



HLC Accreditation 2020-2021

## Evidence Document

---

Academic Affairs


Office of Assessment and Accreditation

---

### Item Writing for the Higher Education Classroom

**Additional information:** Dr. Jason Herron delivered his presentation on assessment theory and best practices at the University's Assessment Symposium in October, 2019. The slides are available at: <https://www.wichita.edu/administration/assessment/documents/HerronSymposiaFall2019.pdf> (Accessed on May 11, 2021.)

---



# Item Writing For the Higher Education Classroom

For cognitive assessments

Dr. Jason P. Herron

Educational Psychology

# Agenda

- Background
- Rationale
- Theory of Assessment
- True or false
- Multiple choice
- Matching
- Essay and short answer

Background



# Why are we discussing item creation?

- There are many components of assessments in higher education (item analysis, theory of testing, selection of level of task type and level, assessing testing galaxy etc.)
- In the end all things related to testing theory lead to creating great items to help assess learner's skill
- It is a good start point to the complex discussion of assessment

# First, What are you assessing?

*Only discussing assessments in the cognitive domain today*

- Cognitive Domain – knowledge-based. Exhibited by intellectual abilities
- Affective Domain – addresses a learner's emotions, beliefs, attitudes, etc.
- Psychomotor Domain – refers to the use of basic motor skills and physical movement.

# Testing Theory...

- The basic principle of testing theory for cognitive measures is to correctly identify classification of students into two group (master vs non master)
- better test differentiation is due to test/item construction
- Items differentiate individually
- Tests differentiate between as a whole
- Poor items mask your ability to determine student level of skill

# Scenario

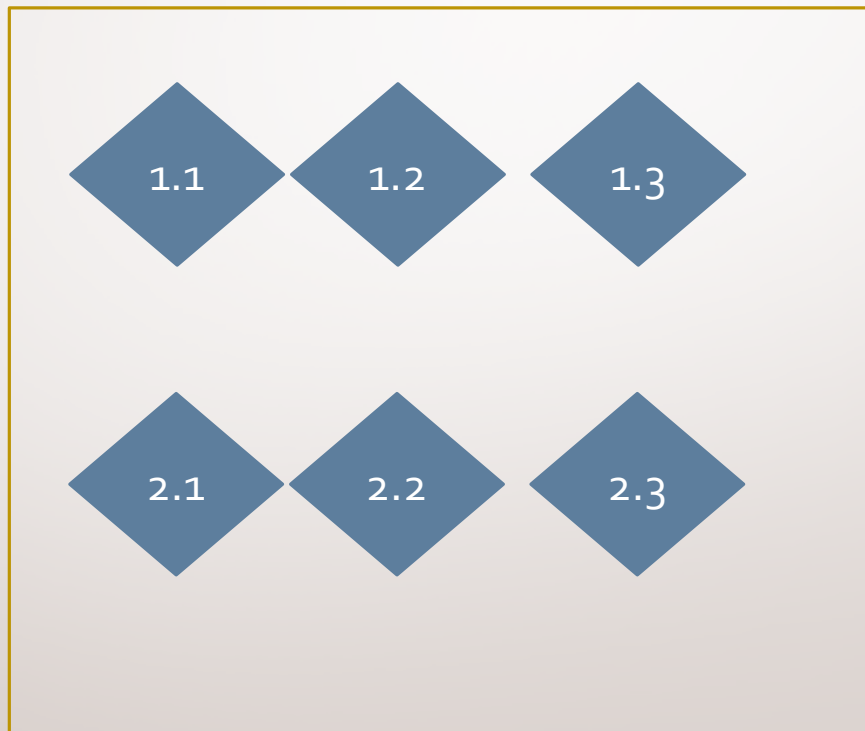
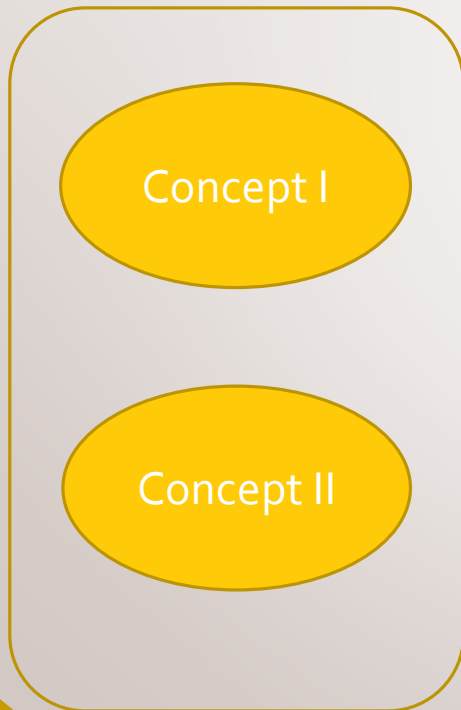
- You are an instructor of a graduate statistics course
- You are creating a examination that is aligned to the first unit of your course
- The unit consists of 2 major components



