


“Essay vs. Multiple-Choice Exams
The Ups and Downs of Each Format”
Mark G. Christensen, Ph.D.



**“Essay vs. Multiple-Choice Exams
The Ups & Downs of Each Format”**

Mark G. Christensen, Ph.D.
National Board of Chiropractic Examiners

Presented at the 2005 CLEAR Annual Conference
September 15, 2005 Phoenix, Arizona

Advantages of Essay Exams

- Relatively easy to write the questions and less time is required for test construction
- Test-takers may express their ideas with fewer restraints and greater reflection
- Test-takers must *supply* answers, not just *select* them, possibly reducing the effects of guessing
- Practical for evaluating small groups of test-takers

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Disadvantages of Essay Exams

- Difficult to score objectively – even if raters are well trained
- Measure only limited aspects of a test-taker’s knowledge
- Labor-intensive & time-consuming to score
- Subject to questions about validity & reliability
- May not actually measure ability to organize and develop ideas

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Advantages of Multiple-Choice

- Greater versatility in measuring rote memory to more complex knowledge
- Large amount of information can be measured in a short time
- Evaluating large numbers of test-takers takes considerably less time
- Scoring is objective and can be reliably completed

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

More M/C Advantages

- Can require test-takers to discriminate among options that vary in correctness
- With 3 – 5 options, the effects of guessing can be reduced
- Particularly amenable to item analysis to detect areas of student weakness, term ambiguity, item difficulty, etc.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Diagnostic Item Analysis

Sample question:
On which bone is the mastoid process located?

Choices	# Who Selected	%	Corr.	Mean	Standard Deviation
A - temporal	672	56.4	0.329	74.1	12.4
B - occipital	405	34.0	-0.190	64.0	11.5
C - zygomatic	57	4.8	-0.193	56.2	11.9
D - parietal	57	4.8	-0.151	58.7	12.1
N - no answer	0	0	0	0	0

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Essay vs. Multiple-Choice Exams
The Ups and Downs of Each Format”
Mark G. Christensen, Ph.D.

Disadvantages of Multiple-Choice

- Questions are time-consuming and difficult to write well.
- Good item writers are rare – it is a special skill that is not easily learned.
- Questions do not evaluate a test-taker’s ability to develop and organize ideas.
- Questions must be pre-tested and reviewed to ensure validity.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Future Possibilities

- Combine multiple-choice and essay testing to capture the advantages of both
- Machine-scored essay exams (using software such as that from IntelliMetric™)

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Speaker Contact Information

Mark G. Christensen, Ph.D.
National Board of Chiropractic Examiners
901 54th Avenue, Greeley, CO 80364
Phone: 970-356-9100 Fax: 970-395-0021
mchristensen@nbce.org
www.nbce.org



Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Essay vs. Multiple-Choice Exams
The Ups and Downs of Each Format”
Mark G. Christensen, Ph.D.

References

- Anastasi, Anne and Susana Urbina. 1997. *Psychological Testing*. New Jersey: Pearson Education.
- Bridgeman, Brent and Rick Morgan. 1996. “Success in College for Students with Discrepancies Between Performance on Multiple-Choice and Essay Tests.” *Journal of Educational Psychology* 88, no. 2: 333-340.
- Davis, B.B. 2001. *Tools for Teaching*. San Francisco, CA: Jossey-Bass.
- Dewey, Russell A. “Writing Multiple Choice Items which Require Comprehension.” Website. <http://www.psychwww.com/selfquiz/aboutq.htm>.
- Miller, Allen H., Bradford W. Imrie, and Kevin Cox. 1998. *Student Assessment in Higher Education: A Handbook for Assessing Performance*. London: Kogan Page.
- Myers, Albert E., Carolyn B. McConville, and William E. Coffman. 1966. “Simplex Structure in the Grading of Essay Tests.” *Educational and Psychological Measurement* 26: 41-54.
- Oosterhof, M. 2001. *Classroom Applications of Educational Measurement*. Upper Saddle River, NJ: Merrill Prentice Hall, Inc.
- Payne, David A. 1997. *Applied Educational Assessment*. Belmont, CA: Wadsworth Publishing Company.
- Sax, Gilbert. 1997. *Principles of Educational and Psychological Measurement and Evaluation*. Belmont, CA: Wadsworth Publishing Company.
- Tozoglu, Dogan, Musavver D. Tozoglu, Ahmet Gurses, and Cetin Dogar. 2004. “The Students’ Perceptions: Essay Versus Multiple-Choice Type Exams.” *Journal of Baltic Science Education* 6, no. 2: 52-59.
- Tuckman, B.W. 1993. “The Essay Test: A Look at the Advantages and Disadvantages.” *Nassp Bulletin* 77, no 55: 20-26.

