Integrating Testing and Learning

Anthony D. Albano, PhD, University of California at Davis
Peggy Brickman, PhD, University of Georgia
Melissa Csikari, MS, Howard Hughes Medical Institute
David Julian, PhD, University of Florida
Rebecca B. Orr, PhD, Collin College
Michael C. Rodriguez, PhD, University of Minnesota

hhmi | BioInteractive

Table of Contents

Section 1: Writing Learning Objectives 1
Why Use Learning Objectives? 1
Backward Design 1
Learning Objectives Are Key to Good Teaching 2
Activity: Recognizing Effective Learning Objectives 2
How Do We Construct Learning Objectives? 3
Higher-Order and Lower-Order Thinking Skills 5

Section 2: General Test Development 7
Assessing Learning Objectives by Constructing a Fair Test 7
Test Blueprints 7
Guidelines for Establishing Target Percentages for Topics and Thinking Skills 8
Steps in Creating a Test Using a Blueprint 8
General Tips for Constructing Effective Tests 11
Activity: Practice Constructing a Test Blueprint 11

Section 3a: Critiquing Test Items 13
Measuring Learning Objectives Using Selected Response Items 13
Activity: Analyze a Learning Objective 14
Item-Writing Checklist[7] 20
Activity: Critiquing Multiple-Choice Items 22

Section 3b: Writing Your Own Test Items 24
Activity: Critique and Edit Your Own Items 24
Activity: Constructing Selected-Response Items 24

Section 4: Test Analysis and Item Review 26
Evaluating Items: Terminology 26
Activity: Examine an Item Analysis Report 27
Resources 29
Next Steps! 30

Appendix 1: Revised Bloom's Taxonomy Action Verbs 31

Appendix 2: Selected-Response Formats 32

Appendix 3: Guidelines for Selected Response Items 33

Appendix 4: Flipped Dichotomous Key to Rank Multiple-Choice Questions 37

Integrating Testing and Learning

Section 1: Writing Learning Objectives[1]

The workshop session will begin with a brief introduction to learning objectives (LO), sometimes referred to as instructional objectives, learning outcomes, or teaching targets. LO are concise but specific statements that identify what instructors expect students to know and be able to do upon completion of a course. LO define the goals and purposes of instruction in terms of core concepts and competencies and can thus be used to organize course content with a focus on learning while also serving as a foundation for lectures, class activities, assignments, and assessments.

Why Use Learning Objectives?

Instructors typically write and organize LO for each unit of instruction in a course. Guidance in developing LO may come from textbooks and other course materials as well as external sources such as professional organizations, state or federal agencies, and the college or university. Learning objectives connect teaching, learning, and assessment. They allow instructors to be systematic about gathering assessment evidence of academic achievement for the purposes of improving student learning and their own teaching.

Backward Design

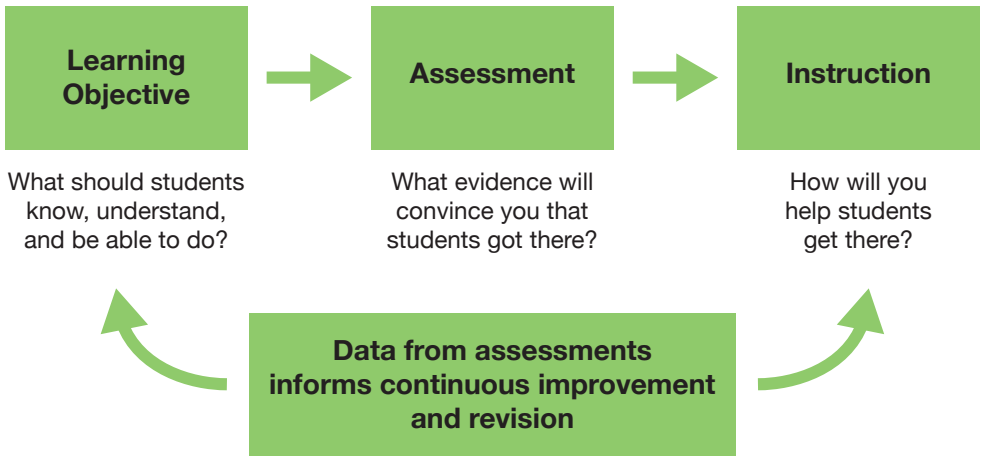


Figure 1—Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Assoc. for Supervision and Curriculum Development

[1] Rodriguez, M., & Albano, A. (2017). *The College Instructor's Guide to Writing Test Items: Measuring Student Learning*. Routledge.

In backward design of course content, the instructor focuses first on the desired learning objective instead of lectures, activities, and resources (figure 1). The backward design approach was developed by Grant Wiggins and Jay McTighe (1998)[2] primarily for K-12 education, but it is now increasingly favored in college STEM instruction.

Backward design follows a three-step process:

1. **Identify the Desired Learning Objectives:** Develop LOs that identify what students should know, understand, and be able to do after instruction.
2. **Plan the Assessment:** Develop assessments and measures that would provide acceptable evidence that students achieved the desired outcomes.
3. **Plan the Learning Experiences and Instruction:** Develop the activities, experiences, materials, and instruction that will help students achieve the desired outcomes.

Learning Objectives Are Key to Good Teaching

- Learning objectives allow instructors to articulate claims they want to make about student learning.
- Learning objectives identify evidence that students can demonstrate their knowledge, skills, and abilities.
- Learning objectives allow instructors to be systematic about developing assessments to collect relevant evidence for the purposes of improving student learning and their own teaching.

Activity: Recognizing Effective Learning Objectives

Effective learning objectives are *behavioral*, *measurable*, and *attainable* and support teaching and learning.

Read these example learning objectives:[3]

Upon successful completion of this unit, you will be able to:

- Draw a pedigree based on information in a story problem.
- Using pedigrees, distinguish between dominant, recessive, autosomal, X-linked, and cytoplasmic modes of inheritance.
- Design genetic crosses to provide information about genes, alleles, and gene functions.

What stands out to you as key/defining features of the learning objectives above?

[2] Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Assoc. for Supervision and Curriculum Development.

[3] Learning Objective taken from Core Concepts and Competencies in Genetics, revised and approved by the Genetics Society of America Education Committee in November 2015. These and others can be found at the CourseSource Website for different disciplinary areas: https://www.coursesource.org/supporters https://www.genetics-gsa.org

Compare the key/defining features you identified on the last page with the recommendations below:

Learning objectives should not be generic or too broad and they should be measurable.

LOs that meet these criteria can also:

- reflect current thinking and expectations of the field,
- communicate what students need to know and be able to do,
- point to essential aspects of the course content and expectations for learning,
- direct students' study behaviors,
- direct instructional focus,
- support the development of effective assessment activities and test items,
- provide a guide for instructional feedback,
- support student self-reflection and self-assessment.

How Do We Construct Learning Objectives?

To develop a learning objective, consider the (1) performance (demonstrable thinking skills or action verb), and (2) content (what the student will work with).

