NATIONAL ATHLETIC TRAINERS' ASSOCIATION
HEALTH CARE FOR LIFE & SPORT

WORLD FEDERATION OF ATHLETIC TRAINING & THERAPY

EducATionalist COMMUNITY

Introduction to Item Writing and Using Microsoft Forms for e-Assessment in AT Education

ITHACA ATHLETIC TRAINING

Paul R. Geisler, EdD, ATC
Professor & Director, AT Education
Ithaca College, NY, USA

Tuesday, August 4 2020
1:00 – 2:00 p.m., EST (New York, USA)
Registration Required

Originally from Lakeville, Massachusetts with lived spells in OH, DC, VA, NC, FL, GA & NY, Paul's been a Certified Athletic Trainer (USA) since 1987. He has 15 yrs. varied clinical experience & 22 yrs. experience as an AT educator/administrator at 2 different institutions.

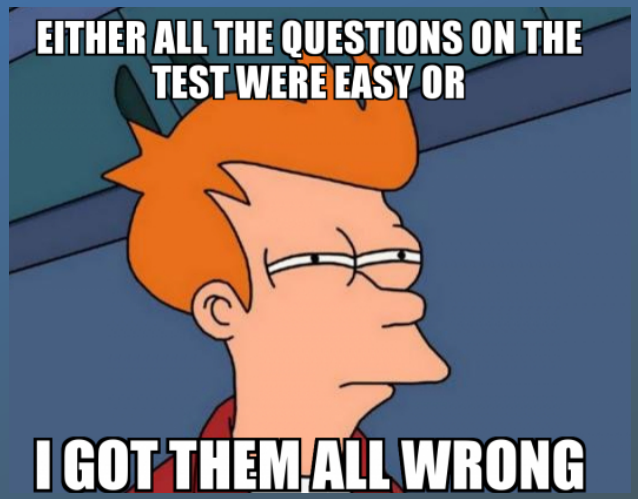
He's deeply interested in best practices in health professions education, clinical reasoning, the development of capability, and authentic assessment.

ITHACA ATHLETIC TRAINING

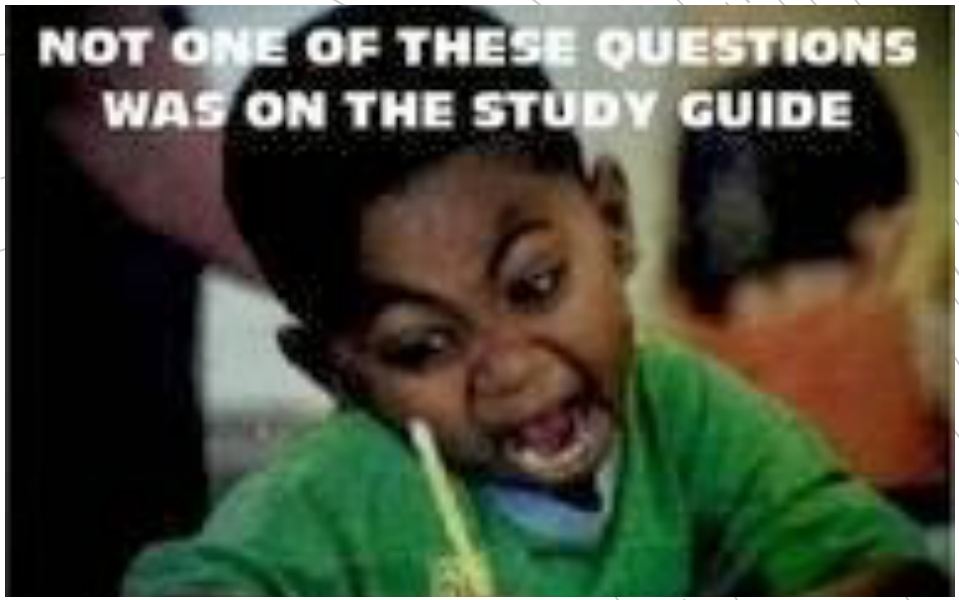
Disclosures & Conflicts of Interest

The content and information presented herein are free from personal conflicts of interest and nothing covered today represents the views of the WFATT, BASRaT or the NATA.

