Constructing Test Questions for the ACT Certifying Exam
Certifying Examination Committee – 2013

1. Testing objectives

Candidates writing the ACT Certifying Exam are expected to do more than just know facts. Questions that test only rote memory for isolated facts without requiring their application are classified as knowledge (or recall) questions. This type of question only evaluates the lowest level of learning outcome in the cognitive domain (see Table 1) and its use should be limited. Instead, preference should be given to questions that require higher-order skills and involve application of knowledge, interpretation of information, and problem-solving, rather than just rote memory of factual information.

Questions that require higher-order skills require the candidate to reach a conclusion, make a prediction, or select a course of action for example. Typically, the candidate would need to be able to synthesize a relatively large volume of information in order to answer these questions. These types of questions also help to identify those candidates who have memorized a substantial body of factual information, but are unable to use that information effectively.

The following pair of questions illustrates the difference between a question assessing recall of an isolated fact and a question assessing application of knowledge:

**Pair 1**
- **Recall question:** What area is supplied with blood by the posterior inferior cerebellar artery?
- **Application of knowledge question:** A 62-year-old man develops left-sided limb ataxia, Horner’s syndrome, nystagmus, and loss of appreciation of facial pain and temperature sensations. What artery is most likely to be occluded?

**Pair 2**
- **Recall question:** Acute intermittent porphyria is the result of a defect in the biosynthetic pathway for:
- **Application of knowledge question:** An otherwise healthy 33-year-old man has mild weakness and occasional episodes of steady, severe abdominal pain with some cramping but no diarrhea, and mild weakness in the upper arms. Other family members have had similar episodes. These findings suggest a defect in the biosynthetic pathway for:

In addition, it is important that questions explore important concepts; testing time should not be wasted with questions assessing knowledge of trivial, “tricky,” esoteric, or simply interesting topics. The focus should be on key concepts and principles that are essential information for all candidates to understand and on specific tasks that candidates must be able to undertake as specialists (e.g., determine the most likely diagnosis, indicate what additional tests should be performed, formulate the next step in management, etc).
<table>
<thead>
<tr>
<th>Taxonomy categories</th>
<th>Abilities related to educational objectives</th>
<th>Sample verbs for stating specific learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge*</td>
<td>Candidate is able to use rote memorization and recall certain facts; test questions focus on identification and recall of information</td>
<td>Cite, define, describe, identify, label, list, match, name, select, state</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Candidate is able to extrapolate and interpret important information and put other’s ideas into their own words; test questions focus on use of facts, rules, and principles</td>
<td>Classify, convert, distinguish between, explain, give examples, illustrate, interpret, match, summarize</td>
</tr>
<tr>
<td>Application</td>
<td>Candidate is able to take new concepts and apply them to another situation; test questions focus on applying facts or principles</td>
<td>Apply, arrange, construct, demonstrate, modify, predict, prepare, produce, relate, show, solve</td>
</tr>
<tr>
<td>Analysis</td>
<td>Candidate is able to take new information and break it down into parts to differentiate between them; test questions focus on separation of a whole into component parts</td>
<td>Analyze, associate, determine, diagram, differentiate, discriminate, distinguish, estimate, infer, order, outline, point out, separate, subdivide</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Candidate is able to take various pieces of information and form a whole creating a pattern where one did not previously exist; test questions focus on combining ideas to form a new whole</td>
<td>Combine, compile, compose, construct, create, design, develop, devise, formulate, integrate, modify, organize, plan, propose, rearrange, reorganize, revise</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Candidate is able to look at someone else’s ideas or principles and see the worth of the work and the value of the conclusions; test questions focus on developing opinions, judgments or decisions</td>
<td>Appraise, assess, compare, conclude, contrast, criticize, discriminate, evaluate, judge, justify, support, weigh</td>
</tr>
</tbody>
</table>

*Lowest level of learning outcome in the cognitive domain; learning at the higher levels is dependent on having attained prerequisite knowledge and skills at lower levels.
2. Guidelines for writing multiple-choice questions

Multiple-choice questions used in the ACT Certifying Exam consist of two basic parts: a problem (stem) and a list of suggested solutions (alternatives). The list of alternatives contains one correct or best alternative (answer) and four incorrect or inferior alternatives (distractors).

In questions of the single-correct-answer variety, all but one of the alternatives are incorrect; the remaining alternative is the correct answer. The candidate is directed to identify the correct answer. In questions of the best-answer variety, the alternatives differ in their degree of correctness. Some may be completely incorrect and some correct, but one is clearly more correct than the others. This best alternative serves as the answer, while the other alternatives function as distractors. The candidate is directed to identify the best answer. Lead-in questions for these questions usually include the terms “best”, “most likely” or “most probable”.

The stem is the foundation of the question. After reading the stem, the candidate should know exactly what the problem is and what he or she is expected to do to solve it. The stem is usually much longer than the alternatives, especially when assessing higher-order skills. Clinical or laboratory vignettes provide a good basis for a question. These questions usually begin with the presenting problem of the animal, followed by other information like signalment, history, physical findings, results of diagnostic tests, initial treatment, subsequent findings, etc. The stem may be in the form of either a direct question (see Pair 1 above) or an incomplete statement (see Pair 2 above).

