

Practical Item Writer Training



Assessment and Accountability
Conference 2009

2009 OEAA Conference



Characteristics of Assessments: Reliability

...the extent to
which the
assessments are
consistent.



2008 OEAA Conference

2



Characteristics of Assessments: Reliability

- Factors affecting Reliability:
 - Consistency in ...
 - Calculator Use
 - "Open" Notes / Book
 - Formula Sheet / Resources
 - Time of Day
 - Day of Week
 - Length of Test
 - Use of Scoring Guides / Rubrics



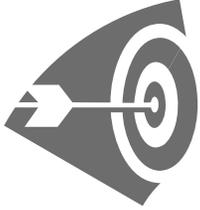
2008 OEAA Conference

3



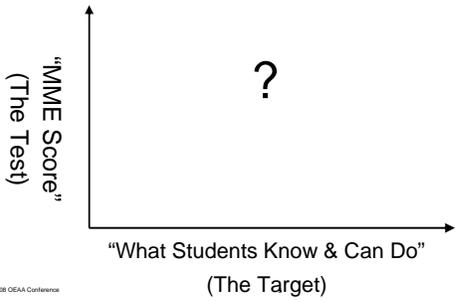
Characteristics of Assessments: Validity

...the extent to which scores correlate with what the test is supposed to measure.



