How to Write Good Multiple-choice Exam Questions

We had no formal training to draw from when we wrote our first high-stakes exam, and most research faculty seem to be in the same position. In the next few pages, we share what we have learned over the past few years about exam-writing, and offer suggestions (with examples) to help you construct exams that effectively assess student learning.

General Philosophy:

A good exam has three major goals:

1. **Good validity**: it matches the learning goals of the course. For example, if you want students to learn how a kinase receptor signaling pathway works, the exam should ask a question that tests their mastery of this concept.

2. **Appropriate difficulty**: it tests reasoning and problem-solving. An exam should be difficult because it requires students to apply class information and analyze data. But many exams are difficult because they are long, confusing, or ask about irrelevant minutiae of the course.

3. **Good discrimination**: it accurately measures levels of understanding. An exam should separate out A students from B students, and students who understand the information from students who do not.

These are worthy goals, but difficult to accomplish by a researcher with heavy demands on their time, little grading manpower, and versions required by multiple sections. Given these constraints we suggest the following steps to make your exam-writing efficient, productive and successful.

Step 1: Articulate learning goals for each lecture

Our old way of writing exam questions was to look through each lecture’s PowerPoint and design questions around material presented. This led us to write more questions associated with recall rather than comprehension or application of the broader concepts. When developing an online version of our class we identified a set of “learning goals” for each lecture. This helped us define the key concepts covered in each lecture and what we wanted students to actually learn.

So, even though it may sound like busy work, try writing 2-4 statements for each lecture that indicate what you want the students to learn. We have found the following format to be most useful: “By the end of this lecture a student should be able to recognize / describe / predict / draw...” For example, we might write, “by the end of this lecture a student should be able to describe the relationship between free and bound ribosomes and the proteins they synthesize.” We have found these learning goals easier to measure than ones that say, “A student should understand...”

We then design exam questions that test acquisition of this learning goal.
Step 2: Write questions to test higher levels of understanding

University biology courses, even at the introductory level, should have learning goals that go beyond knowledge acquisition, so it is important that exams are designed to evaluate higher level learning goals. Questions should assess the ability of students to integrate and apply what they learn, and to analyze new information in the light of what they have been told in lecture. These “higher” level learning goals were first outlined by Benjamin Bloom in 1956, and are now being used quite widely in the field of biology education.

1. A mean near 60% will likely have a distribution that is broad enough to distinguish multiple levels of understanding. If exam is too hard (the mean is 50% or below) this often indicates the exam is confusing or not well matched to what was taught. If the exam is too easy (the mean is 75% or higher) it is difficult to distinguish between A-level and B-level understanding.

2. Because we write questions that require application and analysis, we are careful to go over many samples of these questions in class to give students practice. We also talk about the levels of understanding, and label each question’s level on the answer key.

3. To help us write questions at different levels we find it useful to refer to a sample set of questions we developed on one concept at each level we use on our exams: a) knowledge, b) comprehension, c) application, and d) analysis. Here is our sample set:

a. A “Knowledge” question requires only recognition of information as presented in lecture. Possible answers to do not include significant distractors. We have very few of these questions on our freshman biology exams -- up to 15% at most. Note that the stem forms a complete sentence, with no blank in the middle. We feel missing words in the middle of a stem make it harder for students to retrieve the framework of the question, even if they understand it.

   Small organelles that carry out protein synthesis are called:
   a. ribosomes
   b. lysosomes
   c. nuclear pores
   d. liposomes
   e. nucleoli

b. A “Comprehension” question requires understanding of a concept or of terms. It is still information from lecture, but it is presented in a novel format and possible answers include significant distractors. Note that the Bloom’s level of a question is not a mark of difficulty, but of presentation. All of the information here is from lecture, but because it requires students to recognize that a potassium channel is an integral membrane protein, and apply the steps of transcription and translation, it is difficult and only top students get it correct. We have up to 40% of our exams at this level of learning.

In order to synthesize a potassium channel:
   a. mRNA encoding this protein will be inserted into the ER lumen
   b. a ribosome will transfer the synthesized channel protein to the rER
   c. a ribosome will bind to the mRNA encoding this protein
   d. the protein will diffuse into the lumen of the rER
   e. mRNA encoding this channel will bind to the rER to attract a ribosome
c. An “Application” question asks students to predict what will happen in a new situation, or as the result of perturbing an already-discussed system. This sample question is similar to the comprehension question above, but also requires students to apply translation to neuron anatomy, which were never discussed together in lecture. We aim for about 40% of our exam to be application level.