The purpose of the distractors is to appear as plausible solutions to the problem for those candidates who have not achieved the learning objective being measured by the question. Conversely, the distractors must appear as implausible solutions for those candidates who have achieved the objective; only the answer should appear plausible to these candidates. The “correctness” of the single answer should be unequivocally supported by the most current scientific literature, which should be listed for reference with each question. This is especially important for questions of the best-answer variety.

Good multiple-choice questions are generally more difficult and time-consuming to write than other types of questions. Coming up with plausible distractors requires a certain amount of skill. This skill, however, may be increased through study, practice, and experience.

2.1. General technical considerations

2.1.1. Base each question on a specific problem stated clearly in the stem

A common fault in multiple-choice question writing is to have a brief, meaningless stem with problem definition revealed in the options. In such cases, it can be difficult to see the intent of the question after reading the stem. If the candidate has to infer what the problem is, the question will likely measure the candidate’s ability to draw inferences from vague descriptions rather than his or her achievement of a learning objective.
The basic rule for stem-writing is that candidates should be able to understand the question without reading it several times and without having to read all the options. The central idea or problem should be included in the stem instead of in the options. The stem can be judged to be clearly presenting a problem if it forces the options to be parallel in type of content. The options should be covered to determine if the question is clear and if the candidate could produce an answer based only on the stem.

In the poor example below, the stem fails to present a definite problem and the four options appear to be a hodgepodge of ideas. Clearly, the question in the better example is more focused and the stem poses a clear, definite problem that assesses a single learning objective.

**Poor example**
Which of the following statements concerning electrochemical cells is correct?
*a. There is a spontaneous chemical reaction in each electrochemical cell.*
*b. The e.m.f. of an electrochemical cell is measured in joules.*
*c. The anode is labeled (+) while the cathode is labeled (–).*
*d. The salt bridge provides electrons to complete the circuit.*

**Better example**
What is the main function of the salt bridge in an electrochemical cell?
*a. Supply ions moving to the two half-cells.*
*b. Draw electrons from one half-cell to the other half-cell.*
*c. Keep the levels of solutions equal in the two half-cells.*
*d. Supply electrons to complete the circuit.*

2.1.2. State the question as briefly as possible, avoiding wordiness, undue complexity, and irrelevant information

In higher-order skills questions the stem will normally be longer than in lower-level questions, but should still be brief and straight to the point. Excess irrelevant information in the stem that is not essential to answering the problem increases the reading burden and adds to candidate confusion over what he or she is being asked to do.

**Poor example**
Suppose you are a mathematics professor who wants to determine whether or not your teaching of the unit on probability has had a significant effect on your students. You decide to analyze their scores from a test they took before the instruction and their scores from another exam taken after the instruction. Which of the following t-tests is appropriate to use in this situation?
*a. Dependent samples*
*b. Heterogeneous samples*
*c. Homogeneous samples*
*d. Independent samples*
Better example
When analyzing your students’ pretest and posttest scores to determine if your teaching has had a significant effect, an appropriate statistic to use is the t-test for:
*a. Dependent samples
b. Heterogeneous samples
c. Homogeneous samples
d. Independent samples

2.1.3. Include as much of the item as possible in the stem

Rather than repeating redundant words or phrases in each of the alternatives, place such material in the stem to decrease the reading burden and more clearly define the problem in the stem.

Poor example
If the pressure of a certain amount of gas is held constant, what will happen if its volume is increased?
a. The temperature of the gas will decrease.
*b. The temperature of the gas will increase.
c. The temperature of the gas will remain the same.

Better example
If you increase the volume of a certain amount of gas while holding its pressure constant, its temperature will:
a. Decrease.
*b. Increase.
c. Remain the same.

2.1.4. Avoid the use of negative questions

Negative questions direct the candidate to identify either the alternative that is an incorrect answer, or the alternative that is the worst answer. These questions are usually of the type “Each of the following is correct EXCEPT” or “Which of the following statements is NOT correct?”. These questions suffer from the same problem as true/false questions, i.e. if options cannot be rank-ordered on a single continuum, the candidate cannot determine either the “least” or the “most” correct answer. In addition, for most educational objectives, a candidate’s achievement is more effectively measured by having him or her identify a correct answer rather than an incorrect answer. Just because the candidate knows an incorrect answer does not necessarily imply that he or she knows the correct answer.

Occasionally, negative items are appropriate where knowing what not to do is important. In these situations, negative items must be carefully worded to avoid confusing the candidate. The negative word should be placed in the stem, not in the alternatives, and should be emphasized
by CAPITALS. In addition, each of the alternatives should be phrased positively to avoid forming a confusing double negative with the stem.

2.1.5. Avoid the use of completion-type questions

Completion-type questions are those which the stem is written as an incomplete statement and the candidate is required to identify the correct option to be inserted to complete the statement. This type of question should be avoided because a candidate has to retain the stem in short-term memory while completing the stem with each option. Test anxiety is even higher if the candidate is not a native English speaker. If the completion format is unavoidable, the omission should occur toward the end of the stem rather than in the middle or at the beginning of the stem.

