the kind of thinking you want the students to demonstrate, based on Bloom’s taxonomy (Table 1).
   - Words like discuss and explain can be ambiguous. If you use explain or discuss, then give specific instructions as to what points should be explained or discussed.
   - Commonly used words like define and describe are associated with recall.
   - You can either phrase questions as statements using any of the verbs listed, or as true questions using the type of wording listed under question frame.
   - Note that the table contains examples of question frames for all levels; but in the Bridges Curriculum we aim to use open-ended questions to assess application and above.
   - *If you can answer questions without reading the vignette, they are likely recall only*
3. Since we recognize that content knowledge is important, we encourage you to state explicitly what knowledge students need to demonstrate while they are providing answers to higher-level questions. For example: “What medication would you prescribe to the patient in this vignette? Explain your answer [application] and include a description of the mechanism of action of this drug in your answer [content knowledge].”
4. Create a rubric as you write the question (see pages 5-6) it helps ensure you are truly assessing application of knowledge in addition to content knowledge. You don’t have to think of all the possible answers upfront, but outlining the rubric at time of question writing saves time and confusion with grading later.
5. Formative OEQs will not be graded, but should prepare students for Summative OEQs so should really be very similar. So think about what a rubric might look like even when writing formative OEQs.
6. For each question, be sure to link the following:
   - Learning objective(s)
   - Disciplines addressed
   - Cognitive skill level (Bloom’s level)
III. RUBRIC PRINCIPLES AND GUIDELINES:

- Rubrics are often classified as holistic (to give an overall sense of student performance) versus analytic (to detail where students perform well). The UCSF Bridges rubrics have elements of both, but are mostly holistic. This combination allows for specific feedback to students, but also allows for flexibility for the block directors to decide what constitutes a good answer.

- Open-ended questions in the Bridges curriculum aim to test higher levels of Bloom Taxonomy objectives (apply, analyze, evaluate etc), but we recognize that content knowledge is important. The highest score can only be obtained if an answer is correct both in application and content knowledge. If a student only demonstrates content knowledge this will result in a lower score than demonstration of higher levels of ability according to Bloom’s taxonomy.

- Questions need to be constructed in such a way that the student knows to demonstrate both the content knowledge and application of knowledge

- The rubric divides the scores in three columns, with the highest scores for demonstration of application and other higher level cognitive processes and adequate content knowledge, middle scores for demonstration of application with limited or absent content knowledge OR for demonstration of application and knowledge but with significant errors in the answer, and the lowest scores if content knowledge only is provided. An item that remains unanswered or with an answer that is so limited that the quality cannot be judged will receive a score of 0

- Within each column, there are two score options (1-2, 3-4, 5-6). How to distinguish between these two options will vary by test item, and may not become apparent until you have some student answers to compare. In general, a score of 3 should be given to an answer in which the application is correct, but this seems to be based on guess-work or there are major limitations in the reasoning and/or content knowledge. A score of 4 would be appropriate if the ultimate answer is incorrect due to a misstep along the way, but content knowledge is overall adequate and the reasoning is adequate. To ensure transparency and clarity for the students it is however essential that faculty score answers similarly, and come to an agreement of what constitutes a higher versus a lower score.

- Since not all exams will have the same number of total questions, the total score across questions will be normalized to 100

- In creating the rubric for a specific question, it is good to have a model answer in mind, but the rubric is most useful if it contains the essential elements for each answer category, rather than all potential answer options. Constructing a rubric while creating the question helps ensure that the questions is complete and at the right Bloom’s level.

- In addition to knowledge and analytic thinking, students will be assessed on writing skills, contributing to max 1/6 of the total score. Anchors to help with rating writing skills are included in the rubric, and should be used to adjust the score on each item. If the score for answer content does not align with the corresponding description of writing style, content score should be decreased by 1 point. See the rubric template and example rubrics for further explanation.
Table 2. Rubric Template for scoring of analytic thinking and knowledge

<table>
<thead>
<tr>
<th>Objective/Score criterion</th>
<th>Meets expectations</th>
<th>Borderline achievement of expectations</th>
<th>Does not meet expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic skills and content knowledge</td>
<td>Demonstrate ability to apply/evaluate/analyze/create with appropriate and complete content knowledge (6-5 pts.)</td>
<td>Demonstrate ability to apply/evaluate/analyze/create but limited content knowledge or answer has errors/is incomplete (4-3 pts.)</td>
<td>Demonstrate content knowledge only but does not apply/evaluate/analyze/create (2-1 pts.)</td>
</tr>
<tr>
<td>Writing style component</td>
<td>Explanation of elements of model answer that contribute to the score (knowledge and application), and approximate delineation of 6 vs 5 pt score</td>
<td>Explanation of elements of answer that contribute to the score (has to have some application), and approximate delineation of 4 vs 3 pt score</td>
<td>Explanation of elements of answer that contribute to the score (can be content knowledge only), and approximate delineation of 2 vs 1 pt score</td>
</tr>
<tr>
<td>Organization and clarity of content and explanations are clear with natural flow</td>
<td>Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity</td>
<td>Writing characterized by lack of clarity and difficulty following organization and flow of argument.</td>
<td></td>
</tr>
</tbody>
</table>

0 points will be assigned if the student did not answer the question, or the answer was so limited that no score can be assigned.