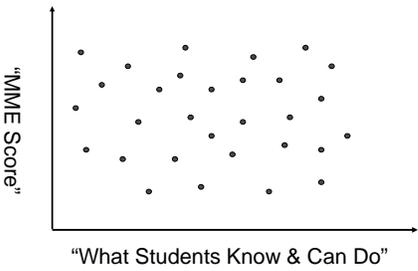
2008 OEA Conference 4

Characteristics of Assessments: Validity

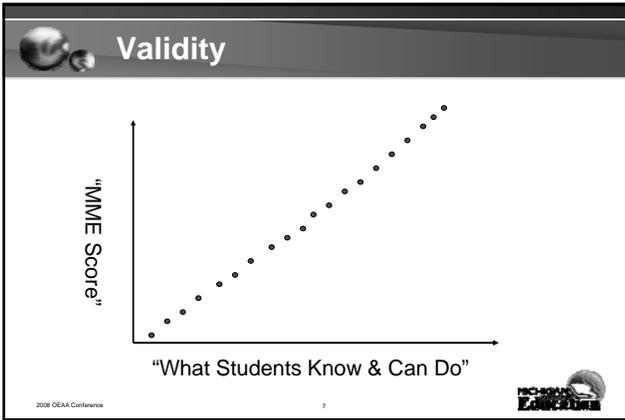


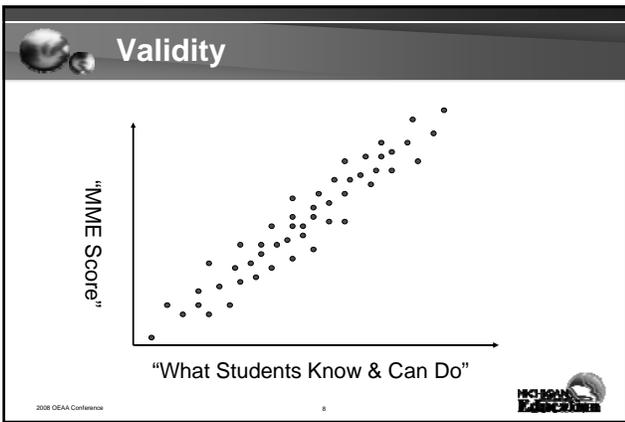
2008 OEA Conference

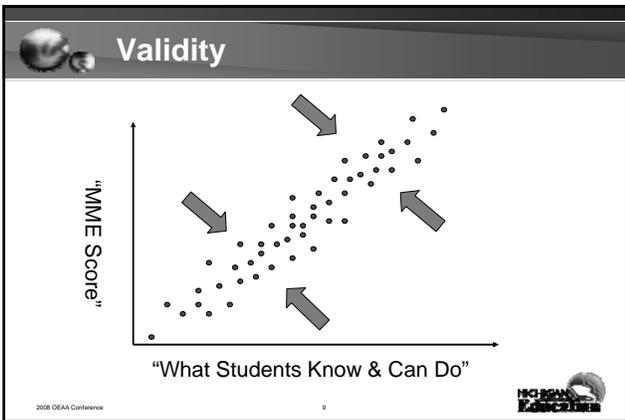
Validity



2008 OEA Conference 6









Basic Item Writing Principles

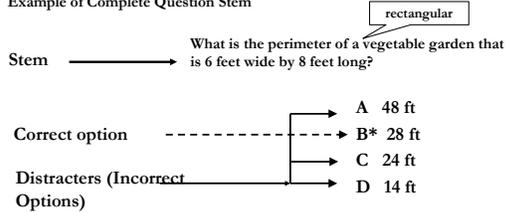
- The following slides give some guidelines and rules that should be considered when writing assessment items





Parts of an Item

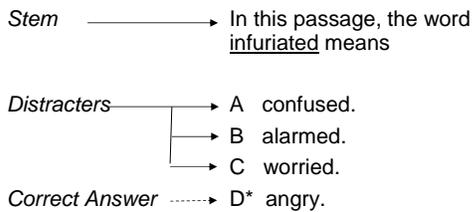
Example of Complete Question Stem





Parts of an Item

Example of Open-Ended Question Stem





Guidelines for Writing Multiple Choice Items

1. Write questions that measure understanding, insight, and higher level skills. Don't be trivial.
2. Use concise, straightforward language.
3. Present the problem or question as clearly as possible in the stem. Don't include unnecessary information.
4. Don't be tricky or cute.
5. Strive to be interesting to students at the target grade level.
6. Avoid bias toward or against any group of individuals.





Guidelines for Writing Multiple Choice Items

7. When calling for a judgment or conclusion, the student must be able to infer the correct answer from the text. The stems of items that call for a judgment or a conclusion should be worded accordingly ("According to the author," or "according to the selection").
8. Don't pose questions such as "What do you think?" or "What would you do?" because any option might be defensible.
9. Avoid using negative stems that ask the students to choose the one wrong answer, such as those that use the terms "except" and "not".





Guidelines for Writing Multiple Choice Items

10. Use options that fit grammatically with the stem.
11. Write options that are parallel in grammatical structure, logic and length.
12. Use distracters that are incorrect but plausible to students at the target grade level.
13. Provide one and only one correct answer, or one answer that is clearly the best.
14. Don't use "all of these" or "none of the above" as an option. This is inconsistent with there being only one correct answer.





Guidelines for Writing Multiple Choice Items

- 15. Numerical options should be presented in ascending or descending order.
- 16. Unnecessary or redundant information should generally not be included in the stem and never in the options.
- 17. Items should include multicultural contexts and names (unfamiliar names should be used in short, easy items and familiar names in hard items).





Guidelines for Writing Multiple Choice Items

- 18. Complete Question stems are preferred.
 - Open-ended stems are acceptable if it is clear from the stem what the item is asking.
 - For example, "The boy in the story lived —" is an unclear stem. Where did the boy live? With whom did he live? When did he live? How did he live? Students cannot determine what the item is asking until they read all the answer choices.
 - On the other hand, "The boy liked to watch ships because he —" is acceptable. The stem clearly asks "why" he liked to watch ships.





Six Criteria for Valid Items

- 1. The **CONTENT** of the item matches the **CONTENT** of the expectation.
- 2. The **PERFORMANCE** required in the item matches the **VERB** of the expectation.
- 3. The item cannot be solved merely by **TEST-WISENESS**.



 **Six Criteria for Valid Items**

- The item addresses **ONLY ONE** content expectation (may not be required in some instances).
- The item addresses important content related to the expectation.
- The **CONTEXT** of the item is **APPROPRIATE** and **ENGAGING**.

Adapted from Theron Blakeslee,
Michigan Mathematics Leadership Academy (MMLA)

2008 OEAA Conference 

 **1. The CONTENT of the item matches the CONTENT of the expectation (i.e., "Alignment")**

Chemistry prerequisite HSCE:

P4.p2D Recognize that the properties of a compound differ from those of its individual elements.