As an example, the relevant content might be "learning objectives."

> Performance: *the student will distinguish*
>
> Content: *effective versus ineffective learning objectives*

The resulting learning objective could be stated as:

> *The student will distinguish ineffective from effective learning objectives based on whether they are behavioral, measurable, and attainable*.

Activity: Construct a Learning Objective

Identify a content-relevant topic for a course you teach and specify the parts of a learning objective:

Topic:

_____

Performance [action verb]:

_____

Content [what students will work with]:

_____

Activity : Review/Critique Learning Objectives

Table 1: Example Learning Objectives

Learning Objectives	Evaluate
1. What role do concentration gradients play in the active and passive transport of substances across the membrane?	Performance: Content:
2. Identify the independent and dependent variables in an experiment.	Performance: Content:
3. Explain the evolution of organisms using the principles of natural selection.	Performance: Content:
4. Provide students with opportunities to learn about contemporary problems in the field of biology.	Performance: Content:
5. Calculate the probability of a particular gamete being produced from an individual, assuming independent segregation.	Performance: Content:

Higher-Order and Lower-Order Thinking Skills

Performance can be conceptualized into actions that require lower-order thinking skills (LOTS) and those that require higher-order thinking skills (HOTS). Tasks may be presented graphically so that their level of complexity is indicated, as shown to the right. The most well-known conceptual organization scheme for measuring higher- and lower-order thinking skills is Bloom's Taxonomy (Figure 2). See Appendix 1 for a list of action verbs for each level of Bloom's Taxonomy.

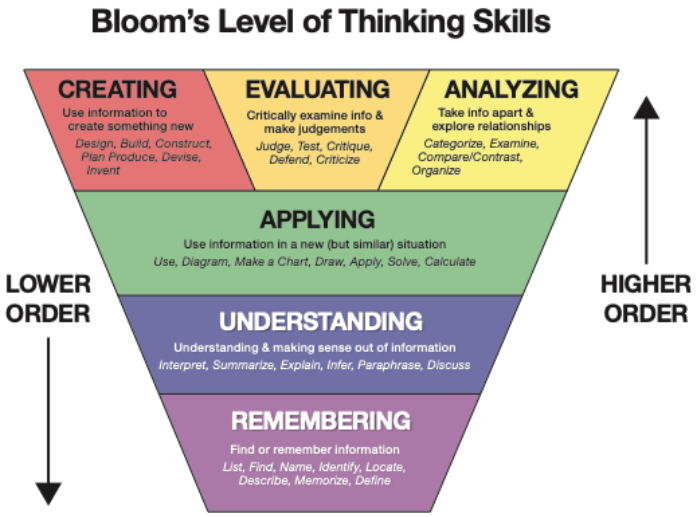


Figure 2—See Appendix 1 of the workbook for a list of action verbs to help you write your own objectives.

LOTS include remembering facts and understanding concepts, whereas HOTS include critical thinking and problem solving. HOTS provide opportunities for skills to be demonstrated, so that questions that assess these skills cannot be answered through memory alone. HOTS are more difficult to teach and learn, but they are considered more valuable, especially in novel situations. Therefore, a general guideline is to emphasize HOTS over LOTS (but this can depend on the content being taught).

When creating learning objectives, the thinking skill level is usually indicated by the verb identifying the behavior to be measured. For example, verbs such as define, label, and list are associated with LOTS, whereas verbs such as create, infer, and model are associated with HOTS.

Activity:

Identify the thinking skills indicated in the learning objectives that you critiqued in Table 1.

Activity: Revise/Create Your Own Learning Objectives

Table 2: Your Learning Objectives

See Appendix 1 for action verbs at different Bloom's levels.

Learning Objective	Bloom's Level	How Would You Assess?
Using pedigrees, distinguish between dominant, recessive, autosomal, X-linked, and cytoplasmic modes of inheritance.	Analyzing	Provide students with pedigrees, and then ask them to characterize the type of inheritance demonstrated in a multiple-choice or short answer question.

Section 2: General Test Development

Now that you have considered the learning objectives that you have developed, you can begin to think about how you plan to assess whether students have mastered these objectives in a way that can also help inform your instruction.

Assessing Learning Objectives by Constructing a Fair Test

In your experience creating tests, you may have a process that you use to choose and organize your question items. You may not have considered creating a systematic map or blueprint to examine and rationalize your choices and how evenly they address each learning objective. The next section will discuss this process.

Test Blueprints

A test blueprint helps you determine how your test addresses the relevant learning objectives and if the test is representative and balanced across learning objectives. If both are true, then you can be more confident that scores reflect the extent to which students achieved the learning objectives, AND that the scores reflect your ability to help students achieve the learning objectives.

Table 3: Generic Test Blueprint

Content Topics or Areas	Thinking Skill Items		Total Items	
	Lower Order Qty	Higher Order Qty	Qty	Target %
Topic 1				
Topic 2				
Topic 3				
Topic 4				
Total Qty				
Target Total % (aspirational)				

Guidelines for Establishing Target Percentages for Topics and Thinking Skills

The blueprint template is designed to plan assessments using the learning objectives to help allocate test items based on the (1) topics and (2) higher and lower order thinking skills.

In order to estimate the amount of content knowledge and skills that your test will measure (target or aspirational totals), it is important to think about:

- The amount of time spent on the topic and skill.
- What is really important for students to know and be able to do in biology as a discipline?
- What is valued and/or emphasized in terms of content standards in the discipline?
- The alignment of the level of thinking skills with what students have worked on in class and practiced in their homework.

Steps in Creating a Test Using a Blueprint

1. For one of your quizzes, tests, or exams, indicate the topics you are assessing. Each topic may include numerous learning objectives. You may not want to write one item for each learning objective. This blueprint helps you outline the learning objectives that you plan to focus on.
2. Indicate the percentage of items that you desire to assess at lower and higher order thinking skills overall and for each content topic (e.g., set your aspirational targets in the last column and your target total in the last row ± 10%).
3. Determine the total quantity of items allowable, given the time constraint.
4. Identify and indicate items for each topic. Collect and write items and distribute them across topics.
5. Add up the total quantity of items within each topic, and then compare your totals to the percentages indicated in the target final row and final column.

