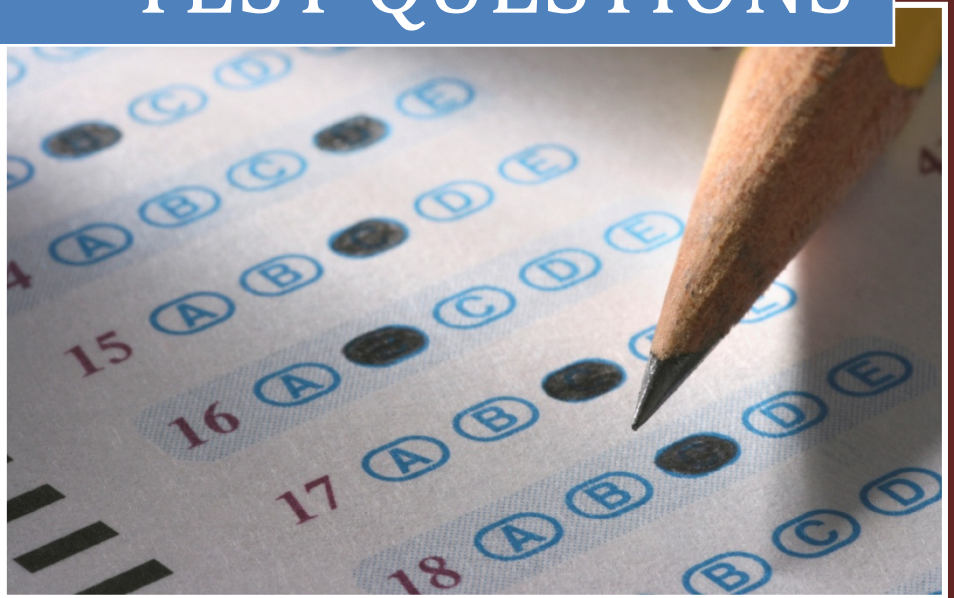


MANUAL FOR THE PREPARATION OF TEST QUESTIONS



National Measurement and Testing, Inc.
P.O. Box 339
Montvale, NJ 07645
866-724-9600
www.nmetest.com

THIS PAGE INTENTIONALLY LEFT BLANK.

MANUAL FOR THE PREPARATION OF TEST QUESTIONS

This document is intended for use by National Measurement and Testing, Inc. and its clients. All rights reserved. Questions and comments should be directed to NMT, Inc.

Contents

I. Maintaining Confidentiality _____	4
II. Basic Characteristics Of Multiple-Choice Items _____	4
III. Types of Multiple-Choice Items _____	5
IV. Guidelines for Multiple-Choice Item Punctuation _____	9
V. Guidelines for Writing Multiple-Choice Items _____	11
VI. Basic Characteristics Of Essay Questions _____	24
VII. Guidelines for Writing Essay Questions _____	24
VIII. Levels of Ability _____	28
IX. Procedures for Generating Items _____	37

I. MAINTAINING CONFIDENTIALITY

Persons engaged in the preparation of questions should bear in mind the confidential nature of this activity. This work should be done in strict privacy, and the utmost care should be taken to prevent as few persons as possible from having access to tentative or final drafts of questions.

II. BASIC CHARACTERISTICS OF MULTIPLE-CHOICE ITEMS

A multiple-choice item is a test question that has a number of alternate response choices, from which the correct or best answer is to be selected. Multiple-choice items typically have 4 or 5 choices, identified by the letters A, B, C, and so on. A multiple-choice item is comprised of the following parts:

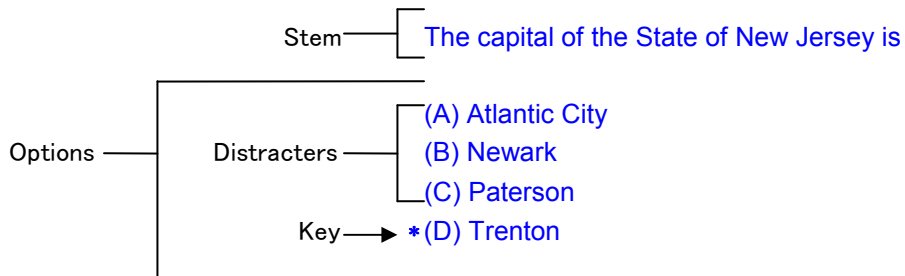
A. Stem: The initial part of the item in which the task or problem is delineated. It may be a question, directions, or an incomplete statement.

B. Options: All the choices in an item.

C. Key: The correct or best option.

D. Distracters: All options other than the key.

Example:



III. TYPES OF MULTIPLE-CHOICE ITEMS

The following are examples of the types of multiple choice items typically employed on tests:

A. Closed Stem Item

This item is characterized by a stem that is a complete sentence which concludes with a question mark. The options may be complete or incomplete sentences.

Example:

Which of the following is generally viewed as a major stumbling block to attaining more international agreements on standards?

- (A) European insistence on adopting the metric system
- (B) Foreign pressure to minimize United States participation
- * (C) Resistance to change on the part of some U.S. industries
- (D) ANSI's refusal to fully participate in international committees

B. Sentence Completion Item

This item is characterized by a stem which is an incomplete sentence. The options represent conclusions to the sentence.

Example:

The movement of substances in and out of the vessels of the cardiovascular system generally occurs within

- (A) arteries
- * (B) capillaries
- (C) veins
- (D) venules

C. EXCEPT Format

This type of item is, in effect, the inverse of the typical multiple-choice item. Rather than being comprised of a correct option and three incorrect options, items of the “EXCEPT” format contain *one* incorrect statement and *three* correct statements. The examinee’s task is to identify the *incorrect* option.

Example:

All of the following are parts of a motor neuron EXCEPT the

- (A) Myelin Sheath
- (B) Neurilemma
- (C) Node of Ranvier
- * (D) Patella

D. MOST/LEAST/BEST Format

This type of item requires the candidate to select an option which is either better or worse than the others. It is important that the basis on which respondents are required to make their evaluations is clearly stated in the stem. Such items should *not* reflect the examiner’s own personal bias. Rather, the correct answer should represent a generally accepted point of view among experts in the field.