2008 OEAA Conference 

 **1. The CONTENT of the item matches the CONTENT of the expectation**

P4.p2D Recognize that the properties of a compound differ from those of its individual elements.

Example: Which of the following is a property of water that differs from its individual elements?

- Water is combustible in air
- Water is metallic
- Water is less dense
- Water is a liquid at room temperature*

(Getting closer...)

2008 OEAA Conference 

Form of Carbon	Charcoal (C)	Carbon Dioxide (CO ₂)
State at Room Temperature	Solid	Gas
Soluble in Water	No	Yes
Combustible in Air	Yes	No

Example: Based on the information in the table above, which is a reasonable hypothesis regarding elements and their compounds? (NAEP item)

- An element retains its physical and chemical properties when it is combined into a compound.
- When an element reacts to form a compound, its chemical properties are changed but its physical properties are not.
- When an element reacts to form a compound, its physical properties are changed but its chemical properties are not.
- Both the chemical and physical properties of a compound are different from the properties of the elements of which it is composed.* (Well-aligned)

2008 OEA Conference 22

2. The PERFORMANCE matches the VERB of the expectation

3. Find x.

Here it is

2008 OEA Conference

2. The PERFORMANCE matches the VERB of the expectation

P4.4A Describe specific mechanical waves (e.g. on a demonstration spring, on the ocean) in terms of wavelength, amplitude, frequency, and speed.

Example: Calculate the speed of a wave with a wavelength of 3.0 m and a frequency of 15 Hz.

- 5.0 m/sec
- 12 m/sec
- 18 m/sec
- 45 m/sec*

(Poor Alignment)

2008 OEA Conference 24

2. The PERFORMANCE matches the VERB of the expectation

Example: The figure above shows some ocean waves. Which of the labeled distances represents the wavelength? (NAEP item)

- a. A*
- b. B
- c. C
- d. D

2008 OEAA Conference (Better Alignment) 25

Speaking of Verbs

Mathematics High School Content Expectations
(all four content strands – Quantitative Literacy, Algebra, Geometry, Statistics & Probability)

“Know”	11%
“Solve”	10%
“Identify”	7%
“Describe”	5%
“Write”	5%
“Interpret”	5%
“Construct”	4%

2008 OEAA Conference (over 35 different verbs in all)

Speaking of Verbs

Science High School Content Expectations
(all four content areas – Physics, Chemistry, Life Science, Earth Science)

“Explain”	27%
“Describe”	18%
“Identify”	8%
“Calculate”	5%
“Predict”	3%
“Compare”	3%
“Recognize”	2%

2008 OEAA Conference (over 20 different verbs in all)

 **3. The item cannot be solved merely by TEST-WISENESS**

An initial population of 300 people grows at 2% per year. What will the population be in 10 years?

- 234
- 265
- 302
- 366*

More examples later...

2008 OEAA Conference 

 **3. The item cannot be solved merely by TEST-WISENESS**

Example: *Nitrogen-fixing bacteria* help cycle nitrogen through ecosystems. How do they do this?

- They change *nitrogen* into forms usable by plants.*
- They convert organic compounds to inorganic compounds during decomposition.
- They release chemical energy during respiration.
- They convert sunlight into chemical energy during photosynthesis.

More examples later...

2008 OEAA Conference 

 **5. ALL the content in the expectation is addressed in the item bank**

B3.4d Describe the greenhouse effect and list possible causes.

Example: Which of the following facts about the earth's temperature is due mainly to the greenhouse effect?

- Earth's average temperature is about 30° C warmer than expected.*
- Earth's oceans warm and cool much more slowly than its land masses.
- Earth's average summer temperature is about 15° C warmer than its winter temperature.
- Earth's temperature at the poles is cooler than at the equator.

2008 OEAA Conference 

5. ALL the content in the expectation is addressed in the item bank

B3.4d Describe the greenhouse effect and list possible causes.