2.1.6. Keep the alternatives homogeneous in content

If the alternatives consist of a potpourri of statements related to the stem but unrelated to each other, the candidate’s task becomes unnecessarily confusing. Alternatives that are parallel in content help the item present a clear-cut problem more capable of measuring the attainment of a specific objective.

The poor example below contains alternatives testing knowledge of state agriculture, physical features, flags, and nicknames. If the candidate misses the question, it does not reveal in which of the four areas the candidate is weak. In the better example, all of the alternatives refer to state agriculture, so if the candidate misses the item, it reveals that the candidate has a weakness in that area.

Poor example
Idaho is widely known as:
*a. the largest producer of potatoes in the United States.
b. the location of the tallest mountain in the United States.
c. the state with a beaver on its flag.
d. the “Treasure State.”

Better example
Idaho is widely known for its:
a. apples.
b. corn.
*c. potatoes.
d. wheat.

2.1.7. Word the alternatives clearly and concisely

Clear wording reduces candidate confusion, and concise wording reduces the reading burden placed on the candidate. The alternatives in the poor example below are rather wordy, and may
require more than one reading before the candidate understands them clearly. In the better example, the alternatives have been streamlined to increase clarity without losing accuracy.

**Poor example**
The term hypothesis, as used in research, as defined as:

a. A conception or proposition formed by speculation or deduction or by abstraction and generalization from facts, explaining or relating an observed set of facts, given probability by experimental evidence or by factual or conceptual analysis but not conclusively established or accepted.

b. A statement of an order or relation of phenomena that so far as is known is invariable under the given conditions, formulated on the basis of conclusive evidence or tests and universally accepted, that has been tested and proven to conform to facts.

*c. A proposition tentatively assumed in order to draw out its logical or empirical consequences and so test its accord with facts that are known or may be determined, of such a nature as to be either proved or disproved by comparison with observed facts.

**Better example**
The term hypothesis, as used in research, is defined as:

a. An assertion explaining an observed set of facts that has not been conclusively established.

b. A universally accepted assertion explaining an observed set of facts.

*c. A tentative assertion that is either proved or disproved by comparison with an observed set of facts.

2.1.8. *Keep the alternatives mutually exclusive*

Alternatives that overlap create undesirable situations. Some of the overlapping alternatives may be easily identified as distractors. On the other hand, if the overlap includes the intended answer, there may be more than one alternative that can be successfully defended as being the answer.

**Poor example**
How long does an annual plant generally live?

*a. It dies after the first year.

b. It lives for many years.

c. It lives for more than one year.

*d. It needs to be replanted each year.

**Better example**
How long does an annual plant generally live?

*a. Only one year.

b. Only two years.

c. Several years.
2.1.9. Avoid the use of imprecise terms

Although imprecise terms (e.g. “usually”, “frequently”, “rarely”) are used in our everyday speech and in our writing, these terms cause confusion when they are used in the text of examination items.

2.1.10. Avoid composing alternatives that are only microscopically distinct

Alternatives should be clearly different unless the ability to make fine distinctions is considered a significant learning objective.

2.1.11. Avoid using the options “all of the above”, “none of the above” or “both A and B”

Avoid using these options as alternatives, since these options make it possible for students to guess the correct answer with only partial knowledge.

2.2. Technical flaws related to testwiseness

Poorly-written questions often contain clues that help students who do not know the correct answer eliminate incorrect alternatives and increase their chance of guessing correctly. It is important to keep the alternatives free from clues as to which response is correct.

2.2.1. Use plausible distractors

For the candidate who does not possess the ability being measured by the question, the distractors should look as plausible as the answer. Unrealistic or humorous distractors are nonfunctional and increase the candidate’s chance of guessing the correct answer. Plausible distractors may be created by using common misconceptions and faulty reasoning or using words that “ring a bell” or that “sound official.” The distractors should be plausible enough to keep the student who has not achieved the objective from detecting them, but not so subtle that they mislead the student who has achieved the objective.

In the poor example below, the options are quite divergent and the correct answer easily identified. Someone who knows relatively little about American history could answer this correctly. The different set of alternatives in the better example makes the question more difficult since all options are plausible answers to someone who has limited knowledge.

**Poor example**
Who was the primary author of the Declaration of Independence?
- a. Abraham Lincoln
  - *b. Thomas Jefferson
- c. King George II
- d. Catherine the Great
**Better example**
Who was the primary author of the Declaration of Independence?

a. George Washington  
b. Thomas Jefferson  
c. James Madison  
d. Benjamin Franklin

2.2.2. Keep the grammar of each alternative consistent with the stem

Because a question writer tends to pay more attention to the correct answer than to the distractors, grammatical errors are more likely to occur in the distractors. Candidates often assume that inconsistent grammar is the sign of a distractor, and they are generally right. In the poor example below, the answer fits better grammatically with the stem than do the distractors. This problem has been solved in the better example by rewording the alternatives.

**Poor example**
Which of the following would do the most to promote the application of nuclear discoveries to medicine?

a. Trained radioactive therapy specialists.  
b. Developing standardized techniques for treatment of patients.  
c. Do not place restrictions on the use of radioactive substances.  
d. If the average doctor is trained to apply radioactive treatments.