Table 4: Steps in Completing a Test Blueprint

Content Topics or Areas		Thinking Skill Items		Total Items	
		Lower Order Qty	Higher Order Qty	Qty	Target %
Topic 1	Step 1	Step 5	Step 5	Step 4	Step 2
Topic 2	Step 1	Step 5	Step 5	Step 4	Step 2
Topic 3	Step 1	Step 5	Step 5	Step 4	Step 2
Topic 4	Step 1	Step 5	Step 5	Step 4	Step 2
Total Qty		Step 4	Step 4	Step 3	
Target Total % (aspirational)		Step 2	Step 2		100%

Table 5: Example Test Blueprint

Content Topics or Areas	Thinking Skills Items		Total Items	
	Lower Order Qty	Higher Order Qty	Qty	Target %
Topic 1: Macromolecules **LO1:** Recognize and differentiate carbohydrates, proteins, and lipids. **LO2:** Identify ingredients that contain carbohydrates, lipids, and proteins.	1 6	0 0	7	40%
Topic 2: Enzymes **LO3:** Describe the roles of substrate, product, and enzyme in an enzyme-catalyzed reaction. **LO4:** If given a table of data showing genotype frequency and disease phenotypes, predict which genes and alleles have a demonstrated link.	1 0	1 3	4	25%
Topic 3: Transcription & Translation **LO5:** Describe how DNA, mRNA, RNA polymerase, promoter, gene, amino acids, and ribosomes are used to make proteins in a cell. **LO6:** Identify the RNA transcript and protein produced from a template of DNA. **LO7:** Identify how a mutation in DNA may or may not affect protein sequence.	1 2 0	1 0 1	6	35%
Total Qty	11	6	17	
Target Total % (aspirational)	65%	35%		100%

General Tips for Constructing Effective Tests

- Cluster questions on individual topics. Better measurement is gained with this strategy versus a "popcorn" effect where the questions are distributed throughout the exam.
- Don't set grades by a bell curve (or shifted curve); this is poor measurement practice. Grades should not be relative to those of other students. Students should be graded based on their performance against (achievement of) the learning objectives.
- Timing: Create tests where at least 95% of your class can complete your test in the given time period. A general guideline is to budget 1.5 minutes/item if your exam is a mixed exam (lower-higher questions). If predominantly higher-level questions, budget 1.5 to 2 minutes/item. Studies have indicated that students are more successful with difficult items if they are front-loaded at the beginning of the test. However, it is not recommended that the first test question be a more difficult item.

Activity: Practice Constructing a Test Blueprint

Using an existing test or thinking about a test that you plan to develop, complete the test blueprint table below with your learning objectives within each topic area and fill in each row of the table indicating how many items of higher and lower order you might imagine using for one topic area.

Steps:

1. Organize the items by content topic (this could be a chapter of your book, a big idea or "Vision and Change" core concept, or another method for organizing your instruction).
2. Assign a thinking skill level to each item.
3. Indicate the number of items for the content and tasks in the blueprint. (Mapping a test that you have already created will not give you targets but actual totals.)
4. Compare the test with your expectations.

Content Topics or Areas	Thinking Skill Items		Total Items	
	Lower Order Qty	Higher Order Qty	Qty	Target %
Topic 1: Learning Objectives:				
Topic 2: Learning Objectives:				
Topic 3 Learning Objectives:				
Topic 4 Learning Objectives:				
Total Qty				
Target Total % (aspirational)				

Section 3a: Critiquing Test Items

Accessibility & Inclusion

You should select items and testing formats that reduce the impact of assumed experiences and knowledge. Items should be reviewed to identify and remove biased use of ethnic, gender, and cultural stereotypes to ensure that items measure learning that is free from confounding bias. (Detailed information can be found in the Handbook of Accessible Instruction and Testing Practices.[4])

Measuring Learning Objectives Using Selected Response Items

In this section, we will focus on writing selected response items for assessing the learning objectives you listed in Table 6. In selected response items, the student chooses a response provided by the teacher or test developer. Example selected response items include; multiple choice, multiple true/false, true/false, matching, and fill-in-the-blank (including numerical response) where the response can be machine graded. These questions are usually more difficult to write, but because they are easier to score, they provide a great way for students to practice and self-assess their progress. However, selected response questions cannot measure all learning objectives well (Table 7).

Table 6: Example of an Item Linked to a Learning Objective

Learning Objective	Assessment Item
Explain the relationship between cancer and mutations in tumor suppressor genes (like p53).	Which of the following ***best*** explains ***why*** mutations in p53 are observed in a large percentage of cancerous tumors? A. The p53 protein can upregulate DNA repair genes. B. Once p53 is mutated, cells are less likely to downregulate the cell cycle. C. Mutation rates are elevated in cancer cells.

[4] Kettler, R. J., Elliott, S. N., Beddow, P. A., & Kurz, A. (2018). *Handbook of accessible instruction and testing practices*. New York, NY: Springer. doi, 10, 978–3.

Table 7: Suitability of Selected Response Items for Measuring Example Learning Objectives

Measured well	Not measured as well
• Analyzing conditions and phenomena • Applying concepts and principles in new situations • Evaluating cause-and-effect associations • Evaluating the relevance of information or evidence • Solving problems • Discriminating between facts and opinions • Interpreting graphical displays • Drawing inferences from data	• Exploring reasoning or thinking • Providing information to support an argument • Organizing thoughts • Performing tasks • Producing innovative ideas • Providing examples

Activity: Analyze a Learning Objective

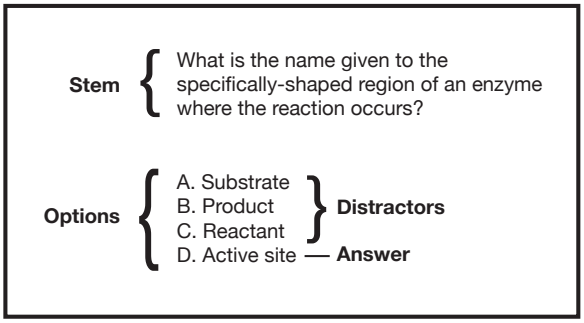
Look at a learning objective that you have written in Table 2 and use Table 7 to determine which could be effectively measured well using selected response items. Discuss your observations with a partner.

Table 8: Advantages and Disadvantages of Selected Response Items[5]

Advantages	Disadvantages
• Allows for broad coverage of content • Can be used to measure many different objectives • Can provide diagnostic information about errors in thinking, misconceptions, etc. • Easy to score; can be machine scored • Efficient for large classes • Able to control difficulty to some degree • Students are familiar with the format	• Challenging to construct high-quality items • Time-consuming to construct well • Not good for measuring innovative thinking, creating something • May be affected by students' reading skills • May encourage overtesting recall • Doesn't reflect real-world problem solving

[5] Rodriguez, M., & Albano, A. (2017). *The College Instructor's Guide to Writing Test Items: Measuring Student Learning*. Routledge

Selected-Response Item Structure



Stem – The problem, sentence portion, or question. Formally called the **stimulus**.

Options – All responses that could be selected (i.e., including the answer and distractors). Sometimes called alternatives or answer choices.

Answer – The correct or best option or combination of options.

Distractors – All incorrect or inferior answer options. "**Attractor**" may be a more useful term, as an effective distractor is one that attracts students with misconceptions or errors in thinking and reasoning.

