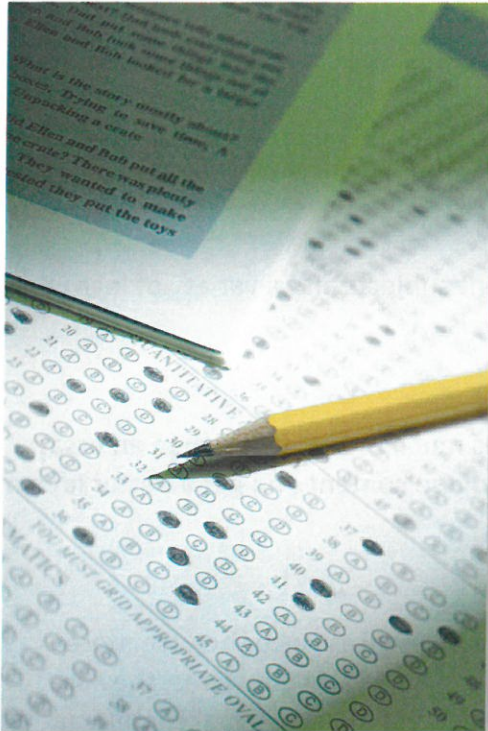


Guidelines for Writing Multiple-Choice Items

DRAFT ONLY

Table of Contents

Introduction.....	3
History of Multiple-Choice Tests	3
Opportunities and Challenges of Multiple-Choice Tests	4
Constructing Multiple-Choice Test Items	5
Other Types of Multiple-Choice Items	8
Common Construction Errors.....	9
Writing Guidelines.....	11
How to Test Higher Cognitive Levels	12
Steps to Writing a Multiple-Choice Item	15
Checklist for Reviewing Multiple-Choice Test Items.....	16
Evaluating Multiple-Choice Tests - Item Analysis	16
Appendix A	19
Bibliography	21



Introduction

Multiple-choice tests have been used for decades. Over the years, they have developed a mixed reputation, with many people believing they are unable to test higher levels of student learning. But this is not the case. If constructed properly, multiple-choice tests can go beyond simple recall and assess higher cognitive levels across a variety of subjects. This guide will discuss the opportunities and challenges associated with multiple-choice tests and provide recommendations aimed at improving the quality of test items. It will also explore methods to test higher cognitive levels and will discuss ways you can evaluate and improve your own multiple-choice items.

History of Multiple-Choice Tests

For centuries, teachers have been assessing their students' knowledge. The testing process has evolved over time as more students enrolled in schools and different methods of measurement and technology were developed.

Educational psychologist Edward Thorndike is reported to have developed prototypes of multiple-choice tests early in the 20th century, before Frederick Kelly, dean of the College of Education at the University of Kansas, introduced the first multiple-choice test in 1914. Lewis Terman adapted that test when he developed the Stanford-Binet Intelligence Test and a host of other exams claiming to measure knowledge, critical thinking and reading comprehension¹.

Over time, as technology improved, multiple-choice tests became a favoured method of assessing students since they could be graded easily and objectively. However, the multiple-choice format doesn't come without its limitations.

Opportunities and Challenges of Multiple-Choice Tests

Opportunities

Versatility

You can use multiple-choice tests to assess different subjects and a variety of learning outcomes, from simple recall to more complex levels.

Validity

Since people can answer multiple-choice questions much more quickly than an essay question, you have time to test a wider range of course content and get a better idea of how students performed overall in the course.

Reliability

Well-written multiple-choice tests can produce reliable scores since they are more objective than short answer and essay questions. Like true-false questions, they are susceptible to guessing, but increasing the number of test items will make it difficult to perform well by blind guessing. For example, if your test includes 15 multiple-choice items of four options each (A,B,C,D), you can expect only 1 out of 8,670 of your students to score 70% or more by guessing blindly (Table 1).

TABLE 1²

Number of 4-Option Multiple-Choice Test Items	Chance of scoring 70% or Higher by Blind Guessing
2	1 out of 16
5	1 out of 64
10	1 out of 285
15	1 out of 8,670
20	1 out of 33,885
25	1 out of 942,651

Efficiency

Multiple-choice tests can be scored in a matter of minutes with the help of a machine, rather than being graded by hand. This quick turnaround time gives students the chance to clarify any issues before the course continues much further.

