

## Section IV: Supply Items

### A. Essential Terminology:

1. **Short Answer Items:** Short answer (or completion) items require students to produce answers in the form of words, phrases, or numbers. Short answer items generally assess knowledge and comprehension, but may be used to measure more complex thinking when responding requires interpretation, solving numerical problems, or manipulating symbols (e.g., chemical equations).
2. **Restricted-Response Essay Items:** The restricted response essay item places strict limits and boundaries on the type of answer required. The specific answer form is generally identified in the question by such words as "list", "define", "describe", or "state reasons." Restricted response essay items are easier to prepare, relate more directly to a specific instructional objective, and are easier to score than extended response essay items. However, restricted response essay items restrain the student's opportunity to demonstrate organization, integration, and creativity skills. Restricted response essay items generally assess the lower complexity outcomes (e.g., knowledge, comprehension, application).
3. **Extended-Response Essay Items:** The extended response essay item provides students with almost unlimited freedom in determining the organization, scope, and form of their answer. The extended response essay item allows for the creative integration of ideas, the evaluation of materials and procedures, and a broad approach to problem solving. However, the extended response essay item is difficult to score reliably. Extended response essay items generally assess complex outcomes (e.g., analysis, synthesis, evaluation).
  1. **Scoring Rubric:** A set of scoring guidelines that describe the characteristics of the different levels of performance used in scoring or judging a performance.
  2. **Holistic Scoring:** The assignment of a score based on an overall impression of a performance or product rather than a consideration of the individual components. The overall judgement is typically guided by descriptions of the various levels of performance or scoring rubrics.
  3. **Analytic Scoring:** The assignment of a score based on sub-scores of individual components of a performance or product. The analytic judgement is typically guided by checklists and rating scales.

## Short Answer Items

### B. Strengths and Weaknesses of Short Answer Items (Gronlund, 1998):

#### Strengths:

1. It is easy to write test items.
2. Guessing is less likely than in selection-type items.
3. This item type is well suited to computational problems and other learning outcomes where supplying the answer is important.
4. A broad range of knowledge outcomes can be measured.

#### Limitations:

- It is difficult to phrase statements so that only one answer is correct.
- Scoring is contaminated by spelling ability when responses are verbal.
- Scoring is tedious and time consuming.
- This item type is not very adaptable to measuring complex learning outcomes.

### C. Guidelines for Constructing Short Answer Items (Hopkins, 1998; Oosterhof, 1999):

**1. The item should measure a meaningful and specific learning outcome.** The content of all questions must be important enough to warrant a question. "Importance" is measured in relation to the specific learning outcomes of the lesson. In addition, the important content must be the focus of the question; thus, the question should not contain irrelevant or meaningless material.

#### Poor Example:

*Specific Learning Outcome:* Identifies and defines relevant terminology.

Pythagoras created a theorem relating the sides of a right triangle called?

#### Better Example:

*Specific Learning Outcome:* Identifies and defines relevant terminology.

What is the name of the theorem that states for a right triangle  $a^2 + b^2 = c^2$  ?

[Pythagorean]

**2. The reading level of the item should be commensurate with the student's reading level.**

The wording and word choice of a question should not be a hindrance to a student's ability to correctly answer the question, given that the student knows the answer.

**Poor Example:**

Ventriloquation is subsumed under the appellation of what Vygotskian process?

**Better Example:**

Ventriloquation is comprised of what Vygotskian process?

[internalization]

**3. The item should require a single, brief response.** The correct response should be a single word or phrase. In addition, the question should be phrased so that there is a single correct response. This can be facilitated by determining the answer first and then constructing the item around the answer.

**Poor Example:**

What is an example of a selection response item type?

**Better Example:**

What is the name of the selection response item type that contains premises and responses?

[MC]

**4. Avoid using textbook language, or sentences directly from a text.** Using a sentence from a text and eliminating a key word leaves the possibility that the student may answer the question correctly based simply on association. Paraphrase statements from texts.

