

Primer: Writing Question for Readiness Assurance Tests

At the heart of the Readiness Assurance Process is a series of readings and a multiple-choice test that covers the important fundamental knowledge that students will need to know to begin the application activities.

Well-constructed multiple-choice questions are not easy to create. In this section we will examine multiple-choice question writing and how to use item analysis to guide your instruction and improve your questions.

Example:

In your argument, you are citing a number of cases from different courts. This is the first time you cite any of these cases. What is the most accurate citation sentence (use your citation manual)?

- A. Wyman v. Newhouse, 93 F.2d 313, 315 (2d Cir. 1937); Henkel Co. v. Degremont, 136 F.R.D. 88, 94 (E.D. Pa. 1991), Willametz v. Susi, 54. F.R.D. 363, 465 (D. Mass. 1972).
- B. Henkel Co. v. Degremont, 136 F.R.D. 88, 94 (E.D. Pa. 1991); Willametz v. Susi, 54. F.R.D. 363, 465 (D. Mass. 1972); Wyman v. Newhouse, 93 F.2d 313, 315 (2d Cir. 1937).
- C. Willametz v. Susi, 54. F.R.D. 363, 465 (D. Mass. 1972); Henkel Co. v. Degremont, 136 F.R.D. 88, 94 (E.D. Pa. 1991); Wyman v. Newhouse, 93 F.2d 313, 315 (2d Cir. 1937).
- D. Wyman v. Newhouse, 93 F.2d 313, 315 (2d Cir. 1937), Willametz v. Susi, 54. F.R.D. 363, 465 (D. Mass. 1972), Henkel Co. v. Degremont, 136 F.R.D. 88, 94 (E.D. Pa. 1991).

Multiple-Choice Questions have a reputation for only testing lower level skills like knowledge and recall. As you can see from the previous question example students are asked to select the citation that is most accurate. All citations have errors and the students are really being asked to “hypothesize” which errors will have the greatest impact on the citations effectiveness. This question is testing at a very high “Blooms” level. Writing questions at higher “Blooms” level is difficult, but NOT impossible.

When selecting topics for your questions you should begin by reviewing the readings, the activities and objectives. If you have developed objectives and constructed a reading guide for your TBL course, this is a good place to start with identifying important concepts to be tested. Reading guides can be 30-40 questions that students try to answer while completing the readings. This reading guide questions can then be used for ideas for question stem development.

When beginning to construct multiple-choice question you should write the stems first. A well-constructed stem is a stand-alone question that can be answered without examining the options.

Stem Construction

- Stems should be stand-alone questions.
- Stems should be grammatically complete.
- Negative stems should be used with caution.
- Do not incorporate words in distractors that could be incorporated in the stem.
- State stems so that one option is indisputably correct.

The wording of the stem and the verbs it contains, determines the overall cognitive level of the question. It can be useful to use Bloom's Taxonomy to help you prepare the stems to test concepts at the appropriate level. In Table 1 there is a list of verbs that relate to Bloom's levels in the cognitive domain, and in Table 2 the Bloom's level are related to possible question stem starts. Writing multiple-choice questions at the higher Bloom's is possible, but can be very difficult and time-consuming.

Table 1: Bloom's Revised Taxonomy – Verbs

Remembering	Understanding	Applying	Analyzing	Creating/Evaluating	
know	restate	translate	distinguish	compose	judge
define	discuss	interpret	analyze	plan	appraise
memorize	describe	apply	differentiate	propose	evaluate
list	recognize	employ	calculate	design	compare
recall	explain	demonstrate	experiment	assemble	value
name	identify	dramatize	compare	construct	select
relate	locate	practice	contrast	create	choose
		illustrate	criticize	design	assess
		operate	solve	organize	estimate
			examine	manage	measure

Table 2: Bloom's Revised Taxonomy Levels - Question Stems

Remembering	
Recalling information, defining, recognizing, listing, describing, retrieving, naming, finding	<p>What is...? How is...? Where is...? When did ... happen? How would you describe...? Can you select...? Why did...?</p>
Understanding	
Explaining ideas or concepts, Interpreting, summarizing, paraphrasing, classifying, explaining	<p>How would you classify...? What facts or ideas show...? Interpret in your own words...? Which statement supports...? How would you summarize...? What is the main idea of...?</p>
Applying	
Using information in another familiar situation, Implementing, carrying out, using, executing	<p>How would you use...? How would you solve... using what you have learned What approach would you use..? What would result if...? What facts would you select to show...?</p>
Analyzing	
Breaking information into parts to explore understandings and relationships, Comparing, organizing, deconstructing, interrogating, finding	<p>How is....related to...? What is the theme of...? Can you list the parts...? What inferences can you make...? What conclusions can you draw...? How would you categorize...? What is the relationship between...? What is the function of...? What ideas justify...?</p>

Creating

Generating new ideas, products, or ways of viewing things, Designing, constructing, planning, producing, inventing. Evaluating, Justifying a decision or course of action, Checking, hypothesising, critiquing, experimenting, judging

What changes would you make to solve....?
How would you improve...?
Can you propose an alternative...?
How would you adapt...to create a different...?
What could be done to minimize....?
Can you formulate a theory...?
Can you predict the outcome if...?
How would you prove...?
Why did they choose...?
Would it be better if....?
How would you evaluate...?
How would you prioritize...?
How would you justify...?