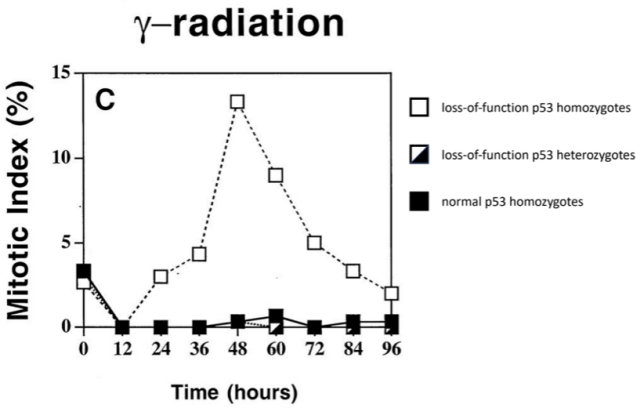
Table 9: Types of Selected-Response Items[6]

Item Type*	Example Item	Recommend
True/False (TF)	Drugs that inhibit DNA replication are more likely to harm a fully grown adult than a developing embryo because it has more cells. a. True b. False	No
Alternate-Choice (AC) (two-option MC item, in which one option is a distractor)	For each of the following pairs of events, select the one that happens first in a typical turn of the cell cycle. 1. a. Cell signaling pathways detect the abundance of nutrients in the environment. b. Initiation proteins bind to the origin of replication. 2. a. Chromosomes are separated into two daughter cells. b. Nuclear membrane breaks down. 3. a. Chromosomes are condensed. b. DNA polymerase gene is upregulated. 4. a. Cell cycle checkpoint signaling pathways respond to mutations in DNA. b. Chromosomes are aligned at the metaphase plate.	Yes

[6] Questions contributed by Alexa Clemons, University of Washington.

Matching	A scientist is investigating five genes that may be involved in the cell cycle. They introduce loss-of-function mutations into each gene individually in one of five populations of human cells. They then record the phenotypes of these cells. Match each observed phenotype (1-5) with the phase of the cell cycle (a-d) that the mutated gene is likely involved in. Cell cycle phases (a-d) may be used more than once or not at all. 1. DNA polymerase gene was transcriptionally downregulated. 2. DNA was not copied. 3. DNA was condensed and aligned at metaphase plate, but no cell division occurred. 4. Cells divided more quickly, even when treated with radiation. 5. Cells divided more quickly, despite low nutrient levels in environment. a. G1 phase b. S phase c. G2 phase d. M phase e. Cannot determine	Yes
Conventional Multiple Choice (MC type A)	Which of the following **best** explains **why** mutations in p53 are observed in a large percentage of cancerous tumors? A. The p53 protein can upregulate DNA repair genes. B. Once p53 is mutated, cells are less likely to downregulate the cell cycle. C. Mutation rates are elevated in cancer cells.	Yes
Complex Multiple Choice (MC type K)	DNA sequencing of cancerous tumors typically shows mutations in more than one cell-cycle gene. Which of the following options explain this observation? A. The function of cell cycle proteins is often redundant with other cell cycle proteins. B. Cancerous cells are less likely to detect and respond to mutations. C. During the cell cycle, DNA is more often condensed. D. All of the above. E. A and B F. B and C G. A and C H. None of the above	No (preferable to use MTF)

Multiple True-False (MTF)	A scientist is investigating five genes that are hypothesized to be involved in the cell cycle. For each gene, the scientist introduced a mutation causing a loss of function of that gene in a population of human cells. The scientist then recorded the resulting phenotype that occurred for each of the five mutated genes studied. For each item, indicate whether the hypothesis are supported (A) or not supported (B) by the observed phenotype. 1.Hypothesis: Gene 1 is a tumor suppressor that normally functions in S phase. Phenotype: DNA polymerase gene was transcriptionally downregulated. 2.Hypothesis: Gene 2 is a proto-oncogene that normally functions in G2 phase. Phenotype: DNA was not copied. 3.Hypothesis: Gene 3 is a proto-oncogene that normally functions in M phase. Phenotype: DNA was condensed and aligned at metaphase plate, but no cell division occurred. 4.Hypothesis: Gene 4 is a proto-oncogene that normally functions in G2 phase. Phenotype: Cells divided more quickly, even when treated with radiation. 5. Hypothesis: Gene 5 is a tumor suppressor that normally functions in G1 phase. Phenotype: Cells divided more quickly, despite low nutrient levels in environment.	Yes

Context-Dependent Item and Item Set	 *Alleles of the* p53 *gene were selectively disrupted in a line of human cells and then monitored after exposure to DNA-damaging gamma (γ) radiation to determine what proportion of the cells entered mitosis (cell division). The shaded squares represent cells with two normal alleles of the* p53 *gene. The half-shaded squares represent cells with one normal and one disrupted allele of the* p53 *gene. (Note that some of the half-shaded squares are covered by the shaded squares.) The unshaded squares represent cells in which both alleles of the* p53 *gene were disrupted. The mitotic index is the proportion of cells undergoing mitosis at a given time.* 1. Which population of cells is likely to have a greater level of expression of primase at time 48h? A. Normal p53 homozygotes B. Loss-of-function p53 homozygotes 2. Which population of cells does the "mitotic index" phenotype of the loss-of-function p53 heterozygotes more closely resemble? A. Normal p53 homozygotes B. Loss-of-function p53 homozygotes 3. Based on the "mitotic index" phenotype, p53 is more likely to be a _____. A. Tumor suppressor B. Proto-oncogene 4. Which of the following would most likely **increase** the mitotic index if present in addition to loss-of-function mutations in both copies of the p53 gene? A. A loss-of-function mutation in a proto-oncogene B. Insufficient nutrients in the environment C. High levels of growth factors 5. If scientists were investigating a gain-of-function mutation in a proto-oncogene, which phenotype would this new phenotype most likely look like? A. Normal p53 homozygotes B. Loss-of-function p53 homozygotes	Yes

* See Appendix 2 for a complete description of each item type.

Item-Writing Checklist[7]

The following recommendations will promote construction of items that

- Remove obstacles that might prevent knowledgeable students from getting the item right.
- Remove clues that might result in the unprepared student getting the item right.

Content Concerns

1. Base each item on one aspect of content and one cognitive task.
2. Use new material and context to elicit higher-order cognitive skills.
3. Keep the content of items independent of one another.
4. Test important content. Avoid overly specific and overly general content.
5. Avoid opinions and trick items.

Formatting

1. Format each item vertically instead of horizontally.
2. Edit and proof the items.
3. Keep the language complexity of items at an appropriate level for the class being tested.
4. Minimize the amount of reading in each item. Avoid window dressing.

Writing the Stem

1. Write the stem as a complete question or a phrase to be completed by the options.
2. State the main idea in the stem clearly and concisely and not in the options.
3. Word the stem positively; avoid negative phrasing.
4. Move any words that are repeated in each option into the stem.

Writing the Options

1. Write as many options as are needed given the topic and cognitive task; three options are usually sufficient.
2. Make all distractors plausible.
3. Make sure that only one option is the correct answer.
4. Place options in logical and numerical order.
5. Vary the location of the correct answer evenly across the options.
6. Avoid using the options none of the above, all of the above, and I don't know.
7. Word the options positively, avoiding negative words such as "not."
8. Avoid giving clues to the right answer:
    a. Keep the length of the options about equal.
    b. Avoid specific determiners including "always," "never," "completely," and "absolutely."
    c. Avoid clang associations, options identical to or resembling words in the stem.
    d. Avoid pairs or triplets of options that clue the test taker to the correct answer choice.
    e. Avoid blatantly absurd, ridiculous, or humorous options.
    f. Keep options homogeneous in content and grammatical structure.