**Poor Example:**

Place the blanks at the \_\_\_\_\_ of the statement.

**Better Example:**

Where should the blanks within a statement be located?

[at or near the end]

**5. The direct-question format is preferable to the incomplete sentence format.** Direct questions are generally more clear than incomplete sentences, and likewise tend to call for a single correct answer. One should switch to an incomplete sentence only when do so increases the clarity of the item.

**Poor Example:**

The concept that knowledge is the result of active cognizing by the individual, and is not the result of passive acquisition is best associated with the \_\_\_\_\_ theory.

**Better Example:**

What theory states that knowledge is the result of active cognizing by the individual, and is not the result of passive acquisition?

[constructivism]

**6. The blank should represent a key word/phrase/concept, not trivia details.** Each item should measure a meaningful and specific learning outcome, triviality should not be a specific learning outcome. In addition, the key word that is left blank should relate directly to the main idea of the question.

**Poor Example:**

Given an observed score of 87% and SEM of 2%, it is reasonable to conclude that the true \_\_\_\_\_ is between 85% and 89%.

**Better Example:**

The true score for an individual, given an observed score of 87% and SEM of 2%, is reasonably between what two bounds?

[85% and 89%]

**7. Use only a single, or at the most two, blanks per item, and locate the blank(s) near the end of the item.** Including more than one blank makes the statement ambiguous and may not provide enough information for the knowledgeable student to answer correctly. In addition, a statement with multiple blanks increases the possibility that unintended answers will effectively answer the question. To increase clarity, when using the incomplete sentence format, place blanks at or near the end of the statement.

**Poor Example:**

The three types of \_\_\_\_\_ items are \_\_\_\_\_, \_\_\_\_\_, and \_\_\_\_\_.

**Better Example:**

Multiple choice, true-false, and matching items are all examples of what item type?  
[selection]

**8. Blanks should be of the same length.** The length of the blanks should not provide clues as to the intended answers.

**Poor Example:**

Each test item should be clearly focused on a \_\_\_\_\_.

**Better Example:**

Each test item should be clearly focused on a \_\_\_\_\_. (all blanks are 10 spaces)  
[specific learning objective]

**9. Item phraseology should not provide clues as to the item's answer.** Items should be phrased such that there are no clues within the item that would lead the uninformed student to the correct answer.

**Poor Example:**

Item types where the student must select the correct answer are called?

**Better Example:**

Item types where the student must choose the correct answer are called?  
[selection]

**10. When the item requires a computational response, the unit of measure and precision of the answer should be specified.** Indicating the desired unit of measure will increase the efficiency of scoring (the scorer will not have to convert between units of measurement), and identifying the precision of the answer will increase the clarity of the question for the student.

**Poor Example:**

How far has Margaret advanced if she walks forward 10 feet, then backward 2 feet, then forward again 4 feet?

[Possible correct answers include 144 inches, 12 feet, or 4 yards.]

**Better Example:**

How many feet has Margaret advanced if she walks forward 10 feet, then backward 2 feet, then forward again 4 feet?

[12 feet]

**D. Checklist for Evaluating Short Answer Items (Nitko, 1996):**

- Does the item assess an important specific learning outcome?
- If possible, is the item written in question form?
- Is the item worded clearly so that answer is a brief phrase, single word, or single number?
- Is the blank or answer space toward the end of the sentence?
- Is the statement paraphrased rather than copied verbatim from learning materials?
- Is the word omitted in a completion item an important word, rather than a trivial word?
- Are there only one or two blanks?
- Are the blanks the same length?
- If appropriate, does the item state the degree of detail, specificity, precision, or units you want the answer to have?
- Does the item avoid grammatical and other irrelevant clues to the correct answer?