If score for content does not align with corresponding description of writing style, content score should be decreased by 1 point. For example, a student may have all the elements required to score 4 points on analytic skills/content knowledge (middle column, borderline achievement), but if the writing style is characterized by lack of clarity (right column) the total score for this item will be decreased by 1 point to 3 points.
**UCSF Bridges Curriculum Faculty Development - Writing Open-ended Questions**

**EXAMPLE QUESTIONS WITH RUBRIC ADAPTED FROM ABC**

**Disciplines: Physiology, Medicine**

**System: Pulmonary**

1. A 58 year-old man presents with hypoxemia 1 week after a myocardial infarction. Vital signs are: BP 75/55, HR 120, RR 22. Physical examination reveals a III/VI holosystolic murmur and clear lungs. A chest x-ray is normal. The following values are obtained: Hemoglobin: 15 grams/100 ml blood ABG drawn on 100% Oxygen: pH 7.39/PaCO2 30 mmHg/PaO2 50/ Saturation 85%. A Pulmonary Arterial catheter is inserted and mixed venous blood is obtained: Mixed venous blood: PmvO2 = 35 with a saturation of 68%.

   **A. Early learner:**
   Calculate the approximate percentage of shunt fraction:
   - Provide the formula used in your calculation (knowledge)
   - Calculate the shunt fraction (apply)

   **Learning objective:** Define shunt and apply the shunt equation

   **B. More advanced learner:**
   You suspect the patient may have a VSD as explanation of his findings. Provide a rationale for this diagnosis based on the both the symptoms and the numbers provided, using appropriate formulas to support your argumentation with calculations as needed.

   **Learning objectives:** Define shunt and apply the shunt equation; Define respiratory failure and distinguish between acute and chronic respiratory failure

<table>
<thead>
<tr>
<th>Objective/Score criterion</th>
<th>Meets expectations</th>
<th>Borderline achievement of expectations</th>
<th>Does not meet expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demonstrate ability to apply/evaluate/analyze/create with appropriate and complete content knowledge (6-5 pts.)</td>
<td>Demonstrate ability to apply/evaluate/analyze/create but limited content knowledge or answer has errors/is incomplete (4-3 pts.)</td>
<td>Demonstrate content knowledge only but does not apply/evaluate/analyze/create (2-1 pts.)</td>
</tr>
</tbody>
</table>
| **1A**                    | Provides the correct formula and calculates the shunt fraction
   - all correct (6)
   - with minor error (5)
   **Writing:** Organization and clarity of content and explanations are clear with natural flow | • Calculates the shunt fraction and provides the formula but has errors in either one (4)
   • Calculates the shunt fraction but does not provide the formula, or both have major errors, eg. completely incorrect formula (3)
   **Writing:** Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity | Provides the formula correctly (2)
   Provides the formula, with minor errors (1)
   **BUT**
   Does not calculate shunt fraction
   **Writing characterized by lack of clarity and difficulty following organization and flow of argument.** |
### 1B

<table>
<thead>
<tr>
<th>Objective/Score criterion</th>
<th>Meets expectations</th>
<th>Borderline achievement of expectations</th>
<th>Does not meet expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide the correct diagnosis, correct argumentation (uses the correct symptoms in argumentation) based on correct knowledge content (correct formula and calculations)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
  • all correct (6)  
  • with minor error (5)  
  Writing: Organization and clarity of content and explanations are clear with natural flow |  
  • Provides the correct diagnosis but argumentation is not correct and/or uses incorrect formula (4)  
  • Provides an incorrect diagnosis with flaws in argumentation or correct diagnosis with major flaws in diagnosis and incorrect formula (3)  
  Writing: Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity | Provides the correct diagnosis and formula (2)  
 Provides either the correct diagnosis or the formula (1)  
 BUT  
 Does not provide argumentation and does not calculate shunt fraction  
 Writing characterized by lack of clarity and difficulty following organization and flow of argument. |

**0 points will be assigned if the student did not answer the question, or the answer was so limited that no score can be assigned.**

**If score for content does not align with corresponding description of writing style, content score should be decreased by 1 point.** For example, a student may have all the elements required to score 4 points on analytic skills/content knowledge (middle column, borderline achievement), but if the writing style is characterized by lack of clarity (right column) the total score for this item will be decreased by 1 point to 3 points.