In TBL, I typically construct the Readiness Assurance Test with 1/3 Bloom's level Knowing (knowledge, recalling facts) and 2/3 Bloom's level simple Application. Once the stem has been constructed, you then create the options (both correct and incorrect answers). Options should focus on testing the understanding of important concepts and testing common misconceptions. Collins (2006) correctly identified the challenge of creating plausible distractors as the most difficult aspect of creating MCQs. She goes on to suggest that the best distractors are: accurate statements that do not meet the full requirements of the problem or incorrect statements that might seem right to the student. Each incorrect option should be plausible but clearly incorrect.

Options Construction

The intended answer is correct and clearly the best.

"All of the above" should be avoided; "none of the above" should be used with caution.

Distractors should be of similar length.

Distractors should be grammatically consistent with stem.

Use parallelism in constructing the distractors.

Numerical answers should be placed in numerical order.

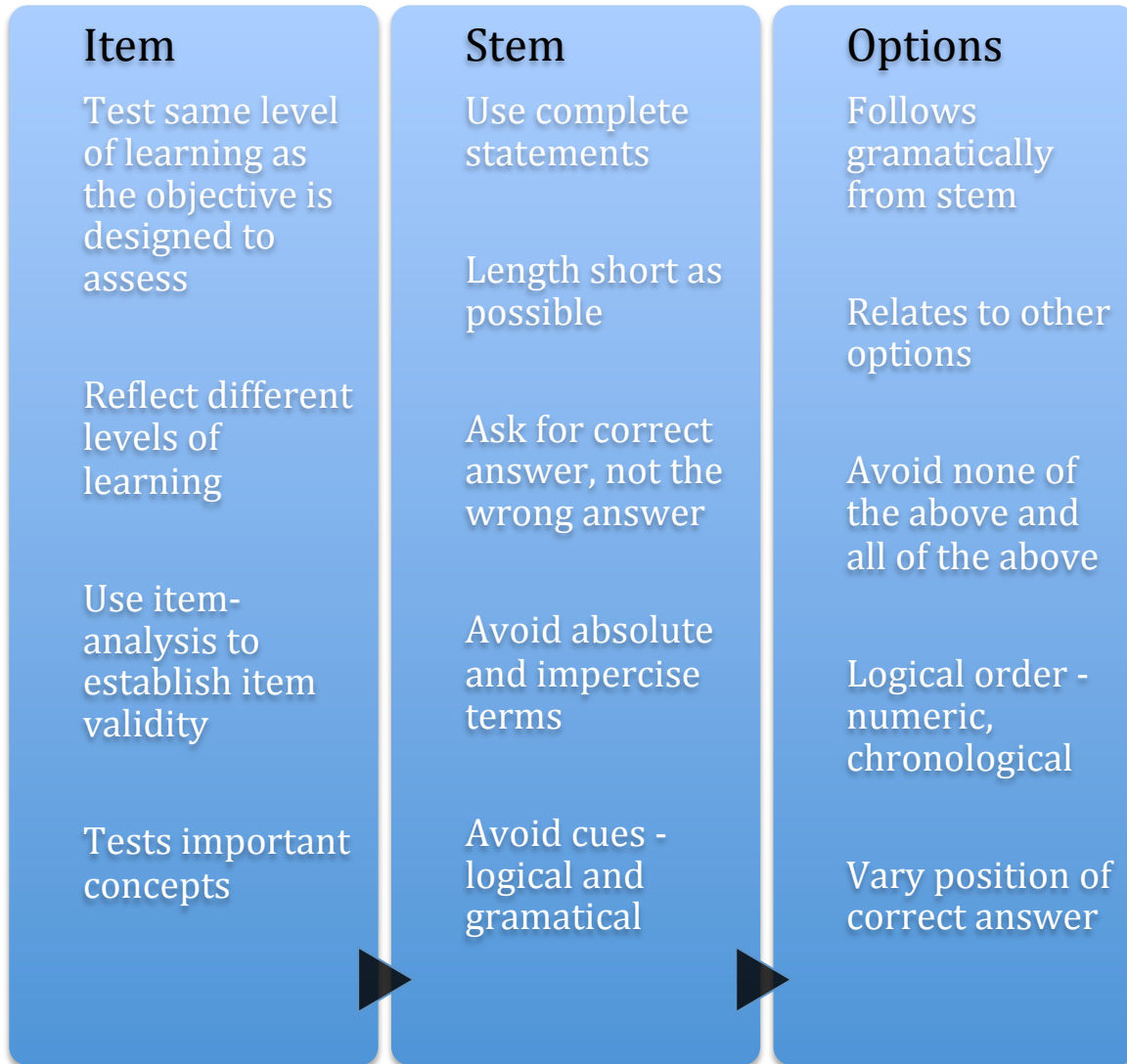
When developing the options it is useful to map the options on a continuum from correct to incorrect. This mapping will allow you to visualize the "correctness" of a given option. If all the distractors except the correct are clustered around the incorrect end of the spectrum then the question will be unambiguous.