Challenges

Versatility

Since multiple-choice tests show a list of potential answers, they are limited in the types of learning outcomes they can assess. For instance, students can't display thought processes or explanations.

Difficulty of Construction

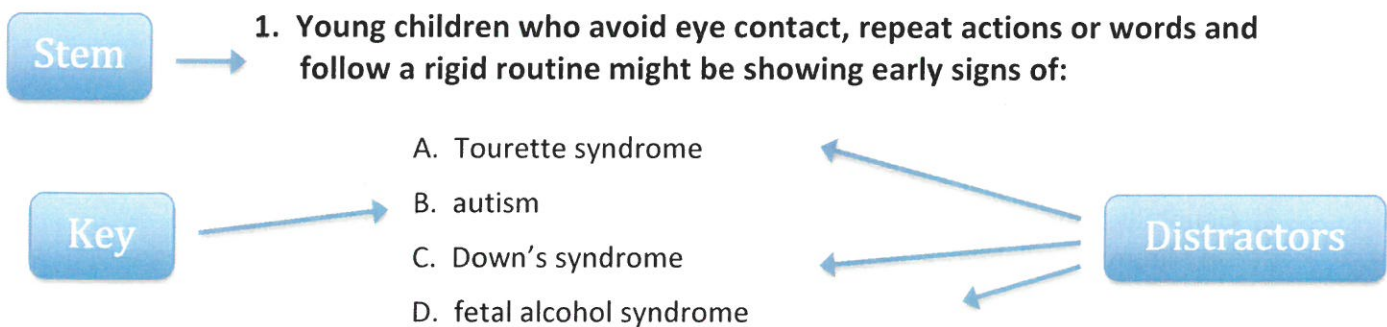
It's generally more difficult and time-consuming to write multiple-choice items compared to other types of test items. It also takes some skill to come up with plausible distractors.

Constructing Multiple-Choice Test Items

Multiple-Choice Item Breakdown

Multiple-choice questions typically have three parts: a stem, the key (the correct answer) and several distractors.

There are a number of ways to design a stem. It can be written either as an incomplete or a complete statement, be brief or lengthy, and it might include numbers, a formula, a chart or pictures. The point is to present a problem that is best answered by one option (the key).



Constructing the Stem

In well-constructed multiple-choice items, the stem presents a self-contained problem with enough information for students to answer without seeing the options. In the

following example, the stem in Item 1 doesn't adequately specify the problem, while the stem in Item 2 is clearer.

Item 1: The Stanley Cup:

- A. is awarded to the playoff champions in the NBA*
- B. was originally commissioned in 1942*
- C. is named in honour of Stanley Kubrick*
- D. was first awarded to the Montreal Amateur Athletic Association*

Item 2: What was the first team to be awarded the Stanley Cup?

- A. Ottawa Senators*
- B. Toronto Maple Leafs*
- C. Winnipeg Victorias*
- D. Montreal Amateur Athletic Association*

Negatively Stemmed Items

It's best to avoid stems that use words such as "not", "except", and "least", but there are exceptions to this rule. For example, if you are testing procedures for a particular treatment, there might be an action that should be avoided. In this case, it would be appropriate to develop a negatively stemmed item to determine if the student knows the danger of performing this action. As is the case with all negatively stated stems, you should capitalize and bold the negative component (i.e., Which of the following is **NOT** an appropriate action when dealing with a spinal injury?).

Constructing the Options

One of the most important things to consider when constructing options is that the distractors must be plausible so that they might appear correct to students who haven't achieved the learning objective. If a distractor is easily eliminated, it will seriously reduce the value of a question.

When deciding on the number of distractors, the main reason to add more options is to reduce the probability of a student selecting the correct answer by chance alone (Table 2). Generally, four-option items are the most effective (A, B, C, D) since a fifth option doesn't improve reliability enough to justify the effort to create it.