2. A 65 year-old man with emphysema is intubated and place on a mechanical ventilator for respiratory failure. Now, 2 days later, he has the following values  

<table>
<thead>
<tr>
<th>pH: 7.32</th>
<th>PaCO2: 50 mmHg</th>
<th>PaO2: 110 mmHg</th>
</tr>
</thead>
</table>

On ventilator settings of  

<table>
<thead>
<tr>
<th>Rate: 12 breaths/minute</th>
<th>TV: 400 ml</th>
<th>FiO2: 50%</th>
</tr>
</thead>
</table>

**A. Early learner**

Identify the acid base abnormality and explain your rationale, comparing the provided values for pH and pCO2 with normal values

  o Provide normal values for pH and pCO2 (knowledge) and the terminology used for deviations in pH with definitions

  o Describe the acid base abnormality in this case and explain your rationale (apply)

**B. More advanced learner:**

How do these numbers help you decide about the patient’s adequacy of gas exchange? Provide your opinion of this patient’s gas exchange, supporting it with a description of the patient’s acid-base status and A-a gradient as compared to normal, providing the values for the numbers that you base your assessment on.
## UCSF Bridges Curriculum Faculty Development - Writing Open-ended Questions

<table>
<thead>
<tr>
<th>Objective/Score criterion</th>
<th>Meets expectations</th>
<th>Borderline achievement of expectations</th>
<th>Does not meet expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective/Score criterion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2A</strong></td>
<td>Describes the acid-base abnormality correctly, explains the rationale with correct definitions, correct normal values</td>
<td>• Describes the acid-base abnormality correctly, but mistakes in definitions and/or normal values (4)</td>
<td>Provides correct definitions/normal values (2)</td>
</tr>
<tr>
<td></td>
<td>• all correct (6)</td>
<td>• Describes the acid-base abnormality incorrectly OR describes acid-base abnormality correctly but does not provide definitions or normal values (3)</td>
<td>Provides definition/normal values, with minor errors (1)</td>
</tr>
<tr>
<td></td>
<td>• with minor error (5)</td>
<td>Writing: Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity</td>
<td>BUT</td>
</tr>
<tr>
<td></td>
<td><strong>Writing:</strong> Organization and clarity of content and explanations are clear with natural flow</td>
<td></td>
<td>Does not describe the acid-base abnormality</td>
</tr>
<tr>
<td></td>
<td><strong>Writing:</strong> Correctly identifies respiratory acidosis correctly calculates the Aa gradient, provides normal values and compares and contrasts these between patient and normal, and provides and explains judgment regarding gas exchange</td>
<td>• Provides and explains judgment regarding gas exchange but uses incorrect calculations and/or normal values (4)</td>
<td>Provides the correct definitions for acid-base abnormalities, correct calculation of A-a gradient, correct normal values</td>
</tr>
<tr>
<td></td>
<td>• all correct (6)</td>
<td>• Provides an incorrect judgment with flaws in argumentation or correct judgment with major flaws in diagnosis and incorrect calculations (3)</td>
<td>• all correct (6)</td>
</tr>
<tr>
<td></td>
<td>• with minor error (5)</td>
<td>Writing: Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity</td>
<td>• with minor error (5)</td>
</tr>
<tr>
<td></td>
<td><strong>Writing:</strong> Correctly identifies respiratory acidosis correctly calculates the Aa gradient, provides normal values and compares and contrasts these between patient and normal, and provides and explains judgment regarding gas exchange</td>
<td><strong>Writing:</strong> Organization and clarity of content and explanations are clear with natural flow</td>
<td>BUT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Writing: Reasonable organization and clarity of content and explanations, with small deficiencies in organization and/or clarity</td>
<td>Does not describe provide and explain a judgment regarding gas exchange</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Writing:</strong> Correctly identifies respiratory acidosis correctly calculates the Aa gradient, provides normal values and compares and contrasts these between patient and normal, and provides and explains judgment regarding gas exchange</td>
<td>Writing characterized by lack of clarity and difficulty following organization and flow of argument.</td>
</tr>
</tbody>
</table>

---

0 points will be assigned if the student did not answer the question, or the answer was so limited that no score can be assigned.

If score for content does not align with corresponding description of writing style, content score should be decreased by 1 point. For example, a student may have all the elements required to score 4 points on analytic skills/content knowledge (middle column, borderline achievement), but if the writing style is characterized by lack of clarity (right column) the total score for this item will be decreased by 1 point to 3 points.