

# Effective Multiple Choice Tests

Edward Prather

Steward Observatory  
Department of Astronomy



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

Multiple choice testing is an efficient and effective way to assess a wide range of knowledge, skills, attitudes and abilities (Haladyna, 1999).



Center for Astronomy Education

» *Dedicated to the professional development of introductory astronomy instructors*

# Consider your answer to these questions about writing multiple-choice questions.

In a multiple-choice question, when is the longest answer the correct answer?

- A. Rarely
- B. Sometimes
- C. It's common for it to be the correct answer, and it's often stuffed with new information that should have gone in the main part of the course but we forgot so now we're putting it in the quiz because we can't possibly leave out the tiniest detail
- D. Occasionally

Adapted from: <http://blog.cathy-moore.com/2007/08/can-you-answer-these-6-questions-about-multiple-choice-questions/>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Consider your answer to these questions about writing multiple-choice questions.

When is it NOT a good idea to avoid negative questions?

- A. Never
- B. Sometimes
- C. Always
- D. What?

Adapted from: <http://blog.cathy-moore.com/2007/08/can-you-answer-these-6-questions-about-multiple-choice-questions/>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors



# Consider your answer to these questions about writing multiple-choice questions.

How often is the correct choice “A”?

- A. Usually
- B. Frequently
- C. Often
- D. Almost never, because if “A” is the right answer, then the learner doesn’t have to read all the other options we spent so much time writing and revising, and where’s the return on investment in that?

Adapted from: <http://blog.cathy-moore.com/2007/08/can-you-answer-these-6-questions-about-multiple-choice-questions/>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Consider your answer to these questions about writing multiple-choice questions.

When is “All of the above” the correct answer?

- A. With alarming regularity
- B. When we try to cover too much in one question
- C. When we use a question to teach instead of assess
- D. All of the above

Adapted from: <http://blog.cathy-moore.com/2007/08/can-you-answer-these-6-questions-about-multiple-choice-questions/>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Consider your answer to these questions about writing multiple-choice questions.

I opened an online course on a topic I know nothing about, clicked through without reading anything, and took the assessment. I passed! What does that suggest?

- A. I am a genius!
- B. The assessment was too easy.
- C. Maybe the online course was too easy, too.
- D. Maybe the course didn't even need to be written.
- E. B, C, and D

Adapted from: <http://blog.cathy-moore.com/2007/08/can-you-answer-these-6-questions-about-multiple-choice-questions/>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# The Montillation of Traxoline

It is very important that you learn about traxoline. Traxoline is a new form of zionter. It is montilled in Ceristanna. The Ceristannians gristerlate large amounts of fevon and then brachter it to quasel traxoline. Traxoline may well be one of our most luclus snausledge in the future because of our zionter lescelidge.

*attributed to the insight of Judy Lanier*



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors



# The Montillation of Traxoline

It is very important that you learn about traxoline. Traxoline is a new form of zionter. It is montilled in Ceristanna. The Ceristannians gristerlate large amounts of fevon and then brachter it to quasel traxoline. Traxoline may well be one of our most luclus snausledge in the future because of our zionter lescelidge.

Directions: Answer the following questions in complete sentences. Be sure to use your best handwriting.

1. What is traxoline?
2. Where is traxoline montilled?
3. How is traxoline quaselled?
4. Why is it important to know about traxoline?



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

Are you really teaching if no  
one is learning?

How would you know?



Center for Astronomy Education

» *Dedicated to the professional development of introductory astronomy instructors*

# A Commonly Held Inaccurate Model of a Student's Conceptual Framework



Bill Watterson,  
*Calvin and Hobbes*

Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# A Commonly Held Inaccurate Model of Teaching and Learning



Bill Watterson,  
*Calvin and Hobbes*

Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors



# Limitations to MC Questions and Tests

- Constructing good items (stem and choices) is difficult and time consuming.
- It is difficult to find/create plausible distractors.
- Ineffective for measuring some types of problem solving and the ability to organize and express ideas. Real-world problem solving involves proposing a solution versus selecting a solution from a set of alternatives.
- There is a lack of feedback on individual thought processes – it can be difficult to determine why individual students selected incorrect responses.
- Often MC questions focus on factual information and fails to assess higher levels of cognitive thinking.
- They place a high degree of dependence on the student's reading ability and the instructor's writing ability. Students can sometimes read more into the question than was intended.
- May encourage guessing.