TABLE 2³

Number of Options	Probability	Decrease in Probability
2	50%	-
3	33%	17%
4	25%	8%
5	20%	5%
6	17%	3%

One of the best ways to come up with plausible distractors is to anticipate the wrong answers. For example:

Calculate the following using BEDMAS: $(5 \times 6) \times 1^2 \div 3 - 8 + 4 = ?$

- A. -110
- B. 6
- C. -2
- D. -30

A common mistake is to confuse the order of operations, so different combinations are used as distractors.

When creating the options, it's important to ensure that you aren't giving away the correct answer by repeating a key word from the stem. It's also important to avoid phrases such as "you would" at the beginning of an option. Instead, use them only at the end of the stem.

Location of Correct Answer

When they are unsure of the correct answer, many students fall back on the saying "When in doubt, pick 'C'." So, it's important to ensure that you haven't favoured "C" as the correct response. Whenever possible, list the options in ascending or descending order. For example, if you're using numbers, they should be placed in serial order.

Use of "All of the Above" or "None of the Above"

You should avoid these options, particularly if you plan on making them the correct answer, because it's very difficult to argue that a set of options is always the case or never the case. Students can usually come up with an option that is more correct than the one you intended to be correct. For example,

Which city is closest to Ottawa?

- A. Vancouver
- B. Edmonton
- C. Toronto
- D. None of the above

If the student selects D, you don't know if they are thinking about Montreal or Calgary.

Many students eliminate these options because of the common belief that they are only used when the person who wrote the question couldn't come up with a valid distractor.

These suggestions, while not exhaustive, should serve as a useful framework for developing multiple-choice test items. Before finalizing your items, refer to the guidelines on page 11 as well as the checklist provided on page 16. You can also test your knowledge using Berk's Test of Testwiseness found in Appendix A.

Other Types of Multiple-Choice Items

So far we have focused on the traditional model of multiple-choice items with a stem and four options. However, other types of multiple-choice items, like case-based models, do exist. Case-based questions present a hypothetical case study that requires students to identify the problem and solution. These types of questions allow you to test complex thinking and a student's ability to apply knowledge.



Another big method moving forward is the key features format. This is a relatively new concept that highlights the fact that not all steps in resolving a problem are equally important, so it's best to focus on evaluating the critical or challenging steps – the key features⁴.

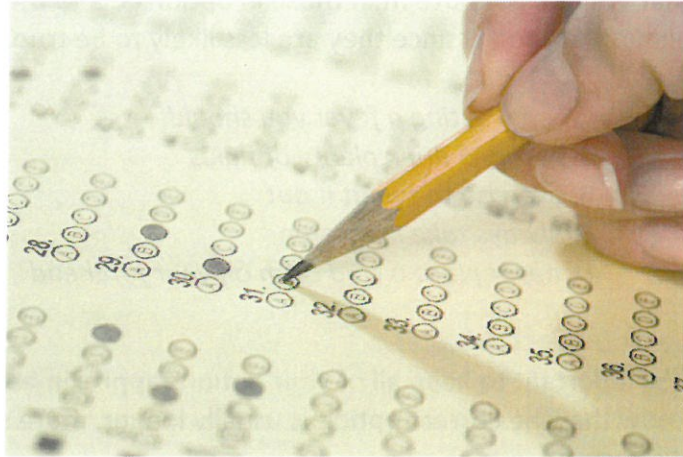
Key features problems are frequently used to assess clinical decision-making skills. While you need specific knowledge to solve a problem, key features takes it a step further by getting a student to apply that knowledge in a scenario to solve a problem.

A general rule to keep in mind when developing key features problems is that if the question can be answered without reference to the clinical scenario, then it is probably not measuring clinical decision-making.

For example, key features problems shouldn't assess a student's ability to "describe features" of a specific illness (a knowledge issue); rather, they should assess their ability to "recognize" a specific illness in a patient (a clinical reasoning issue) and to "prescribe" appropriate therapeutic measures (a clinical decision issue).

Common Construction Errors

There are two main types of construction errors – testwise flaws and irrelevant difficulty. Testwise flaws make it easier for some people to answer the question correctly just based on the fact that they have test-taking skills. Irrelevant difficulty does exactly what it says – it makes the question difficult for reasons that are unrelated to the whole point of the assessment.