Example:

Which of the following company departments is LEAST likely to participate in a value analysis team?

- (A) Engineering
- (B) Manufacturing
- * (C) Marketing
- (D) Quality Control

E. Master List Format

Master Lists are characterized by several different stems that refer to the same options. The options should be parallel—that is, they should list similar objects or events. Widely dissimilar options are not appropriate.

Example:

Questions 1 to 3 refer to the following parts of the nervous system:

- (A) Axon
- (B) Dendrite
- (C) Ganglia
- (D) Myelin

Choose from the above selections the option which is BEST described by each of the following. A choice may be used once, more than once, or not at all.

1. Groups of the nerve cell bodies in the peripheral nervous system (Key=C)
2. A fatty substance that covers most axons (Key=D)
3. The portion of the neuron in which the nerve impulse is generated (Key=B)
4. A substance formed by that spiraling of the Schwann Cell (Key=D)

F. Case Study Format

Case study items consist of one or more paragraphs describing a scenario, followed by several multiple choice questions (typically 3 to 6) based on that situation. An example of a case study in the area of procurement appears below:

Example:

Questions 1 to 3 refer to the following situation:

EFG, Inc. agrees to purchase electronic equipment from DurTech, Inc. The equipment costs \$25,000. During the purchase negotiations, DurTech's sales representative tells EFG's buyer that the equipment possesses certain capabilities not mentioned in any of DurTech's literature, and not expressed by DurTech in any other manner. EFG files suit against DurTech, claiming that the subsequent performance of the equipment disproved the representations made by DurTech's salesperson.

1. EFG most probably will claim that the statements made by DurTech's representative constitute which of the following?

- (A) A breach of the Law of Agency
- * (B) An express warranty
- (C) Oral misstatements
- (D) Unauthorized statements

2. DurTech, Inc. will most likely claim that the sales representative's statements do not constitute evidence because they were

- * (A) excluded from the contract
- (B) misunderstood by the buyer
- (C) solicited by EFG's representative
- (D) conveyed without legal counsel present

3. Which of the following would have been the BEST course of action for EFG to take prior to signing the contract with DurTech?

- * (A) Obtaining the salesperson's statements in writing
- (B) Seeking references from several DurTech customers
- (C) Reviewing the DurTech website for similar claims
- (D) Determining the extent of the salesperson's product expertise

An important consideration in the construction of case studies is that the items should be *dependent upon* the case. That is, the case should contain information needed to answer the question. If the candidate is able to answer the question without the aid of the case, there is little reason to employ it as part of a case study; rather, the item could be used as an independent, stand-alone question.

IV. GUIDELINES FOR MULTIPLE-CHOICE ITEM PUNCTUATION

A. If the optional responses complete the stem, no punctuation is necessary at the end of the options.

Example:

In manufacturing costs systems, variance is the difference between actual cost and

- (A) direct cost
- * (B) fixed cost
- (C) standard cost
- (D) variable cost

B. If the optional responses form an answer to the question as stated in the stem, and if they are in themselves complete sentences, a period is entered at the end of each option.

Example:

A supply manager for DEF, Inc. buys a piece of equipment from supplier XYZ. The equipment is returned to XYZ for adjustment. XYZ makes the necessary adjustments, but then sells the equipment to another buyer, who purchases it in good faith. XYZ then files for bankruptcy. Should DEF seek to retrieve the equipment, which of the following would be TRUE?

- (A) DEF is entitled to retrieve the equipment, because it was bought in good faith from the seller.
- (B) DEF is entitled to retrieve the equipment, because the seller was not bankrupt at the time of the original sale.
- * (C) DEF is not entitled to retrieve the equipment, because the seller sold it to a good-faith purchaser for value, and that purchaser prevails over the true owner.
- (D) DEF is not entitled to retrieve the equipment, because the manager's company ceased to be the "true owner" when it returned the equipment for repair.

C. If the optional responses are incomplete statements or phrases, a period is not used.

Example:

If a patent infringement claim arises, a supply manager is required to notify which of the following?

- (A) The corporation's counsel
- (B) The Court of Patent Appeals
- * (C) The seller of the claim
- (D) The U.S. Patent Office

V. GUIDELINES FOR WRITING MULTIPLE-CHOICE ITEMS

Writing multiple choice test questions is a task that involves considerable knowledge and skill, not just in terms of proficiency with the content areas of the test, but also an awareness of the potential weaknesses in the questions themselves, irrespective of content.

A. Keep the reading and vocabulary level of the item as simple as possible.

It is not appropriate (except in the case of an examination designed specifically to test for abilities in this area) to use exceedingly obscure or difficult words in test items, as verbal ability should not be confounded with the candidate's knowledge of the subject matter.

Example:

Poor:

Punctilious planning of a negotiation session would be LEAST efficacious in accomplishing which of the following?

- * (A) The circumvention of unnecessary costs
- (B) The adduction of a more effective polemic
- (C) The anticipation of counterarguments
- (D) The melioration of relations between bargaining parties

Better:

Careful planning of a negotiation session would be LEAST effective in accomplishing which of the following?

- * (A) The avoidance of unnecessary costs
- (B) The presentation of a more effective argument
- (C) The anticipation of counterarguments
- (D) The improvement of relations between bargaining parties

However, it may be appropriate to test for *technical vocabulary* if it is an objective of instruction or relevant to the topic at hand.

B. Items should not deal with trivial material. Item writers will at times attempt to make items more difficult by basing questions on obscure or trivial detail. Be certain that the item you write deals with important material and not minutia.

Example:

Poor:

In 1929, the death rate for males from accidents of all types per 100,000 persons in the age group 15 to 20 was

- (A) 62.0
- * (B) 62.4
- (C) 62.8
- (D) 63.1

Better:

Of the following, the leading cause of death by accidents for males aged 15 to 20 is fatality due to

- (A) airplane accidents
- * (B) automobile accidents
- (C) sports accidents
- (D) train accidents

C. Be sure that all options in the item are plausible. An important aspect of multiple-choice items is that the incorrect options be plausible, but nonetheless be incorrect. In the example below, the stem informs the examinee that the answer is a United States President. Since Alexander Hamilton was never a U.S. President, the examinee will immediately be able to eliminate Option B, thereby increasing his/her chance of guessing the correct answer.