**Better example**
Which of the following would do the most to promote the application of nuclear discoveries to medicine?

a. Adding trained radioactive therapy specialists to hospital staffs.  
b. Developing standardized techniques for treatment patients.  
c. Removing restrictions on the use of radioactive substances.  
d. Training the average doctor to apply radioactive treatments.

In the poor example below, testwise candidates would eliminate “a” and “c” as options because they do not follow grammatically or logically from the stem. Testwise candidates then have to choose only between “b“ and “d”.

**Poor example**
A man that is found lying unconscious on the sidewalk is brought to the emergency. After ascertaining that the airway is open, the first step in management should be intravenous administration of:

a. examination of cerebrospinal fluid  
b. glucose with vitamin B1 (thiamine)  
c. CT scan of the head  
d. phenytoin

**Better example**
A man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. After ascertaining that the airway is open, the first step in management should be intravenous administration of:

a. penicillin  
b. glucose with vitamin B1 (thiamine)  
c. diazepam  
d. phenytoin

2.2.3. *Keep the alternatives parallel in form*

If the answer is worded in a certain way and the distractors are worded differently, the candidate may take notice and respond accordingly. The answer in the poor example below stands out because it does not include the identical wording underlined in each of the distractors. The answer is less obvious in the better example because the distractors have been reworded to be more parallel with the answer.

**Poor example**
You have just spent ten minutes trying to teach one of your new employees how to change a typewriter ribbon. The employee is still having a great deal of difficulty changing the ribbon, even though you have always found it simple to do. At this point, you should:

a. **tell the employee** to ask an experienced employee working nearby to change the ribbon in the future.

b. **tell the employee** that you never found this difficult, and ask what he or she finds difficult about it.

*c. review each of the steps you have already explained, and determine whether the employee understands them.

d. **tell the employee** that you will continue teaching him or her later, because you are becoming irritable.

**Better example**
You have just spent ten minutes trying to teach one of your new employees how to change a typewriter ribbon. The employee is still having a great deal of difficulty changing the ribbon, even though you have always found it simple to do. At this point, you should:

a. ask an experienced employee working nearby to change the ribbon in the future.

b. mention that you never found this difficult, and ask what he or she finds difficult about it.

*c. review each of the steps you have already explained, and determine whether the employee understands them.

d. tell the employee that you will continue teaching him or her later because you are becoming irritable.

2.2.4. *Keep the alternatives similar in length*

An alternative noticeably longer or shorter than the other is frequently assumed to be the answer, and not without good reason. A correct answer that is longer, more specific, or more
complete than other alternatives is a common flaw. Notice how the answer stands out in the poor example below. Both the answer and one of the distractors have been reworded in the better example to make the alternative lengths more uniform.

**Poor example**
Which of the following is the best indication of high morale in a supervisor’s unit?
a. The employees are rarely required to work overtime.
*b. The employees are willing to give first priority to attaining group objectives, subordinating any personal desires they may have.
c. The supervisor enjoys staying late to plan the next day.
d. The unit gives expensive birthday presents to each other.

**Better example**
Which of the following is the best indication of high morale in a supervisor’s unit?
a. The employees are rarely required to work overtime.
*b. The employees willingly give first priority to attaining group objectives.
c. The supervisor enjoys staying late to plan for the next day.
d. The unit members give expensive birthday presents to each other.

2.2.5 Avoid the use of specific determiners

When words such as “never”, “always”, and “only” are included in distractors in order to make them false, they serve as flags to the alert candidate. In the poor example below, the underlined word in each of the distractors is a specific determiner. These words have been removed from the better example by rewording both the stem and the distractors.

**Poor example**
To avoid infection after receiving a puncture wound to the hand, you should:
a. always go to the immunization center to receive a tetanus shot.
b. be treated with an antibiotic only if the wound is painful.
*c. ensure that no foreign object has been left in the wound.
d. never wipe the wound with alcohol unless it is still bleeding.

**Better example**
To avoid infection after receiving a puncture wound to the hand, you should always:
a. go to the immunization center to receive a tetanus shot.
b. be treated with an antibiotic if the wound is painful.
*c. ensure that no foreign object has been left in the wound.
d. wipe the wound with alcohol unless it is still bleeding.

2.2.6. Avoid including keywords in the alternatives

When a word or phrase in the stem is also found in one of the alternatives, it tips the candidate off that the alternative is probably the answer. In the poor example below, the underlined word
“journal” appears in both the stem and the answer. This clue has been removed from the better example by replacing the answer with another valid answer that does not include the keyword.

**Poor example**
When conducting library research in education, which of the following is the best source to use for identifying pertinent journal articles?
- a. A Guide to Sources of Educational Information
- *b. Current Index to Journals in Education*
- c. Resources in Education
- d. The International Encyclopedia of Education

When conducting library research in education, which of the following is the best source to use for identifying pertinent journal articles?
- a. A Guide to Sources of Educational Information
- *b. Education Index*
- c. Resources in Education
- d. The International Encyclopedia of Education

2.2.7. **Avoid textbook, verbatim phrasing**

If the answer has been lifted word-for-word from the pages of the textbook, the candidate may recognize the phrasing and choose correctly out of familiarity rather than achievement. Another common flaw is stating correct options in textbook language and distractors in everyday language. Note how the correct answer in the poor example below stands out as it is a familiar definition straight out of the textbook, and the distractors are in the teacher’s own words.