# Exam I

Unit I major components

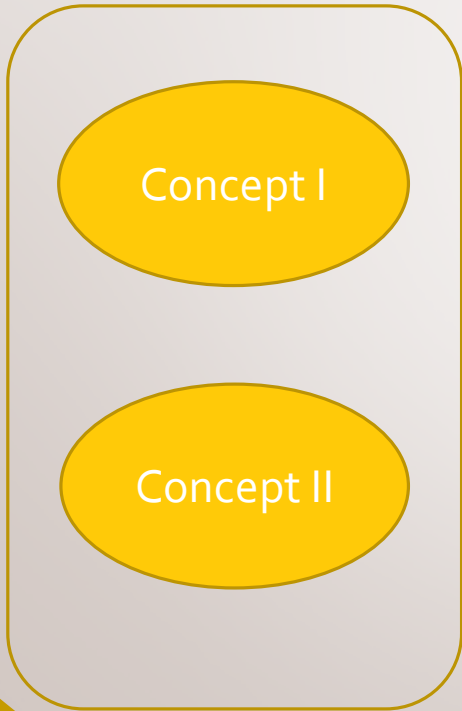
Individual Items



# Scenario

- You are an instructor of a graduate statistics course
- You are creating a examination that is aligned to the first unit of your course
- The unit consists of 2 major components

What happens to my ability to classify students if one or more of my items is broken\*?



\*A broken item is one that is unable to discriminate between classifications of student for one or more reasons

# Exam I

What happens to my ability to classify students if I add more good (well constructed) items?

*We give the masters additional chances to exhibit their mastery. The inverse is also true for the non-masters*

Concept II



# Caveat about adding additional items

- This principle only works to a certain extent, at a certain point adding additional items increases cognitive to an extent to which you destroy any benefit to creating additional items
- A 20 item assessment may differentiate between masters and non masters well
- A 500 item assessment will likely stop differentiating after 50 or so items

# Fundamental Principles for Crafting Items

If you follow these guidelines your  
assessments will be better as a whole

# General Test Item guidelines Design each assessment task (item) to:

- Focus entirely on an important learning target
- Discriminate: The correct answer should be obvious to the master
- Neither prevent nor inhibit a student's ability to demonstrate attainment of the learning target
- Avoid any unintended cues (sequencing of correct responses, phrasing)



# Design each assessment task to:

- Each item should ask a specific question or ask for a specific task.
- A single concept should be tested in selection-type items (e.g., true/false, MC). Why?
- The language of the item should be clear and concise.
- Now, let's discuss specific item types





True/false items



# Guidelines: True-False Items

## ***State items positively:***

T F It was not unnecessary to include the control group in the research study.

Better:

T F It was necessary to include the control group in the research study.

*A master will see this and start over thinking*

## Guidelines: True-False Items

**Statements should be unequivocally true or false:**

T F The President of the United States is elected to office.

Better:

T F The most common method of becoming President of the United States is through election.

# Guidelines: True-False Items

## ***Avoid specific determiners:***

T F Intel CPU's are always faster than competitor CPU's.

Better:

T F Intel's "Pentium dual-core" CPU is faster than the fastest Mac dual-core CPU.

# Guidelines: True-False Items

***Include only one central idea in each statement:***

T F The correlation coefficient is a measure of the relationship between two variables and indicates the cause of the relationship.

Better:

T F The correlation coefficient is a measure of the relationship between two variables.

T F The correlation coefficient indicates the cause of the relationship between two variables.

# Guidelines: True-False Items

***Keep statements short, using simple language:***

T F While it may be difficult to determine the exact pH value of a solution, it is possible to determine if a solution is acid by the red color formed on litmus paper inserted into the solution.

Better:

T F Litmus paper turns red in an acid solution.

# Guidelines: True-False Items

***Statements of opinion should be attributed to a source:***

T F “Mean Girls” is the best movie of all times.

