

## Assessing Student Learning

Some common terms associated with assessment include formative, summative, formal, informal and authentic. It is important to understand the differences among these terms.

- **Summative Assessment** is the generally the last assessment given during a unit of study. It measures overall learning of learning objectives. It can take the form of a test, project, writing, presentation, or performance.
- **Formative Assessment** is the assessment that happens throughout a unit of study. Teachers use the information gained from student performance on formative assessments to adjust instruction and plan for future lessons. Students can also use the information gained from these assessments about their own performance to focus their study and practice.
- **Formal Assessment** is preplanned and contains a consistent and valid way to measure student learning. Formal assessments can be large or small in scale and include quizzes, performances, projects, presentations, writings, discussions, debates....
- **Informal Assessment** is the type of assessment teachers do all the time. It is usually based on classroom observations and done on the spot. It is formative in nature, offering immediate support or correction to students.
- **Authentic Assessment** is often misunderstood as educators have mistakenly begun to call anything that isn't a test authentic assessment. True authentic assessment is what its name says, authentic. It provides students real situations in which to demonstrate their learning. For example, students demonstrate their ability to evaluate political platforms, they read and watch campaign speeches and debates, research the issues from all sides and present their conclusions in orally or in writing.

# Multiple Choice Assessment

A multiple choice question contains 3 parts:

Stem—asks the question, sets the task, states the problem

Alternatives—suggested responses

Keyed Alternative—the correct response

Distractors—incorrect alternatives (also called foils)

## Crafting Items

Items can be crafted to determine if a student is at a certain place on the continuum of learning.

For Example:

1. In what year did the United States enter WWI?
  - A 1776
  - B 1812
  - C 1917
  - D 1981
  
2. In what year did the United States enter WWI?
  - A 1901
  - B 1917
  - C 1941
  - D 1950
  
3. In what year did the United States enter WWI?
  - A 1913
  - B 1915
  - C 1917
  - D 1917

## What MC Items Can Do

They can directly measure students' ability to

- Discriminate and make correct choices
- Comprehend concepts
- Infer and reason
- Compute
- Interpret new data
- Apply information

They can indirectly measure students' ability to

- Recall information
- Articulate explanations
- Give examples
- Solve problems
- Display patterns of reasoning
- Construct things

## Advantages and Criticisms of MC Items

### Advantages include

- Can be used to assess greater variety of learning targets
- Do not require students to write out answers
- Focus on reading and thinking
- Less chance to guess correctly
- Distractors may give insight into difficulties students are facing

### Criticisms include

- Students can't be creative or express own ideas
- Brighter students may be penalized because they may see possibilities not recognized by average students
- Give impression that there is a closed body of knowledge
- Interpretations are decontextualized
- May shape education in undesirable ways

## Types of MC Items

### ❖ Correct Answer

For example

Who invented the sewing machine?

- A Fulton
- B Howe
- C Singer
- D Whitney

### ❖ Best Answer

For example

What was the purpose of the Marshall Plan?

- A military defended western Europe
- B reestablish business and industry in western Europe
- C settle US's differences with Russia
- D directly help the hungry and homeless in Europe

### ❖ Multiple Response

For example

What factors are principally responsible for the clotting of the blood?

- A contact of blood with a foreign substance
- B contact of the blood with injured tissue
- C oxidation of hemoglobin
- D presence of unchanged prothrombin

❖ **Incomplete Statement**

For example

Millions of dollars of corn, oats, wheat, and rye are destroyed annually in the United States by...

- A mildews
- B molds
- C rusts
- D smuts

❖ **Negative Statement**

For example

Which of these is NOT true of viruses?

- A Viruses live only in plants and animals.
- B Viruses reproduce themselves.
- C Viruses are composed of very large living cells.
- D Viruses can cause diseases.

❖ **Substitution**

For example

He passed the test therefore he was admitted into graduate school.

The underline section of this sentence should be replaced with which one of the following?

- A , therefore,
- B ; therefore;
- C : therefore,
- D ; therefore,

❖ **Incomplete Alternative**

For example

A test questions that yield consistent results is said to be

- A reliable
- B standardized
- C true
- D valid

❖ **Combined Response**

For example

In what order should these sentences be written to make a coherent paragraph?