**Poor example**
The term operant conditioning refers to the learning situation in which:
- a. a familiar response is associated with a new stimulus.
- b. individual associations are linked together in sequence.
- *c. a response of the learner is instrumental in leading to a subsequent reinforcing event.*
- d. verbal responses are made to verbal stimuli.

**Better example**
The term operant conditioning refers to the learning situations in which:
- a. a familiar response is associated with a new stimulus.
- b. individual associations are linked together in sequence.
- *c. the learner’s response leads to reinforcement.*
- d. verbal responses are made to verbal stimuli.

2.2.8. **Beware of convergence strategy**

This strategy can be used by the testwise candidate when the correct answer includes the most elements in common with the other alternatives. This flaw is less obvious than the others, but it
occurs frequently and is worth noting. The flaw is seen in several forms. The underlying premise is that the correct answer is the option that has the most in common with the other options; it is not likely to be an outlier. For example, in numeric options, the correct answer is more often the middle number than an extreme value. In double options, the correct answer is more likely to be the option that has the most elements in common with the other distractors.

For example, if the alternatives are “Pencil and pen”, “Pencil and highlighter”, “Pencil and crayon”, “Pen and marker,” the correct answer is likely to be “Pencil and pen” (i.e., by simple count, “Pencil” appeared 3 times in the options, “Pen” appeared twice and other elements each appeared only once). While this might seem ridiculous, this flaw occurs because question writers start with the correct answer and write permutations of the correct answer as the distractors. The correct answer is, therefore, more likely to have elements in common with the rest of the options; the incorrect answers are more likely to be outliers as the question writer has difficulty generating viable distractors.

In the example below, the testwise candidate would eliminate “anionic form” as unlikely because “anionic form” appears only once; that candidate would also exclude “outside the nerve membrane” because “outside” appears less frequently than “inside”. The candidate would then have to decide between alternatives “b” and “d”. Since three of the five alternatives involve a charge, the testwise candidate would then pick option “b”.

Local anesthetics are most effective in the:

a. anionic form, acting from inside the nerve membrane
*b. cationic form, acting from inside the nerve membrane
c. cationic form, acting from outside the nerve membrane
d. uncharged form, acting from inside the nerve membrane
e. uncharged form, acting from outside the nerve membrane

2.2.9. Present the answer in each of the alternative positions in a random order

Many testers have a tendency to avoid placing the answer in the first or last alternative positions, preferring instead to “bury the answer in the middle.” This tendency, however, is not unknown to certain candidates, who generally select one of the alternatives in the middle if they are unsure of the answer. In this case, the unprepared but clever candidate increases his or her chance of obtaining a higher score.

The easiest method of randomizing the answer position is to arrange the alternatives in some logical order. The following table gives examples of three logical orders. The best order to use for a particular item depends on the nature of the item’s alternatives.
### Logical order

<table>
<thead>
<tr>
<th>Logical order</th>
<th>Example</th>
</tr>
</thead>
</table>
| Numerical     | a. 1939  
|               | b. 1940  
|               | c. 1941  |
| Alphabetical  | a. Changing a from 0.01 to 0.05  
|               | b. Increasing the spread of the exam scores.  
|               | c. Reducing the size of the treatment effect.  |
| Sequential    | a. Heating ice from -100°C to 0°C.  
|               | b. Melting ice at 0°C  
|               | c. Heating water from 0°C to 100°C  |

### 2.3. Multiple-choice item templates *(From the National Board of Medical Examiners)*

The overall structure of a question can be depicted by a question template. You can typically generate many items using the same template.

- A (patient description) is unable to (functional disability). Which of the following is most likely to have been injured?
- A (patient description) has a (type of injury and location). Which of the following structures is most likely to be affected?
- A (patient description) has (history findings) and is taking (medications). Which of the following medications is the most likely cause of his (one history, PE or lab finding)?
- A (patient description) has (abnormal findings). Which [additional] finding would suggest/suggests a diagnosis of (disease 1) rather than (disease 2)?
- A (patient description) has (symptoms and signs). These observations suggest that the disease is a result of the (absence or presence) of which of the following (enzymes, mechanisms)?
- A (patient description) has (symptoms, signs, or specific disease) and is being treated with (drug or drug class). The drug acts by inhibiting which of the following (functions, processes)?
- A (patient description) has (abnormal findings). Which of the following (positive laboratory results) would be expected?
- Following (procedure), a (patient description) develops (symptoms and signs). Laboratory findings show (findings). Which of the following is the most likely cause?
- A (patient description) dies of (disease). Which of the following is the most likely finding on autopsy?
- A patient has (symptoms and signs). Which of the following is the most likely explanation for the (findings)?
- A (patient description) has (symptoms and signs). Exposure to which of the (toxic agents) is the most likely cause?
- Which of the following is the most likely mechanism of the therapeutic effect of this (drug class) in patients with (disease)?
A patient has (abnormal findings), but (normal findings). Which of the following is the most likely diagnosis?