Better:

T F According to political commentator Andrea Tantaros, “Mean Girls” is the best movie of all times.

# Writing T/F Items

- Use instructional materials as the source of ideas.
- Write items in pairs, one T and one F.
  - Example pair:
    - The use of optional rather than required essay items reduces the ability to make norm-referenced interpretations. (T)
    - The use of optional rather than required essay items enhances the ability to make criterion-referenced interpretations. (F)
  - Example pair:
    - The reliability advantage of using required versus optional essay items is due to differences in score variability. (T)
    - The reliability advantage of using required versus optional essay items is due to differences in test length. (F)



# Writing T/F Items

- Test an important proposition/idea.
  - Words like *some*, *usually*, *all*, or *never* should be avoided in writing T/F items. (F)
  - Two pitfalls should be avoided in writing T/F items. (F)
- Require understanding, not just memory.
  - If the hypotenuse of an isosceles right triangle is seven inches long, each of the two equal legs must be more than five inches long. (F)
  - The square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides. (T)

Don't forget: The truthfulness and falseness of statements should be obvious to an expert.

# Writing T/F Items

- Avoid unintended clues; make wrong answers attractive to those who lack command of the knowledge.
  1. By adding more solute, a saturated solution can be made supersaturated. (F)
  2. A supersaturated solution contains more solute per unit than a saturated solution. (T)
- To reduce ambiguity, introduce a comparison within the item.
  1. Open-book tests tend to be inefficient. (?)
  2. Open-book tests tend to be less efficient than closed-book tests. (T)

# Multiple choice items



# Parts of Multiple Choice Items

- STEM
  - Part of the MC item that asks the question.
- KEY
  - Correct option
- DISTRACTERS
  - Incorrect options

# Guidelines: Multiple Choice Items

- Provide plausible but incorrect distracters.
- Format the item vertically, not horizontally. (Avoid cognitive overload)
- Keep items independent of one another.
- Avoid textbook verbatim phrasing.
- Balance/randomize the correct option position.



Example Items

# Guidelines: Multiple-Choice Items

*The stem should introduce a clear problem or task:*

## Unclear

A table of specifications...

- a. indicates how a test will improve learning.
- b. provides more balanced content sampling.
- c. arranges objectives in order of importance.
- d. specifies the method used to score a test.

## Better

What is the main advantage for using a table of specifications to prepare an achievement test?

- a. It reduces the amount of time required.
- b. It improves the sampling of content.
- c. It makes the construction of items easier.
- d. It increases test objectivity.

# Guidelines: Multiple-Choice Items

*Avoid extraneous clues:*

The Judiciary Committee's impeachment deliberations resulted in a resolution of

- no impeachment.
- the majority voted for three articles.
- a sharp division between the parties.
- one article cited obstruction of justice.

**Better**

Which of the following would be the best resource on testing student learning?

- Educational Psychology, Gage and Berliner.
- Measurement and Evaluation in Teaching, Gronlund.
- Psychological Testing, Anastasi.
- Handbook of Research on Teaching, Wittrock.



# Guidelines: Multiple-Choice Items

*The stem should contain elements common to each alternative:*

Which of the following statements best represents the views of Microsoft's Bill Gates?

- a. A computer is a tool.
- b. A computer is a profit making device.
- c. A computer is an extension of the software programmer.
- d. A computer is best used by a knowledgeable programmer.

Which of the following statements best represents the views of Microsoft's Bill Gates?

A computer is:

- a. a tool.
- b. a profit making device.
- c. an extension of the software programmer.
- d. best used by a knowledgeable programmer.

# Guidelines: Multiple-Choice Items

*The Alternatives should be plausible:*

Which composer wrote the opera "Madama Butterfly?"

- a. Elton John
- b. Trace Adkins
- c. Beyonce
- d. Giacomo Puccini

Which composer wrote the opera "Madama Butterfly?"

- a. Wolfgang A. Mozart
- b. Gian-Carlo Menotti
- c. Giuseppe Verdi
- d. Giacomo Puccini



# Guidelines: Multiple-Choice Items

- Multiple-choice items can be used effectively with young children
- Multiple-choice items can be used to measure analysis and evaluation of complex ideas

They are considered the most flexible item type because they can be used for different levels of Bloom.