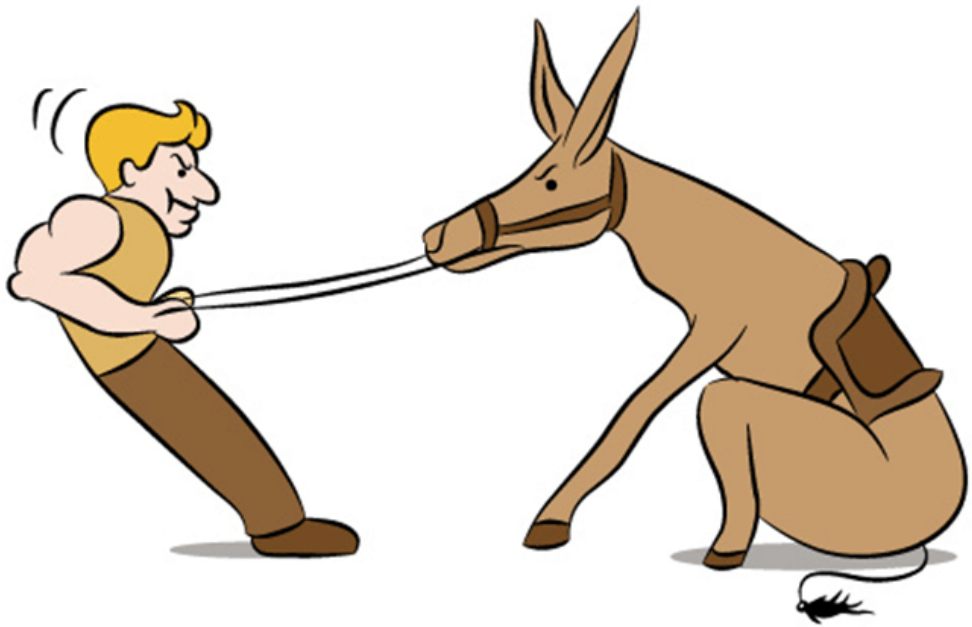
Today's Learning Objectives



1. Better understand the parts, principles and mechanics of *Item Writing*
2. Realize the value and utility of well constructed MCQs for assessment
3. Appreciate that higher level thinking, reasoning and application *can* be assessed with sound MCQs
4. Understand basics of test Psychometrics
5. Be able to use *MS Forms™* to administer high quality assessments in the context of AT Education