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Strengths of MC Questions and Tests

- Learning outcomes from simple to complex can be measured.
- Incorrect alternatives provide diagnostic information.
- Scores are less influenced by guessing than true-false items.
- Scoring is easy, objective, and reliable.
- Can cover a lot of material very efficiently
- Item analysis can reveal how difficult each item was and how well it discriminates between the strong and weaker students
- Performance can be compared from class to class and year to year



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Helpful Hints to Constructing Questions and Tests?

- Base each item on an educational or instructional objective of the course, not trivial information.
- Try to write items in which there is one and only one correct or clearly best answer.
- The phrase that introduces the item (stem) should clearly state the problem.
- Test only a single idea in each item.
- Be sure wrong answer choices (distractors) are at least plausible.
- Incorporate common student naïve ideas or reasoning difficulties in distractors.
- The position of the correct answer should vary randomly from item to item.
- Include from three to five options for each item.
- Avoid overlapping alternatives
- The length of the response options should be about the same within each item (preferably short).



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Helpful Hints to Constructing Questions and Tests?

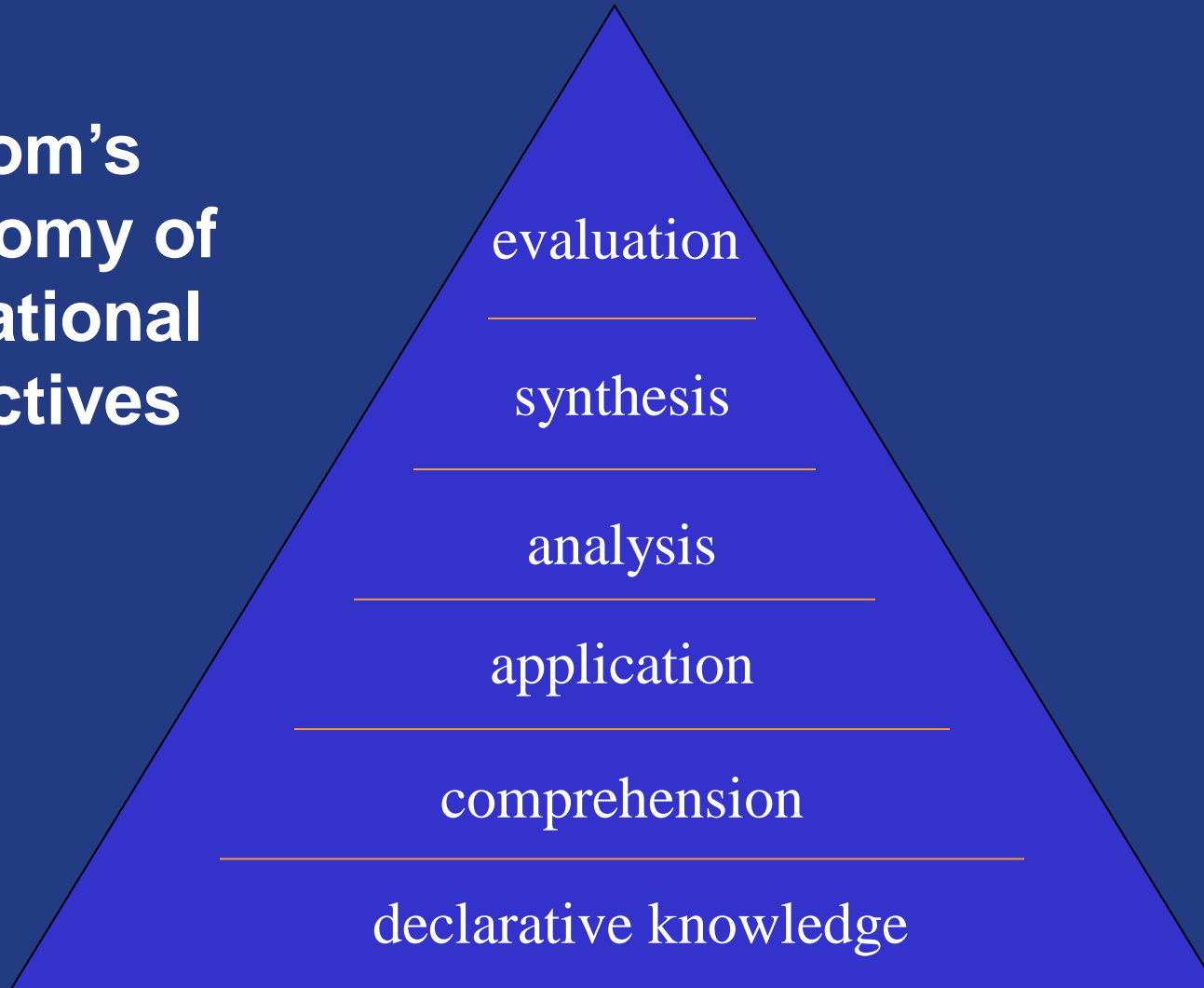
- There should be no grammatical clues to the correct answer.
- Word the stem positively; avoid negative phrasing such as “not” or “except.” If this cannot be avoided, the negative words should always be highlighted by underlining or capitalization: Which of the following is NOT an example .....
- Avoid the excessive use of “All of the above” and “None of the above” in the response alternatives. In the case of “All of the above”, students only need to know that two of the options are correct (in a four or more option question) to determine that “All of the above” is the correct answer choice. Conversely, students only need to eliminate one answer choice as implausible in order to eliminate “All of the above” as an answer choice. Similarly, with “None of the above”, when used as the correct answer choice, information is gained about students’ ability to detect incorrect answers. However, the item does not reveal if students know the correct answer to the question.