Example:

Poor:

Which of the following U.S. Presidents instituted the "New Deal?"

- (A) Calvin Coolidge
- (B) Alexander Hamilton
- * (C) Franklin Delano Roosevelt
- (D) George Washington

D. Be sure all of the options are parallel. That is, make certain the options are relatively homogeneous, and that no option can be eliminated simply because it is irrelevant or different in content or focus from the other options. In the example below labeled “Poor”, Options A, B, and D refer to ways in which the automobile industry could be *classified*, while Option C refers to a possible *activity* of the automobile industry. Option C therefore stands out as a different option, and is likely to be eliminated by the candidate for this reason.

Example:

Poor:

In the United States, the automobile industry
can BEST be described as

- (A) a monopoly
- * (B) an oligopoly
- (C) expanding
- (D) a trust

Better:

In the United States, the automobile industry
can BEST be described as

- (A) a monopoly
- * (B) an oligopoly
- (C) a cartel
- (D) a trust

E. The correct answer should not be controversial. That is, the answer should be a statement on which experts would agree, not a viewpoint that is a matter of debate. The item in the example below is a case in point.

Example:

Poor:

When evaluating the performance of a department employee, the factor to which the manager should give the MOST weight is the employee's

- (A) work quality
- (B) dependability
- * (C) cooperativeness
- (D) knowledge of job activities

The appraisal of an employee requires consideration of the specific situation and the particular job's responsibilities. In its present form, it is unlikely that universal agreement could be obtained on the correctness of this item.

This is not to say that one should avoid testing for topics that are a matter of debate or controversy. The appropriate context for such questions would more likely be in an essay format that allows candidates the opportunity to describe, evaluate and contrast different viewpoints with regard to an issue. It would *not* be appropriate to use a fixed-response question to force the candidate into selecting an option that states a matter of opinion as the correct answer. Rather, the test developer should be reasonably certain that the correct answer to a question is one in which experts will agree.

F. Do not write "trick" or catch questions. This type of item can be characterized as one in which subtle but crucial details have been altered in order to lead the examinee into selecting the incorrect option. This type of item tends to work against the *more able* examinee who tries to focus on the major implications of an item, rather than on each word.

Example:

Poor:

Which of the following precipitated an economic depression in the United States?

- (A) The Cuban Missile Crisis of 1962
- (B) The Oil Crisis of 1973
- (C) The Stock Market Crash of 1928
- * (D) None of the above

In the question above, the answer was keyed “D” rather than “C”, because the Stock Market Crash occurred in the year 1929, not 1928. However, the question conveys the idea that the examinee is to identify the Depression as the result of the Stock Market Crash of 1929. If the purpose of the item was to identify the year in which the Crash occurred, then it should be rephrased accordingly. The item in its current form misleads the better examinee who focuses on the meaning of the entire item, rather than on each detail.

G. Avoid using two options to describe the same idea. In the example below, options “A” and “C” use different words to describe the same concept. Since the examinee already knows that only one answer can be correct, upon recognition of the fact that two options describe the same idea, the examinee will immediately be able to eliminate both Options A and C as incorrect, thereby greatly increasing the odds of guessing the correct answer.

Example:

Poor:

A measure of central tendency based on the value that separates the upper 50% of the group from the lower 50% is called the

- (A) average
- (B) mean
- * (C) median
- (D) mode

H. The problem must be clearly and unambiguously posed in the stem. In the poor example below, the stem does not indicate the problem posed by the item. It only tells the reader that the item is related to the hypothalamus. To determine what the actual problem is, we must read each of the options. In effect, this item is nothing more than four true/false items with the words “the hypothalamus” in common.

Example:

Poor:

The hypothalamus

- (A) lies below the cerebellum
- (B) is connected to vision
- * (C) regulates body temperature and appetite
- (D) is part of the cardiovascular system

Better:

Which of the following parts of the brain are involved with regulating body temperature and appetite?

- (A) Cerebellum
- * (B) Hypothalamus
- (C) Pons
- (D) Thalamus

Ideally, an item should be written such that after reading the stem, the examinee can formulate or calculate the answer in his/her mind (or on paper if the item involves computations), and *select* the correct answer from the options provided.

I. Keep as much of the item in the stem as possible. This is partly related to the concerns mentioned in Section H above. But it is also out of efforts to economize space and reading time. The answer choices should generally be shorter than the stem. Any words or phrases that appear in all of the options should be included within the stem.

Example:

Poor:

As nutrients are transferred in a food pyramid, the potential energy

- * (A) will decrease in the final organism
- (B) will increase in the final organism
- (C) will remain the same in the final organism
- (D) will vary in the final organism

Better:

As nutrients are transferred in a food pyramid, the potential energy in the final organism will

- * (A) decrease
- (B) increase
- (C) remain the same
- (D) vary

J. Avoid using negatives such as “no” or “not” in an item. Too many of these can lead to double negatives and unnecessary confusion. Be particularly careful to avoid using negatives within the options.

Example:

Poor:

Which of the following is NOT a safe driving practice on a road covered with snow?

- (A) Not accelerating too quickly from a standing position
- * (B) Jamming the brakes when one is skidding
- (C) Slowing down gradually to a stop
- (D) Steering in the direction of one's skid

Better:

All of the following are safe driving practices on roads covered with snow EXCEPT

- (A) accelerating slowly from a standing position
- * (B) jamming the brakes when one is skidding
- (C) slowing down gradually to a stop
- (D) steering in the direction of one's skid

K. The stem should only contain material needed to explain the problem.

Long, wordy stems with unnecessary material will reduce the effectiveness of the test by increasing the reading load. Moreover, the inclusion of superfluous material increases the likelihood of giving the examinee clues to other test questions. In the poor example below, the item begins with a sentence that informs the examinee that although the composition of the air we breathe is relatively constant, it can vary by zone and by altitude. This sentence may consequently give away the answers to other questions on this topic.