(See Appendix 3 for a more complete explanation of each guideline listed above.)

Activity: Critiquing Multiple-Choice Items

Use the item-writing checklist to critique the items in Table 10. Back up your claim of a "bad question" with the item-writing guidelines.

Table 10: Selected-Response Item Critique

Item	Critique
Which of the following statements about the structure of proteins is **TRUE**? A. Beta pleated sheet is one type of tertiary structure. **B.** Alpha helix is one type of secondary structure. C. It consists of two or more polypeptide chains joined together. D. Gamma folded mattress is one type of secondary structure. E. Secondary structures of proteins are held together by covalent bonds.	
Insulin maintains blood sugar levels by removing ____ and storing it as ____. A. glycogen, glucagon **B.** glucose, glycogen C. glycogen, glucose D. glucagon, glucose	
Which observation was made first? A. DNA is a double helix. **B.** Cork is made up of cells. C. ATP synthase spins in the mitochondrial membrane. D. Hox genes control development in vertebrates and invertebrates.	

Starch and glycogen are similar because these molecules are both A. in plants **B.** polymers of the sugar molecule glucose linked by glycosidic bonds and they serve as an energy source C. disaccharides D. structural	
A mutation in the operator of the lac operon prevents binding of the repressor protein. Which of the following statements about this lac operon and the *E. coli* that possesses it are TRUE? 1. In an environment where resources are limited, this *E. coli* cell will have a selective advantage over others. 2. The structural genes *lac* Z, *lac* Y, and *lac* A will be transcribed continuously. 3. This *E. coli* cell will be unable to transport lactose across its plasma membrane. 4. Given an environment where there is very little glucose and there is no lactose, this *E. coli* cell will probably not leave many descendants, if any. 5. The cell will have to rely on tryptophan as a source of energy rather than lactose. A. Statement 1. B. Statements 2 and 4. C. Statements 2, 4 and 5. D. All of the statements above are true.	

Which of the following does not accurately define a difference between the products of mitosis and meiosis? A. The end product of mitosis is two *2n* daughter cells with unduplicated chromosomes, whereas the end product of meiosis is four *n* cells with unduplicated chromosomes. **B.** The end product of mitosis is two *n* daughter cells with duplicated chromosomes, whereas the end product of meiosis is four *2n* cells with duplicated chromosomes. C. The end product of mitosis is two *n* daughter cells with unduplicated chromosomes, whereas the end product of meiosis is four *n* cells with unduplicated chromosomes. D. The end product of mitosis is two *4n* daughter cells with unduplicated chromosomes, whereas the end product of meiosis is four *n* cells with unduplicated chromosomes. E. The end product of mitosis is two *2n* daughter cells with duplicated chromosomes, whereas the end product of meiosis is four *n* cells with unduplicated chromosomes.	

Section 3b: Writing Your Own Test Items

Activity: Critique and Edit Your Own Items

Use the item-writing guidelines and work to improve one or two of your own items.

Activity: Constructing Selected-Response Items

Step 1: Refer to the test blueprint that you started to determine where you would like to create items for a specific topic, learning objective, and cognitive level.

Step 2: Item Writing. Create the stem for 1–3 HOTS selected response questions. The stem of an item includes the initial description, context, and question statement that a student refers to when choosing an option. Then, create the answer options for the stems you have written.

See the Item-Writing Checklist on the previous page for instructions for writing the stem and answer options. Appendix 3 contains complete descriptions.

Higher Order Learning Objective:
Stem:
Answer Options:

Higher Order Learning Objective:
Stem:
Answer Options:

Step 3: Present your options to a peer with the learning objective in mind.

Step 4: Revise your question based on feedback.

Step 5: Share your take-aways.

Step 6: Reflection. How could you continue to do this type of work outside of a workshop setting?

Section 4: Test Analysis and Item Review

Reviewing student item responses and evaluating test performance can help improve your ability to write effective test items in general and also specifically improve individual test items that you have given students and want to give again in the future.

Although many institutions provide item analysis services for tests administered on campus, a number of free item analysis programs are available to compute these statistics (refer to the resource section below). Some of these statistics, such as item discrimination, should not be computed by hand. Others, such as frequencies of distractor responses, can easily be estimated by keeping a tally or using a spreadsheet to count response options.

Evaluating Items: Terminology

- **Item Difficulty Index: indicates** the percentage of test takers responding correctly to the question

$$\text{item difficulty} = \frac{\text{\# of correct responses}}{\text{Total number of test takers}} = \frac{41}{50} = 0.82\ (82\%)$$

  Typical flagging criteria: an item difficulty index <.20 or >.90 indicates that additional attention may be warranted. This may indicate a problem with the item if it does *not* make sense that an item is so easy or is so difficult, or it may provide information about the success of instructional activities. Note that the item difficulty index is frequently also referred to as the "p value" (where p refers to performance

- **Item Discrimination:** measures the extent to which the test item differentiates between students with low performance and those with high performance on the whole test. The accuracy of this metric is highest with >80 test takers within your sample.

  Point Biserial Correlations. This is a type of item discrimination score that indicates the correlation between answering a specific test item correctly and the overall test performance. The range is -1.00 to 1.00. A positive correlation means that students who performed well on the test as a whole tended to answer the specific item correctly, a negative correlation means that students who did well on the test tended to answer the item incorrectly, and a correlation of zero indicates that there was no relationship between answering this item and the overall test score. However, this correlation analysis requires some variability in the data.. If everyone got the item correct, the correlation would be zero (because there is no variation to estimate the association). Items that are very easy or very difficult will have lower correlations because there is very little variability in item scores.

- **Distractor Analysis:** frequency of responses to each item answer option. This can provide a great deal of information about student misconceptions, errors in thinking and problem solving. Each distractor should be selected among your students. If distractors aren't selected at least some of the time, then they should be reconsidered and replaced or deleted.

