(Re) Writing strong Multiple-Choice Questions (MCQs)

Sarah Jacobs, M.Ed.

OHSU Teaching and Learning Center

Objectives

At the end of this session, participants will be able to:

- Understand the anatomy of an effective MCQ
- Critique sample MCQs
- Apply reliability and validity measurements to rewriting MCQs

Which of the following is most important to you?

- A. Revising existing MCQs
- B. Writing new MCQs
- C. General knowledge about test statistics and question-writing
- D. All of the above
- E. Something else



Anatomy of an MCQ

Anatomy of a Multiple-choice Question

Distractors

Patients with congenital adrenal hyperplasia present with excessive circulating levels of

STEM

- a) ACTH ← Key
 b) Aldosterone (correct answer)
- c) BAM22 _
- d) Cortisol <

e) CXCR7

OPTIONS OR ALTERNATIVES

Reliability and Validity









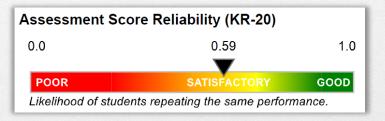
Reliable, Not Valid

Both Reliable & Valid

Image from: https://commons.wikimedia.org/wiki/File:Reliability_and_validity.svg

Reliability and Validity – how are they measured?

- Reliability is measured by:
 - Overall exam KR 20 score
 - Individual questions Point biserial
- Validity is a little more complicated



	Diff(p)	Upper	Lower	Disc. Index	Point Biserial ≥ .30 = Very Good .2029 = Good .0919 = Marginal <.09 = Poor		
Item #	≥75 = Easy ≤25 = Difficult	% of	lowest % of students	Closer to 1 the better			
1	0.85	85.71%	95.45%	-0.1	-0.01		

Correcting Item-Writing Flaws

(Ali & Ruit, 2015)



Functioning Distractors

>5% selection frequency

Functioning distractors discriminate the good test-taker from someone who knows the content.



Point Biserial (Item Reliability)

Item Analysis – what to review?

- ✓ Question Performance/Difficulty/Diff(p) < 0.70
- ✓ Discrim < 0.25
- ✓ Significant # students answered the same incorrect answer
- ✓ Point biserial < 0.2
- ✓ Student queries

What would you flag for review?

Seq #	Correct	Upper 27%	Lower 27%	Point BiSerial	Disc Index	Response Frequencies				
	1					A	В	С	D	E
9	66%	93%	37%	0.49	0.56	7	21	*102	24	
10	91%	100%	81%	0.18	0.19	*140	12	2	0	
11	99%	100%	95%	0.16	0.05	0	*152	1	1	
12	95%	100%	86%	0.25	0.14	4	*146	1	3	
13 📗	43%	56%	26%	0.25	0.30	80	8	0	*66	
33	25%	49%	9%	0.34	0.40	*39	7	101	7	
34	83%	98%	63%	0.47	0.35	19	*128	4	3	
35	99%	98%	100%	-0.09	-0.02	*49	*103	2	0	
36	95%	100%	93%	0.11	0.07	8	*86	*49	*11	
37	98%	100%	93%	0.18	0.07	*151	1	1	1	
38	42%	64%	35%	0.24	0.30	1	24	64	*65	

Technical Item Flaws and Testwiseness

(Case & Swanson, 2001)

 Students can answer questions based on testtaking skills alone



Testwiseness: Grammatical Cues

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

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

Testwiseness: Logical Cues

A subset of the options are collectively exhausted

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

Testwiseness: Absolute Terms

Terms such as "always" or "never" are used in the options

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

Testwiseness: Long Correct Answer

Correct answer is longer, more specific, or more complete than other options

Secondary gain is

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

Testwiseness: Word repeats

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

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

Testwiseness: Word repeats

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

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

Testwiseness: Convergence Strategy

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

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

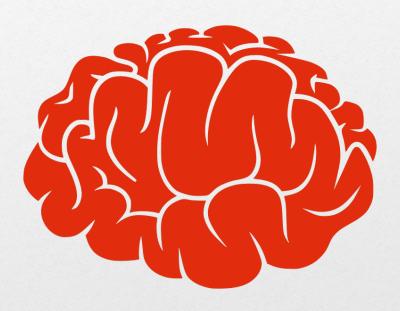
Testwiseness: Convergence Strategy

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

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

Activity: Name that strategy!