Example:Poor:

The composition of the air we breathe is relatively constant throughout the world, although there can be some slight variations depending upon zone and altitude. Which of the following is the predominant gas which makes up our atmosphere?

- (A) Carbon Dioxide
- (B) Helium
- * (C) Nitrogen
- (D) Oxygen

Better:

Which of the following is the predominant gas which makes up our atmosphere?

- (A) Carbon Dioxide
- (B) Helium
- * (C) Nitrogen
- (D) Oxygen

L. Avoid using “clang associations”. Clang associations are words or phrases that appear in both the keyed answer and the stem. They provide clues as to the correct answer. In the example below, the words “nervous” in the stem and “nerve” in Option B are clang associations.

Example:

Poor:

The synapse is the point at which nervous impulses are

- (A) registered by the brain stem
- * (B) passed between nerve cells
- (C) sensed by the spinal cord
- (D) interpreted by the cerebrum

M. Avoid the use of “specific determiners” in options. Options that contain words such as “all”, “always”, or “never” are likely to be *incorrect*. On the other hand, options that contain “sometimes”, “usually”, “maybe”, or “under certain conditions” are likely to be *correct*. The test-wise examinee (i.e., a person adept at taking tests) will be aware of this and could identify correct and incorrect statements on this basis alone.

In the example below, Options B and D use the words “always” and “never” respectively, thereby indicating that they are likely to be incorrect, while Option C’s use of the term “usually” serves as a clue that this statement is likely to be correct.

Example:

Poor:

The principle reason why smoking is banned in many public places is because

- (A) those who smoke will create temptation in those trying to quit
- (B) nonsmokers will always find smoking offensive
- * (C) second-hand smoke is usually a health hazard
- (D) smoking is never socially acceptable

N. Keep all of the options the same length. An option that is considerably longer than the other options is probably the correct one, due to the need to include qualifications and limitations to make the statement unequivocally true. Keep all options at the same relative length, to avoid giving clues to the test-wise examinee.

Example:

Poor:

The “Hawthorne Effect” refers to a distortion of research results due to

- (A) bias in the observer
- (B) the effect of randomization
- * (C) a positive response that results from the attention that the participants receive
- (D) a violation of ethical standards

O. Be sure that all of the options, and not just the correct one, are grammatically consistent with the stem. Grammatical inconsistency is an obvious sign that an option is not correct. In the example below, Option A can be eliminated for this reason.

Example:

Poor:

The Moon makes a complete revolution around the earth once every

- (A) annually
- (B) month
- * (C) twenty-seven days
- (D) week

P. Use “None of the Above” with caution. “None of the Above” (or “None of These”) is an acceptable option to use within a test item, but only under certain conditions. In general, one should use “None of the Above” only when the keyed answer can be classified as unequivocally right or wrong. In the example below, the use of “None of the Above” is improper because the keyed answer is not absolutely correct. It can be argued in this instance that “None of the Above” is actually the best answer, as another therapeutic mode may be more effective than the three presented in the options.

Example:

Poor:

What is the BEST course of action to take for a patient who is depressed and suicidal?

- (A) Biofeedback
- * (B) Electroconvulsive Therapy
- (C) Free Association
- (D) None of the Above

In the example below, the use of “None of the Above” is acceptable, since the answer is unequivocally correct. If one chooses to use “None of the Above” as an option, one should try to include several items with “None of the Above” as the correct option.

Example:

Better:

If $3x + 12 = 24$, then x equals

- (A) 3
- * (B) 4
- (C) 12
- (D) None of the Above

Q. Do not use “All of the Above”. The use of “All of the Above” is improper for three reasons: First, it is inconsistent with instructions that say that only one option is correct. If “All of the Above” is the correct answer, *all* of the options are correct. Second, when “All of the Above” is used, it is very often the correct option, which serves as a clue to the test-wise examinee. Third, an examinee who can recognize two of the options as being correct can logically conclude that “All of the Above” is the correct answer, without knowing anything about the third option.

As an alternative to “All of the Above”, you may use options that represent various combinations of answers. *Be sure however, that the combinations for all options represent plausible answers.*

Example:

Poor:

Sensory nerve fibers associated with taste travel to the brain via which of the following nerves?

- (A) Facial
- (B) Glossopharyngeal
- (C) Vagus
- * (D) All of the Above

Better:

Sensory nerve fibers associated with taste travel to the brain via which of the following nerves?

- (A) Glossopharyngeal only
- (B) Vagus and Facial only
- (C) Glossopharyngeal and Vagus only
- * (D) Glossopharyngeal, Vagus and Facial

R. Keep the items independent. That is, do not require a correct answer to one item as a condition for solving the next item. Beyond important statistical reasons for keeping items independent, it is also a matter of fairness to give all examinees an equal chance at answering each item correctly.

Example:

Poor:

The President of the United States
during the Civil War was

- * (A) Abraham Lincoln
- (B) Andrew Johnson
- (C) James Buchanan
- (D) Ulysses Grant

The party affiliation of this President was

- (A) Democrat
- (B) Federalist
- * (C) Republican
- (D) Whig

VI. BASIC CHARACTERISTICS OF ESSAY QUESTIONS

Essay questions provide a direct measure of complex learning outcomes that cannot be measured by other means. Such learning outcomes include objectives that ask learners to analyze, synthesize, and evaluate content, as opposed to the mere recall of facts. However, the use of essay questions does not guarantee the measurement of complex achievement. While essays are no doubt easier to construct than objective test items, a hastily constructed essay will be no more effective in testing for high-level objectives than will a multiple-choice item.

The major problem with essays is that they require more subjective, less efficient, and less reliable scoring. Studies have shown that even the same teacher can score the same essay differently at different times. The reliability of essay scoring can be greatly increased if the item writer follows a number of scoring rules. Be aware, however, that the implementation of these rules will require more time for essay scoring, not less. Also, the use of essays will in all likelihood restrict the amount of content to be covered in an examination.