Testwise Issues

Grammatical Cues

When distractors don't follow grammatically from the stem, they can easily be eliminated. This is one of the most frequent and easily avoided errors. Since item writers pay more attention to the correct answer, grammatical errors are more likely to occur in the distractors. Looking at the following example, test-wise students would eliminate options B to D because they don't follow grammatically or logically from the stem.

A baby baboon is referred to as an:

- A. *infant*
- B. *cub*
- C. *baby*
- D. *pup*

Logical Cues

When a subset of options is collectively exhaustive, testwise students know that one option must be correct since they include all possibilities. This often happens with options that use the terms "increases", "decreases" and "remains the same". In the following example, options A to C include all possibilities, so some students would know that one of those options must be correct.

The term hypoglycemic refers to someone who has:

- A. *low blood sugar*
- B. *high blood sugar*
- C. *normal blood sugar*
- D. *low blood pressure*

Absolute Terms

Using “always” or “never” as options or including verbs in the options rather than the stem contributes to this flaw. In the following example, options A and B contain terms that are less absolute than those in options C and D, so testwise students would eliminate C and D since they are less likely to be true.

When treating a fever you should:

A. rest and drink plenty of fluids

B. exercise to sweat it out

C. always take ibuprofen

D. never place a cold cloth on your forehead

Long Correct Answer

It’s important to keep all of your options approximately the same length. Most people know that the correct option is usually longer, more specific or more complete than the distractors.

Word repeats

This issue occurs when a word or phrase is included in both the stem and the correct answer. Sometimes a word is repeated in a metaphorical sense. For example, a stem might mention stomach, while the correct answer begins with the prefix “gastro-”.

Convergence strategy

The underlying premise of this flaw is that the correct answer has the most in common with the other options. Students assume that the correct answer is not likely to be an outlier since the person who created the test likely started with the correct answer and wrote variations of it as distractors. For example, if the options are “dog and cat”; “dog and horse”; “dog and rabbit” and “cat and mouse,” the correct answer is likely to be “dog and cat” since “dog” appeared three times, “cat” appeared twice and the other elements each appeared only once.

Irrelevant Difficulty

Options are long and complicated

Long and complicated options can take a significant amount of time to read, which can shift what’s being measured from content knowledge to reading speed.

Numeric data is inconsistent

Numeric options should be listed in serial order and in a single format (i.e., as a single term or as a range).

Frequency terms in the options are vague

Research has shown that vague frequency terms such as “rarely” or “usually” are interpreted differently depending on the individual, so it is best to be more specific.

Writing Guidelines

While many instructors often find it challenging to write good multiple-choice items, there are a number of techniques that can help. Below is a taxonomy of item writing rules developed by Haladyna and Downing⁵ after a review of 46 authoritative textbooks and other sources in educational measurement. They provide a comprehensive set of guidelines that will improve your multiple-choice items.

Procedural

- Use either the best answer (each answer has an advantage, but one of them is best) or the correct answer (only one correct answer) format
- Choices should be listed vertically, not horizontally
- When writing items, use an active, rather than a passive voice
- Minimize the amount of time it takes to read each item
- Avoid giving unintended cues, such as making the correct answer longer than the distractors
- It's best to avoid trick items
- Use good grammar, punctuation and spelling consistently
- Have your questions peer-reviewed and allow time for revisions

Content

- Each item should be based on the objectives of the course, not trivial information
- Avoid overly specific knowledge when developing items or items based on opinions
- Each item should focus on a single problem. Avoid combining multiple ideas into a single question
- Keep items independent of each other. Avoid curing one item with another
- Keep the vocabulary consistent with the examinees' level of understanding
- Avoid textbook verbatim phrasing when developing the items
- Use the author's examples as a basis for developing your items
- Use multiple-choice to measure higher level thinking

Stem Construction

- The stem should be written as a question
- Ensure that the stem is written clearly and that the examinee will know exactly what is being asked
- Include the central idea and most of the phrasing in the stem
- Avoid excessive verbiage
- The stem should be written with positive phrasing; avoid negating phrasing such as "not" or "except". If this can't be avoided, the negative word should always be highlighted by underlining, bolding or capitalizing it