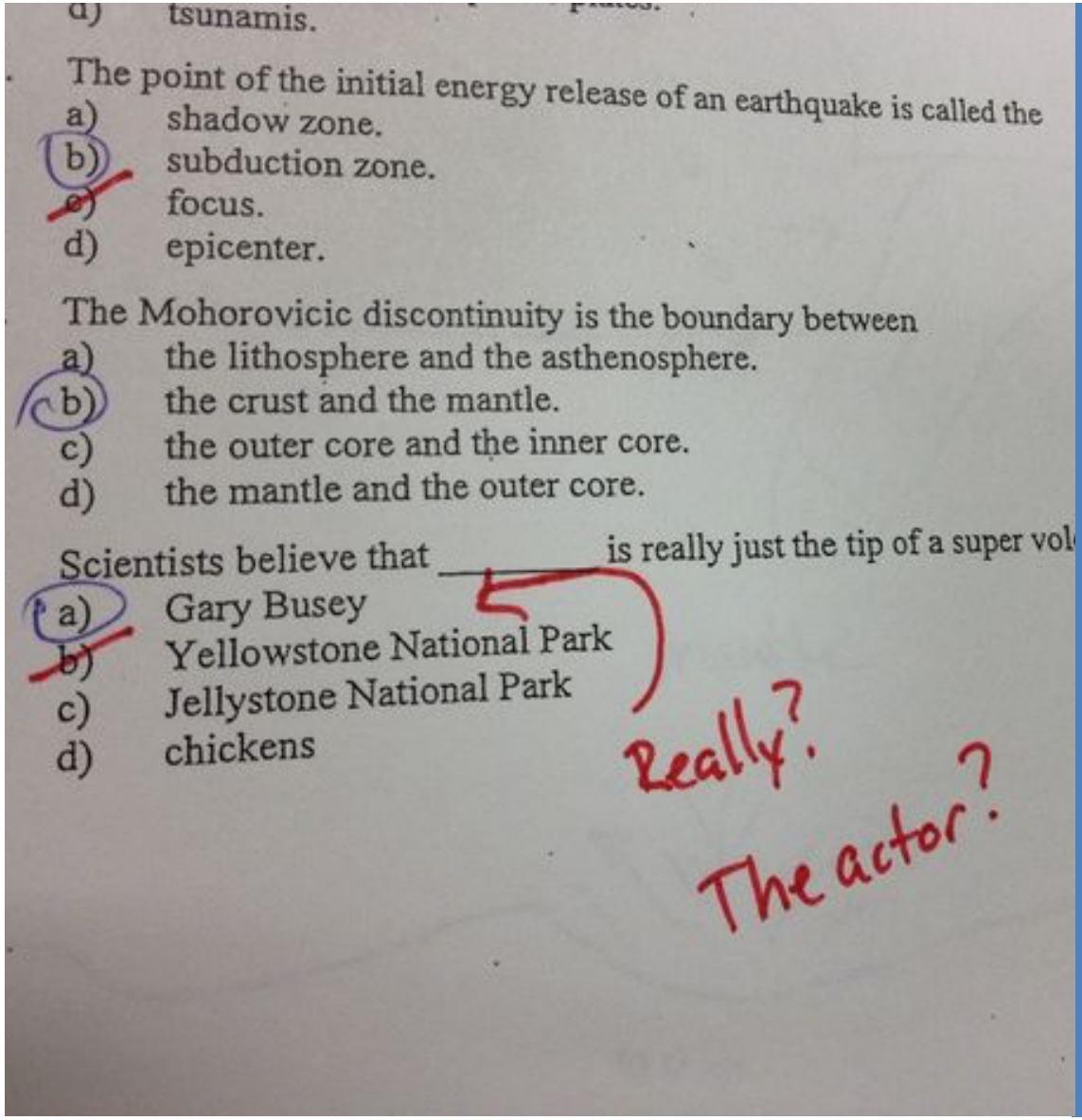


Quick Strike Poll #1
(3 ?s)



- Prioritize clinical reasoning and transferability based pedagogy/assessment
- Avoided MCQs
  - "Can't assess CR & CDM!"
  - "Exams take too much time to make"
- Long, open response, problem-based exams
- Lamented graded time