VII. GUIDELINES FOR WRITING ESSAY QUESTIONS

As in the case of multiple choice questions, there are numerous pitfalls to the development of good essay questions. Some important guidelines are presented herewith:

A. Use essay questions only for testing learning outcomes that cannot be satisfactorily measured with objective (e.g. multiple choice) test items.

Other things being equal, it is usually more practical to test material using an objective response item than it would be using an essay question. Besides the obvious reasons of quicker response time, objectivity, and fairness and convenience in scoring, multiple choice items also allow for greater sampling of topics.

B. Be clear in your own mind what mental processes you want the examinee to use in order to answer the question. For example, if your objective is to have an examinee *think critically* about an argument in a particular subject area, one could consider as evidence of this ability the examinee's capacity to evaluate ideas, to evaluate data and information, to check statements against data, and to determine how well facts support a particular conclusion. Once you have decided on the abilities to be appraised, you are in a better position to select, adapt, or create stimulus materials.

C. Write the essay in a manner such that what is to be tested is clear and unambiguous. The examinees' performance on an essay should *not* be a reflection of how well they can interpret the question. The task outlined in the essay question should be understood in the same way by all examinees. If not, a wide variety of answers will result, making them difficult if not impossible to rate according to a uniform standard.

Example:

Poor:

Discuss the Democratic and Republican Parties.

Better:

Compare the current policies of the Democratic and Republican Parties with regard to the issue of taxes. Support your statements with facts and examples wherever possible.

In the poor example above, the question is so vaguely worded that it is likely to yield a wide range of responses. What is meant by “discuss”? Does this mean the examinee is to outline and compare the different points promulgated by the two political parties? Is the examinee to evaluate and critically analyze each argument? The improved version presents the examinee with a clearly defined task, yet still provides a means for the examinee to select and organize an answer.

D. Be careful with the words you use to start the essay question. Try to use words that will require exam takers to select, organize, and utilize their knowledge. Avoid words that encourage the examinee to merely reproduce information.

Poor starting words

“Who...”
 “What...”
 “When...”
 “Where...”
 “List...”

Good starting words

“Compare...”
 “Contrast...”
 “Give reasons for...”
 “Explain why...”
 “Explain how...”
 “Illustrate...”
 “Predict what would happen if...”
 “Give examples of...”
 “Critically evaluate...”
 “Describe the influence of...”

E. Indicate the approximate time needed for each question. There is a tendency for question writers to underestimate the amount of time needed to complete an essay. Gauge the amount of time it will take for the *slower*

examinees to answer the question, and indicate this figure to the examinees as the estimated time for completing the question.

F. Avoid the practice of providing a “choice” of questions. Essay exams often provide examinees with more questions than they are expected to answer, and permit them to choose a given number of questions (e.g. 3 out of 5) when taking the test. Although this practice is likely to be favored by examinees, it has little psychometric value because examinees are obviously taking different tests, and cannot truly be compared.

G. Prepare a “model answer” for each question. This should include an outline of the major points to be covered and the characteristics of the answer to be evaluated. The great advantage of preparing a model answer is that it provides a common basis for evaluating examinee responses, thereby increasing the likelihood that standards for scoring will remain stable. Preparing a model answer at the time of test construction will also address the need for formulating beforehand the mental processes you want the examinees to utilize when answering the questions (see item “B” above). Difficulty in creating a model answer should serve as an indication that the question is likely to be inappropriate or faulty.

H. Determine beforehand the scoring procedure to be used for each item. The item writer should determine beforehand the total number of points to be granted for each question, and should indicate this information to the examinees next to the test questions. There are generally two methods for scoring essays:

Point method – Essay answers are rated against a model answer and are given points in accordance with how well the examinee’s answer compares to that model;

Rating method – Essays are placed into separate piles representing degree of quality, and credit is determined for each pile. For instance, if an essay question is worth 10 points, one could place the essay papers into five separate piles, and assign the highest quality papers 10 points, the next set 8 points, the next 6, and so on (3 to 5 categories are usually sufficient).

I. Evaluate all of the answers to one question before going on to the next question. This will decrease the likelihood of grading papers according to shifting standards, as well as reduce the likelihood of a “halo effect”, whereby the answers to the first few questions are likely to influence one’s judgment on the remaining answers.

Another point to keep in mind is that a paper with average answers is likely to appear to be of much higher quality when it follows a failing paper than when it follows one with near perfect answers. One way to deal with this problem is to score all the answers to the first question, shuffle the papers, score the answers to the second question, and repeat the process until all of the papers have been scored.

J. Decide beforehand how you plan to deal with factors that are irrelevant to the learning outcomes being measured. A number of factors that graders tend to consider when evaluating answers to essay questions are actually irrelevant to the purposes of the test. Chief among these are grammar, spelling, handwriting and punctuation. Though one obviously desires to see examinees demonstrate good writing and spelling skills, those factors in themselves are not reflective of the examinee's knowledge of the subject matter addressed by the essay question. Therefore, to maintain objectivity, the essay grader needs to make certain that such factors are *not* being considered in scoring.

If, on the other hand, one wishes to give some weight to grammar, spelling, handwriting, and the like, the exam developer should inform the examinees of the intent to do so on the test form, and allot points for these factors within the scoring scheme. For example, if an essay question is worth 10 points, you may wish to reserve 1 or 2 of the 10 points for such factors as grammar or written presentation, irrespective of the correctness of the actual answer.

VIII. LEVELS OF ABILITY

This section presents a set of sample items written at progressively difficult levels of learning or ability. The levels are taken from Benjamin Bloom's *Taxonomy of Educational Objectives: Cognitive Domain* (1954), a frequently-used resource in the field of testing. The items below cover the various levels outlined in the *Taxonomy*, and provide some perspective on the different ways in which test questions can be written.

A. Recall

This is the lowest, or easiest level of learning. Items written at this level are those that demand the recall of ideas, material, or phenomena related to the topic of interest. The major psychological process used to correctly answer such items is the examinee's *memory* of the material.