What strategy can be used to answer this question?

During the comprehensive periodontal evaluation of a new patient you note multiple sites in all four quadrants that probe 6-8 mm, have moderate interproximal bone loss, subgingival calculus, and demonstrate bleeding on probing. A review of the patient's health history reveals the patient has type 2 diabetes mellitus. The patient is taking oral medications for diabetes and high cholesterol.

Based on the above case scenario, please construct a proper periodontal treatment plan by selecting the best treatment option for the problem.

Subgingival calculus:

- A. Medical consultation
- B. Oral hygiene instructions
- C. Scaling and root planning
- D. Reevaluation
- E. Referral to a periodontist (CORRECT ANSWER)

What strategy can be used to answer this question?

During the comprehensive periodontal evaluation of a new patient you note multiple sites in all four quadrants that probe 6-8 mm, have moderate interproximal bone loss, subgingival calculus, and demonstrate bleeding on probing. A review of the patient's health history reveals the patient has type 2 diabetes mellitus. The patient is taking oral medications for diabetes and high cholesterol.

Based on the above case scenario, please construct a proper **periodontal** treatment plan by selecting the best treatment option for the problem.

Subgingival calculus:

- A. Medical consultation
- B. Oral hygiene instructions
- C. Scaling and root planning
- D. Reevaluation
- E. Referral to a **periodontist**

Word repeat between stem and correct answer.

Technical Item Flaws and Irrelevant Difficulty

(Case & Swanson, 2001)

 Make the question difficult for reasons irrelevant to the focus of the assessment



Irrelevant Difficulty: Options long, complicated or double

Stem contains extraneous reading

Options long and/or complicated

Peer review committees in HMOs may move to take action against a 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, and 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/herself and to confront witnesses, and the ability to present evidence in defense.

Irrelevant Difficulty:
Numeric data not
stated consistently

Numeric options should be listed in a single format

Following a second episode of infection, 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%

Irrelevant Difficulty: Frequency Terms are Vague

Frequency terms used in the options are vague, such as rarely, usually

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

Irrelevant Difficulty:
Language not
parallel, options in
non-logical order

Options are long and the language makes it difficult to determine which is the most correct.

In a vaccine trial, 200 2-year-old boys were given a vaccine against a certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. Which of the following statements concerning these results is correct?

- A. No conclusion can be drawn, since no follow-up was made of nonvaccinated children
- B. The number of cases (ie, 30 cases over five years) is too small for statistically meaningful conclusions
- C. No conclusions can be drawn because the trial involved only boys
- D. Vaccine efficacy (%) is calculated as 85-15/100

Original

In a vaccine trial, 200 2-year-old boys were given a vaccine against a certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. Which of the following statements concerning these results is correct?

- A. No conclusion can be drawn, since no follow-up was made of nonvaccinated children
- B. The number of cases (ie, 30 cases over five years) is too small for statistically meaningful conclusions
- C. No conclusions can be drawn because the trial involved only boys
- D. Vaccine efficacy (%) is calculated as 85-15/100

Suggested rewrite

In a vaccine trial, 200 2-year-old boys were given a vaccine against a certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. For which of the following reasons can no conclusion be drawn from these data?