Example: Which of the following atmospheric gases is NOT considered a "greenhouse" gas?

- a. oxygen*
- b. carbon dioxide
- c. water vapor
- d. ozone

2008 OEAA Conference 31



6. The CONTEXT of the item is APPROPRIATE and ENGAGING

OXON HILL, Md. -- The school system in Prince George's County is apologizing to parents after a teacher gave students a math test filled with inappropriate questions.

School administrators confirmed that a 10th-grade geometry teacher at Oxon Hill High School administered the test.

Officials said the test contained phrases like "Jose has two ounces of cocaine," "Willie gets \$200 for a stolen BMW," and "Raul gets six years for murder."

nbc4.com September, 2004

2008 OEAA Conference 32



6. The CONTEXT of the item is APPROPRIATE and ENGAGING

The Danish astronomer Ole Roemer (1644-1710) was the first person to determine that light traveled with a measurable speed. As Earth revolved around the sun, he noticed a variation in the time Io emerged from behind Jupiter, and calculated that light took 22 minutes to cross a diameter of Earth's orbit. If the diameter of Earth's orbit is 3.0×10^{11} m, what value would Roemer have calculated for the speed of light?

- a. 4.4×10^{-9} m/sec
- b. 2.2×10^8 m/sec*
- c. 3.0×10^8 m/sec
- d. 1.3×10^{10} m/sec

(Inappropriate and confusing context)

2008 OEAA Conference 33



6. The CONTEXT of the item is APPROPRIATE and ENGAGING

Some items require a context,...

31 During football season, Ronald recorded his body weight.

Ronald's Body Weight	
1 st week	215 pounds
2 nd week	222 pounds
3 rd week	219 pounds
4 th week	226 pounds
5 th week	223 pounds
6 th week	230 pounds

According to the pattern, how much would Ronald weigh in the 11th week?

A 227 pounds
 B 235 pounds
 C 238 pounds
 D 245 pounds

2008 OEA Conference 34

6. The CONTEXT of the item is APPROPRIATE and ENGAGING

"When items are written to particular content statements, such as the history and nature of science, they may be framed in these contextual components of science content."

Example: Ernest Rutherford found that when he fired alpha particles at a thin gold foil, some were scattered at large angles. What caused this scattering?

a. The gold's positive atomic nuclei attracted the negatively charged alpha particles.
 b. The gold's negative atomic nuclei repelled the negatively charged alpha particles.
 c. The gold's negative atomic nuclei attracted the positively charged alpha particles.
 d. The gold's positive atomic nuclei repelled the positively charged alpha particles.*

2008 OEA Conference Science Framework for the 2009 NAEP

Word economy and formatting

Instead of this formatting,

Jorge needs to evaluate the expression $\frac{-2(4-3)+8-4}{5}$ in order to finish his math homework. What should his answer be?

What about this?

Jorge needs to evaluate the expression below in order to finish his math homework.

$$\frac{-2(4-3)+8-4}{5}$$

What should his answer be?

Something to consider: To provide access to more students, take formulas, expressions, lists, etc. buried in text out of the text and center on a single line.

2008 OEA Conference 35



Word economy and formatting

Carol asked her 4 best friends, Alice, Clark, Ruth, and Matt, if she could measure their heights in inches. She found that Alice was 58 inches tall; Clark was 62 inches tall; Ruth was 60 inches tall; and Matt was 61 inches tall. To the nearest inch, what was the mean height of Carol's 4 friends?

What about this instead?

Carol measured the heights of each of 4 friends. Her results are shown below.

Alice: 58 inches Ruth: 60 inches

Clark: 62 inches Matt: 61 inches

To the nearest inch, what is the mean height of Carol's 4 friends?

Use a chart, a bulleted list, a graphic, etc. to break up reading.





Guidelines

1. Item "stem" should be **longer** (a complete sentence, if possible), the "options" should be **shorter**.





Guidelines

Example: Hanna scored 570 on a standardized exam. Her score exceeded the scores of 95,000 of the 125,000 who took the exam. Therefore

- a. her percentile rank was 6.0
- b. her percentile rank was 24.0
- c. her percentile rank was 57.0
- d. her percentile rank was 76.0*

Change to





Guidelines

Example: Hanna scored 570 on a standardized exam. Her score exceeded the scores of 95,000 of the 125,000 who took the exam. What was her percentile rank?