Example:

The type of document used to enter into an interim agreement pending a definitive contract, so as to permit the start of delivery, is called a

- (A) customized contract
- * (B) letter of intent
- (C) purchase order
- (D) standard contract

To correctly respond to the item above, the examinee has to recall the definition of a letter of intent.

B. Comprehension

The second level of learning deals with items that test for comprehension. Items in this category require the examinee to grasp the *meaning* of the material presented in some novel way. An item testing for comprehension describes some principle or fact in words different from those used in textbooks, and often uses some novel situation as a way to present an idea. In order to get the item right, the examinee must *recognize the principle* demonstrated in the problem; memory alone will not be sufficient for determining the correct answer.

Example:

Which of the following is/are LEAST useful in evaluating a supplier's financial position?

- (A) Current balance sheet
- (B) Dun & Bradstreet reports
- * (C) Market forecasts
- (D) Monthly income statements

In order to answer the item above correctly, the examinee must have knowledge of the data typically provided by each of the sources listed in the options, and must also be able to recognize which of that information is applicable to evaluating a firm's financial position and which is not.

Example:

A buyer orders paint from a supplier and specifies that the paint must dry tack-free in four hours, and dry hard in twelve hours. These specifications are examples of

- (A) engineering specifications
- * (B) performance specifications
- (C) restrictive specifications
- (D) technical specifications

To answer the above item correctly, the examinee must not only know the definition of performance specifications, but must also be able to recognize circumstances that illustrate this definition.

Other questions dealing with this level of testing begin with statements such as:

"Which of the following BEST explains...?"

"Which of the following BEST illustrates...?"

"What trend do you predict in...?"

"Which of the following is an example of...?"

C. Application

Application items measure the understanding of ideas or content to the point where the examinee can use his/her knowledge in an entirely new situation.

Example:

A customer buys 100 shares of DEF stock at 32, and at the same time writes a DEF Jan 40 call at 11. At expiration, DEF must be selling at what price in order for the customer to break even?

- (A) 11
- * (B) 22
- (C) 32
- (D) 40

The problem presented in the item above is a novel situation; in this case, it pertains to the break-even price on a listed option. Rather than relying on memory or comprehension to answer the question, the examinee is required to draw on his/her knowledge and experience to identify a solution to the problem.

Other questions dealing with this level of testing begin with statements such as the following:

"Which situation would require the use of...?"

"What is the BEST procedure for dealing with...?"

"What steps should be followed in applying...?"

The Application level can also be tested through the use of essay questions. An example appears below:

Example:

A customer buys 100 shares of DEF stock at 32, and at the same time writes a DEF Jan 40 call at 11. Compute the price at which DEF must be selling at expiration in order for the customer to break even.

It is important that the Application item presents a specific situation that the examinee has not previously encountered, and cannot solve on the basis of general knowledge alone. When writing an item of this type, you should try to recall or develop specific situations that are unique and unfamiliar to the examinees.

D. Analysis

Analysis involves the ability to break down material into its component parts, so that the organizational structure can be understood. It involves the ability to recognize these parts and the relationships between them, and to identify the principles involved. Items at this level typically ask an individual to differentiate, discriminate, distinguish, infer, and evaluate the relevancy of data. Analysis items go beyond comprehension and application because they require an understanding of content and the structural form of material.

Example:

A firm has a contract with a supplier for the delivery of \$300,000 worth of supplies. The shipment arrives at the buyer's site on June 3rd. The buyer accepts the shipment. Two days later, the receiving clerk tells the firm that the supplies were not what the firm originally wanted, and that the goods should be sent back. The seller refuses to take back the supplies. Is the buyer entitled to return the shipment under the Uniform Commercial Code?

- (A) Yes, the buyer has the right to do so
- (B) Yes, but only because they were rejected within 5 days of receipt
- (C) Yes, but the buyer must pay a penalty
- * (D) No, but the goods can be revoked if they later show defects

In this item, the examinee is presented with a novel situation, and is asked to determine which data, if any, are relevant to the Uniform Commercial Code. Furthermore, the candidate is asked to distinguish between those circumstances and alternatives that are addressed by the Code.

The Analysis level can also be tested through the use of essay questions:

Example:

Statement of Income
JKL Manufacturing Company
January 1 – December 31, 20xx

Net Sales	\$2,600,000
Cost of goods sold	- 1,828,500
Selling, general, and administrative expenses	- 246,400
Depreciation	- <u>67,200</u>
Operating income	\$ 457,900
Other income	+ <u>11,250</u>
Total income (EBIT)	\$ 469,150
Interest on bonds	- 52,000
Taxes	- <u>100,500</u>
Net income	\$ 316,650
Preferred dividends	- <u>7,650</u>
Earnings available for common stock	\$ 309,000

Analyzing the information provided above, determine if JKL can sufficiently provide for payment on its outstanding bonds. Explain why or why not.

E. Synthesis

Another level of learning known as Synthesis involves the creation or formulation of information from a variety of sources. An example of an essay question testing at the Synthesis level appears below:

Example:

TUV, Inc. is a medium-sized firm that has implemented Manufacturing Resource Planning (MRP) for the past decade. Stock-keeping units exceed 25,000, and the number of transactions processed is substantial. As TUV's supply manager, you have been asked by senior management to prepare monthly reports of inventory and purchasing effectiveness, showing quantitative contributions to the organization's business plan. As currently organized, the data within the system are not useful, and thus senior management wants a proposed design for a new monthly report. Time and budgetary resources for instituting the changes are limited, so the new reports need to be comprehensive and yet void of superfluties.

Design an information system for TUV in accordance with senior management's directive. Construct a brief proposal for this revision, including the data to be included in the reports, justifications for including the data, and how it will contribute to the organization's business plan.

In this example, the examinee must draw upon his/her knowledge of information systems and inventory management, and put various pieces of information together in order to form a new whole. The key aspect of problems testing for Synthesis is that the examinee is developing completely new patterns or structures. Accordingly, questions in this category stress creativity and extended thinking.

Problems and exercises at the Synthesis level do not lend themselves to the multiple-choice format; thus no such examples are presented here.