Do the questions you use intellectually challenging your students or simply asses their factual knowledge?

## Bloom's Taxonomy of Educational Objectives



# Bloom's Taxonomy

## Described in terms of Students' abilities

- **Knowledge** – Recognizes students' ability to use rote memorization and recall certain facts.
- **Comprehension** – Involves students' ability to read course content, extrapolate and interpret important information and put other's ideas into their own words.
- **Application** – Students take new concepts and apply them to another situation.
- **Analysis** – Students have the ability to take new information and break it down into parts to differentiate between them.
- **Synthesis** – Students are able to take various pieces of information and form a whole creating a pattern where one did not previously exist.
- **Evaluation** – Involves students' ability to look at someone else's ideas or principles and see the worth of the work and the value of the conclusions.

### Reference:

Hellyer, S. A., *Teaching Handbook for University Faculty. Chapter 1: Course objectives.* Indiana University Web site:  
<http://www.iupui.edu/~profdev/handbook/chap1.html>



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Illustrative Action Verbs for Defining Objectives using Bloom's Taxonomy

- **Knowledge** - Cite, define, identify, label, list, match, name, recognize, reproduce, select, state
- **Comprehension** - Classify, convert, describe, distinguish between, explain, extend, give examples, illustrate, interpret, paraphrase, summarize, translate
- **Application** - Apply, arrange, compute, construct, demonstrate, discover, modify, operate, predict, prepare, produce, relate, show, solve, use
- **Analysis** - Analyze, associate, determine, diagram, differentiate, discriminate, distinguish, estimate, infer, order, outline, point out, separate, subdivide
- **Synthesis** - Combine, compile, compose, construct, create, design, develop, devise, formulate, integrate, modify, organize, plan, propose, rearrange, reorganize, revise, rewrite, tell, write
- **Evaluation** - Appraise, assess, compare, conclude, contrast, criticize, discriminate, evaluate, judge, justify, support, weigh

## Reference:

- Gronlund, N. E. (1998). *Assessment of student achievement*. Boston: Allyn and Bacon.



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

# Test Blueprint

Creating a test blueprint, also known as test specifications, consists of a matrix, or chart, representing the number of questions you want in your test within each topic and level of objective.

The blueprint identifies the objectives and skills that are to be tested and the relative weight on the test given to each.

The blueprint can help you ensure that you are obtaining the desired coverage of topics and level of objective.