As the options begin to cluster at the correct end of the continuum the stem would need to include words like: which is MOST significant? what is MOST important? what would be the BEST solution? These kinds of questions will require finer discrimination by the students. (but can also lead to problems.....use with caution)



Table 3: guidelines for Writing MCQs – adapted from Collins (2006)



What is Item-Analysis?

Item Analysis is a process that analyzes each item on a test to determine the proportions of students selecting each answer. It can be used to evaluate student strengths and weaknesses; may point to problems with the test's validity and to possible bias.

Q	% Correct			Discrim	Tally Frequency				
	Whole Group	Upper 25%	Lower 25%		A	B	C	D	E
1	81	97	56	.46	8	6	79	1	3
2	55	53	59	-.06	1	27	9	54	7
3	99	100	96	.18	0	96	0	1	1
4	100	100	100	.00	98	0	0	0	0
5	89	87	74	.14	5	4	2	87	0
6	89	87	74	.20	18	42	9	13	16

In table above is a typical item-analysis chart. It consists of two major portions: an analysis section that shows the success frequency for bottom 25% students contrasted with the success frequency for the top 25% of students and a tally frequency section that indicates the frequency of all students selecting each distractor. The % correct across different subsets of students can be used to determine the “discrimination” of the question. Discrimination Index represents the ability for the question to discriminate between students who performed well overall on the test (i.e. studied and prepared) and students that did not perform well. As the discrimination approaches zero and becomes negative; students that didn't prepare are more often answering the question correctly than students that did prepare!

You hope to create questions that can discriminate between the students who prepared and those that did not.

Discrimination Index - ranges from -1 to +1 with positive values indicating a positive correlation between success on the question and score on the test. Negative values indicate that students who scored well on the test scored poorly on this question.

$$Index = \frac{H - L}{Nm}$$

where

H = number of students who responded correctly in upper subset of class

L = number of students who responded correctly in lower subset of class

Nm = number of students in either subset

Using the Item-Analysis

There are two useful ways to use the item analysis in your TBL course.

If you use a test sheet scanner in the classroom, you can use the tally frequency portion of the item-analysis to identify questions with which students had difficulty. Understanding where students are having difficulty can help the instructor provide appropriate and timely clarification on the troublesome concepts.

The item analysis results are more commonly used to revise and improve your multiple-choice questions in the off-season. By reviewing the discrimination of a particular question you can understand how the question performed in the test.

In the example below, there is a low discrimination (meaning that students who didn't prepare performed as well or better, than students that prepared). This question should be revised.

Q	% Correct			Discrim	Tally Frequency				
	Whole Group	Upper 25%	Lower 25%		A	B	C	D	E
2	55	53	59	-.06	1	27	9	54	7

In this example below, there is a high level of discrimination (meaning that students who prepared performed better than students who did not prepare). For this question no revision is needed.

Q	% Correct			Discrim	Tally Frequency				
	Whole Group	Upper 25%	Lower 25%		A	B	C	D	E
1	81	97	56	.46	8	6	79	1	3

Typically, when discrimination falls below .6, question writers will re-examine the question and how it has performed to determine if revision is required.

By examining the tally section (labeled "frequency") you can identify distractors that were never selected. This might point to the need to re-write this distractor to be more attractive and thereby require students to more accurately discriminate between all options.

Last few pieces of advice, preparing questions in isolation can be more difficult, write your questions in a group, and test your questions on a colleague or your TA. Ask questions that examine the relationship between topics. Why ask what A is, then what B is, when you could simply ask about there relationship and implicitly know that the student must know A and B to understand there relationship.

This MCQ primer is based on the wonderful writings of Jeanette Collins in her article "Writing Multiple-Choice Questions for Continuing Medical Education.", Woodfords and Bancroft's "Multiple-Choice Questions not Considered Harmful", and the National Medical Examiners Board book "Constructing Multiple Choice Test Questions for the Basic and Clinical Sciences."

The question example on page one was provided by Sophie Sparrow and Margaret McCabe from the Pierce Law Center in Concord, New Hampshire.