General Option Development

- Use as many functional distractors are possible
- Place options in logical or numerical order
- Ensure your options are independent and do not overlap
- Keep all options in an item homogenous in content
- Keep the length of each option fairly consistent
- Avoid, or use sparingly, the phrases “all of the above” and “none of the above”
- Avoid the phrase “I don’t know”
- Phrase options positively
- Ensure each option is grammatically correct as it relates to the stem so as not to give a clue
- Avoid specific determinates, such as “never” and “always”

Correct Option Development

- Make sure the correct option appears about the same number of times in each possible position for a set of items
- Ensure there is only one correct option

Distractor Development

- Use plausible distractors
- Incorporate students’ common errors in the distractors
- Avoid technically-phrased distractors. Instead use familiar phrases
- Use true statements that do not correctly answer the item
- Avoid using humor when developing options
- Avoid distractors that can clue test-wise examinees
- Distractors that aren’t chosen by any examinees should be replaced

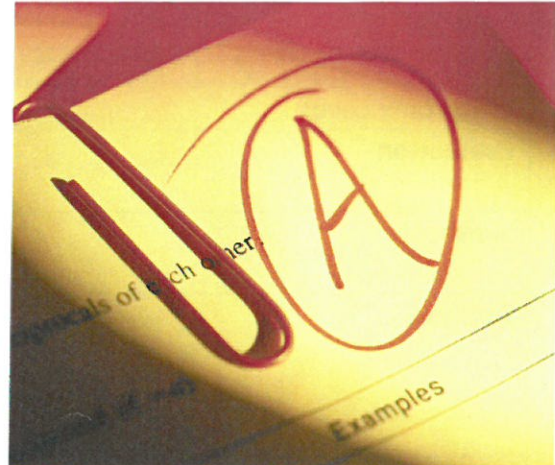
How to Test Higher Cognitive Levels

What is higher-level thinking?

Benjamin Bloom⁶ described six levels of cognitive behavior, starting from the most basic – knowledge – at the bottom to the most complex – evaluation – at the top:

- Evaluation
- Synthesis
- Analysis
- Application
- Comprehension
- Knowledge

J. P. Guilford offered another way of looking at cognition with his description of convergent and divergent thinking. When it comes to assessments, convergent thinking means there is a preexisting correct answer, while divergent thinking means that someone must take existing knowledge and create new knowledge (there is no preexisting right answer). Verbs related to convergent thinking include *select*, *identify*, *calculate*, *label* and *diagnose*, while the verbs for divergent thinking include *create* and *compose*.



Mapping Guilford's concepts onto Bloom's taxonomy, convergent thinking applies to Bloom's first four levels of cognitive behavior (from knowledge to analysis), and divergent thinking applies to the top two levels (synthesis and evaluation). See Table 3.

TABLE 3'

Bloom's Taxonomy with representative verbs	
Evaluation	<ul style="list-style-type: none"> • Critique • Summarize
Synthesis	<ul style="list-style-type: none"> • Organize • Design
Analysis	<ul style="list-style-type: none"> • Compare • Categorize
Application	<ul style="list-style-type: none"> • Organize • Solve
Comprehension	<ul style="list-style-type: none"> • Distinguish • Match
Knowledge	<ul style="list-style-type: none"> • Identify • Label

Writing higher-order multiple-choice items

Tying Guilford's concepts into Bloom's taxonomy suggests that it's possible to write multiple-choice questions for Bloom's first four levels of cognitive behaviour because they require a predictable or calculable answer. Since the top two levels – synthesis and evaluation – are divergent, they are better tested with essay style questions since there is no predetermined correct answer.

Using the cold and flu context as an example, we can see how thinking skills progress (Table 4).

TABLE 4⁸

Sample behaviours for each of Bloom's levels	
Taxonomy Level	Sample Behaviour
Evaluation	<ul style="list-style-type: none"> Assess the effectiveness of that protocol
Synthesis	<ul style="list-style-type: none"> Develop a new protocol for treating the cold
Analysis	<ul style="list-style-type: none"> Compare and contrast progression of cold and flu, or Determine if a patient as a cold or the flu
Application	<ul style="list-style-type: none"> Describe the standard process for determining if a patient has a cold or the flu
Comprehension	<ul style="list-style-type: none"> Match symptoms with their associated ailments
Knowledge	<ul style="list-style-type: none"> Identify three symptoms of a cold

Specific Techniques

In order to test higher cognitive levels, you can ask students to distinguish whether statements are consistent with a principle, concept or rule. Basically, instead of stating a concept and asking students to identify characteristics, you can flip the question by presenting the characteristics in the stem and asking students to identify the concept. In the example below, Item 1 assesses recall, while Item 2 asks the students to understand the potential complication and plan measures for early identification of the problem⁹.