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors



# Test Blueprint

Bloom Level	Topic A	Topic B	Topic C	Topic D	
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	0	1	0	1	2 (5%)
Evaluation	1	0	1	0	2 (5%)
TOTAL	10 (25%)	10 (25%)	10 (25%)	10 (25%)	40



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

How long does it take the Moon to complete a full cycle of phases?

A. 27.3 days

B. 28 days

C. 29.5 days

D. 30 days

Approximately how long does it take the Moon to complete a full cycle of phases?

A. one day

B. one week

C. one month

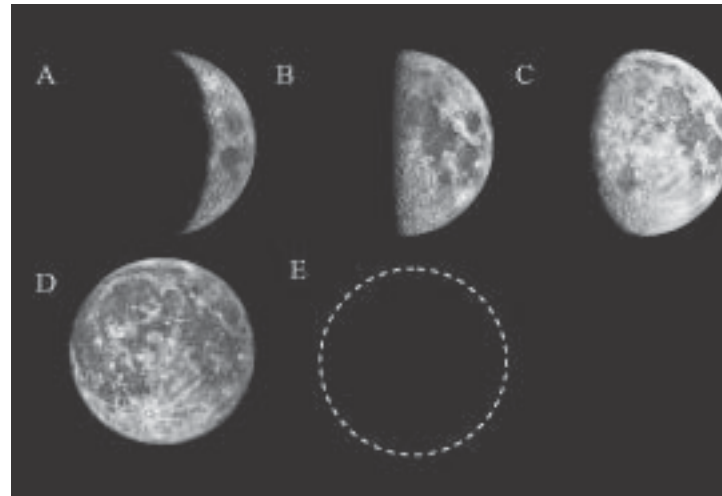
D. one year

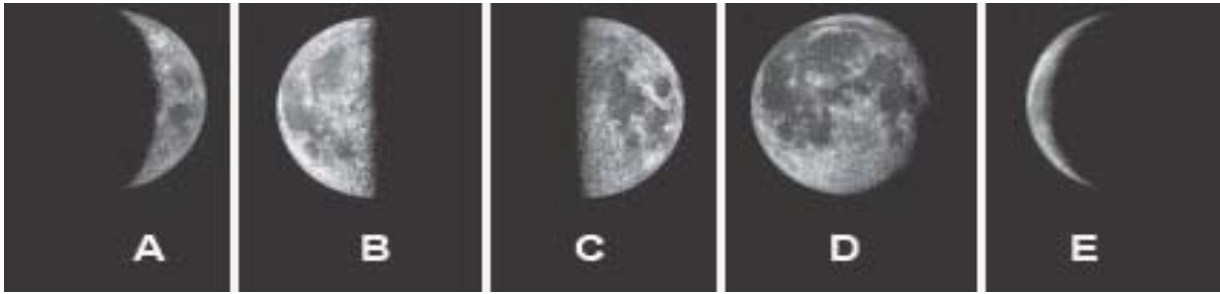
If the Moon is in the Waxing Gibbous phase today, approximately how long will it be until the Moon is in the Waxing Crescent phase?

- a day
- one week
- three weeks
- two weeks
- a month



You look to the western horizon as the Moon is setting and discover that it is in the full moon phase. Earlier that same day when the moon was first rising, which of the moon phases shown below would the Moon have looked like?





If the moon is in the waxing gibbous phase today, how many of the moon phases shown above (A-E) would the moon go through during the next 11 days.

- A. only one
- B. Two
- C. Three
- D. more than three
- E. none

Which of the following groups of moon phases can be above the horizon at 4pm?