## Essay Items

### E. Strengths and Weaknesses of Essay Items (Gronlund, 1998):

#### Strengths:

- The highest level learning outcomes (analysis, synthesis, evaluation) can be measured.
- Preparation time is less than that for selection-type items.
- The integration and application of ideas is emphasized.

#### Limitations:

- There is an inadequate sampling of achievement due to time needed for answering each question.
- It is difficult to relate to intended learning outcomes because of freedom to select, organize, and express ideas.
- Scores are raised by writing skill and bluffing and lowered by poor handwriting, misspelling, and grammatical errors.
- Scoring is time consuming, subjective, and tends to be unreliable.

### F. Additional Limitations (Hopkins, 1998):

- **Reader/Scorer Unreliability:** There is great variability within a reader's score. This variability is both between different readers, and between the same reader evaluating the same essay at different times. For example, the Educational Testing Service (1961) has 53 scorers evaluate 300 essays written by college students on a 9 point scale. 34% of the essays received at least one score at each of the 9 ratings, another 37% received at least one score at 8 of the 9 ratings, and another 23% received at least one score at 7 of the 9 ratings.
- **Halo Effect:** There is a tendency for one's view of the person to influence the scoring of that person's essay. If a scorer has positive feelings regarding the individual, that individual's score will tend to be elevated. Essays should be read as anonymously as possible to avoid the halo effect.
- **Item-to-Item Carryover:** A scorer's opinion/rating/score of an individual's previous essay question influences the scoring of subsequent essay questions. Scorers should rate a particular question for all individual's before moving on to the next question. This will help to alleviate, though not eliminate the carryover effect.
- **Test-to-Test Carryover:** A rating of an essay is influenced by the rating of the essay immediately preceding it. If the previous essay is scored less than favorably, then the subsequent essay will tend to be scored more favorably. If the previous essay is scored more favorably, then the subsequent essay will tend to be scored less favorably.

- **Order Effects:** Essays read earlier in a scoring session tend to be scored more favorably, while essays read later tend to be scored less favorably.
- **Mechanics Effects:** Scorers ratings are influenced not only by content, but also by spelling, grammar, punctuation, length of answer, and handwriting.

### G. Guidelines for Constructing Essay Items (Hopkins, 1998):

**1. The item should measure a meaningful and complex learning outcome.** The content of all questions must be important enough to warrant a question. "Importance" is measured in relation to the specific learning outcomes of the lesson. In addition, an essay item should be used to assess complex outcomes, not simple outcomes that could be better assessed through objective items.

**Poor Example:**

List the four causes of World War II.

**Better Example:**

Analyze the United States' and Britain's motivations for entering World War II, and in two pages describe the similarities and differences between those motivations.

**2. Make sure the essay question is carefully focused on the complex learning outcome.** The essay question must be structured to direct the student toward the performance of the desired complex learning outcome. However, an essay question that is too highly structured will not allow the student the freedom to create, organize and present - eliminating the potential strength of the item.

**Poor Example:** (too focused)

Identify the three primary leaders of the civil rights movement in the 1960's, and one contribution made by each.

**Better Example:**

Briefly describe the main leaders of the civil rights movement in the 1960's and elaborate upon their contributions to the movement (3 pages/20 minutes).

**3. The use of optional questions ("Answer 5 of the 7 essay questions.") should be avoided.**

When optional questions are used, students do not all take the same test. Students will tend to select those questions for which they are the most prepared. In addition, students may not be assessed on a skill/knowledge component that is essential if they are able to avoid particular questions.

**4. The content and length of essay question responses should be conveyed to the student.**

Providing some structure as to the intent of the essay question will assist students in constructing a meaningful answer. In addition, providing some indication as to how much time should be spent on various essay questions will allow students to better structure their time. In general, since essay questions are designed to assess *thinking*, these time limits should be generous.

**Poor Example:**

Synthesize the various types of constructivism and describe a unified set of pedagogical principles, providing a concrete exemplary application of each.