- a. 6.0
- b. 24.0
- c. 57.0
- d. 76.0*

MEAP HST in Mathematics Released Items 2004





Guidelines

2. Grammar & Vocabulary: Keep the grammar (e.g. verb tense) consistent.





Guidelines

Example: A certain species of bird can be brown or white. The white color is a recessive trait, while the brown color is a dominant trait. When two brown birds mate, is it possible for them to have white offspring?

- A. No, because both parents will have only genes for being brown.
- B. Yes, because offspring color does not depend on the genes of the parents.
- C. Yes, because both parents may have and pass on the gene for being white.*
- D. No, because the parents passed on only the dominant trait to their offspring.





Guidelines

2. Grammar & Vocabulary: Don't repeat a word in an option that was used in the stem.





Guidelines

Example: According to the *addition rule*, the probability that *either* event 'A' ($P(A)$) or event 'B' ($P(B)$) will occur is equal to

- a. $P(A) + P(B)^*$
- b. $P(A) - P(B)$
- c. $P(A) \times P(B)$
- d. $P(A) / P(B)$





Guidelines

2. Grammar & Vocabulary: Keep the readability appropriate.





Guidelines

A1.2.9 Know common formulas (e.g. slope, distance between two points, quadratic formula, compound interest, distance = rate x time), and apply appropriately in contextual situations. (Algebra II)

Example: A credit union returns 5.5% per annum compounded quarterly on a 15-month CD. If \$10,000 is deposited and the interest is accrued, what is the balance in the account after one year?

(Any troublesome vocabulary here?)





Guidelines

3. (a) Avoid the use of absolute terms, e.g. “always” or “never.”





“Always” & “Never” – Exception

Example: The Earth’s Moon is

- A. always much closer to the Sun than it is to the Earth
- B. always much closer to the Earth than it is to the Sun*
- C. about the same distance from the Sun as it is from the Earth
- D. sometimes closer to the Sun than it is to the Earth and sometimes closer to the Earth than it is to the Sun

Science Framework for the 2009 NAEP





Guidelines

3. (b) Avoid the use of “all of the above,” or “none of the above” in the options.





Guidelines

Example: What is the percent composition of carbon in carbon dioxide, CO₂?

- | | | |
|-----------------------|-------------|--------|
| a. 12% | | a. 12% |
| b. 32% | Change to → | b. 27% |
| c. 44% | | c. 32% |
| d. none of the above* | | d. 44% |





Guidelines

4. Make all the options the same length, with similar detail.





Guidelines

Example: A local car dealership wants to know how many people hear their advertisements on radio. Which method provides the most valid results?

- A. Survey the next 20 customers
- B. Survey all the people living within 1/2 mile
- C. Survey a large random sample of people living within the listening range of the radio station*
- D. Survey customers at a nearby auto repair shop





Guidelines

5. Present numerical options consistently.

- Usually in ascending, or descending order.
- Don't mix specific values with ranges, etc.





Guidelines

Example: In certain breeds of dogs, deafness is due to a recessive allele (d) of a particular gene, and normal hearing is due to its dominant allele (D). What is the probability that the offspring of a normal heterozygous (Dd) dog and a deaf dog (dd) will have normal hearing?

- | | | |
|------------------|-------------|---------|
| A. 100% | | A. 0% |
| B. 0.25 | Change to → | B. 25% |
| C. less than 1/3 | | C. 50%* |
| D. One half* | | D. 100% |





Guidelines

6. Make sure there is only ONE correct response.





Guidelines

Example: If Karen were to measure the length of each of the butterflies in her collection, which would be the best measurement unit to use?