Personal Backdrop



RL Ebel

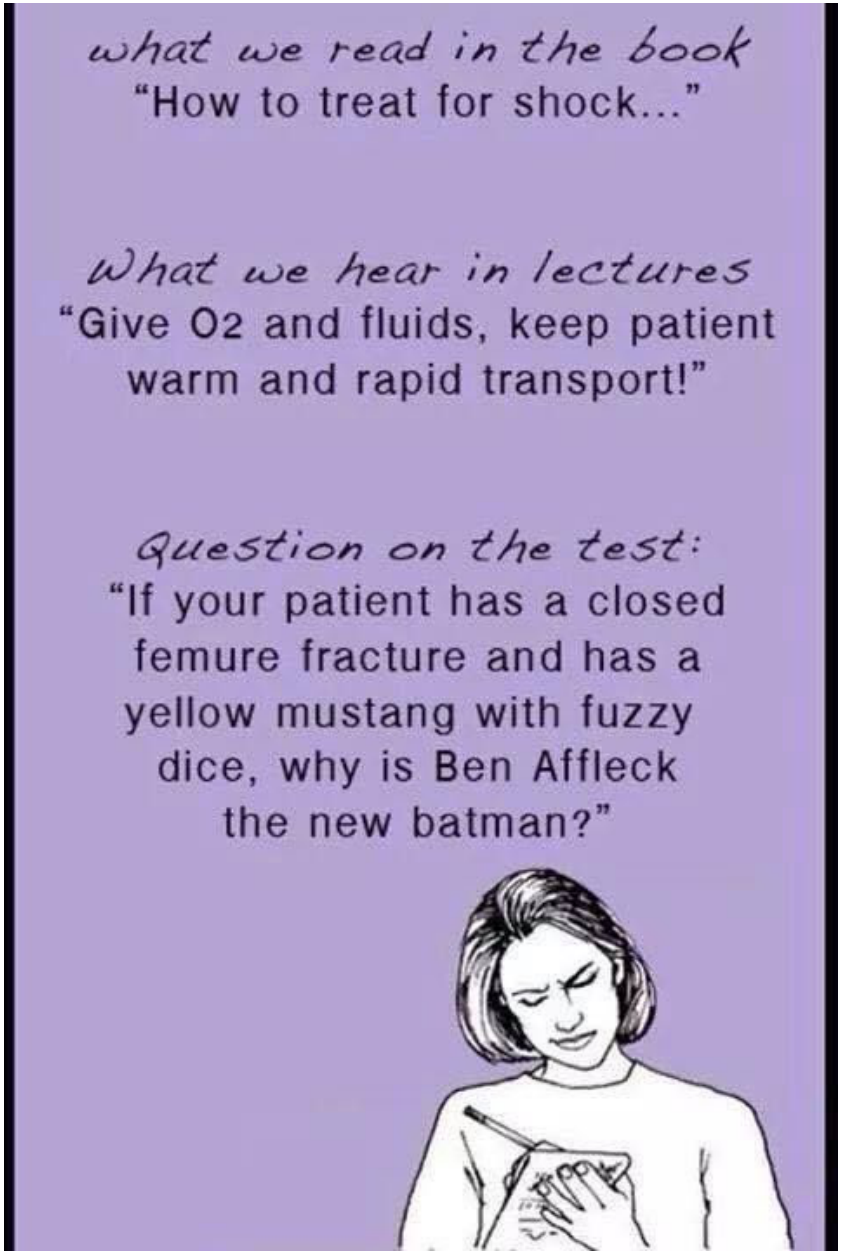
(1951, p. 185)