Activity: Examine an Item Analysis Report

Standard Item Analysis Report: Exam 2

Course #: BIOL 1000 Instructor: XXXXXXXXX

Course Title: General Biology Description: Section 1A

Total Possible Points: 17.00 Median Score: 12.14 Highest Score: 17.00

Standard Deviation: 2.59 Mean Score: 12.39 Lowest Score: 6.00

Students in This Group: 74 Reliability Coefficient (KR20): 0.63

Student Records Based On: All Students

No.	Correct Group Responses	Discrimination Index	Correct Answer	Response Frequencies * indicates correct answer					Non Distractor
				A	B	C	D	E	
1	66.22%	0.37	B	3	*49	6	13	3	
2	60.81%	0.34	C	16	13	*45			
3	89.19%	0.47	C	1	0	*66	0	7	BD
4	78.38%	0.33	B	1	*58	8	5	2	
5	81.08%	0.27	D	1	0	3	*60	10	B
6	81.08%	0.46	A	*60	10	1	3		
7	87.84%	0.42	A	*65	9	0	0		CD
8	48.65%	0.60	B	6	*36	0	11	21	C
9	74.32%	0.45	A	*55	2	0	1	15	C
10	74.32%	0.22	B	1	*55	5	13		
11	91.89%	0.29	B	4	*68	0	2	0	CE
12	40.54%	0.36	C	17	22	*30	2	3	
13	29.73%	0.59	A	*22	9	2	40		
14	62.16%	0.46	D	1	10	17	*46		
15	97.30%	0.03	B	2	*72	0	0		CD
16	91.89%	0.29	D	6	0	0	*68		BC
17	83.78%	0.29	A	*62	0	6	6		B

Using the item analysis report on the previous page, answer the questions below.

1. What is the most difficult item? What proportion of students got this item correct?

2. What is the easiest item? What proportion of students got this item correct?

3. Which questions would you revisit or examine? Based on what criteria?

4. Which item has the lowest discrimination index? What does this indicate?

Resources

Item Analysis Software

ZipGrade: https://www.zipgrade.com/aboutus/

Iteman: https://assess.com/iteman/

GradeCam: https://gradecam.com/

Resource Readings

Rodriguez, M., & Albano, A. (2017). *The College Instructor's Guide to Writing Test Items: Measuring Student Learning*. Routledge.

Learning Objective for Core Concepts and Competencies from Vision and Change can be found at the CourseSource Website for different disciplinary areas: https://www.coursesource.org/supporters

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS One*, 12(10), e0186419.

Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E., Blankenbiller, M., & Brownell, S. E. (2016). Cognitive difficulty and format of exams predicts gender and socioeconomic gaps in exam performance of students in introductory biology courses. *CBE—Life Sciences Education*, 15(2), ar23.

Haladyna, Thomas M. "Developing Test Items for Course Examinations." *Ideaedu.org*, June 2018, Haladyna, http://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/IDEA_Paper_70.pdf

Suskie, Linda. "Making Multiple Choice Tests More Effective." *Schreyer Institute for Teaching Excellence*, 25 Apr. 2017, http://www.schreyerinstitute.psu.edu/pdf/MakingMultipleChoiceTestsMoreEffective.HandoutsForWeb.pdf

"Guidelines for Writing Learning Objectives." *Ucdenver.edu*, http://www.ucdenver.edu/academics/colleges/medicalschool/education/degree_programs/MDProgram/administration/curriculumoffice/Documents/CUSOM_Learning-Objectives-Guidelines.pdf.

Next Steps!

Tackle Your Learning Objectives:

- Start to break down your course into modules (a class or a week) and work on the learning objectives for one module at a time; revise learning objectives as needed.
- Choose some of your previous test questions and identify how each helped you assess a specific learning objective.

Revise One Test at a Time:

- Create a Test Blueprint: does your test assess your learning objectives?
- Use the item analysis: which multiple-choice items need revising?
- Revise the items with the Item Checklist in hand. Swap your revised items with a participant from this workshop/colleague at home who is interested in improving their tests and share feedback.

Appendix 1: Revised Bloom's Taxonomy Action Verbs

I. Remember	II. Understand	III. Apply	IV. Analyze	V. Evaluate	VI. Create
Exhibit memory of previously learned material by recalling facts, terms, basic concepts, and answers.	Demonstrate understanding of facts and ideas by organizing, comparing, interpreting, giving descriptions, and stating main ideas.	Solve problems in new situations by applying acquired knowledge, facts, techniques, and rules in a different way.	Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations.	Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.	Compile information in a different way by combining elements in a new pattern or proposing new solutions.
Arrange Copy Define Describe Identify Label List Locate Match Name Outline Quote Recall Recite Record Repeat Recognize Reproduce Retell Select State Tabulate Tell Visualize	Classify Describe Differentiate Discuss Distinguish Explain Extend Generalize Give an example Group Illustrate Indicate Infer Interpret Organize Order Paraphrase Report Restate Review Rewrite Select Show Summarize Translate	Calculate Chart Choose Compile Compute Construct Demonstrate Diagnose Interpret Modify Predict Prepare Relate Show Solve Teach Transfer Use Write	Appraise Break Down Categorize Classify Compare Conclude Connect Contrast Correlate Criticize Deconstruct Deduce Diagram Discriminate Dissect Evaluate Map Outline Prioritize Role-play Separate Subdivide Survey Test	Argue Assess Choose Consider Convince Criticize Critique Debate Decide Defend Editorialize Find errors Grade Judge Justify Persuade Rate Rearrange Reorganize Recommend Reframe Score Support Weigh	Adapt Anticipate Assemble Collaborate Combine Compose Construct Design Develop Devise Express Facilitate Formulate Hypothesize Infer Integrate Intervene Invent Negotiate Originate Plan Prepare Produce Propose Report Revise Simulate Speculate Structure Validate Write

Adapted from Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*, Abridged Edition. Boston, MA: Allyn

Appendix 2: Selected-Response Formats

There are seven item formats:

1. **Conventional Multiple Choice (MC type A):** The quintessential MC format, and the most common format of standardized tests.
2. **Complex Multiple Choice (MC type K):** MC item in which only a specific combination of answer options is correct (e.g., options A and C). Complex MC items are difficult to write, are typically longer than most other MC formats, and require more administration time. MTF format is the preferred replacement format.
3. **Alternate Choice (AC):** Essentially a two-option MC item, in which one option is a distractor. Easier to write and administer than conventional MC and may perform as well. Can be used to test a variety of content and cognitive behaviors.
4. **True-False (TF):** Also referred to as two-choice and binary choice. Includes variations in which a statement is given and the student chooses an answer from yes or no, right or wrong, correct or incorrect, fact or opinion, or any other bipolar set of answer options that can be keyed right or wrong. Accumulating research evidence has identified problems with the use of TF items, but the format continues in classroom testing.
5. **Multiple True-False (MTF):** A hybrid of MC and TF in which the answer alternatives follow a leading question or scenario, and the student evaluates each answer option as true or false. This format has some very positive characteristics, including being easier to write than conventional MC. Guessing is a problem, but this is reduced with a larger quantity of questions. However, the type of cognitive behavior may be lower level than found in conventional MC items.
6. **Matching:** The matching format requires a set of answer options followed by a set of matching stems (statements, questions, or phrases). May be a useful format, but it is uncommon in standardized tests and is not yet strongly supported by research.
7. **Context-Dependent Item** and **Item Set:** Usually has a scenario, vignette, table, chart, graph, reading passage, or other visual material followed by a single item. Requires considerable space in a test and longer administration time. A variation is the item set (also known as testlets, interpretive exercises, context-dependent item sets, and superitems). These formats may be preferable for testing the application of knowledge and skills to a more complex set of behaviors (e.g., mathematical problem-solving ability, reading comprehension, writing skills) in response to stimulus material (e.g., a picture, chart, graph, or vignette).