Sample lead-ins and option lists

Which of the following is (abnormal)?
Options sets could include sites of lesions; list of nerves; list of muscles; list of enzymes; list of hormones; types of cells; list of neurotransmitters; list of toxins, molecules, vessels, spinal segments.

Which of the following findings is most likely?
Options sets could include list of laboratory results; list of additional physical signs; autopsy results; results of microscopic examination of fluids, muscle or joint tissue; DNA analysis results; serum levels.

Which of the following is the most likely cause?
Options sets could include list of underlying mechanisms of the disease; medications that might cause side effects; drugs or drug classes; toxic agents; hemodynamic mechanisms, viruses, metabolic defects.

Which of the following should be administered?
Options sets could include drugs, vitamins, amino acids, enzymes, hormones.

Which of the following is defective/deficient/nonfunctioning?
Options sets could include list of enzymes, feedback mechanisms, endocrine structures, dietary elements, vitamins.

3. Guidelines for writing essay questions

Essay questions are intended to generate evidence of higher-order thinking processes and profound understanding of valuable content through well-developed written reasoning. Essay questions are generally better than multiple choice questions to assess the complexity of candidates’ thought processes and ability to think critically or to demonstrate their understanding by taking factual information and applying it to solve problems.

The first step in formulating essay questions is determine whether an essay question is the most appropriate format for the type of learning outcome to be assessed. Essay questions should be used when the learning outcomes involve abilities like: (1) evaluate information, (2) analyze evidence, (3) articulate explanations, (4) display thought processes, (5) furnish information, (6) perform a specific task, and (7) provide examples.

An essay question does not automatically assess higher-order thinking skills. Essay questions often simply assess recall. Also, if an essay question is meant to assess higher-order thinking but then responses are scored in a way that only rewards recall ability, that essay is not assessing higher-order thinking. Compare the following two examples:
Example 1
What are the major advantages and limitations of essay questions?

Example 2
Given their advantages and limitations, should an essay question be used to assess candidates’ ability to create a solution to a problem? In answering this question provide brief explanations of the major advantages and limitations of essay questions. Clearly state whether or not you think an essay question should be used and explain the reasoning for your judgment.

Example 1 assesses recall of factual knowledge, whereas Example 2 requires more of the candidates. It not only requires candidates to recall facts, but also to make an evaluative judgment, and to explain the reasoning for the judgment. Example 2 requires more complicated thinking than Example 1.

3.1. General technical considerations

3.1.1. Clearly define the intended learning outcome to be assessed by the question

Specific intended learning outcomes are crucial to designing effective essay questions. If the expected outcome to be assessed lacks clarity and specificity, the essay question meant to assess candidates’ achievement of the outcome will likely assess something other than what is intended. It is important to thoughtfully and carefully select the most appropriate “directive verb” to indicate what thought processes and actions students must exhibit to provide evidence that learning has occurred. Compare the following two examples:

Example 1
Students will appreciate the process of cell division.

Example 2
Given a chart illustrating the process of cell division, students will compare and contrast each major step in the process.

Example 1 is too general to provide clear guidance in writing an essay question. The directive verb, “appreciate” is vague and difficult to assess. It is especially difficult to judge or assign a grade to candidates’ “appreciation”. Example 2 is more useful for guiding the development of an essay question.

3.1.2. Clearly define the task of the question aligning it to the intended learning outcome

With some essay questions, candidates can feel like they have an infinite supply of lead to write a response on an indefinite number of pages about whatever they feel happy to write about. This can happen when the essay question is vague or open to numerous interpretations. Effective essay questions provide candidates with an indication of the types of thinking and
content to use in responding to the question. Thus, effective essay questions provide a well-defined task for students.

The task (or prompt) is the core of the essay question, akin to the stem of a multiple-choice question. The task should provide candidates with an indication of the types of thinking they should use, the content they should reason with, and the performance they should exhibit when responding to the question. In essence, an effective essay question produces valid evidence of the degree to which candidates have achieved the intended learning outcome.

A clearly defined task is composed of a directive verb and the object of that verb, which should elicits the desired thought processes in the minds of the candidates. The following example demonstrates the importance of carefully choosing directive verbs to align the essay question with the intended learning outcome.

**Example**

*Intended learning outcome:*
Analyze the impact of America at war on the American economy.

*Less effective question:*
Describe the impact of America at war on the American economy.

*More effective question:*
Analyze the impact of America at war on the American economy by describing how different effects of the war work together to influence the economy.

According to definition, “analyze” means to break material into its constituent parts and to determine how the parts relate to one another and to an overall structure or purpose. In the less effective example, candidates are asked to describe the impact. To describe requires students to give an account of the impact of America at war on the American economy, but it does not require students to make an analysis based on the different effects of the war and how they work together to affect the economy. The more effective question does make this distinction for candidates, thereby providing guidance concerning the task of analyzing. Thus, directive verbs must be carefully selected to clearly reflect the task required of students and to be aligned to intended outcomes.