- A. No follow-up was made of nonvaccinated children
- B. The number of cases was too small
- C. The trial involved only boys
- D. [Insert new option]

Irrelevant Difficulty: None of the above

Can be problematic when options are not absolutely true or false

The diagnosis of a large ovarian cyst is most strongly suggested by

A. an anterior dullness, lateral tympany

B. a decreased peristalsis

C. a fluid wave

D. a shifting dullness

E. none of the above

Irrelevant Difficulty:
Stems are tricky or
unnecessarily
complicated

Arrange the parents of the following children with Down's syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. The karyotypes of the daughters are:

I: 46, XX, -14, +T (14q21q) pat

II: 46, XX, -14, +T (14q21q) de novo

III: 46, XX, -14, +T (14q21q) mat

IV: 46, XX, -21, +T (14q21q) pat

V: 47, XX, -21, +T (21q21q) (parents not karyotyped)

A. III, IV, I, V, II

B. IV, III, V, I, II

C. III, I, IV, V, II

D. IV, III, I, V, II

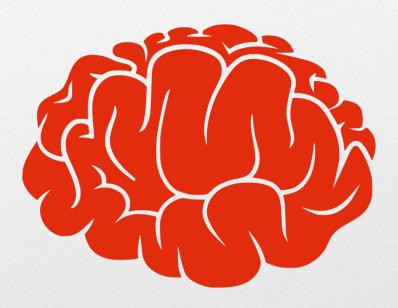
E. III, IV, I, II, V

Writing MCQs: What to Avoid

- Testwiseness:
 - Logical cues
 - Absolute terms
 - Long correct answer
 - Word repeats
 - Convergence strategy

- Irrelevant Difficulty:
 - Long, complicated options and/or stems
 - Inconsistent data and/or language
 - Vague terms
 - Options are in a nonlogical order
 - "None of the above" is used as an option
 - The answer to an item is "hinged" to the answer of a related item

Activity: What is wrong with this question?

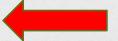


Activity: What is wrong with this question?

Who received a Nobel Prize for discovering the structure of DNA?

- A. Francis Crick
- B. James Watson
- C. Rosalind Franklin
- D. A and B
- E. B and C





Unnecessarily complicated answer choices.

Activity: What is wrong with this question?

How many chromosomes are found in a typical human cell?

A. 12

B. 18

C. 32

D. 46

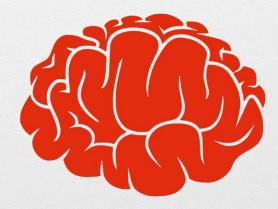
E. 54

Answers overlap: A cell that has 46 chromosomes also has 32, 18 and 12 chromosomes, though these aren't the complete count.

Activity: Critique Questions

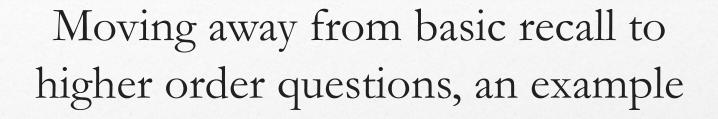
Using the activity handout:

- 1. Get into groups 2-4
- 2. You will be assigned a question number
- 3. Review statistics
- 4. Review for test-writing flaws
- 5. Suggest rewrite
- 6. Share results!



How to Write an MCQ

- Test main concepts, not trivia
- Use your objectives!
- Pose a clear question
- "Cover the options" rule
- Avoid item-writing flaws
- Answers should be homogenous
- Move away from basic recall and toward higher order thinking



Basic Science Recall:

What area is supplied with blood by the posterior inferior cerebellar artery?

Basic Science Application of Knowledge: A 62-year-old man develops left-sided limb ataxia, Horner's syndrome, nystagmus, and loss of appreciation of facial pain and temperature sensations. What artery is most likely to be occluded?

Objectives

At the end of this session, participants will be able to:

- Understand the anatomy of an effective MCQ
- Critique sample MCQs
- Apply reliability and validity measurements to rewriting MCQs

Questions?



References

- Ali, S. H., & Ruit, K. G. (2015). The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspect Med Educ Perspectives on Medical Education*, 4(5), 244-251. doi:10.1007/s40037-015-0212-x
- Brame, C., (2013) Writing good multiple choice test questions. Retrieved August 22, 2016 from https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/.
- Case, S. M., & Swanson, D. B. (2001). Constructing written test questions for the basic and clinical sciences. Philadelphia: National Board of Medical Examiners.
- Phelan, C., & Wren, J. (2005). Exploring Reliability In Academic Assessment. Retrieved August 23, 2016, from https://www.uni.edu/chfasoa/reliabilityandvalidity.htm