We also like to write questions that contain enough information in the stem to help students mentally gather the material they will need to solve the problem. A difficult question like this is preferable to a difficult question that asks “Which of the following is false?” which doesn’t give students any direction as to what learning they are supposed to demonstrate.

d. An “Analysis” level question requires interpretation of data and selection of the best conclusion. Most often these problems have a data set or a figure to read.

These questions are difficult to write, and so only around 5% of our exam consists of these. We hope to get better at writing them and increase this sort of thinking in our students. One of our important learning goals is making students successful at reading graphs.

One of the difficulties we have is that we prefer to have questions that have real biology in them, rather than questions that say “imagine an alien that has DNA made out of arsenic.” We try to maintain a file of ideas as we read journal articles through the year, which helps some.

e. Starting in Fall 2011, we began working on adding “Synthesis/ Evaluation” questions to our exams, which is considered the next higher Bloom’s level. These questions present students with experimental data, and then ask them which hypothesis the data supports. All options tend to be biologically true in these, but only one is relevant to the data. These questions will ideally contain data from published research. This is only doable for most faculty if, when they area reading literature they would read anyway, set up at least a
framework of the question every time they see data that would successfully test the learning goals in their course. An example of a synthesis/evaluation question is given here.

We have made these question levels, along with sample questions, available as a one-page download on our website.

<table>
<thead>
<tr>
<th></th>
<th>Amount of DNA</th>
<th>Amount of Ribosomes</th>
<th>Amount of nuclear mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated cell</td>
<td>Normal</td>
<td>Normal</td>
<td>Reduced</td>
</tr>
<tr>
<td>Untreated cell</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Which hypothesis is supported by these data?

a. The drug blocks nuclear pores
b. The drug blocks transcription of DNA
c. The drug blocks the translation of protein

4. We also recommend **exams be cumulative**. This reinforces the idea that students should be learning core concepts that are broadly applicable and require critical thinking skills rather than just memorizing facts. For team taught courses this requires coordination with other members of the faculty team to write questions that span multiple topics.

**Step 3: Avoid common multiple-choice exam errors**

Guidelines for exam-writing are found in many textbooks and in the literature. Some are common-sense (use good grammar), but others have been studied experimentally. Exam writers call a multiple choice exam question an *item*, consisting of the *stem* (the question), and the *options*. One option is correct, the others are *distractors*. Here are the most common guidelines, with our philosophy and the research results indicated where present:

1. Avoid complex (called K-type) questions (“a and b,” “a and c”). Research has indicated these questions are more difficult (fewer students get them correct) but not more discriminating (good students do not do better compared to poor students).
2. Make only one answer correct, and the others clearly and defensibly wrong.
3. We see pedagogical value in writing a complete question stem that sets up the biology and asks the question at the end. This anchors the student to the learning goal we are testing. Exam guidelines also consistently recommend complete stems rather than:
   a. incomplete stems (“In mitochondria…”)
   b. blanks (“In the mitochondria, a high concentration of ___ is in the stroma.”)
   c. unfocused stems (“Which of the following is true?”)
Research results are mixed, however, about whether a non-complete stem in an otherwise clear question increases difficulty or reduces discrimination.
4. Avoid the use of “all of the above” or “none of the above.” Research indicates using “none of the above” often increases the item difficulty, and using “all of the above” often reduces the item discrimination.
5. Questions only need three options (1 correct, 2 distractors). This surprised us, as it would seem to increase the success of guessing. But very few students guess randomly and generally choose within the three best choices. As it is very difficult to write plausible distractors, research supports writing only two and instead increasing the total number of questions.  

6. Keep the stem concise and options short. Research shows that long, complicated questions are particularly difficult for slower readers and English language learners. Higher Bloom’s level questions do require text to set up the biology, but we minimize extra verbiage in the stem and we try to keep each option one line long or less.  

7. We recommend only sparing use negative questions (“Which is NOT” or “Which is FALSE”). These may be difficult for students to logically follow. If you do use one, be sure to capitalize NOT. Research shows no strong effect of negative question types, as long as there are no double negatives or other confusions.  

8. Avoid absolute terms like “always” or “never.” We found we were using these to make otherwise true options wrong, as there is rarely a clear “always” or “never” in science. We also avoid vague terms like “rarely” or “usually.”  

9. Be careful that the longest answer isn’t the correct answer. This is a common cue that students will use to guess more accurately. There is evidence that flawed questions make an exam less valid – meaning they do not test as effectively for comprehension of science. This is why we encourage faculty to keep these guidelines in mind.  