**Better Example:**

Synthesize the various types of constructivism and describe a unified set of pedagogical principles, providing a concrete exemplary application for each. (5 pages/ 40 minutes)

**5. For general assessment of achievement, several shorter questions are usually preferable to fewer longer questions; while for complex thinking, a few longer questions is preferable to several shorter questions.** One of the main limitations of essay questions is its inability to provide extensive sampling of the achievement domain. Thus, when assessing the achievement use a greater number of shorter items to increase the sampling. However, if one is interested mainly in thought construction and presentation, then fewer questions with more time to think and construct would be appropriate.

**H. Checklist for Evaluating Essay Items (Gronlund, 1998):**

1. Is this type of item appropriate for measuring the intended learning outcome?
2. Does the item task match the learning task to be measured?
3. Is the question designed to measure complex learning outcomes?
4. Does the question make clear what is being measured and how the answer will be evaluated?
5. Has terminology been used that clarifies and limits the task (e.g., "describe," not "discuss")?
6. Are all students required to answer the same questions?
7. Has an ample time limit been indicated for each question?
8. Have adequate provisions been made for scoring answers?

## I. Guidelines for Grading Essay Items (Gallagher, 1993; Stiggins, 1997):

- **Score each response anonymously.** In order to reduce the halo effect and carryover effects, each item should be scored without knowledge of the writer's name. As one comes to know one's students, this may not be possible, but it should be maintained whenever possible.
- **Score essay items by item, not by student.** Scoring one item at a time, across all individuals, helps to increase scoring consistency and decrease item-to-item carryover effects.
- **Score all of one item in one sitting without interruption.** Even when scoring items by item, all of one item should be scored at the same time to reduce variability in scoring.
- **Decide on a holistic or analytical method of scoring the essays.** A holistic evaluation generally provides a "general impression" score to the essay. This general impression can be structured somewhat by developing a series of criteria that the scorer keeps in mind when scoring the essay. The analytical evaluation is generally comprised of a series of criteria that are assigned point values. The essay is then scored on these various criteria, with the scores then aggregated for an overall score.
- **For restricted response essay items, prepare a model answer before scoring the responses.** There are two variations of the model answer strategy. Firstly, write a model answer and assign point values to various characteristics within the model answer. Secondly, write a model answer that would yield a perfect score. Then write a model answer that would yield a good, but not perfect score. Then write a model answer that would yield a satisfactory, but not good score. Finally, write a model answer that would yield a poor, but not good score. Use these models in assigning student essays scores.
- **For extended response essay items, prepare a rating system before scoring the responses.** Extended response essay items are generally scored using a rating system that is based on several, 3-5, criteria. The criteria are described at various levels of proficiency (as in the second method of RRE items), and these various levels are then assigned point values. These criteria and levels are then used to assess the essay items.
- **Decide how factors irrelevant to the learning outcome being assessed are to be handled before scoring the responses.** How will non-content/non-achievement factors, such as grammar and spelling, be dealt with within the grading scheme? If spelling and grammar are essential to the learning objectives, then they should be included in the scoring rubric. If spelling and grammar are not essential to the learning objective, that is, if the objective is content oriented, then spelling and grammar should not be included in the scoring rubric. Often, there are two scoring systems used simultaneously, a content rubric and a grammar rubric.

- **Check scoring guides, model answers and rubrics, against a few real responses to make any last minute adjustments.** Scoring rubrics should be modified based on a quick reading of several responses. This process provides a reality check to make sure the rubric contains the essential elements needed to effectively assess the items, and to make sure that the rubric is of the correct level of complexity.

**Restricted Response Essay Item:**

1. Briefly discuss the essential tenets of constructivism from a Vygotskian perspective.

*Answer Type I (Model + Points):*

Vygotsky was a socio-cultural theorist. He believed that a person's cognitive structures existed on a social level first (intermental), and then through internalization on a cognitive level (intramental). Vygotsky stressed the social origin of knowledge, but he also proposed that psychological components arising from these social origins were fairly representative of these experiences.