Matching items



# Guidelines: Matching Items

- Use homogeneous options and items.
- Use unequal numbers of options and items (avoid process of elimination).
- Indicate that options may be used more than once or not at all.
- Use the longer statements as the “premise” of the item and shorter statements as response options.
- Arrange options systematically (e.g., alphabetically, conceptually, chronologically).
- The directions should inform students how to respond; telling what is in the columns on the right and left.



Example Items

# Guidelines: Matching Items

*What is wrong with the following matching item?*

For each phrase on the left, choose the option on the right that goes with it.

- |                   |   |
|-------------------|---|
| ___1. New Orleans | A. Largest river in the U.S.            |
| ___2. Glacier     | B. The French Quarter                   |
| ___3. Erie        | C. One of the Great Lakes               |
| ___4. Mississippi | D. A large body of ice<br>found on land |

# Guidelines: Matching Items

## A Better Example:

On the line to the left of each historical event in Column A, write the letter from Column B which identifies the time period during which the event occurred. The dates may be used once, more than once, or not at all.

Column A

Column B

B

1. Boston Tea Party

A. 1765-1769

A

2. The Stamp Act

B. 1770-1774

C

3. Battle of Lexington

C. 1775-1779

B

4. First Continental Congress

D. 1780-1784

E. 1785-1789





Short answer items

# Guidelines: Completion/Short-Answer Items

*Items should clearly state the task:* Use a direct question; use incomplete statement only when it is more concise.

The validity of IQ tests is \_\_\_\_\_.

**Better**

What is the average correlation between IQ and school achievement?  
\_\_\_\_\_

# Guidelines: Completion/Short-Answer Items

*Put answers at or near the end in a single column on the right*

\_\_\_\_\_ is the number of nerve cells in the brain of the average Spider monkey.

**Better**

How many nerve cells are found in the brain of the average Spider monkey? \_\_\_\_\_

# Guidelines: Completion/Short-Answer Items

*Use only one blank per item*

The \_\_\_ score is a \_\_\_\_\_ that indicates a student's scholastic aptitude.

**Better**

The standard score that indicates a student's scholastic aptitude is the \_\_\_\_\_ score. \_\_\_\_\_

# Guidelines: Completion/Short-Answer Items

*Blanks should be of uniform size to avoid giving clues*

The second stage in Piaget's theory of cognitive development is the \_\_\_\_\_ stage.

The fourth stage in Piaget's theory of cognitive development is the \_\_\_\_\_ operational stage.

A cognitive structure which governs action is called a \_\_\_\_\_.



# Guidelines: Completion/Short-Answer Items

*Be sure the item assesses something meaningful*

The difference between preoperational and concrete operational \_\_\_\_\_ lies in their ability to reason logically. \_\_\_\_\_



# Essay Items



# Guidelines: Planning & Development of Essay Items

- Be sure the essay questions match a clearly stated objective.
- Do not use essays to measure "Knowledge Level" objectives.
- Specify limitations (constraints): time, length, point values.
- *Can I write a list? No you cant. Why?*
- **Essay items are easy to construct, but difficult to score reliably and fairly**





# Guidelines: Planning & Development of Essay Items

- **Specify a clear task: *extended-response items***
  - appropriate for evaluating writing skill, ability to synthesize or organize
  - inappropriate for testing content achievement
  - keep in mind the importance of writing skill - consider teaching students how to:
    - \* compare and contrast
    - \* analyze or evaluate
    - \* organize ideas