“Presentation Title”

Speaker Name



Developing MCQs that Test Higher Order Thinking Skills

George Gray

ACT, Inc.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

or

Creating Sterling Silver Multiple Choice Items: an Alchemist's Approach

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, AZ

“Presentation Title”

Speaker Name

Recall or Higher Level Concept

- How was the content learned by the examinee?

How many arteries and veins are present in the umbilical cord?

- A. 1 artery, 1 vein
- B. 1 artery, 2 veins
- C. 2 arteries, 1 vein
- D. 2 arteries, 2 veins

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Recommendations

- Use a practice-based scenario that requires a decision.
or
- Base the item on a visual.
or
- Base the item on the understanding of a calculation.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Presentation Title”

Speaker Name

Practice-based Scenario

Which of the following children would need to be evaluated immediately in an emergency department?

- A. A child who presents with a loud bark-like cough without swallowing problems
- B. A child who presents with a fever of 102°F for the past 24 hours
- C. A child who has green watery-type stools for the past 24 hours
- D. A child who has a low blood pressure, a pulse of 60 beats per minute and petechiae on the face and trunk

©Commission on Graduates of Foreign Nursing Schools

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Real Situations vs. Window Dressing

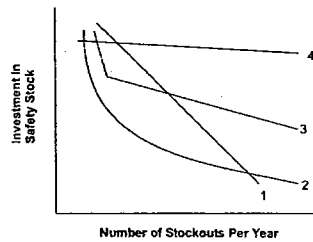
- How much information should be included in the stem?
- Does description of a “real” situation add window dressing?

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Presentation Title”

Speaker Name

Use a Visual



Which of the curves in the graph above shows the relationship between investment in safety stock and the number of stockouts per year?

- A. 1
- B. 2
- C. 3
- D. 4

©APICS - The Association for Operations Management

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Understanding of Calculations

The volume of a gas is 8 liters. Both the pressure on the gas and the absolute temperature are doubled. What is the new volume of the gas?

- A. 16 liters
- B. 12 liters
- C. 8 liters
- D. 4 liters

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Presentation Title”

Speaker Name

A company markets a family of products. Customer orders are filled by assembling and then decorating base units. The decorations vary from simple to complex, some of which require special tooling. Which of the following manufacturing process combinations is most appropriate for the manufacturer?

- A. Line for assembly and batch for decorating.
- B. Batch for assembly and line for decorating.
- C. Line for both assembly and decorating.
- D. Batch for both assembly and decorating.

© APICS - The Association for Operations Management

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Speaker Contact Information

George Gray
ACT, Inc.
101 ACT Drive
Iowa City, IA 52243
Phone: 319-337-1168, Fax: 319-341-2551
E-mail: george.gray@act.org
Website: www.act.org

ACTTM
INFORMATION FOR LIFE'S TRANSITIONS

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, AZ



“Re-thinking the Shunned Multiple Choice Item Types”

Steve Nettles, EdD
Applied Measurement Professionals, Inc.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

MC Item Types

- Positively Worded Stems
 - What is $2 + 2$?
- Negatively Worded Stems
 - All of the following are true EXCEPT
 - Which of the following is NOT ...
- Scenario
- Complex Multiple Choice
 - K-Type
 - Generic

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Shunned” Items

- CMC
 - Appear to be more difficult
 - Candidates can use partial knowledge to answer the items
 - Have lower discrimination
 - Difficult to construct and edit
 - Take up more space on the page
 - Requires more reading time

Haladyna (1994). *Developing and Validating MC Test Items*

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Shunned” Items

- Negatively Worded Stems
 - Use of negative words have negative effects on candidates and their responses to items
 - Sometimes difficult to understand the meaning of negatively phrased items
 - Candidates may read through the *not* and forget to reverse the logic of the relationship being tested

Haladyna (1994). *Developing and Validating MC Test Items*

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

K-Type

A child suffering from an acute exacerbation of rheumatic fever usually has

1. An elongated sedimentation rate
 2. A prolonged PR interval
 3. An elevated antistreptolysin O titer
 4. Subcutaneous nodules
- A. 1, 2, and 3 only
B. 1, 3 only
C. 2, 4 only
D. 4 only
E. All are correct

(From: Constructing Written Test Questions for the Basic and Clinical Sciences, Case & Swanson, 1996, NBME, Philadelphia, PA)

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Generic CMC

Which of the following is considered a descriptive statistic?

1. mean
 2. F-ratio
 3. median
 4. correlation
- A. 1 and 2 only
B. 1 and 3 only
C. 2 and 4 only
D. 3 and 4 only

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Scenario

A cactus removal service provides the following data for the month of September:

Week	No. of Jobs	Weekly Revenue
1	25	\$10,800
2	20	\$8,000
3	45	\$45,700
4	19	\$9,500

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Scenario Items

What was the total revenue for the month of September?