Similarly, the object of the directive verb must be carefully written. Just as it is important to select the right verb, it is important to delimit the scope of the object of that verb. Delimiting the scope of the task helps to avoid responses containing ideas unrelated to the essay question and extreme subjectivity when scoring responses. The following example illustrates the process of limiting the scope of the task for a given essay question.
Example of an evolving essay question that becomes more focused

Less focused essay question:
Evaluate the impact of the Industrial Revolution on England.

More focused questions:
Evaluate the impact of the Industrial Revolution on the family in England.
Evaluate the impact of the Industrial Revolution on the role of fathers in poor communities of England based on whether or not the Industrial Revolution improved fathers’ abilities to provide the material necessities of life and education and training for their children.

In these examples, five essay questions are provided. Example 1 has little focus. The directive verb is “evaluate” and the object of the verb is “the impact of the Industrial Revolution on England.” Very little guidance is given to candidates about the task of evaluating and the scope of the task. A student reading Example 1 may ask the following: the impact on what in England? The economy? Foreign trade? A particular group of people? Evaluate based on what criteria? The significance of the Revolution? The quality of life in England? Progress in technological advancements? After reading the question, the candidate would not have a clear idea of what exactly he/she would be expected to do in his/her evaluation.

Example 2 delimits the task for candidates by specifying a particular unit of society in England affected by the Industrial Revolution (family). Example 3 is even more focused than Example 2 because candidates are asked to focus on a subunit of the family (fathers) and a specific community of families in England (poor communities). With Example 4, the task is further delimited by giving candidates a criterion for evaluating the impact of the Industrial Revolution (whether or not fathers’ abilities to provide for their children in two different ways improved because of the Industrial Revolution); in this example the task is focused and delimited by clarifying for candidates what must be done to “evaluate.”

Failure to establish adequate and effective limits for the candidate response to the essay question allows candidates to set their own boundaries for their response, meaning that candidates might provide responses that are outside of the intended task, that are too long, or that only address a part of the intended task. If the candidate’s failure to answer within the intended limits of the essay question can be ascribed to poor or ineffective wording of the task, the tester is left with unreliable and invalid information about the candidate’s achievement of the intended learning outcome and has little or no basis for grading the responses. Therefore, teachers are responsible for writing essay questions in such a way that they provide students with clear boundaries for responses. In short, educators should avoid indeterminate questions.

A question is indeterminate if it is so unstructured that candidates can redefine the problem and focus on some aspect of it with which they are thoroughly familiar, or if experts in the subject matter cannot agree that one answer is better than another. One way to avoid
indeterminate questions is to stay away from vocabulary that is ambiguous. For example, teachers should avoid using the verb “discuss” in an essay question. This verb is simply too broad and vague and therefore fails to provide adequate guidance for students as to how to respond to the essay question.

3.1.3 Presenting the task

In essay questions, the task can be presented either in the form of a direct question or an imperative statement. If written as a question, then it must be readily translatable into the form of an imperative statement. The example bellow illustrates the same essay question twice, once as a question and once as an imperative statement. Both essay questions elicit the same performance from the student (compare and contrast processes based on cost). Whether essay questions are written as imperative statements or questions, they should be written to align with the intended outcome and in such a way that the task is clear to the students.

Examples

Question:
How are the processes of increasing production and improving quality in a manufacturing plant similar or different based on cost?

Imperative statement:
Compare and contrast the processes of increasing production and improving quality in a manufacturing plant based on cost.

3.1.4 Situate the task within a problem situation

Once the task of the essay question is clearly defined and the content that candidates are to use in accomplishing the task has been delimited, the essay question is ready to be situated in a problem. The ability to situate well-written tasks within problems is what makes testers effective at writing essay questions. Both the tasks and the problems are key elements of essay questions.

The problem in essay questions includes the unsettled matter or undesirable state of affairs that needs to be resolved. The purpose of the problem is to provide the students with a context within which they can demonstrate the performance to be assessed. In cases where the intended outcome to be assessed requires complex or critical thinking, often a unique or novel problem situation is developed.

The use of clinical or laboratory vignettes, herd health data, reproductive performance information, and graphs are excellent options for presenting problem situations for essay questions. Although candidates can be asked to fulfill tables in essay questions, this kind of task usually only assesses the ability to recall information, which is the not what essays are generally
intended for. Therefore, tables should preferably be only part of an essay question and their used should be limited in the exam.

3.1.5. Use structured essay prompts

In addition, it is also usually more appropriate to present the problem and then divide the question into separate components by using a structure assay prompt. Dividing the question allows assessment of various learning outcomes in a specific area and helps the candidates to be more organized, hence making the process more efficient. If multiple questions are not identified, some candidates may inadvertently omit some parts, especially when time constraints are great. It also makes the grading process easier because it encourages organization in the responses.

Example of structured essay prompts

You are presented with three essay questions to review for the ACT Certifying Exam.
Essay question 1....
Essay question 2...
Essay question 3...
What are the intelligence factors assessed by these items?
What is the mental ability that each item type is testing?
What labels from Bloom’s taxonomy best describe your factors?
Why are these factors important measures of intelligence?