*Scoring Method I: Characteristics and Points (Analytic)*

Total Points: 5

3 points: Discussion of the two levels of knowledge acquisition - social & cognitive.

1 point: Emphasis on the social origin of knowledge.

1 point: Acknowledgement of the mental representations.

*Scoring Method II: Answer Versions with Points (Analytic)*

5 Point Answer: Vygotsky was a socio-cultural theorist. He believed that a person's cognitive structures existed on a social level first (intermental), and then through internalization on a cognitive level (intramental). Vygotsky stressed the social origin of knowledge, but he also proposed that psychological components arising from these social origins were fairly representative of these experiences.

3 Point Answer: He believed that a person's cognitive structures existed on a social level first (intermental), and then through internalization on a cognitive level (intramental). **OR** Vygotsky stressed the social origin of knowledge, but he also proposed that psychological components arising from these social origins were fairly representative of these experiences.

1 Point Answer: Vygotsky stressed the social origin of knowledge. **OR** He proposed that psychological components arising from social origins were fairly representative of these experiences.

**Extended-Response Essay Item:**

2. Briefly discuss the essential tenets of constructivism from a Vygotskian and Piagetian perspective. Contrast their two views and generate a single common pedagogy.

*Scoring Criteria:*

Adequate discussion of Vygotsky's tenets.

Discussion of the social differences.

Adequate discussion of Piaget's tenets.

Rationale for generating the common pedagogy.

Discussion of the mental similarities.

Efficacy of the common pedagogy.

**Extended-Response Essay Item:**

3. Briefly discuss the essential tenets of constructivism from a Vygotskian and Piagetian perspective. Contrast their two views and generate a single common pedagogy.

**Complex Scoring Rubric**

<b>Vygotskian Tenets:</b>	10 points
a. is Vygotsky's social epistemology made explicit	
b. is internalization fully explored	
c. is the ZPD well integrated	
d. is the role of language well integrated	
<b>Piagetian Tenets:</b>	10 points
a. is Piaget's genetic epistemology made explicit	
b. is adaptation well integrated	
c. are assimilation and accommodation well explored	
d. is equilibrium discussed as foundational	
<b>Theoretical Contrast:</b>	15 points
a. are theoretical similarities made explicit	
b. are theoretical differences made explicit	
c. are similarities/differences meaningful	
d. are similarities/differences justified	
<b>Common Pedagogy:</b>	25 points
a. are the principles independent of each other	
b. does the pedagogy reflect both Vygotsky's and Piaget's tenets	
c. are the principles clearly stated and efficacious	
d. do the principles form a unified perspective on teaching & learning	
<b>Organization, Writing, Grammar:</b>	10 points
a. does the introduction set forth the organization and does the conclusion bring closure	
b. do the expressed ideas follow a logical progression without major gaps in the ideas	
c. does word choice facilitate understanding	
e. is correct grammar used	

**J. Student Bluffing and other Anomalies** (Gronlund, 1998):

5. **Bluffing:** Bluffing comprises a series of strategies that are used by students to artificially inflate their scores on essay exams. Bluffing is most effective when the teacher has not adequately prepared scoring rubrics

- 1. Students may repeat the question in statement form, paraphrasing, with an indication/evaluation of the topic's importance.**

*Essay Question:* Briefly describe the primary advantages of selection type items in assessment and evaluation.

Selection type items, a foundational and excellent form of assessment and evaluation, have many advantages relevant to their use and application as compared to other types of assessment and evaluation that may be used by educators at all levels including K-12, college, and university.

- 2. Students may relate the topic of the essay to a topic that is more well known by the student and then proceed to elaborate on the better-known topic rather than the original topic of the essay.**

*Essay Question:* Briefly describe the primary advantages of selection type items in assessment and evaluation.