Item 1

A nurse should monitor a client who has had a thyroidectomy for signs of:

- A. laryngeal nerve damage*
- B. increasing intracranial pressure
- C. carotid artery distension
- D. hypercalcemia

Item 2

Which of these nursing measures should be carried out at regular intervals when caring for a client during the immediate postoperative period after a thyroidectomy?

- A. Asking the client to speak*
- B. Checking the client's pupillary response
- C. Palpating the client's carotid arteries
- D. Instructing the client to flex and extend the neck

Asking learners to interpret the meaning of charts and graphs is another way to increase the level of thinking.

Note that even if your question is written at a higher level, if you use examples from reading assignments or presentations, then students may be doing nothing more than recall.

Test Blueprints

Before creating your first item, you need to identify which learning objectives you want to test and build a test blueprint to ensure that you're covering all the topics and objectives you want to measure. The blueprint consists of a chart that displays the number of questions you want to test within each topic. It identifies the objectives and skills that you're looking to test and the relative weight that you're giving to each. See Table 5¹⁰ for a generic example.

TABLE 5

	Topic A	Topic B	Topic C	Topic D	TOTAL
Knowledge	1	2	1	1	5 (12.5%)
Comprehension	2	1	2	2	7 (17.5%)
Application	4	4	3	4	15 (37.5%)
Analysis	3	2	3	2	10 (25%)
Synthesis		1		1	2 (5%)
Evaluation			1		1 (2.5%)
TOTAL	10 (25%)	10 (25%)	10 (25%)	10 (25%)	40

Steps to Writing a Multiple-Choice Item

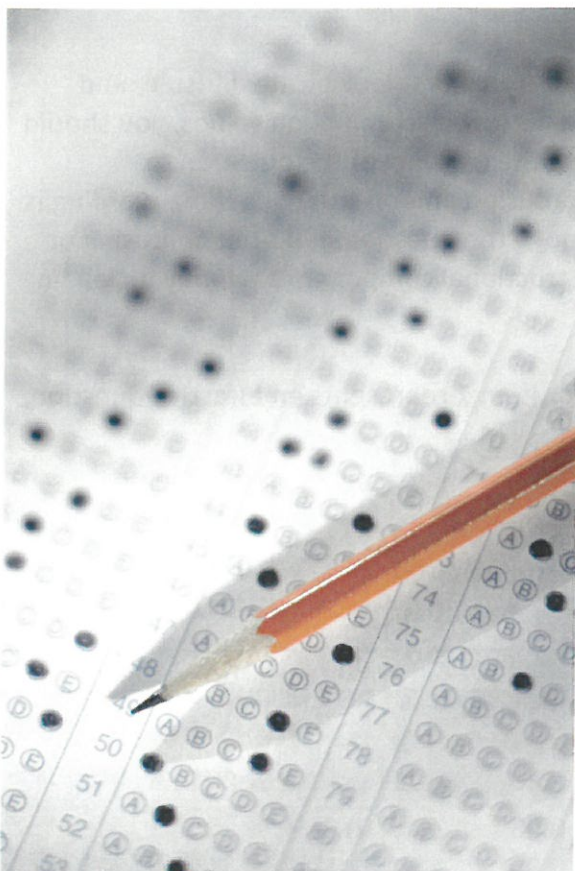
1. Identify the learning objective you're assessing
2. Determine if it's appropriate to use a multiple-choice item to measure the objective (i.e., use Bloom's Taxonomy and decide which levels of knowledge you want to test)
3. Construct the stem
4. Determine the correct answer
5. Create plausible distractors
6. Review the multiple-choice item (i.e., content, formatting, grammar, etc.)