- A. \$18,500
- B. \$74,000**
- C. \$96,000
- D. \$108,000

Which week was the MOST productive in terms of amount of money per job?

- A. 1
- B. 2
- C. 3**
- D. 4

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Four Groups

- Three tests are used for certification; one used for licensing;
- Groups have item statistics based on approximately 200- 4000 candidates.
- Classical item statistics were calculated for each group for each item type
 - Item difficulty (p-value)
 - Item discrimination (point biserial correlation)

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Group A (N varies from 200-400)

- Positive Stem (N=220)
 - Mean P = .80; mean PB = .22
- Negative Stem (N=21)
 - Mean P = .87; mean Pb = .23
- CMC Stem (N=109)
 - Mean = .81; mean PB = .22
- Scenario Stem (N=19)
 - Mean = .75; mean PB = .22

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Group B (N varies from 400-1700)

- Positive Stem (N=411)
 - Mean P = .77; mean PB = .22
- Negative Stem (N=26)
 - Mean P = .74; mean Pb = .22
- CMC Stem (N=103)
 - Mean = .76; mean PB = .21
- Scenario Stem (N=17)
 - Mean = .76; mean PB = .27

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Group C (Mean N 2000-4000)

- Positive Stem (N=168)
 - Mean P = .70; mean PB = .27
- Negative Stem (N=9)
 - Mean P = .76; mean Pb = .27
- CMC Stem (N=17)
 - Mean = .72; mean PB = .23

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Group D
(N varies from 200-1400)

- Positive Stem (N=380)
 - Mean P = .75; mean PB = .20
- Negative Stem (N=10)
 - Mean P = .90; mean PB = .14
- CMC Stem (N=114)
 - Mean = .69; mean PB = .16

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Conclusions

- Positively worded stems always the safest approach
- Scenario based items may be superior for some groups
- CMC items present mixed results
- Negatively worded stems should be avoided

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Speaker Contact Information

Steven Nettles, EdD
Vice-President, Research and Development
Applied Measurement Professionals, Inc.
8310 Nieman Road,
Lenexa, KS, 66214
phone 913-495-4405
snettlles@goamp.com
<http://www.goamp.com>



Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Test Wiseness: How to Avoid Common Pitfalls in Item Writing”

Carol Morrison Featherman, PhD



Test Wiseness: How to Avoid Common Pitfalls in Item Writing

Carol Morrison Featherman, PhD
National Board of Medical Examiners

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

A special thank you to Dave Swanson, Deputy VP of Professional Services at the NBME, for providing insights and examples cultivated from his many years of working with test committees

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Flaws Related to Testwiseness

- Flaws related to testwiseness make it easier for some examinees to answer the question correctly, based on their test taking skills alone
- Probability of answering a question correctly should relate to the examinee's level of expertise on the topic being assessed rather than his/her test taking-strategies

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Grammatical Cues

- One or more distractors don't follow grammatically from the stem

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Grammatical Cues - Example

A 60-year-old man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. After ascertaining that the airway is open, the first step in management should be intravenous administration of

- A. examination of cerebrospinal fluid
- B. glucose with vitamin B₁ (thiamine)
- C. CT scan of the head
- D. phenytoin
- E. diazepam

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Logical Cues

- A subset of the options are collectively exhaustive

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Test Wiseness: How to Avoid Common Pitfalls in Item Writing”

Carol Morrison Featherman, PhD

Logical Cues - Example

Crime is

- A. equally distributed among the social classes
- B. overrepresented among the poor
- C. overrepresented among the middle class and rich
- D. primarily an indication of psychosexual maladjustment
- E. reaching a plateau of tolerability for the nation

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Absolute Terms

- Absolute terms such as “always”, “never”, “totally”, “absolutely”, and “completely” are so extreme that they are seldom the correct answer

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Absolute Terms - Example

In patients with advanced dementia, Alzheimer’s type, the memory defect

- A. can be treated adequately with phosphatidylcholine (lecithin)
- B. could be a sequela of early parkinsonism
- C. is never seen in patients with neurofibrillary tangles at autopsy
- D. is never severe
- E. possibly involves the cholinergic system

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Long Correct Answer

- The correct answer is noticeably longer than the other options

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Long Correct Answer - Example

Secondary gain is

- A. synonymous with malingering
- B. a frequent problem with obsessive-compulsive disorder
- C. a complication of a variety of illnesses and tends to prolong many (>3) of them
- D. never seen in organic brain damage

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Word Repeats

- A word or phrase is included in the stem and in the correct answer

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Test Wiseness: How to Avoid Common Pitfalls in Item Writing”