- 1 A sharp distinction must be drawn between table manners and sporting manners.
- 2 This kind of handling of a spoon at the table, however, is likely to produce nothing more than an angry protest against squirting grapefruit juice about.
- 3 Thus, for example, a fly ball caught by an outfielder in baseball or a completed pass in football is a subject for applause.
- 4 Similarly, the dexterous handling of a spoon in golf to release a ball from a sand trap may win a championship match.
- 5 But a biscuit or a muffin tossed and caught at table produces scorn and reproach.

- A 1, 2, 3, 4, 5
- B 1, 3, 5, 4, 2
- C 1, 5, 3, 4, 2
- D 2, 5, 4, 3, 1

## Some Common Flaws in MC Items and How to Correct Them

### Incomplete Stem

*Poor*— ☹️

W.E.B. DuBois

A actively pressed for complete political participation and full rights for African Americans.

B taught that the immediate need was for African Americans to raise their economic status by learning trades and crafts.

C emphasized helping African Americans through the National Urban League.

D founded the Association for the Study of Negro Life and History.

*Better*— 😊

Which of the following comes closest to expressing W.E.B. DuBois' ideas about priorities of activities of African Americans during the 20<sup>th</sup> century?

A African Americans should first improve their economic condition before becoming fully involved in politics.

B African Americans should postpone the fight for equal access to higher education until their majority acquire salable trade skills.

C African Americans should withdraw from white society to form a separate state in which they have complete political and economic control.

D African Americans should become active, seeking out complete citizenship and full political participation immediately.

### Options in the Middle of the Stem

*Poor*— 😞

Before the Civil War, the South's

- A emphasis on staple-crop production
- B lack of suitable supply of raw materials
- C short supply of personnel was one of the major reasons manufacturing developed more slowly than it did in the North.

*Better*— 😊

Before the Civil War, why did manufacturing develop more slowly in the South than in the North?

- A The South emphasized staple-crop production.
- B The South lacked a suitable supply of raw materials.
- C The South had a short supply of capable personnel.

### Unnecessary Wordiness and Complexity

*Poor*-- 😞

Given the present day utilization of the automobile in urban settings, which of the following represents an important contribution of Garrett A. Morgan's genius?

- A automobile safety belts
- B crosswalk markers
- C traffic lights
- D vulcanized rubber tires

*Better*-- 😊

Which of the following did Garret A. Morgan invent?

- A automobile safety belts
- B crosswalk markers
- C traffic lights
- D vulcanized rubber tires



### **Extraneous Information**

*Poor--* ☹

There are 10 preservice teachers in the School of Education who recently registered for the college-sponsored weight-loss program. At the beginning of the program, each was weighed, and 10 had a mean weight of 139.4 pounds. Suppose there were three men in this group, and that their mean weight was 180 pounds. What was the mean weight of the women at the beginning of the program?

- A 115.0 pounds
- B 122.0 pounds
- C 140.0 pounds
- D 159.7 pounds

*Better--* ☺

Ten persons have a mean weight of 139.4 pounds. The mean weight of three of them is 180 pounds. What is the mean weight of the remaining seven persons?

- A 115.0 pounds
- B 122.0 pounds
- C 140.0 pounds
- D 159.7 pounds

### Negatively Phrased Stem

*Poor--* ☹️

Sometimes a teacher finds it necessary to use a mild form of punishment. When this occurs, which of the following should not happen?

- A Children should not believe all of their behavior is bad.
- B Children should understand the reason why they are being punished.
- C Children should understand that the teacher, not them, controls when the punishment will end.

*Better--* 😊

Sometimes a teacher finds it necessary to use a mild form of punishment. When this occurs, it is important that the children understand

- A that it may be a while before they can earn back the teacher's respect
- B the reason why they are being punished.
- C the teacher, not the children, controls when the punishment will end.

### Links or Cues Following Items

*Poor--* 😊

1. The perimeter of a rectangle is 350 centimeters. The length of the rectangle is 3 centimeters longer than the width. What is the width?

- A 18.7 cm
- B 86.0 cm
- C 89.0 cm
- D 116.7 cm

2. What is the area of the rectangle described in item 1?

- A 1050 sq cm
- B 7396 sq cm
- C 7654 sq cm
- D 8188 sq cm

*Better--* 😊

The width of a triangle is 4 centimeters and the length is 3 centimeters. What is the area?

- A 9 sq cm
- B 12 sq cm
- C 16 sq cm
- D 17 sq cm

### **Heterogeneous Alternatives**

*Poor--* ☹️

What is the official state bird of North Dakota?

- A Prairie Rose
- B Bismarck
- C Meadow Lark
- D Missouri River

*Better--* 😊

What is the official state bird of North Dakota?

- A Barn Owl
- B Meadow Lark
- C Robin
- D Swallow

### **Words Repeated in Each Alternative**

*Poor--* ☹️

Which of the following is the best definition of seismograph?

- A an apparatus for measuring sound waves
- B an apparatus for measuring heat waves
- C an apparatus for measuring earthquake waves
- D an apparatus for measuring ocean waves

*Better--* 😊

What type of waves do a seismograph measure?

- A sound waves
- B heat waves
- C earthquake waves
- D ocean waves

## Some Other Tips for Writing Effective MC Items

- Arrange alternatives in a meaningful order.
- Make sure alternatives are grammatically correct.
- Don't cue correct response with indefinite pronoun or s/v agreement, etc.
- Make sure alternatives don't overlap or mean the same thing.
- Avoid none and all of the above.

### Checklist for Quality Multiple Choice Items

- ✓ Does the item assess an important aspect of the unit's instructional targets?
- ✓ Does the item match your assessment plan?
- ✓ Does the stem ask a direct question or set a specific problem?
- ✓ Is the item based on a paraphrase rather than words lifted directly from the textbook?
- ✓ Are the vocabulary and sentence structure at a relatively low level?
- ✓ Is each alternative plausible?
- ✓ If possible, is every distractor based on a common student error or misconception?
- ✓ Is the correct answer to this item independent of the correct answers of other items?
- ✓ Are all of the alternatives homogeneous?
- ✓ Did you avoid using all or none of the above?
- ✓ Is there only one correct or best answer for each item?

# True-False Items

True-False items present a proposition that the student must judge.

## There are Six Varieties of True-False Items

### 1. True-False

T F The sum of all the angles in any four-sided closed figure equals 360 degrees.

### 2. Yes-No

Y N Is it possible for a presidential candidate to become President of the United States without obtaining a majority of votes cast on Election Day?

### 3. Right-Wrong

R W  $5 + 3 \times 2 = 16$

### 4. Correction

Read each statement below and determine if it is correct or incorrect. If it is incorrect, change the underlined word or phrase to make the statement correct.

C I The new student, who we met today, came from Greece.

### 5. Multiple True-False

Under the Bill of Rights, freedom of the press means that newspapers

1. have the right to print anything they wish without restrictions. T F
2. Can be stopped from printing criticisms of the government. T F
3. have the right to attend any meeting of the executive branch of the federal government. T F

### 6. Yes-No with Explanation

Guided reading is an effective approach to early reading instruction. Yes or No

Explanation:

## Pros and Cons of True-False Items

- ✓ Certain aspects of materials lend themselves to verbal propositions.
- ✓ They are relatively easy to write.
- ✓ They can be scored easily and objectively.
- ✓ They can cover a wide range of content in a relatively small number of questions.

But if not carefully written...

- ✓ They can only assess specific facts, which are often trivial.
- ✓ They can be ambiguous.
- ✓ They can be answered correctly by random.
- ✓ They encourage students to accept oversimplified statements of truth.

## Writing Effective True-False Items

1. Remove ambiguity. Your false item is not worth using if you can only write a true version of it by adding the word *not*.

2. If your statement asks students to make evaluative judgments, make it into a comparative statement.

T F Compared to strictly phonics-based approach to reading remediation, a holistic approach has been found to be more successful with most students.

3. Write false statements that reflect the actual misconceptions held by students who have not achieved the learning targets.

T F Not correcting a child who has miscued is detrimental to their continuing development as a reader.

4. Multiple choice items may be rewritten into true-false items using the foils from the multiple choice options.

## Checklist for Effective True-False Items

1. Does the item assess an important aspect of the unit's instructional targets?
2. Does the item assess important ideas, knowledge, or understanding (rather than trivia, general knowledge, or common sense)?

T F George Washington had wooden teeth.

T F George Washington actively participated in the Constitutional Convention.

3. Is the statement either definitely true or definitely false?
4. Is the statement paraphrased rather than copied directly from the learning materials?
5. Are the word lengths of true statements about the same length as word lengths of false statements?

T F The Red River does not flow northward.

T F The Red River flows southward.

6. Did you avoid presenting items in a pattern (TTFTTF...)?
7. Is the item free of clues that give the answer away?

T F It is always better...

T F Frequently, it is better....

(students have become wise to the fact that certain qualifiers are often false, so use it in reverse to determine those who have mastered the material. Avoid words like this with younger students as they do not attend to such cues often until after 9<sup>th</sup> grade.)

8. If the statement presents an opinion, have you stated the source of the opinion?
9. Does the item focus on only one important idea?

T F The Apple River flows west to join the Snake River where they form the Missouri River.

T F The Apple River and the Snake River empty into the Missouri River.



# Matching Format

*Use matching format only when you have several multiple choice items that require repeating identical sets of alternatives.*

**Matching format has three parts:**

Directions—

Premises—

Responses—

## **Advantages and Criticisms of Matching Format**

### **Advantages**

- Space-saving
- Objective
- Can be appropriate at all grade levels

### **Criticisms**

- Tends to focus on rote memorization
- Misuse by teachers

## **Possible Applications for Matching Format**

Accomplished	Persons
Noted Events	Dates
Definitions	Terms and Phrases
Examples	Rules, principles, classifications
Concepts	Symbols
Titles	Authors/Artists
Foreign Words	English correspondence
Uses and Functions	Parts and machines
Names of Objects	Pictures of Objects

## Tips for Crafting Effective Matching Format Items

- Make the matching items homogeneous.
- Explain completely the intended basis for matching.  
For example

*Poor--* ☹

Match column A with column B. Write your answer on the blank provided.

*Better--* ☺

Column A lists the parts of a plant cell. For each cell part, choose from Column B the main purpose of that cell part. Write the letter of that purpose on the blank provided.

- All responses should function as plausible options for each premise.
- Avoid using incomplete sentences as premises.  
For example

*Poor--* ☹

_____ 1. Most normally green plants lose their color when	a. through their stomata.
_____ 2. The common characteristic of a flowering plant is	b. grown in the dark.

Etc.

- Avoid perfect matching.
- Longer phrases appear in premise list.
- Identify premises with numbers and responses with letters.

### **Checklist for Quality Matching Format**

- ✓ Does the exercise assess an important aspect of the unit's instructional targets?
- ✓ Does the exercise match your assessment plan?
- ✓ Does every premise and response belong the same category?
- ✓ Do your directions clearly state the basis you intend the students to use?
- ✓ Does every response function as a plausible alternative to each premise?
- ✓ Are there fewer than 10 responses in the exercise?
- ✓ Did you avoid perfect matching?
- ✓ Are the longer statements in the premise list?
- ✓ Are the premises numbered and the responses lettered?

# Short Answer

## Three Types of Short Answer

### 1. Question-

What is the capital city of North Dakota? Bismarck

### 2. Completion-

The capital city of North Dakota is Bismarck.

### 3. Association-

On the blank next to each symbol write the name for it.

Omission

Repeat

Pause

Insertion

Corrected

## Pros and Cons of Short Answer

- ✓ Easy to construct
- ✓ Can be scored objectively
- ✓ Lowers probability of correctly guessing

But...

- ✓ Can't anticipate all possible responses
- ✓ Spelling errors, grammar and legibility complicate scoring
- ✓ Possibility exists for partially correct answers



## Writing Short Answer Items

**Question** is usually the best choice for format.

The author of *Where the Red Fern Grows* was \_\_\_\_\_.

Who is the author of *Where the Red Fern Grows*? \_\_\_\_\_



\*The first example may illicit responses such as “a story writer,” “a man,” “an American” or “was buried in 1984.”

**Key Idea**----Focus the item on the specific knowledge sought. While most short answer items test **lower-order thinking** such as recall and comprehension, **higher level thinking** can be reached including the following:

- Ability to make simple **interpretations** of data and **application** of rules such as counting the number of syllables in a word, **demonstrating** knowledge of the place value system, **analyzing** and **evaluating** information to make a judgment such as with symptoms of a disorder.
- Ability to solve numerical problems in science and mathematics.
- Ability to balance equations in science and mathematics.

$$a^2 + b^2 = c^2$$

## Checklist for Effective Short Answer Questions

1. Does the item assess an important aspect of the unit's instructional targets?
2. If possible, is the item written in question format?
3. Is the item worded clearly so that the correct answer is a brief phrase, single word, or single number?

Why are scoring guides recommended for use with essay tests?

The main reason for using a scoring guide with an essay test is to increase the \_\_\_\_\_ of scoring.

Where is Pittsburgh, Pennsylvania, located?

Pittsburgh, Pennsylvania is located at the confluence of what two rivers?