- a. meter
- b. centimeter*
- c. millimeter*
- d. kilometer





Depth of Knowledge

- 1. Recall
- 2. Skills and Concepts
- 3. Strategic Thinking
- 4. Extended Thinking

Can be applied to Standards (Content Expectations), AND to Assessment Items

(Dr. Norman Webb, University of Wisconsin, 1997, 2002)





Depth of Knowledge

Level 1 – Recall

The recall of information (fact, definition, or term), or performing a simple procedure (a “recipe”), or applying a simple algorithm or formula. Requires only a rote response, a well-known formula, or following a well-defined procedure that typically involves only **one step**. Key words include “**identify**,” “**recognize**,” “**use**,” “**calculate**” and “**measure**.”

A student answering a Level 1 item either knows the answer or does not. The answer does not need to be “solved” or “figured out.”

(Verbs like “describe” and “explain” can be used at different levels depending on the complexity of what’s being “described” or “explained.”)





Depth of Knowledge

B2.1A Explain how cells transform energy (ultimately obtained from the sun) from one form to another through the processes of photosynthesis and respiration. **Identify the reactants and products in the general reaction of photosynthesis.**

Level 1 Example: Which of the following is a product of the overall process of photosynthesis?

- a. Carbon dioxide
- b. Protein
- c. Glucose*
- d. Water





Depth of Knowledge

E4.p2A Describe the composition and layers of the atmosphere. (*prerequisite*)

Level 1 Example: Air is made up of many gases. Which gas is found in the greatest amount?

- a. Nitrogen*
- b. Oxygen
- c. Carbon Dioxide
- d. Hydrogen

Science Framework for the 2009 NAEP





Depth of Knowledge

Level 2 – Application

This level is **more complex** and involves mental processing beyond simply recalling or reproducing a response. Items require students to make some decisions and typically involve **more than one step**. Key words and activities include “**classifying**,” “**organizing**,” “**estimating**,” or “**interpreting or comparing data**” in tables graphs or charts.





Depth of Knowledge

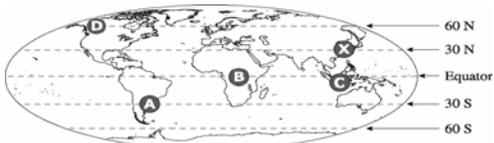
Level 2 Examples include:

- Explain the relationship between facts or variables;
- Describe examples and non-examples of science concepts;
- Select a procedure and perform it;
- Formulate a routine problem given data & conditions; and
- Organize, represent and interpret data.





Depth of Knowledge



The diagram above shows a map of the world with the lines of latitude marked. Which of the following places marked on the map is most likely to have an average yearly temperature similar to location X?

- A. **A**
- B. **B**
- C. **C**
- D. **D**

From Science Framework for the 2009 NAEP





Depth of Knowledge

Level 3 – Strategic Thinking

This level is more demanding and requires planning, using evidence, and **complex and abstract** reasoning. In most instances, requiring students to ***explain their thinking*** is Level 3. Students are asked to draw conclusions, cite evidence, develop logical arguments, solve complex problems, explain concepts and **justify their response**.





Depth of Knowledge

Examples include:

- **Identify research questions and design investigations;**
- **Solve complex, non-routine problems;**
- **Develop a scientific model; and**
- **Form conclusions from experimental data.**





Depth of Knowledge

Level 3 Example: The main reason for Earth's temperature being hotter in summer than in winter is:

- The earth's distance from the sun changes.**
- The sun is higher in the sky.***
- The distance between the northern hemisphere and the sun changes.**
- Ocean currents carry warm water north.**





Depth of Knowledge

Level 4 – Extended Thinking

This level requires complex reasoning, experimental design, and planning usually over **extended periods of time**. Students are asked to make connections *within* or *among* content areas. Level 4 tasks are typically assessed locally and often involve performance or open-ended assessments. **Many on-demand instruments will not include any items at Level 4.**





Depth of Knowledge

Examples:

- Based on provided data from a complex experiment that is novel to the student, deduce the fundamental relationship between several controlled variables
- Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions.





Contact Information

- Tom Wessels, Director
– Grand Traverse Regional Math and Science Center
twessels@tbaids.k12.mi.us
- Bill Brown, Independent Contractor
bbcllc@sbcglobal.net