Checklist for Reviewing Multiple-Choice Test Items¹¹

1. Is this the most appropriate type of item to use?
2. Do the items present a meaningful problem?
3. Are the distractors written clearly, without any irrelevant material?
4. Are the stems stated in a positive terms? Or if negative terms are used, has the negative wording been given special emphasis (bold, capitalized)?
5. Are the distractors grammatically consistent with the stem and homogenous in content?
6. Are the distractors similar in length and form?
7. Are the distractors plausible?
8. Does the item include as many functional distractors as are feasible?
9. Is there only one correct and best answer?
10. Are the items free of clues to the correct answer?
11. Are the distractors listed in alphabetical or numerical order if applicable?
12. Have “none of the above” and “all of the above” been avoided as distractors?
13. Has the answer been randomly assigned to one of the alternative positions?
14. Is the grammar, punctuation and spelling correct in the stem, distractors and key?
15. Have the items been reviewed by another individual?

Evaluating Multiple-Choice Tests – Item Analysis

Now that you know the basics of writing a multiple-choice item, it's important to discuss how you will evaluate your tests. An item analysis report is the best way to do this. It provides a statistical summary of how people performed on the test overall, including the mean, the minimum and maximum score and the test's reliability. It also provides information on each question, including the distribution of choices. Two of the most important statistics to consider are question difficulty and discrimination.



Question Difficulty

The question difficulty statistic shows the percentage of students who selected the correct response. The most effective questions are those that are answered correctly by about half the students. Most questions range in difficulty from easy (.90) to very difficult (.40). If a question's difficulty is outside of this range, it probably won't be a very effective evaluation.

It's important to keep tests balanced in terms of question difficulty. A few relatively easy questions are important to verify that some objectives were mastered, while a few very difficult questions are important to challenge the best students.

Instructors are often fascinated by the choices students make among the distractors. The pattern of incorrect

responses can help you identify the confusion students are encountering, which can be helpful the next time you teach that particular concept.

Discrimination

This statistic indicates how well the item discriminated between the students who achieved the highest overall scores on the test and those with the lowest total scores. The discrimination index (item effectiveness) is a correlation that describes the relationship between a student's response to a single question and their total score on the test.

The most effective questions have moderate difficulty and high discrimination values. The higher the discrimination value is, the more effective an item is in discriminating between students who perform well on the test and those that don't. Questions with low or negative discrimination values need to be reviewed very carefully for confusing language or an incorrect key. If you don't find confusing language, then you should review the course design for that topic.

Reviewing an item analysis report

When reviewing your item analysis report, you should first look at the difficulty and discrimination of each question. If questions have low discrimination values, you should review the distribution of responses along with the wording of the question to determine what might be causing potential confusion among students. If the wording is confusing, you can either alter the text or remove the question. If it's not confusing or faulty, then you need to try to identify the instructional component that may be leading to student confusion.

In the following examples, the item analysis for Item 1 represents an effective question, while Item 2 is less effective.

Item 1

Prop. Answering Correctly: 0.74

Discriminating Index: 0.42

Point Biserial: 0.40

Alt.	Total Ratio	Low (27%)	High (27%)	Point Biserial	Key
A	0.74	0.44	0.86	0.40	*
B	0.09	0.17	0.03	-0.13	
C	0.04	0.06	0.08	-0.03	
D	0.13	0.33	0.03	-0.38	
E	0.00	0.00	0.00	0.00	
Other	0.00	0.00	0.00		

Item 2

Prop. Answering Correctly: 0.18

Discriminating Index: -0.03

Point Biserial: -0.01

Alt.	Total Ratio	Low (27%)	High (27%)	Point Biserial	Key
A	0.34	0.25	0.44	0.23	?
B	0.18	0.22	0.19	-0.01	*
C	0.12	0.06	0.17	0.10	?
D	0.36	0.47	0.19	-0.29	
E	0.00	0.00	0.00	0.00	
Other	0.00	0.00	0.00		

Final Note

Creating effective multiple-choice questions that test different levels of student learning takes patience, time and skill. Take the knowledge that you have acquired from this guide to fine-tune your items and see what the multiple-choice format can do for your assessments.

Appendix A

Berk's Test of Testwiseness¹²

Directions: Each of the following items contains at least one flaw. Use this flaw to choose the correct answer. You will be able to answer each item without any content knowledge.