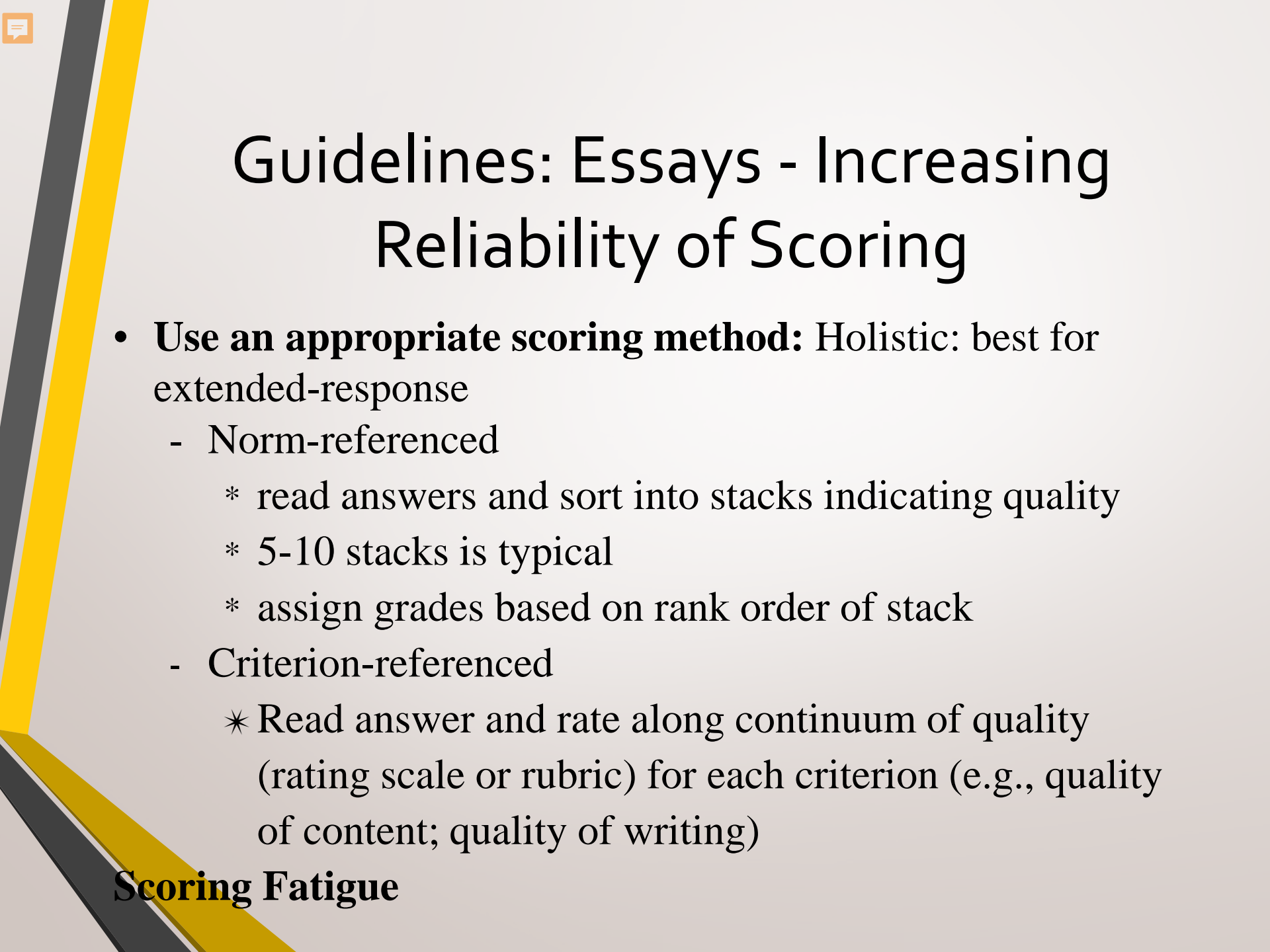
# Guidelines: Planning & Development of Essay Items

- **Specify a clear task: *restricted-response items***
  - appropriate for testing content achievement
  - the question dictates the structure/organization of the answer
  - if you have a specific answer in mind, ask a specific question
- **Ask everyone the same questions**



# Guidelines: Essays - Increasing Reliability of Scoring

- **Use an appropriate scoring method:**
  - Analytic: best for restricted-response
    - \* a checklist of key ideas and point values
    - \* a model answer with key ideas and point values



# Guidelines: Essays - Increasing Reliability of Scoring

- **Use an appropriate scoring method:** Holistic: best for extended-response
  - Norm-referenced
    - \* read answers and sort into stacks indicating quality
    - \* 5-10 stacks is typical
    - \* assign grades based on rank order of stack
  - Criterion-referenced
    - \* Read answer and rate along continuum of quality (rating scale or rubric) for each criterion (e.g., quality of content; quality of writing)

**Scoring Fatigue**

# Guidelines: Essays - Increasing Reliability of Scoring

## **Strategies for increasing reliability:**

- Maintain student anonymity while grading
- Evaluate everyone's answer to one item before grading another
- Minimize bias - keep previously scored items out of sight
- Establish a consistent policy for dealing with irrelevant or incorrect but incidental responses
- Re-read papers (especially the ones read early) to increase consistency of grading



# Questions?

Jason P. Herron

Hubbard Hall Rm 313

[Jason.herron@Wichita.edu](mailto:Jason.herron@Wichita.edu)