Step 4: Use a teaching TA as a proofreader  

Another set of eyes will catch many mistakes in your exam draft. Before you photocopy your exam, arrange to have one or more experienced TAs (ideally ones who are teaching discussion sections) read through an un-keyed copy of the exam. Ask each TA to:  

1. Mark the right answers  
2. Make notes where questions or answers are confusing  
3. Make grammatical corrections  
4. Indicate how long it took them to take the exam  
5. Indicate the expected difficulty of the exam  

Sit down with the TAs and discuss their answers and comments. Our TAs have saved us from many, many exam errors. And when they miss a question we can ask them why. Sometimes they didn’t understand the concept either, which lets us know we should teach this topic in a different way in subsequent years.  

Step 5: Make an annotated key  

We have found making a high-quality exam very difficult. It is easy to accidentally write a question that is ambiguous or has two right answers… or no right answers. We catch many of our mistakes if we write the key before printing the exam. We catch even MORE errors if we
write an annotation for each question after we write the question. This is the key we post for the students:

Anne, a TA in Bio 93, studies limb regeneration in the axolotl, a freshwater salamander from Mexico that can regrow lost or injured body parts. Human cells are cultured in DMEM, a solution of salts, sugars and nutrients. However, axolotl cells must be cultured in DMEM that has been diluted to 60% strength. A possible reason for this is that axolotl cells:

a. will decrease in volume in 100% DMEM
b. have more aquaporin channels than human cells
c. contain more cholesterol
d. have a higher solute concentration inside than human cells
e. are hypertonic to 100% DMEM

Level 3 application. Lecture 3. You have not heard about this particular animal before but this is a basic osmosis question. Axolotl are fresh water animals and from the question you learn that cells from this animal must be cultured in 60% DMEM. Therefore they must have a lower solute concentration in their cells than human cells which are grown at 100% DMEM. The cells are thus hypotonic to 100% DMEM and will lose water and decrease in volume in 100% DMEM.

We have found it helpful to include the lecture number and the Bloom’s level of the question (we talk about these levels in class). The annotation walks the students through the problem, explains why the right answer is right and the wrong answers are wrong. Writing this out forces us to justify each choice before the exam is given, and helps us be sure that the question is as clear as we can make it. After the exam the students can go read over the biology behind each question and possibly learn from the key. And they are much less likely to email us with random questions about the exam.

Step 6: Build a file of good exam questions

Some of your exam questions will work very well, while others will end up being confusing or ambiguous. Don’t let the quarter end without figuring out which questions were which. When your TA grades the multiple choice scantrons, they will get a report that indicates for each question:

- The % of students who answered correctly
- The % of TOP students who answered correctly (top 25% of scorers)
- The % of BOTTOM students who answered correctly (bottom 25% of scorers)
- The number who chose each option (correct and distractors)

1. Have your TA write the scantron info onto a clean version of the exam (or type it into a document that also contains the questions), and give it to you.
2. Do 15 minutes of processing while your teaching and testing are still fresh.
   - Mark questions as too hard (0-40% correct) appropriately hard (40-70% correct) and easy (70-90% correct). Think about how this corresponded to class material. Did students meet your learning goals?
   - Circle questions that had high discrimination (%TOP students answered at least 1.5 times better than the %BOTTOM students) ². These should be kept.
• Cross out questions with poor discrimination values (%TOP less than 1.5 times higher than %BOTTOM). Why do you think these questions did not work? Can they be fixed, or does your teaching need to be adjusted?
• Scan the # who chose each distractor for each item. How many students chose each distractor? If your distractors are often ignored by students (chosen by fewer than 5%), then you can focus on making fewer, more plausible distractors next time.

3. Keep these write-ups in an exam file, and pull them out when you get ready to write your next exam as examples of good questions that can be modified by changing the example, including a graph with different data, etc.
4. Make a short list of topics you would like to give more focus to the next time you teach, and place the list where you will see it when you prep next time.

Do you have more questions about exam writing? Suggestions? Please email us at dkodowd@uci.edu

Further reading on crafting exam questions:
• Short pdf tutorial (from chemistry) http://www3.fed.cuhk.edu.hk/chemistry/files/constructMC.pdf
• Biology MC questions that test higher-order learning: http://www.lifesied.org/cgi/content/full/7/4/368
• Many more sources at http://cte.umdnj.edu/student_evaluation/evaluation_constructing.cfm

References: