

Best Practices in Writing and Using Assessments

NCTM Annual Conference
April 7, 2017

4/7/2017



Measuring the Power of Learning.®

Best Practices in Writing and Using Assessments

Robin O'Callaghan

rocallaghan@ets.org

Daniel Klag

dklag@ets.org



Purpose of the presentation

- Share best practices used in developing standardized assessments
- Help teachers write classroom tests to measure what students know and can do
- Help teachers to understand the assessments students take outside of the classroom

Overview

- What is a test “item”?
- Item types
- General guidelines for item writing
- Audience participation
- Guidelines for test assembly



What is a test “item”?

“A test item in an examination of mental attributes is a unit of measurement with a stimulus and a prescriptive form for answering; and, it is intended to yield a response from an examinee from which performance in some psychological construct (such as knowledge, ability, predisposition, or trait) may be inferred.” (Osterlind, 1990.)

Key points of the definition:

- Unit of measurement
- A stimulus and a prescriptive form for answering
- The response is interpreted in terms of learning about examinees’ performance in a particular psychological construct

Item Types

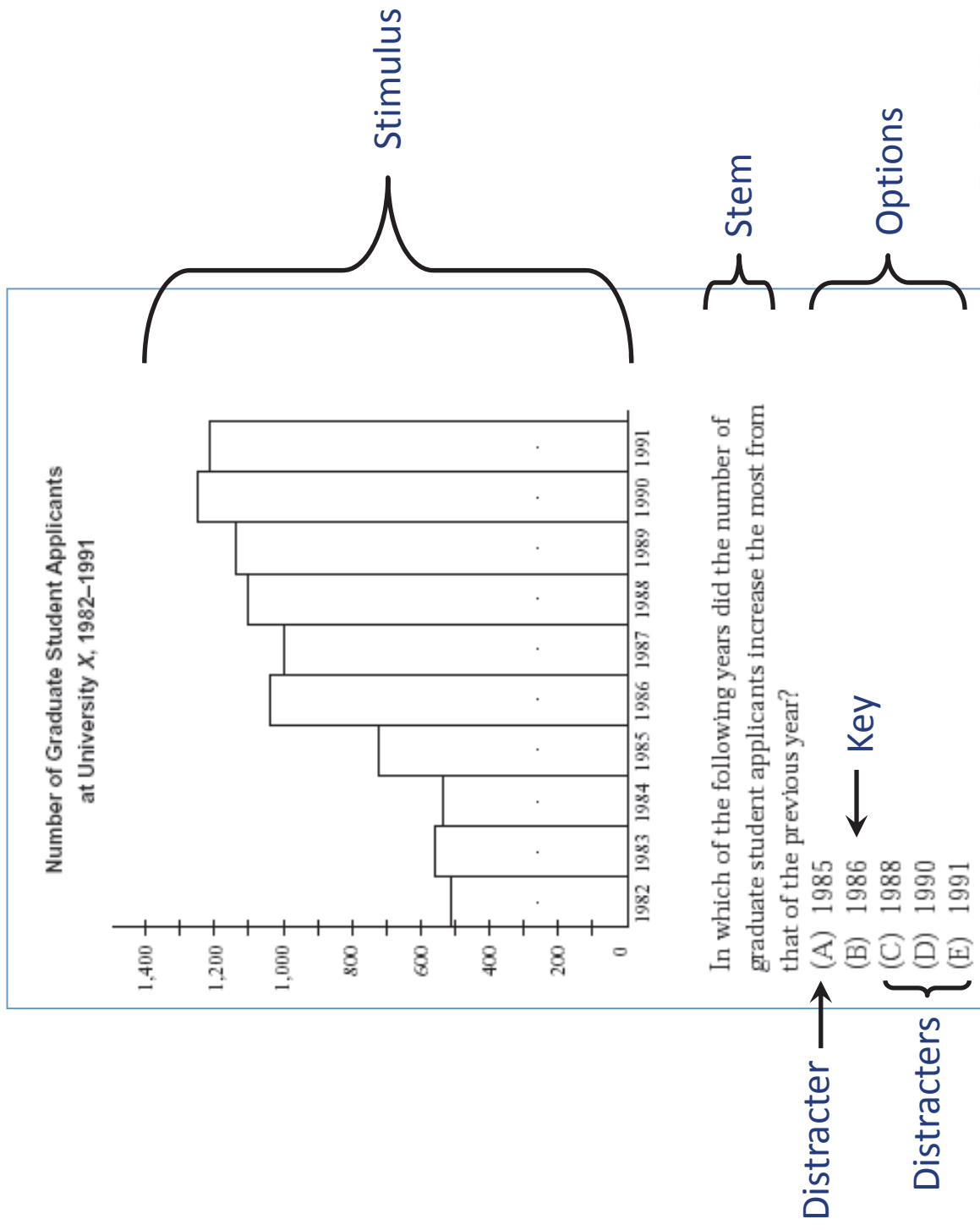


Item Types

- Multiple choice
- Numeric entry
- Constructed response
- Other technology-enhanced item types



Multiple Choice



Multiple Choice

Advantages:

- Format is simple and understandable
- Easy to score
- Can target common misconceptions
- Can ask many questions in a limited time

Disadvantages:

- Subject to guessing
- Can be back solved
- Can be difficult to write good distracters

Numeric Entry

The function f has the property that $f(3x) = 3f(x)$ for all x .

If $f(12) = 6$, what is the value of $f(4)$?

Enter your answer in the box.

$$f(4) =$$

Numeric Entry

Advantages:

- Nearly impossible to guess
- Cannot be backsolved
- No need to write options

Disadvantages:

- Must have numerical answer
- Cannot rely on “best answer”
- Potential rounding errors

Constructed Response Items

- Used to address skills that cannot be directly measured with other item types (e.g. modeling, argumentation)
- Requires a more in-depth response than machine-scored items
- Requires more time to administer and more time to score
- Can show evidence of partial understanding
- Requires a clear rubric for scoring

Other technology-enhanced item types

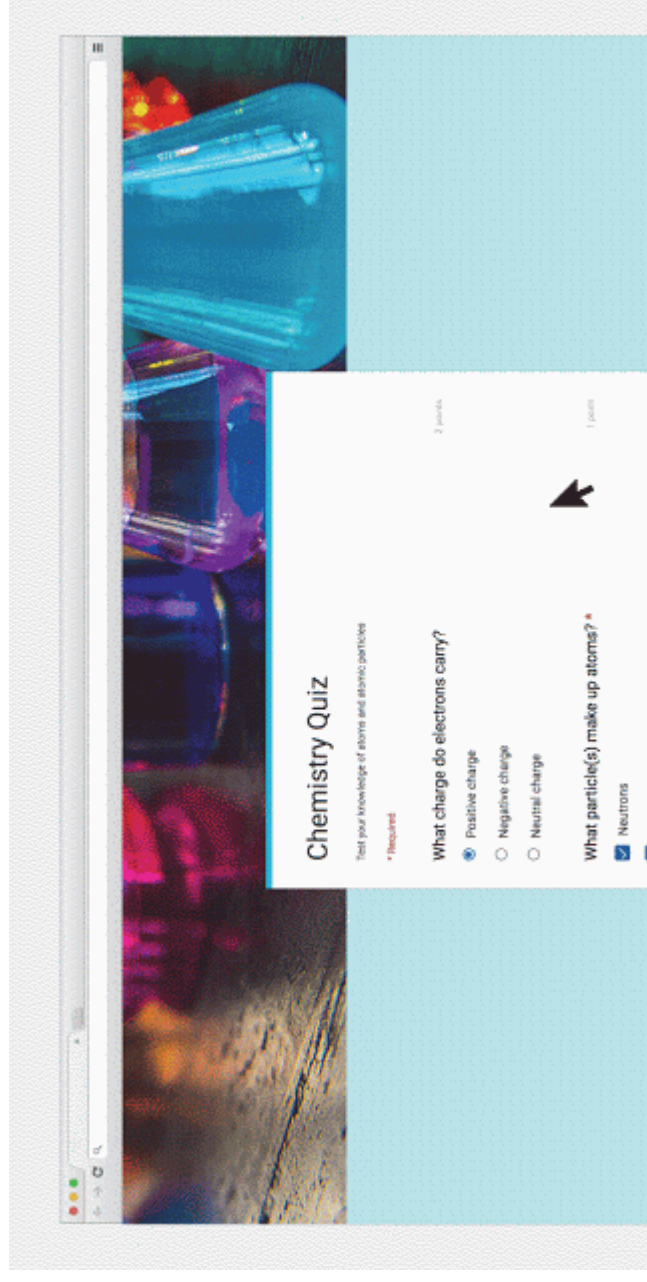
- Multiple-Selection Multiple Choice
- Drag-and-drop
- Drop-down lists
- Number line
- Coordinate graphing
- Bar Graphing

Considerations when using technology-enhanced items

- Limit the number of types of interactions on a given test
- Some interactions can be complicated
 - These items tend to be time-consuming
 - We want to assess students' proficiency with mathematics, not their proficiency with the tool

Administering your own computer-based assessments

- Try using Google Forms at forms.google.com
- Includes:
 - Single-selection multiple choice
 - Multiple-selection multiple choice
 - Drop-down lists
 - Grid



The background of the slide is a solid dark blue. It features several geometric elements: a large orange square in the top-left corner; a smaller blue square below it, slightly to the right; a horizontal white line extending from the left edge, passing through the blue square; a vertical white line extending from the bottom edge, passing through the orange square; a medium-sized orange square in the bottom-right area; and a dark red square in the bottom-right corner, partially overlapping the orange square.

General Guidelines for Item Writing

Avoid construct-irrelevant variance

- The **construct** is defined as the universe of content or skills to be measured. Elements that prevent or interfere with the test measuring the intended construct are described as **construct-irrelevant**.
- Assessments should provide an opportunity for all students to show what they know, measuring the construct without any irrelevant or distracting elements.

Avoid construct-irrelevant variance

- Avoid specialized terminology that is not relevant to the content being tested, e.g., sports or military terminology, farming terms, and religious terminology
- Avoid terminology, subject matter, or graphics that might be problematic for students with visual or auditory disabilities, e.g., decorative rather than informative illustrations, letters that look similar used as labels
- Be sensitive to differences with respect to special populations, e.g., English language learners and students with disabilities

Avoid construct-irrelevant variance

- Avoid material that is unnecessarily controversial or that may be upsetting to students
- Avoid material that may cause a student to have an emotional reaction (positive or negative) that might interfere with answering the question, including topics that
 - are emotionally charged for children (e.g., death, divorce, violence, natural disasters)
 - could be considered offensive (e.g., drinking, gambling, smoking)
 - are controversial in some communities (e.g., immigration, prayer in school)
 - encourage inappropriate behavior (e.g., cheating, fighting, lying)
- Remember that taking a test is stressful enough; we don't need to add additional negative feelings to the experience!

Focus on one central idea

$$(x - 3)(x + 5) = x^2 + bx + c$$

In the equation above, what are the values of b and c ?
Enter your answers in the boxes.

$b =$, $c =$

Focus on one central idea

The number 60 has how many factors?

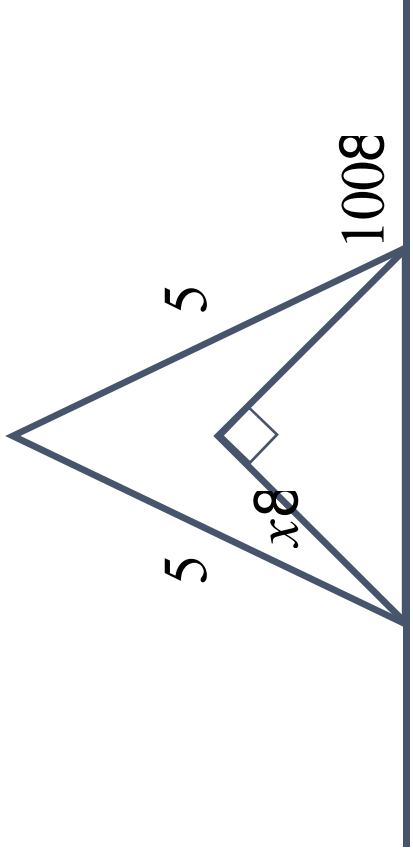
- A) 6
- B) 8
- C) 10
- D) 11
- E) 12

Be clear and concise

$$v(t) = 5,000(0.93)^t$$

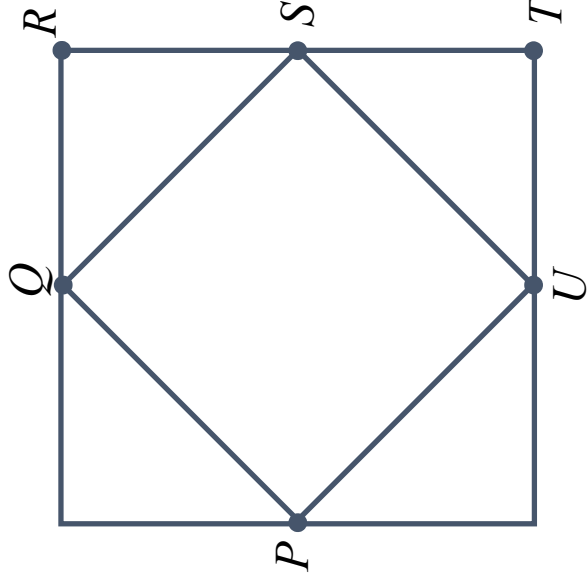
The value, v , in dollars, of a machine t years after it was purchased is given by the function defined above. Two years after it was purchased, the value of the machine was p percent of its purchase price. What is the value of p ?

Be clear and concise



In the figure above, what is the value of x ?

Avoid ambiguities



In the figure above, P , Q , S , and U are the midpoints of the sides of a square. Which of the following points is the same distance from P as it is from Q ?

- A) R
- B) S
- C) T
- D) U

Choose the right item type

3, 8, 13, 18, ...

The first term in the sequence above is 3, and each term after the first term is 5 more than the preceding term. Which of the following numbers is a term in the sequence?

- A) 998
- B) 999
- C) 1,000
- D) 1,001
- E) 1,002

Choose the right item type

- Distracters can model misconceptions.

If c is a positive number, which of the following expressions is equivalent to $(4 + c)^2$?

- A) $4 + c^2$
- B) $16 + c$
- C) $16 + c^2$
- D) $16 + 4c + c^2$
- E) $16 + 8c + c^2$

Choose the right item type

- Avoid using multiple choice in a question that can be back solved from the answer options.

If 4 times a number n is increased by 7, the result is 9. What is the value of n ?

- A) 0.4
- B) 0.5
- C) 1
- D) 4
- E) 16

Choose the right item type

- Avoid using numeric entry in a case where there are only 5 numeric possibilities

When the integer n is divided by 5, the remainder is 4.
What is the remainder when $3n$ is divided by 5 ?

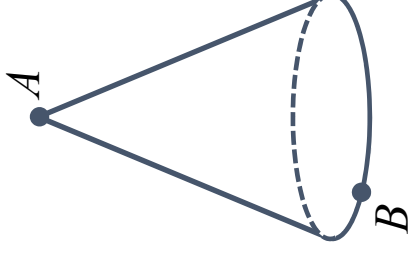
Enter your answer in the box.

Choose the right item type

- Avoid using multiple choice if the answer can be guessed by making a visual estimate.

In the figure shown, point A is the vertex of a cone and point B lies on the circumference of the base of the cone. The cone has a height of 12 centimeters and a diameter of 10 centimeters. What is the distance, in centimeters, from point A to point B ?

- A) 7
- B) 10
- C) 12
- D) 15





Audience Participation

Example

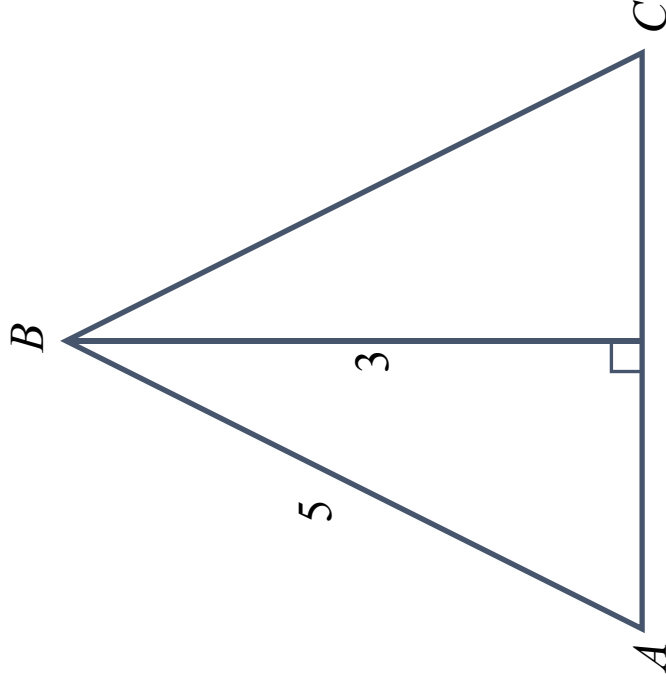
If $3x + 6 = 24$, what is the value of x ?

- A) 3
- B) 6
- C) 10
- D) 12

Example

In the figure shown, triangle ABC is equilateral. What is the area of triangle ABC ?

- A) 5.0
- B) 7.5
- C) 10.0
- D) 13.0



Example

John went to Oregon for his vacation. It was 8:00 p.m. when John arrived at his destination. He left at 8:00 that morning and traveled 500 miles, at an average speed of 50 mph, taking only one break after driving 250 miles. What was the duration of the hiatus?

Example

Jane drove from 8:00 a.m. to 8:00 p.m. with one stop to rest.
She drove 500 miles at an average speed of 50 miles per hour.
How long did she stop to rest?

Example

If a , b , and c are consecutive integers whose sum is 21, what is the value of b ?

- A) 5
- B) 6
- C) 7
- D) 8

Example

Last semester, Luis scored 88 on his first mathematics test and 96 on his last three mathematics tests. What was his average mathematics grade for the semester?

- A) 90
- B) 92
- C) 94
- D) 96

Example

In a certain contest, Fred must select any 2 of 5 different gifts offered by the sponsor. From how many different combinations of two gifts can Fred make his selection?

- A) 7
- B) 10
- C) 15
- D) 20

Example

<u>Column A</u>	<u>Column B</u>
$(-2)^{100}$	$(-2)^{99}$

- A) The quantity in Column A is greater.
- B) The quantity in Column B is greater.
- C) The two quantities are equal.
- D) The relationship cannot be determined from the information given

Example

<u>Column A</u>	<u>Column B</u>
$(-2)^{98}$	$(-2)^{99}$

- A) The quantity in Column A is greater.
- B) The quantity in Column B is greater.
- C) The two quantities are equal.
- D) The relationship cannot be determined from the information given

The background of the slide is a solid dark blue. It features several geometric elements: a large orange square in the top-left corner; a smaller blue square just below and to the right of the orange one; a medium orange square in the bottom-right area; and a dark red square in the bottom-right corner, partially overlapping the orange one. Thin white lines intersect to form a grid-like structure, with one line running vertically through the orange square in the bottom-right and another running horizontally through the blue square in the top-left.

Guidelines for Test Assembly

Guidelines for Test Assembly

- Write items with a view to maximizing the discriminating power of the items.
- For a typical test, have items that range in difficulty.
- The greatest concentration of items in a test is usually in the 30% to 70% correct range.
- Have items on the same topic at different levels of difficulty.

Questions?



Robin O'Callaghan
Daniel Klag

rocallaghan@ets.org
dklag@ets.org