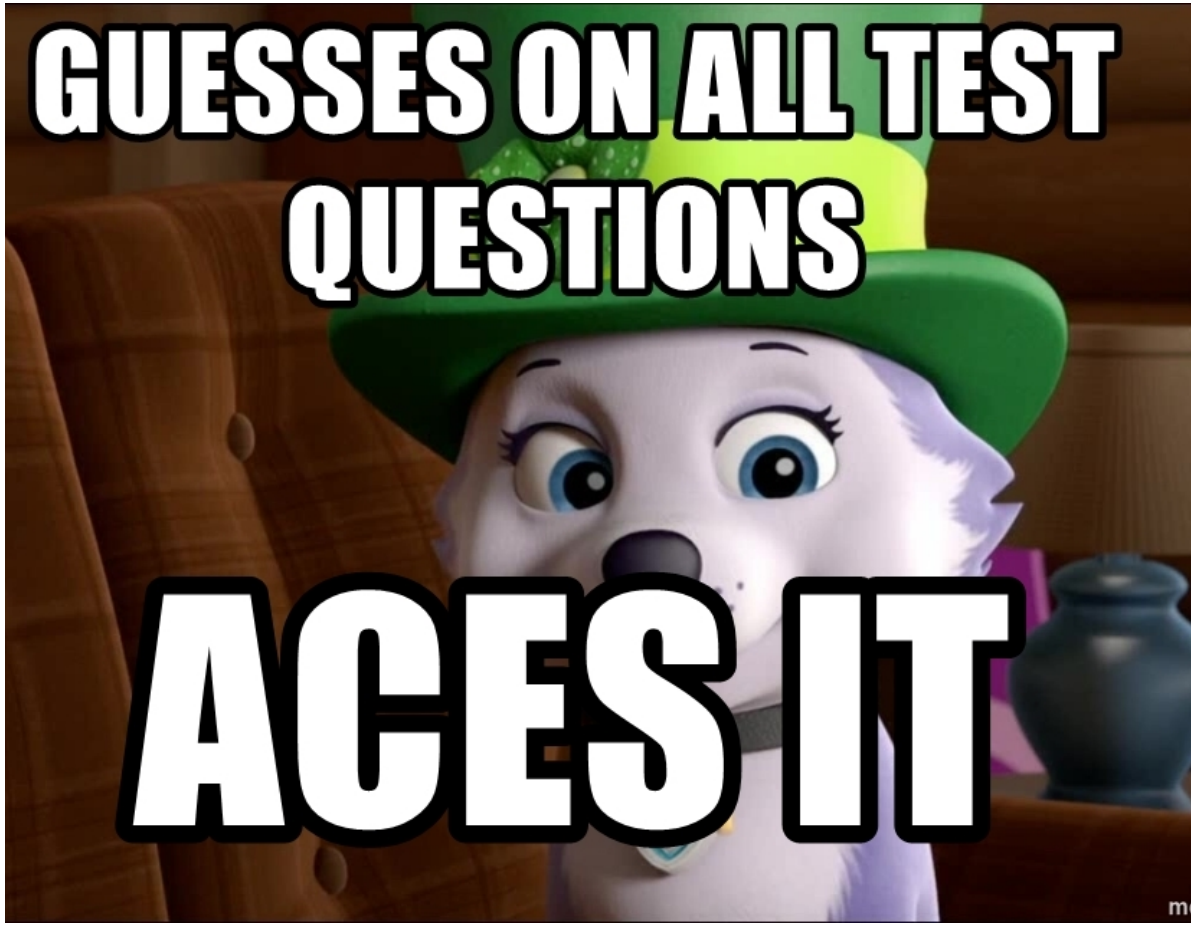
"Item writing is an art. It requires an uncommon combination of special abilities. It is mastered only through extensive and critically supervised practice. It demands, and tends to develop, high standards of quality and a sense of pride in craftsmanship."

Ebel, R. L. (1951). Writing the test item. In E. F. Linquist (Ed.), Educational measurement (pp. 185–249). Washington, DC: American Council on Education.

Item Writing

- **Content Validity** = Most Important Element
  - Does test/question measure AND sample relevant learning objectives or outcomes?
  - "Increase the signal, decrease the noise"
- **Construct Validity** = Does test/question measure an underlying cognitive trait?
  - e.g., clinical decision-making or reasoning


what we read in the book
"How to treat for shock..."
What we hear in lectures
"Give O2 and fluids, keep patient warm and rapid transport!"
Question on the test:
"If your patient has a closed femure fracture and has a yellow mustang with fuzzy dice, why is Ben Affleck the new batman?"