A. Full, Waning Crescent, and Waxing Gibbous

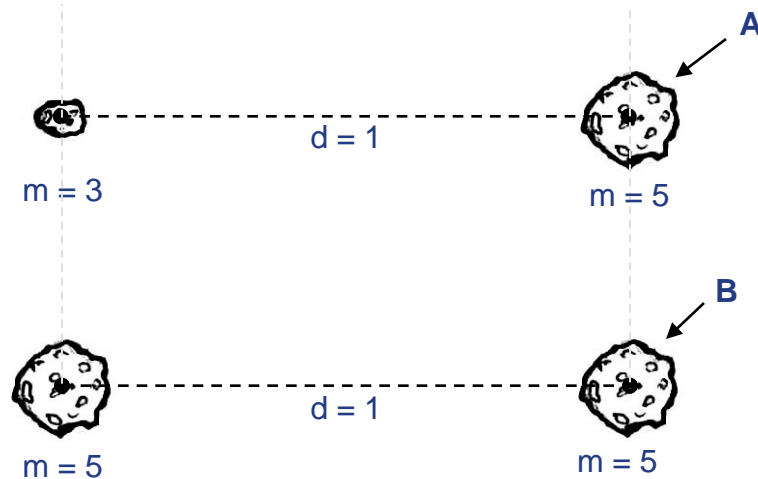
B. New Moon, First Quarter, and Waxing Gibbous

C. Waxing Gibbous, Full Moon, Waning Gibbous

D. Waxing Crescent, Third Quarter, Waxing Gibbous

E. None. The moon is only visible above the horizon during the night time.

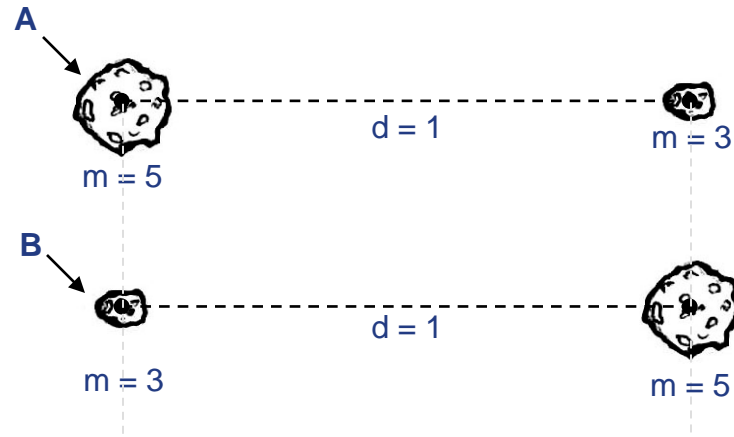
Which is greater, the gravitational force exerted on asteroid "A" or the gravitational force exerted on asteroid "B".



- A. The force exerted on asteroid A is greater.
- B. The force exerted on asteroid B is greater.
- C. The same force is exerted on asteroid "A" and on asteroid "B".

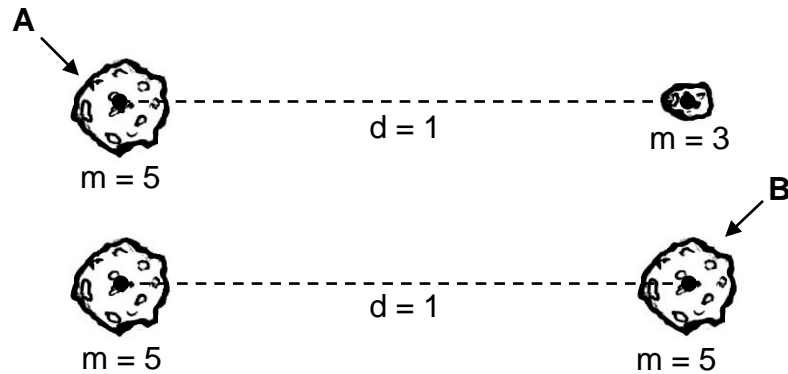


Which is greater, the gravitational force exerted by asteroid "A" on its "partner" asteroid or the gravitational force exerted by asteroid "B" on its "partner" asteroid.



- A. The force exerted by A on its partner asteroid is greater.
- B. The force exerted by B on its partner asteroid is greater.
- C. The same force is exerted by asteroid "A" and asteroid "B".

Which would experience a greater acceleration, as a result of the gravitational force exerted on it, asteroid “A” or asteroid “B”?



- A. Asteroid A would experience the greater acceleration.
- B. Asteroid B would experience the greater acceleration.
- C. The asteroids will experience the same acceleration.

Which of the descriptions given below best describes the sign a person will have if born on that day?