1. Which synthetic fabric, chemically known as "polyacrylonitrtyl," was discontinued by DuPont?
 - A. Dacron
 - B. Acrylon
 - C. Rayon
 - D. Pylon
 - E. Who cares?
2. Which of the following is an "onomatopoeia"?
 - A. hiss
 - B. fizz
 - C. kerplunk
 - D. kerplop
 - E. all of the above
3. Which country in Eastern Europe has a chronic shortage of vowels?
 - A. Saskachawan
 - B. South Dakota
 - C. Bolivia
 - D. Trskmykczstkygistan
 - E. Nambia
4. A group of *frogs* is called an
 - A. gaggle
 - B. army
 - C. pod
 - D. mob
 - E. flink
5. In the TV series, *The Incredible Hulk*, when someone enraged Dr. David Banner, what did he say?
 - A. "De plane! De plane!"
 - B. "What 'choo talkin' bout, Dudley?"

- C. "I've got dandruff older than you!"
 - D. "Don't make me angry. You wouldn't like me when I'm angry."
 - E. "Y'all come back now, ya hear?"
6. Who were the first two "founding fathers" to sign the *Declaration of Independence*?
- A. Thomas Jefferson and John Hancock
 - B. Charles Thomson and Tony Bennett
 - C. Ronald Reagan and John Hancock
 - D. Charles Thomson and John Hancock
 - E. Oscar Madison and Charles Thomson
7. Who was the only "Angel" to appear on the entire run of *Charlie's Angels*?
- A. Jenna Elfman
 - B. Jaclyn Smith
 - C. Roma Downey
 - D. Halle Berry
 - E. Calista Flockhart
8. Which of the following is a sign that you're NOT young anymore?
- A. You choose your cereal for the fiber, not the toy.
 - B. You play connect the dots on your liver spots.
 - C. Getting the mail is one of the highlights of your day.
 - D. You sprinkle tenderizer on your applesauce.
 - E. You look both ways before crossing a room.
9. What is one of only two Disney cartoon movies in which both human parents are present and don't die?
- A. A Beautiful Mind
 - B. The Lion King
 - C. The Lord of the Rings
 - D. Peter Pan
 - E. Titanic
10. In a typical European city, there are
- A. always one big ole famous cathedral near the center.
 - B. exactly six blocks of ritzy, overpriced "boutiques."
 - C. two square miles of quaint, ancient, cobblestone roads, called the "Old City."
 - D. only one famous bridge that's supposed to be romantic.
 - E. one hill with a breathtaking view, always closed for repairs.

Bibliography

1. The Center for Teaching and Faculty Development, San Francisco State University. (n.d.). *Origins and Purposes of Multiple Choice Tests*. Retrieved January 7, 2013, from <http://ctfd.sfsu.edu/teaching-practices/origins-and-purposes-of-multiple-choice-tests.htm>
2. Burton, S. J., Merrill, P. F., Sudweeks, R. R., Wood, B. (1991). How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty. *Brigham Young University Testing Services and The Department of Instructional Science*. <https://testing.byu.edu/handbooks/betteritems.pdf>
- 3 & 11. Marini, A., Violato, C. *Guidelines for Writing Multiple-Choice Test Items*.
4. Medical Council of Canada (2012). *Guidelines for the Development of Key Feature Problems & Test Cases*. Retrieved January 7, 2013, from http://www.mcc.ca/pdf/CDM_Guidelines_Eng_2012.pdf
5. Haladyna, T.M. & Downing, S.M. (1989). A Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2, 37-50.
6. Bloom, B.S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.
- 7 & 8. Dickinson, M. (2011). Writing Multiple-Choice Questions for Higher-Level Thinking. *Learning Solutions Magazine*. <http://www.learningsolutionsmag.com/articles/804/writing-multiple-choice-questions-for-higher-level-thinking>
10. Zimmaro, D.M. (2004). *Writing Good Multiple-Choice Exams*. The University of Texas at Austin Division of Instructional Innovation and Assessment. <http://www6.cityu.edu.hk/edge/workshop/seminarseries/2010-11/Seminar03-WritingGoodMultipleChoiceExams.pdf>