F. Evaluation

This is the highest level of learning, and it involves the ability to judge the value of material for a given purpose, based on defined criteria. Items in this category include those that ask individuals to appraise, conclude, support, compare, contrast, interpret, and summarize information.

Example:

Troy Toys is a retailer operating in several cities. The individual store managers deposit daily collections at a local bank in a non-interest bearing checking account. Twice per week, the local bank issues a depository transfer check (DTC) to the central bank at headquarters. The controller of the company is considering using a wire transfer instead. The additional cost of each transfer would be \$25. Collections would be accelerated by two days, and the annual interest rate paid by the central bank is 7.2% (.02% per day). Assuming a 360-day year, which of the following actions should you recommend in this situation?

- (A) Do not use the wire transfer at all
- (B) Use the wire transfer if the amount transferred is \$173 or above
- * (C) Use the wire transfer if the amount transferred is \$62,500 or above
- (D) Use the wire transfer if the amount transferred is \$125,000 or above

In this problem, the examinee is confronted with a novel situation, and is asked to compute the costs to the firm for the different alternatives. Furthermore, the candidate is asked to compare, contrast, and appraise the results, and determine the most economically feasible alternative.

Example:

A home services organization has been using the straight line method in calculating the depreciation expenses of its equipment. The firm takes one-half year's depreciation in both the year of acquisition and the year of disposal. Based on recently acquired information, the firm's assistant controller has altered the estimated useful lives of the equipment. The corresponding changes in depreciation result in a change of a small profit for the year into a loss. The assistant controller is asked by his supervisor, the controller, to reduce by half the total depreciation expense for the current year. The assistant controller believes that he is faced with an ethical conflict.

Questions 1 to 4 refer to the following options:

- (A) Appropriate as an immediate action
- (B) Appropriate action only after other alternatives have been tried
- (C) Not an appropriate action
- (D) Not enough information given to evaluate this action

Given the scenario described above, evaluate each of the following courses of action taken in response to this situation, and select the option that BEST characterizes each action. A choice may be used once, more than once, or not at all.

1. The assistant controller reports the problem to the Board of Directors. (Key=B)
2. The assistant controller reports the matter confidentially to the local news media. (Key=C)
3. The assistant controller reports the problem to the vice president of finance. (Key=D)
4. The assistant controller discusses the problem with a colleague in the department. (Key=A)

In this exercise, an examinee is presented with a real-life, novel situation, and is asked to evaluate each of the various courses of action one could take. Item #1 was given "B" as the correct option, because while it is an appropriate response, the assistant controller may be able to resolve the situation by less drastic means first. Item #2 is a totally inappropriate response that will generate unwanted and unnecessary publicity for both the firm and the assistant controller. Item #3 *may* be an appropriate response, but it is not clear that the vice president of finance is next within the reporting line to the controller. Item #4 is an appropriate action for the assistant controller to take - the colleague is close to the problem, and may offer a suggestion that will help to prevent the situation from getting out of hand. In each case, the examinee is asked to make a *judgment* on the appropriateness of the various actions to the situation.

An example of an essay question testing at the Evaluation level is as follows:

Example:

Critically evaluate the advantages and disadvantages of the Electoral College system within the United States. Explain whether or not you believe this system should continue to be employed, and provide justification for your position.

IX. PROCEDURES FOR GENERATING ITEMS

The development of good test items, particularly multiple-choice items, is a challenging process that not only requires extensive knowledge of the area to be tested, but also considerable creativity. At times, item writers may feel they've reached a "mental block" when it comes to developing suitable questions. They may know the material, but cannot think of an original or suitable question that will adequately test for it. The purpose of this section is to present three methods for generating test items. These procedures are taken from the educational measurement and job analysis literature, and have been used with a variety of tests, as well as for related functions such as questionnaire development. If utilized properly, they can help the writer create a number of good items.

One may find these methods useful in situations where the material does not readily lend itself to the formation of test questions. However, not all exam content areas will lend themselves to these procedures. The writer is encouraged to employ any approach or resource which helps him/her to develop meaningful material.

A. Prose Transformation

In this method, items are created by transforming sections of prose, including sentences or groups of interrelated sentences. The rationale for this method is that it will match the content of the test to content of the textbooks and courses in a very direct way. This method is employed as follows:

Step #1: Screen the prose for instructionally relevant key sentences. Most prose will include at least some useless or irrelevant material. It is therefore important that the writer recognize and screen out this material beforehand.

Step #2: Select the important sentences. The identification of important sentences will guard against the testing of trivial points. Important sentences can be identified in 3 ways:

A) Selection by subject matter experts. That is, a subject matter expert can highlight the important sentences in relation to the learning objectives of the material.

B) Writing of summary sentences. Important sentences can be identified by writing summary sentences that abstract the important ideas from the prose passage.

C) Identification of key words. The important or key words of a passage (usually nouns) can be highlighted. Rare nouns in a passage will often turn out to be the important terms the examinee needs to know.

Step #3: Transform the sentences. The transformation of important sentences into test questions involves five steps:

A) Copy the sentences onto a worksheet or card.

B) Clarify references to other sentences. That is, substitute all pronouns that refer to previous sentences (e.g., this, that, they, it) with the appropriate noun phrase.

C) Simplify the sentences. Screen out any extraneous clauses or sentences so that only the segment with the key noun remains.

D) Replace the key noun or phrase with who, what, where, why, when, or which.

E) Rewrite the sentence into a question.

Step #4: Construct the item distracters. The distracters for an item can be obtained from a list of related key nouns or phrases that the examinee may tend to confuse with the correct answer.

An example of an item created with the prose transformation method appears below. The passage is taken from a business text.

Example:

COMMON AND PREFERRED STOCK

There are two basic types of stock: common stock and preferred stock. Common stock is a certificate showing ownership in a corporation. All shares of common stock are equal in value and all common shareholders enjoy the same rights. **Common stock is voting stock and common shareholders have a residual right to a corporation's earnings. They have a right to earnings that remain (residual earnings) after the corporation has met the prior claims of bondholders and preferred shareholders.** The actual payment of a common stock dividend does not occur until the board of directors declares a common stock dividend. If the corporation goes bankrupt, the common shareholders are the last to receive any proceeds from the sale of the corporation's property. Creditors, bondholders, and preferred shareholders share in the proceeds before the common shareholders. Thus common shareholders are the residual owners of a corporation.