The preferred item formats are conventional MC (type A), alternate choice (AC), multiple true-false (MTF), matching, and the context-dependent item and item set formats. Avoid true-false (TF) and complex MC (type K) formats.

Appendix 3: Guidelines for Selected Response Items

These guidelines are best understood within the context of an actual educational test or assessment. So before using them to develop your own items, you should first establish your:

1. Purpose for the test or assessment you are developing,
2. Outline specifying the structure and format of your test, and
3. Learning objectives that meet your test purposes and fit within your outline.

The overarching goal of the guidelines is to develop items that measure the target construct for a test or assessment while minimizing the impact of potentially irrelevant constructs.

Content

An item should be carefully constructed to measure only important and relevant content, as specified within the corresponding learning objective. Any content unrelated to the target learning objective, including opinions, misinformation, trivial information, or content or tasks pertaining to different learning objectives, should be avoided.

Item content refers generally to the information, materials, and cognitive task addressed within an item. Different items assessing different content may require different levels of cognitive demand, from simple recall of knowledge to understanding and using complex skills. Cognitive tasks are sometimes referred to as knowledge, skills, and abilities (KSA).

1. **Focus on one type of content and cognitive demand.** Effective SR items typically focus on a single task requiring a specific cognitive demand. An item assessing multiple types of content or multiple levels of cognitive demand becomes more complex and thus more difficult for the student. This added complexity may be unrelated to what we're trying to measure. As a result, students may respond incorrectly not because they lack the targeted KSA, but because they're unable to disentangle and interpret the item content.

   For example, a question on photosynthesis targeted to fourth graders should only assess the intended task, whether it involves the definition of a term like chlorophyll or the interpretation of results from an experiment. Photosynthesis, chlorophyll, and the experiment details constitute item content. Defining and interpreting constitute cognitive tasks. The question then should not require understanding of other contexts, such as solar power, or other terms, such as mitosis, unless they are relevant and specified within the learning objective.

   Note that a good learning objective lends itself to specific content and level of cognitive demand.

2. **Keep content unique and independent across items.** In addition to targeting specific content from the learning objective, this content should be independent of other items. When content for one item depends on another, students can use this dependence to deduce the correct response without actually having the targeted KSA.
3. **Assess important content, avoiding content that is trivial, overly specific, or overly general.** Trivial, specific, or general content often appears when an item does not align well with its learning objective, or when the learning objective itself is flawed. A good learning objective

focuses on specific information or generalities only when they are important to the course or curriculum.

4. **Use novel material and applications to engage higher-level thinking and depth of knowledge.** Items tend to assess low levels of cognitive demand when they include basic content that is more likely to be recalled or recognized by students. Comprehension and understanding can be assessed by incorporating novel contexts, examples, and demonstrations, where students have to apply KSA in a situation they may not have encountered.
5. **Avoid referencing unqualified opinions.** Opinions that are not qualified can be difficult or impossible to answer objectively. Consider a question about the best book of all time. Without reference to a clear metric for what constitutes "best," such as best according to a particular reviewer, or best based on total number of copies sold, no response choice can be identified as most correct. Items referencing opinions, whether qualified or not, may also violate guideline three by assessing trivial, overly specific, or overly general content.
6. **Avoid trick items that intentionally mislead students.** Tricky items have no place in assessments that students are expected to take seriously. Items should be difficult only because they involve challenging content and tasks, not because they introduce misleading or deceptive information.

Format and Style

Guidelines on formatting and style cover best practices for the wording and presentation of item content.

7. **Format the item vertically instead of horizontally, with options listed vertically.** Presenting an item and response options horizontally may save space, but it can lead to difficulties aligning response options with their labels. This tends to be problematic for younger test takers. To avoid confusion, SR items are typically presented vertically.
8. **Edit and proof all items, including for correct grammar, punctuation, capitalization, and spelling.** Items should be proofread and edited for correctness and clarity of wording. Misspellings, grammatical errors, and poor word choice can be distracting at the least. If possible, items should be reviewed by peers with expertise in the subject matter.
9. **Minimize the amount of reading required by each item.** Reading load should be minimized for items that are not intended to measure reading ability or related constructs. By removing unnecessary text, we can focus more on important content while reducing testing time. Items that involve scenarios or examples often include superfluous details and descriptions that can be eliminated without changing the nature of the item. Any information that is not essential to identifying the correct response should be removed.
10. **Keep vocabulary and linguistic complexity appropriate for the target construct and target population.** Our target construct and population should lead us to an appropriate level of linguistic complexity when writing an item. The appropriate level of complexity will depend on the age and reading ability of students, and the role of reading ability within the learning objectives. Reading ability should not interfere with item content. Knowledge of difficult vocabulary should only be required when it is specified within the learning objectives.

The Stem

The stem of an SR item includes the initial description, context, and question statement that a student refers to when choosing an option. Whereas guidelines on item content, formatting, and style pertain to the stem and response options, these two guidelines pertain specifically to writing the item stem.

11. **Include the main idea for the item within the stem instead of the options.** The stem should usually include the majority of the information in an item. Long and wordy response options are more difficult to compare and contrast, especially when it's unclear what question they are addressing. An effective SR item stem can typically stand alone as a constructed-response item. So if you remove the response options from an SR item, with minimal rewording of the stem, a student should still be able to respond.
12. **Word the stem positively, avoiding negative phrasing such as NOT or EXCEPT.** Negatives in an item stem can be easily overlooked, resulting in an incorrect response even for students with the required KSA. Negatives can also increase cognitive load. It is usually best to phrase the stem positively rather than negatively. When the use of negatives cannot be avoided, they should appear capitalized and boldface.

The Response Options

Writing response options can be the most challenging part of constructing an SR item. Incorrect options, also known as distractors, can be especially difficult to write. When an item is found to be too easy, often it is either because the correct option stands out from the rest, or because the incorrect options can be easily discounted.

13. **Use only options that are plausible and discriminating.** Three options are usually sufficient. Good response options follow directly from the question or segue provided in the stem. If the stem asks for the most appropriate interpretation of a statistical result, each option should be worded as a plausible interpretation. If the stem describes a classroom assessment procedure, and then asks for the term that correctly defines it, each response should be a term related to assessment. Discriminating items are ones that distinguish well between students of high and low ability. Students with the required KSA should select the correct option, and other students should select the incorrect options. When this trend is reversed or not apparent, the item provides less useful information about the construct. Although increasing the number of options will tend to make an item more difficult, you shouldn't feel bound to write four or more options per item. The number of options can differ by item, and three options may suffice.
14. **Ensure that only one option per item is the correct or keyed response.** An SR item that asks for a single response may end up having more than one correct answer. This can happen when the stem does not provide sufficient information for disqualifying distractors as incorrect. The problem is best avoided through proofreading and peer review.
15. **Vary the location of the correct response according to the number of options.** Patterns in the placement of correct responses should be avoided. Instead, the placement of the correct choice should be varied across items, ideally with an even distribution.
16. **Put options in logical or numerical order when possible.** When responses are all numeric, they should be formatted similarly, for example, with the same number of decimal places. They should also be presented in ascending or descending numerical order. Text responses that capture labels or terms should be ordered alphabetically when possible.
17. **Keep options independent; options should not overlap in content.** Overlapping or dependent response options typically result in more than one option being correct. For example, consider a

question asking for the correct scientific classification of a domestic cat. The options Felis and Mammalia would both be correct, as one is a subset of the other. The question would be improved by writing mutually exclusive options. The stem itself should also be clarified to ask for a specific level of classification.