3.2. Grading essay questions

Reliability is the characteristic of a set of measurements that relates to the accuracy of the scores. Scorer reliability estimates indicate the extent to which the ratings of one grader would be replicated by another. When scorers are in agreement about the numbers to be assigned to each of several essay responses, the scorer reliability is high. Scores of the same response might disagree about the score to be assigned due to different interpretation of the essay question, different perception of an "ideal response" or differences in how much weight is put on factors such as grammar, spelling, and organization.

Scorer reliability increases when the analytical scoring method is used. In analytical scoring, essential parts of an ideal response are identified and the assigned score is based upon the number of parts included in the response. Each part is evaluated individually. This scoring method requires that the tester develop an ideal response and create a scoring key or guide. The scoring key provides an absolute standard for determining the total points awarded for a response. Candidate responses are compared to the scoring standard and not to the responses of other candidates. Identifying the essential parts of the ideal response and assigning them scores greatly facilitates the preparation of the scoring guide that is used at the day of the examination. Using a bullet point format to write (review) the ideal response greatly facilitates the process of identifying the essential parts of the response and creating score guides.
Consider the following example:

*Identifying essential parts of the ideal response and assigning scores*

Compare the evaluation of libido and copulatory ability during breeding soundness examination in bulls and stallions. Assume that semen collection in bulls is performed by electroejaculation and semen collection in stallions is performed with an artificial vagina. (2 pt total)

- semen collection by electroejaculation in bulls does not allow the evaluation of libido and copulatory ability. (0.5 pt)
- Copulatory ability needs to be assessed with a teaser cow/heifer if there is a concern about the ability of the bull to breed normally. (0.5 pt)
- In contrast, libido and copulatory ability are readily evaluated during semen collection with artificial vagina in stallions. (0.5 pt)
- Stallion age, experience, familiarity with the environment, and time of the year are some factors that need to be considered when making judgments about libido and copulatory behavior. (0.5 pt)

Prior to grading, scorers must also agree on the score that would be assigned to answers that are not included in the scoring guide but are considered relevant and whether or not points will be deducted from incorrect statements. Assigning scores to “overall quality” of the answer should be avoided as this is not part of the analytical method of grading and tend to decrease score reliability. In addition, since handwriting quality, spelling ability, and proper use of grammar and punctuation are not the focus of instruction, they should not be a focus of grading.

3.3. Checklist for writing and reviewing essay questions

- Could the item be better assessed with a different kind of assessment?
- Is the intended learning outcome clearly defined?
- Is the essay question aligned with the intended learning outcome?
- Does the essay question contain a clear and delimited task and a specific problem situation?
- Is the task presented to candidates reasonable?
- Is the question worded and structured in such a way that it will be clear to the candidates what they are expected to do?
- Is the essay question adequately structured with clear prompts?
- Do the candidates know how many points each essay prompt is worth?
- Have you written a model answer or an outline of major points that should be included in the answer? Is the model answer aligned with the intended learning outcome and the essay question?
3.4. Direct verbs and associated mental tasks

**Analyze:** break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose.

**Apply:** decide which abstractions (concepts, principles, rules, laws, theories, generalizations) are relevant in a problem situation. Use the selected abstraction to solve the problem.

**Attribute:** determine a point of view, bias, value, or intent underlying presented material.

**Classify:** determine which category belongs to something.

**Compare:** identify and describe points of similarity.

**Contrast:** bring out the points of difference.

**Create:** put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure.

**Criticize:** make judgments as to the correctness, faults, or merits of an item or issue; criticism may approve or disapprove.

**Critique:** detect consistencies and inconsistencies between a product and relevant external criteria; detect the appropriateness of a procedure for a given problem.

**Define:** give the meaning of a word or concept; place it in the class to which it belongs and distinguish it from other items in the same class.

**Describe:** give an account of; tell or depict in words; represent or delineate by a word picture.

**Design:** devise a procedure for accomplishing some task.

**Develop:** bring to a more advanced, effective, or usable state; produce.

**Differentiate:** distinguish relevant from irrelevant parts or important from unimportant parts of presented material.

**Explain:** make clear the cause or reason of something; construct a cause-and-effect model of a system; tell "how" to do; tell the meaning of.

**Evaluate:** make judgments based on criteria and standards; determine the significance, value, quality, or relevance of; give the good points and the bad ones; identify and describe advantages and limitations.

**Generate:** come up with alternative hypotheses, examples, solutions, proposals, etc. based on criteria.

**Identify:** recognize as being a particular person or thing.

**Illustrate:** use a word picture, a diagram, a chart, or a concrete example to clarify a point.

**Infer:** draw a logical conclusion from presented information.

**Interpret:** give the meaning of; change from one form of representation (e.g. numerical) to another (e.g. verbal).

**Justify:** show good reasons for; give your evidence; present facts to support your position.

**List:** create a series of names or other items.

**Predict:** know or tell beforehand with precision of calculation, knowledge, or shrewd inference from facts or experience what will happen.

**Propose:** offer for consideration, acceptance, or action; suggest.

**Recognize:** locate knowledge in long term memory that is consistent with presented material.

**Recall:** retrieve relevant knowledge from long-term memory.

**Summarize:** sum up; give the main points briefly.
4. References