Preferred stock is a certificate that also shows ownership in a corporation. Preferred shareholders usually cannot vote their shares. But they do enjoy certain preferences. As we have seen, they have a right to receive the dividend indicated on their stock certificates before common shareholders receive any dividends. This dividend is also not owed until declared by the corporation's board of directors.

Key Words: "Common shareholders...right to earnings...after prior claims of bondholders and preferred shareholders"

Sample Test Item:

Before paying dividends to its common shareholders, the corporation must first meet the prior claims of which of the following?

- (A) The Board of Directors
- (B) Bondholders only
- (C) Preferred shareholders only
- * (D) Bondholders and preferred shareholders

The primary drawback with this approach is that it can lead to trivial items geared to unimportant ideas, and the verbatim recall of text. The intent of this method, however, is not to transform any sentence into a question. It is instead designed to help the item writer focus on important ideas within a reading passage that can be the source of good items.

B. Item Forms

In this method, items are developed by a specific operational definition. This procedure is taken from methods used to develop mathematics curriculum materials, and is probably best suited for quantitative items. The item form method generates items using a fixed syntax containing one or more variable elements with a set of replacement values for those elements. The procedure is to develop the item form, and by substituting values within the variable elements of the form, individual items are generated. A simple kind of item form is one employing a "skeleton" of an item, with blank spaces in which words or numbers are inserted.

Example:

If it takes an employee 30 hours to produce one unit of a specific item, what would be the average labor hours per unit after the production of the 8th unit, assuming that an 80% learning curve is in effect?

- (A) 12.30 hours
- * (B) 15.36 hours
- (C) 19.21 hours
- (D) 24.00 hours

The item above represents a form in which the number of hours to produce one unit, the number of units produced, and the learning curve percentage can all vary. Replacing these variables with different elements will produce an unlimited

number of new items. For the item above, the "skeleton" sentence would be the following:

If it takes an employee X hours to produce one unit of a specific item, what would be the average labor hours per unit after the production of the Y th unit, assuming that a $Z\%$ learning curve is in effect?

To produce new items, the variables X , Y , and Z can be replaced with a variety of combinations:

X hours	Yth unit	Z%	Key
10	80	2	8.00 hours
10	80	4	6.40 hours
20	80	8	10.24 hours
20	80	16	8.19 hours
10	85	2	8.50 hours
10	85	4	7.23 hours
20	85	8	12.28 hours
20	85	16	10.44 hours

This method can also be used to generate incorrect options that an examinee is likely to select. The item forms method can save a lot of time and result in the creation of many new items with considerably less effort than it would take to produce the same number of questions without a form.

C. Critical Incident Technique

Items of the application type present "real life" situations that a person working in a profession would be confronted with during the work day. Such items present certain scenarios that call for an application of the examinee's knowledge. In order to develop situations that can serve as items, one may wish to employ a job analysis method known as the "Critical Incident" technique. This approach works as follows:

Think of a situation or event that was critical to an incumbent's job performance, that involved either particularly effective (or successful) types of behavior, or particularly ineffective (or unsuccessful) types of behavior. In recalling this situation, consider:

- 1. What led up to the incident;**
- 2. Exactly what the job incumbent did that was so effective or ineffective;**
- 3. The perceived consequences of the behavior, and whether or not the consequences were within the job incumbent's control.**

Bear in mind that this technique can draw upon instances of poor performance as well as good performance (as the saying goes, “Bad decisions make good stories”). An example of the critical incident technique applied to the development of an item for the accounting profession is presented below:

Example:

1) What led up to the incident. XYZ, Inc. purchased a piece of manufacturing equipment for its main plant. The equipment itself costs \$555,000. The freight charges for the equipment were \$2,500, paid by XYZ. Six temporary workers were hired to install the equipment, at a cost of \$6,000. In order to install the machine, the old equipment had to be removed at a cost of \$800 (the original cost of the old machine was \$275,000, purchased 10 years before). The accountant calculated the cost basis of the new machine to be \$563,500.

2) Exactly what was done that was so ineffective. The accountant failed to consider the \$800 cost for removing the old piece of equipment.

3) The perceived consequences of the behavior, and whether or not the consequences were within control. The cost principle states that the value of an asset should include all costs necessary to put the asset into service. The \$800 to remove the old piece of equipment should therefore have been added to the cost basis of the new equipment, making the correct cost \$564,300.

Sample Test Item:

A firm purchases a piece of capital equipment for \$555,000. The machine is replacing an earlier model that the firm purchased 10 years ago for \$275,000. The firm pays shipping costs of \$2,500 for the new machine, and \$6,000 in labor costs to have it installed. It also spends \$800 to have the old machine removed. What is the cost basis for the new machine?

- (A) \$555,000
- (B) \$557,500
- (C) \$563,500
- * (D) \$564,300

Notice that this question contains data that are irrelevant to solving the problem. Such superfluous information can serve as a “red herring” and is acceptable to include within the item.

Besides Application level items, the critical incident technique is useful for the development questions that require the examinee to make professional judgments or evaluations. As stated earlier in this manual, the correct answer should be one in which experts will agree, and not the mere bias or preference of the item writer.

D. A Note about Distracters

The most difficult aspect of formulating multiple choice items is not in writing the correct answers, but in the creation of plausible *incorrect* options (i.e., distracters). Distracters cannot, of course, be correct statements, yet they must be plausible enough to examinees whose knowledge is incomplete or faulty. A useful technique for producing distracters is to try to anticipate the types of errors that examinees are likely to make if they are asked to formulate rather than select the answers, and then formulate the distracters based on those errors. Typically, an effective distracter will leave out important facts or information, or will contain information that may be related but not required to answer the question.

THIS PAGE INTENTIONALLY LEFT BLANK.



National Measurement and Testing, Inc.
P.O. Box 339
Montvale, NJ 07645
866-724-9600
www.nmetest.com