18. **Avoid using the options "all of the above," "none of the above," and "I don't know."** Research shows that these response options are difficult to implement and tend to lower the quality of SR items. Alternatives to "all of the above" and "none of the above" are multiple true/false items and items where students select all that apply. These SR item types are more complicated to score but can provide more reliable results.
19. **Phrase options positively, avoiding negatives such as NOT.** Response options, like the SR stem, should be phrased positively whenever possible. The rationale is the same as with the guideline for positive phrasing in the stem. Negatives can be easily overlooked and can make it more difficult to compare and contrast response options.
20. **Make options as similar as possible in length, grammatical structure, and content.** Teachers often provide more justification for the correct response than for incorrect ones, whether to ensure that the correct response is completely correct, or because the incorrect options are more difficult to write. The longer response option stands out from the rest and tends to be right. Students can pick up on this cue and respond correctly without the requisite KSA. In addition to being similar in length, options should also be similar in grammatical structure and content.
21. **Avoid specific determiners, including "always," "never," "completely," and "absolutely."** Extreme determiners like these more often signal incorrect responses than correct ones. Students can recognize that less extreme options are more likely correct.
22. **Avoid repetition and association, where options are identical to or resemble parts of the stem.** Response options are more likely correct when they contain the same or similar wording as in the stem. For example, an item could ask for the type of reliability that is measured within a single test form. Internal consistency may attract more attention than Stability simply because of the association between the terms within and internal.
23. **Avoid groups of options that are similar in structure or terminology.** Subsets of item responses that share the same wording can be distracting. In some cases, they can also appear to include or exclude the correct option, depending on what features make them similar. Groupings among options should be avoided in favor of consistency in length, structure, and content across all options.
24. **Make all incorrect options plausible.** Base incorrect options on typical errors of test takers. This extends guideline 13 by focusing specifically on incorrect options, which should be designed to capture common misconceptions and errors among students. When distractors are written to target common student errors rather than unrealistic responses, they can provide more diagnostic information to inform future teaching and learning. Items with distractors based on common student errors can also be more discriminating. Options should not be used if they could easily be discounted by students. Absurd options are sometimes used for humor or to round out the number of options so it matches other items. Neither of these reasons warrants their use.
25. **Avoid the use of humor.** In smaller scale testing, such as low-stakes classroom assessment where teachers are familiar with their students, humor may not be a problem and may reduce test anxiety for some students. However, it may end up being a distraction for others, which could have implications for fairness. Given the trend toward higher stakes in testing, even with teacher-made assessments, it is recommended that humor be avoided.

Appendix 4: Flipped Dichotomous Key to Rank Multiple-Choice Questions[7]

DRAFT

Step 1

Are students exercising scientific judgment about the strengths and weaknesses of a data set or an experimental design or technique? Are they being asked to determine whether conclusions are consistent with the data?

Examples of key evaluation skills: drawing conclusions about the strength of support for a hypothesis; critiquing the rigor of an experimental design; deciding on the relative merits of different experimental approaches to the same question; understanding which types of evidence are more convincing than others and why. Checking, critiquing. Making a judgment based on research. Making judgments based on internal evidence or external criteria.

Yes: Rank as Evaluating (HOCS) (5)

No: Go to Step 2.

Step 2

Do students have to decide what data are important to solve a problem—for example by picking out relevant from irrelevant information?

Yes: Score as Analyzing (HOCS) (4)

No: Go to Step 3.

Step 3

Are students making inferences from data?

Examples of key skills in data interpretation and analysis: drawing conclusions from graphs—including pedigrees and phylogenetic trees—interpreting standard error bars or the results of statistical tests, interpreting differences in gels, explaining the biological significance of a calculated result, explaining the relationships among terms in an equation, reading micrographs.

Yes: Score as Analyzing (HOCS) (4)

No: Go to Step 4.

[7] Kelly Hennessy, University of Washington BERG group.

Step 4

Are students being asked to diagnose the relationships among multiple concepts or three or more pieces of information—for example by predicting the outcome of a major change or perturbation to a system (multiple changes in variables or conditions or single change with multiple consequences)?

Examples of key higher-order analytical skills: understanding how all of the components of a system relate to each other and to the process as a whole; diagnosis. Determining how parts relate to each other and to an overall structure or purpose through differentiating and organizing or ordering. Breaking down parts of a concept for deeper analysis. Breaking down objects or ideas into simpler parts and seeing how the parts relate and are organized. What do you see as other possible outcomes? What are some of the problems of…? Can you compare your… with that present in…?

Yes: Score as Analyzing (HOCS) (4)

No: Go to Step 5.

Step 5

Are students using a concept or several pieces of information to predict an outcome under new conditions—for example in response to a single, simple change or perturbation in a system? Example of applications tasks: transcribing DNA or translating an mRNA; simple manipulation of variables in an equation to calculate a result. Determining, discovering, expressing, predicting. Executing, implementing. Involves some level of implementation or to follow a procedure. Applying knowledge to actual situations. What do you think will be the end result? What more information can you gather on…? How does this connect with…? What do you think will happen when…?

Yes: Score as Applying — (LOCS/HOCS) (3)

No: Go to Step 6.

Step 6

Are students demonstrating conceptual understanding by matching examples to concepts, recognizing examples of concepts, linking structure and function, distinguishing related ideas or structures, or putting data into a different form? Are they redescribing data to demonstrate they understand what the data represent?

Examples of specific understanding tasks: associate, classify, compare/contrast, describe, indicate, observe, explain, summarize, infer. Grasping the meaning of information.

Yes: Score as Understanding (LOCS) (2)

No: Go to Step 7.

Step 7

Does the question rely on recall but include distractors that represent common misconceptions, so that students have to make distinctions among terms or concepts that are easily conflated?

Yes: Go to Step 8.

No: Score as Remembering (LOCS) (1)

Step 8

Could students rely only on memorized information to answer this question? Stated another way, does the answer just require the ability to recall?

Examples of specific recall tasks: list, recognize, define, name, identify, describe, match, select. Students identify parts of graphs, list or identify variables. Students list characteristics, label diagrams. Remembering previously learned information.

Yes: Score as Remembering (LOCS) (1)

No: Go back through each category or refer to category descriptions to see which fits the best.