Carol Morrison Featherman, PhD

Word Repeats - Example

A 58-year-old man with a history of heavy alcohol use and previous psychiatric hospitalization is confused and agitated. He speaks of experiencing the world as unreal. This symptom is called

- A. depersonalization
- B. derailment
- C. derealization
- D. focal memory deficit
- E. signal anxiety

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Convergence Strategy

- The correct answer includes the most elements in common with the other options

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Convergence Strategy - Example

Local anesthetics are most effective in the

- A. anionic form, acting from inside the nerve membrane
- B. cationic form, acting from inside the nerve membrane
- C. cationic form, acting from outside the nerve membrane
- D. uncharged form, acting from inside the nerve membrane
- E. uncharged form, acting from outside the nerve membrane

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Convergence Strategy - Example

Local anesthetics are most effective in the

- A. anionic form, acting from inside the nerve membrane
- B. cationic form, acting from inside the nerve membrane
- C. cationic form, acting from outside the nerve membrane
- D. uncharged form, acting from inside the nerve membrane
- E. uncharged form, acting from outside the nerve membrane

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Convergence Strategy - Example

Local anesthetics are most effective in the

- A. anionic form, acting from inside the nerve membrane
- B. cationic form, acting from inside the nerve membrane
- C. cationic form, acting from outside the nerve membrane
- D. uncharged form, acting from inside the nerve membrane
- E. uncharged form, acting from outside the nerve membrane

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Convergence Strategy - Example

Local anesthetics are most effective in the

- A. anionic form, acting from inside the nerve membrane
- B. cationic form, acting from inside the nerve membrane
- C. cationic form, acting from outside the nerve membrane
- D. uncharged form, acting from inside the nerve membrane
- E. uncharged form, acting from outside the nerve membrane

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Test Wiseness: How to Avoid Common Pitfalls in Item Writing”

Carol Morrison Featherman, PhD

Issues Related to Irrelevant Difficulty

- Some characteristics of items that are unrelated to the content being assessed make them more difficult than they should be
- Irrelevant difficulty contributes to measurement error

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Options are Long, Complicated or Double - Example

Peer review committees in HMOs may move to take action against physician's credentials to care for participants of the HMO. There is an associated requirement to assure that the physician receives due process in the course of these activities. Due process must include which of the following?

- A. Notice, an impartial forum, council, a chance to hear and confront evidence against him/her.
- B. Proper notice, a tribunal empowered to make the decision, a chance to confront witnesses against him/her, & a chance to present evidence in defense.
- C. Reasonable and timely notice, impartial panel empowered to make a decision, a chance to hear evidence against himself and confront witnesses, and the ability to present evidence in defense.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Numeric Data are Not Stated Consistently - Example

Following a second episode of salpingitis, what is the likelihood that a woman is infertile?

- A. Less than 20%
- B. 20 to 30%
- C. Greater than 50%
- D. 90%
- E. 75%

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Frequency Terms in the Options are Vague - Example

Severe obesity in early adolescence

- A. usually responds dramatically to dietary regimens
- B. often is related to endocrine disorders
- C. has a 75% chance of clearing spontaneously
- D. shows a poor prognosis
- E. usually responds to pharmacotherapy and intensive psychotherapy

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Related Answers

- The answer to one question should not depend on the answer to another question; items should not cue other items
- Item response theory models assume local independence – examinee's performance on one item on a test does not affect his/her responses to any other items on the test

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Summary of Issues Related to Testwiseness

- **Grammatical cues:** one or more distractors don't follow grammatically from the stem
- **Logical cues:** a subset of the options is collectively exhaustive
- **Absolute terms:** terms such as "always" or "never" are in some options
- **Long correct answer:** correct answer is longer, more specific, or more complete than other options
- **Word repeats:** a word or phrase is included in the stem and in the correct answer
- **Convergence strategy:** the correct answer includes the most elements in common with the other options

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

“Test Wiseness: How to Avoid Common Pitfalls in Item Writing”

Carol Morrison Featherman, PhD

Summary of Issues Related to Irrelevant Difficulty

- Options are long, complicated, or double
- Numerical data are not stated consistently
- Terms in the options are vague (e.g., “rarely,” “usually”)
- The answer to an item is related to the answer of another item

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

References

- Case, S.M. & Swanson, D.B. (2002). *Constructing written test questions for the basic and clinical sciences, (3rd ed. revised)*. Philadelphia: National Board of Medical Examiners.
<http://www.nbme.org/about/itemwriting2.asp>
- Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum.

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona

Speaker Contact Information

Carol Morrison Featherman, PhD
National Board of Medical Examiners
3750 Market Street
Phone: 215-590-9745, Fax: 215-590-9453
cmorrison@nbme.org
<http://www.nbme.org/>

Presented at the 2005 CLEAR Annual Conference
September 15-17 Phoenix, Arizona