- A. Taurus is high in the southern sky at sunset.
- B. Aquarius is on the eastern horizon at sunrise.
- C. Scorpius is on the western horizon at noon.
- D. Leo is high in the southern sky at midnight.



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

1. The promiscuous use of sprays, oils, and antiseptics in the nose during acute colds is a pernicious practice because it may have a deleterious effect on:

- A. the sinuses
- B. red blood cells
- C. white blood cells
- D. the olfactory nerve

2. Frequent use of sprays, oils, and antiseptics in the nose during a bad cold may result in:

- A. the spreading of the infection to the sinuses
- B. damage to the olfactory nerve
- C. destruction of white blood cells
- D. congestion of the mucous membrane in the nose





1. In 1965, the death rate from accidents of all types per 100,000 population in the 15-24 age group was:

- A. 59.0
- B. 59.1
- C. 59.2
- D. 59.3

2. In 1965, the leading cause of death per 100,000 population in the 15-24 age group was from:

- A. respiratory disease
- B. rheumatic heart disease
- C. accidents
- D. cancer



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

1. About how many calories are recommended daily for a 14-year old who is 62 in. tall, weighs 103 lbs., and is moderately active?

- A. 1,500
- B. 2,000
- C. 2,500
- D. 3,000

2. About how many calories are recommended daily for a 14-year old who is 62 in. tall, weighs 103 lbs., and is moderately active?

- A. 0
- B. 500
- C. 2500
- D. 10,000



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors

**Item Difficulty:** computed as the proportion of students who got the item correct. So a low value means a hard question and a high value means an easy question on a scale from 0.00 to 1.00. It is best to have a mix of difficulties: some hard ones to challenge top students; some easy ones so low-performing students will persist; and the bulk of items at a moderate difficulty level. The average item difficulty across a test, for most multiple-choice types, should be between 0.6 and 0.7 (Gronlund & Linn, 1990).

**Item Discrimination:** How well does the item differentiate among students who have mastered the content and students who have not? This is calculated either as a correlation coefficient between the item score and the total score or as a proportion of high-scoring student who got the item right to low-scoring students who got the item right. Either way, it is expressed on a scale from -1.00 to +1.00. Negative 1 means all low scorers got the item right and all high scorers got the item wrong. Given that we want students who have mastered the content getting each item correct, that's bad. A positive 1 means the item worked exactly as it should. A zero means the item doesn't distinguish between mastery and non-mastery students. That's also bad. There are a number of statistical issues at work that cause +1.00 to be a rare occurrence, so that a reasonable expectation for item discrimination indices is between 0.3 and 0.5 (Oosterhof, 2001).

**Distracter Analysis:** this approach looks to see who is choosing each option for an item. Usually the examinees are divided into low-scoring and high-scoring groups, and the proportion of each choosing each option is reported. High-scoring students will usually pick the correct response and low scoring students will usually pick a distracter. If the opposite happens, that's a cue to revise the item. Something about a distracter is attracting high performance students (Oosterhof, 2001).

**Reliability Coefficient:** Reliability is the degree of accuracy present in the score. For multiple-choice tests, this is indexed using a reliability coefficient like Cronbach's Alpha. Both range from 0.00 to 1.00 with higher values indicating higher reliability. Though there are not strict cutoffs for "acceptable" reliability coefficients, 0.60 to 0.70 are generally considered acceptable lower values (Oosterhof, 2001). Having most item discrimination values at or near .50, having a variety of item difficulties, and having more items will all increase estimates of reliability (Gronlund, 1988; Nitko, 2001; Oosterhof, 2001). In essence, you are calculating internal correlations of student responses between individual items.



A. The promiscuous use of sprays, oils, and antiseptics in the nose during acute colds is a pernicious practice because it may have a deleterious effect on:

- A. the sinuses
- B. red blood cells
- C. white blood cells
- D. the olfactory nerve

B. Frequent use of sprays, oils, and antiseptics in the nose during a bad cold may result in:

- A. the spreading of the infection to the sinuses
- B. damage to the olfactory nerve
- C. destruction of white blood cells
- D. congestion of the mucous membrane in the nose



Center for Astronomy Education

» Dedicated to the professional development of introductory astronomy instructors