Selection type items, a foundational and excellent form of assessment and evaluation, have many advantages relevant to their use and application as compared to other types of assessment and evaluation that may be used by educators at all levels including K-12, college, and university. Include in these excellent items are multiple choice items, matching items, and true-false items. Multiple choice items consist of a stem, a basic statement of the problem, and a series of alternatives. The alternatives consist of the correct answer and distracters. The distracters should be plausible, and may often contain frequent errors committed by unknowledgeable students. Multiple choice items are an effective means for assessing both simple and complex outcomes.

**3. Students may use relevant terms/concepts within their essay without applying the terms and, perhaps, without really understanding the terms/concepts.**

*Essay Question:* Briefly describe the primary advantages of selection type items in assessment and evaluation.

Selection type items, a foundational and excellent form of assessment and evaluation, have many advantages relevant to their use and application as compared to other types of assessment and evaluation that may be used by educators at all levels including K-12, college, and university. Include in these excellent items are multiple choice items, matching items, and true-false items. Multiple choice items consist of a stem, a basic statement of the problem, and a series of alternatives. The alternatives consist of the correct answer and distracters. The distracters should be plausible, and may often contain frequent errors committed by unknowledgeable students. Multiple choice items are an effective means for assessing both simple and complex outcomes. Multiple choice and other selection items must be evaluated carefully for factors related to reliability, validity, and bias. Likewise when considering the advantages of items one must always be cognizant of sampling, the specific learning objectives/outcomes, and performance constraints.

**4. Students may include the names of relevant experts without understanding their contribution or meaning (i.e., name dropping).**

*Essay Question:* Briefly describe the primary advantages of selection type items in assessment and evaluation.

Selection type items, a foundational and excellent form of assessment and evaluation, have many advantages relevant to their use and application as compared to other types of assessment and evaluation that may be used by educators at all levels including K-12, college, and university. Include in these excellent items are multiple choice items, matching items, and true-false items (Gronlund, 1998). Multiple choice items consist of a stem, a basic statement of the problem, and a series of alternatives (Nitko, 1996). The alternatives consist of the correct answer and distracters. The distracters should be plausible, and may often contain frequent errors committed by unknowledgeable students (Linn & Gronlund, 1995). Multiple choice items are an effective means for assessing both simple and complex outcomes (Oosterhof, 1998). Multiple choice and other selection items must be evaluated carefully for factors related to reliability, validity, and bias. Likewise when considering the advantages of items one must always be cognizant of sampling, the specific learning objectives/outcomes, and performance constraints. Gronlund (1998), Nitko (1996), and Oosterhof (1998) have all emphasized the advantages of selection type items.

**5. Students may include the teacher's basic beliefs in answers.**

*Essay Question:* Briefly describe the primary advantages of selection type items in assessment and evaluation.

Selection type items, a foundational and excellent form of assessment and evaluation, have many advantages relevant to their use and application as compared to other types of assessment and evaluation that may be used by educators at all levels including K-12, college, and university. Included in these excellent items are multiple choice items, matching items, and true-false items (Gronlund, 1998). Multiple choice items consist of a stem, a basic statement of the problem, and a series of alternatives (Nitko, 1996). The alternatives consist of the correct answer and distracters. The distracters should be plausible, and may often contain frequent errors committed by unknowledgeable students (Linn & Gronlund, 1995). Multiple choice items are an effective means for assessing both simple and complex outcomes (Oosterhof, 1998). Multiple choice and other selection items must be evaluated carefully for factors related to reliability, validity, and bias. Likewise when considering the advantages of items one must always be cognizant of sampling, the specific learning objectives/outcomes, and performance constraints. Gronlund (1998), Nitko (1996), and Oosterhof (1998) have all emphasized the advantages of selection type items. One should, however, be careful not to use various types of items simply to fulfill a "test taking style" requirement. The data on test taking styles are limited at best, and misleading at worst. Finally, all uses of selection type items, based on their advantages, should be appropriate and defensible.