4. Is the blank or answer space toward the end of the sentence?

\_\_\_\_\_ is the name of the capital city of North Dakota.

The name of the capital city of North Dakota is \_\_\_\_\_.

5. Is the item paraphrased rather than a sentence copied from the learning materials?
6. If the item is completion, is the omitted word an important word rather than a trivial word?
7. Are there only one or two blanks?
8. Are the blanks all the same size?
9. Avoid grammatical or other clues.

A specialist in urban planning is called an\_\_\_\_\_.

A specialist in urban planning is called a(n)\_\_\_\_\_.

# Essay Assessment

## Two Types of Essay Format

### 1. Restricted Response-

Limit both the content and form of responses often restricting response to only certain aspects or components of complex learning. (Reliability and validity are higher for this format as all students interpret the question in the same way and are answering within focused parameters.)

Write a brief essay comparing and contrasting the poetry of Robert Frost and Gary Soto according to their use of (a) rhyme scheme, (b) imagery, and (c) analogies.

### 2. Extended Response-

Students are free to express their own ideas and interrelationships among their ideas and to use their own organization of their answers. Usually no single answer is correct. This can be used to judge general writing ability and subject-matter knowledge.

Students are provided with a prompt...

Devise a plan to determine whether the Democrats and Republicans are evenly distributed throughout the city, or whether the supporters of each party are concentrated in certain wards.

Throughout the story, *Because of Winn-Dixie*, we hear of Opal's mother and how her abandoning of Opal and her father affected them. Imagine what would happen if Opal's mother walked into the party. Write about it as if you were Kate DiCamillo.

## Pros and Cons of Essay Assessment

- ✓ Can achieve higher order thinking
- ✓ Values, and provides practice for effective communication
- ✓ Supports organizational skills

But...

- ✓ Poor scoring reliability
- ✓ Scoring is time-consuming
- ✓ Poor student communication skills may obstruct ability to assess actual student knowledge

## Checklist for Effective Essay Assessment

1. Does the essay assess an important aspect of the unit's instructional targets?
2. Does the essay require your students to apply their knowledge to a new or novel situation?
3. When viewed in relation to the other items on the test, does this contribute to the range of content and thinking skills represented?
4. Is the prompt focused?
5. Is the task appropriate for the educational level of your students?
6. To get a good mark on the item is the student required to demonstrate more than recall?
7. Is the prompt worded in a way that leads the students to interpret assignment in the way you intended?
8. Does the wording make clear the amount of writing, purpose for writing, and time to devote to this question?
9. If the essay prompt asks for an opinion, is it clear that assessment will be based on logic and evidence in support of their argument?

Poor: Name the principles that determined postwar American foreign policy.

Better: Describe three principles on which American postwar foreign policy was based between 1945 and 1960; illustrate each of the principles with two actions of the executive branch of government.

Poor: Why does an internal combustion engine work?

Better: Explain the functions of fuel, carburetor, distributor, and the operation of the cylinder's components in making an internal combustion engine run.



## Scoring Essays

There are two types of Scoring Rubrics

1. Holistic-

These focus on overall process or product as a whole, without judging the components separately. They are usually appropriate when performance tasks require students to create some sort of response and where there is no definitive correct answer. The score reports overall quality, proficiency, or understanding of the specific content and skills. The nature of this assessment is summative.

<u>Score</u>	<u>Description</u>
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

## 2. Analytic-

These focus on separate, individual components of the product or performance, and then sum the individual scores to obtain a total score. They are usually preferred when a focused type of response is required in which there may be one or two acceptable responses and creativity is not an essential feature of the students' responses. An analytic rubric creates a profile of specific student strengths and weaknesses. Formative feedback is generally the goal of this type of rubric.

**Table 2:**

*Template for analytic rubrics*

	<b>Beginning 1</b>	<b>Developing 2</b>	<b>Accomplished 3</b>	<b>Exemplary 4</b>	<b>Score</b>
<b>Criteria #1</b>	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
<b>Criteria #2</b>	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
<b>Criteria #3</b>	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
<b>Criteria #4</b>	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	

## References

- Gronlund, N. E. (2003). *Assessment of student achievement* (7<sup>th</sup> ed.). Boston: Allyn and Bacon.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment and Evaluation, Research Evaluation*, 7 (25). Retrieved March 12, 2003, from <http://ericae.net/pare/getvn.asp?v=7&n=25>
- Nitko, A. J. (2004). *Educational assessment of students* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Education.