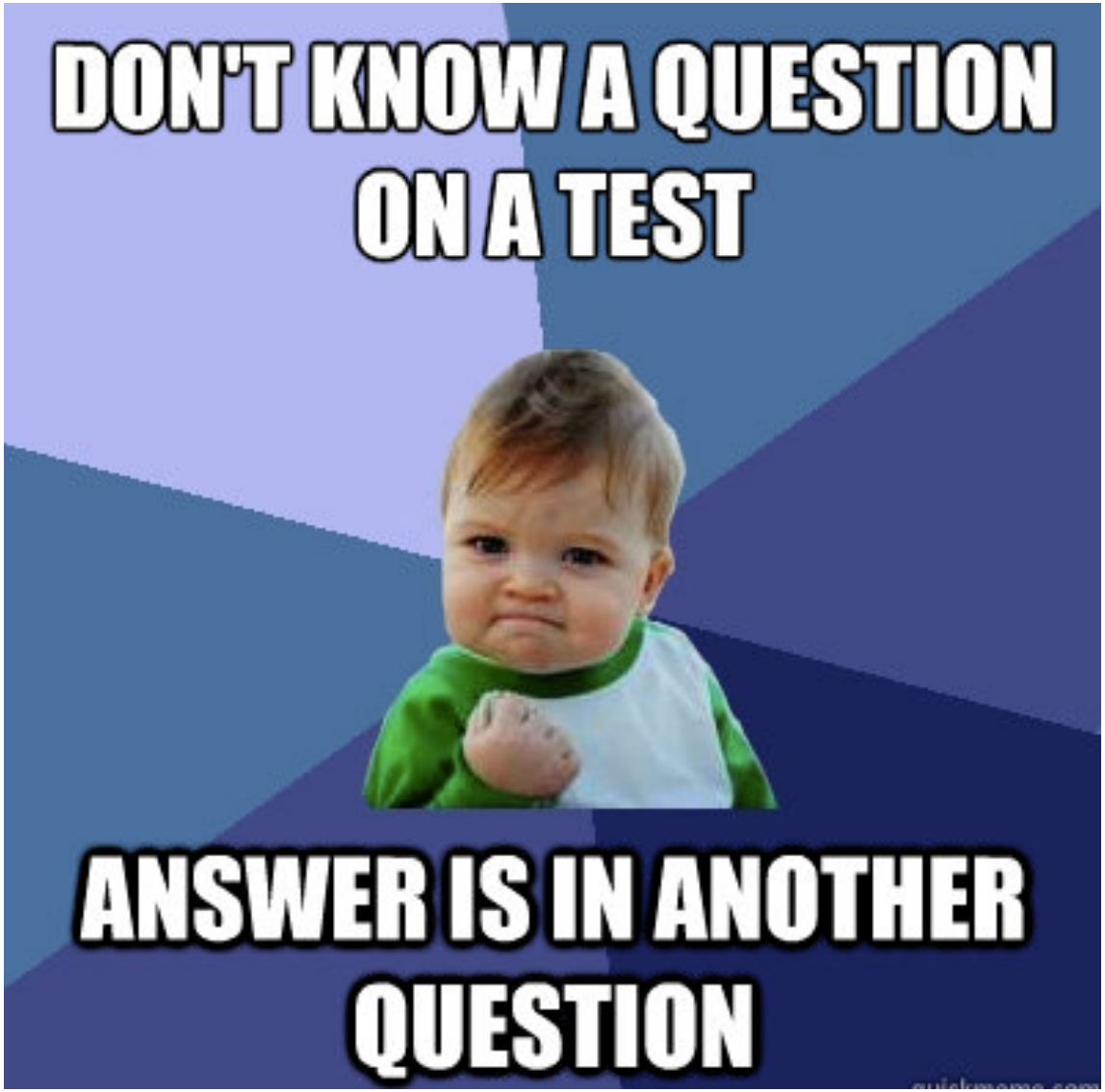


Item Writing

Discrimination allows you to distinguish students' knowledge levels:

- Good test takers vs. high performers
- Guessers vs. knowers
- If best performing students get question "x" incorrect, question "x" is probably poorly written or conceived
- If poor performing students get question "y" correct more than higher performing students, question "y" is poorly written or conceived

Quick Strike Poll #2 (3 ?s)



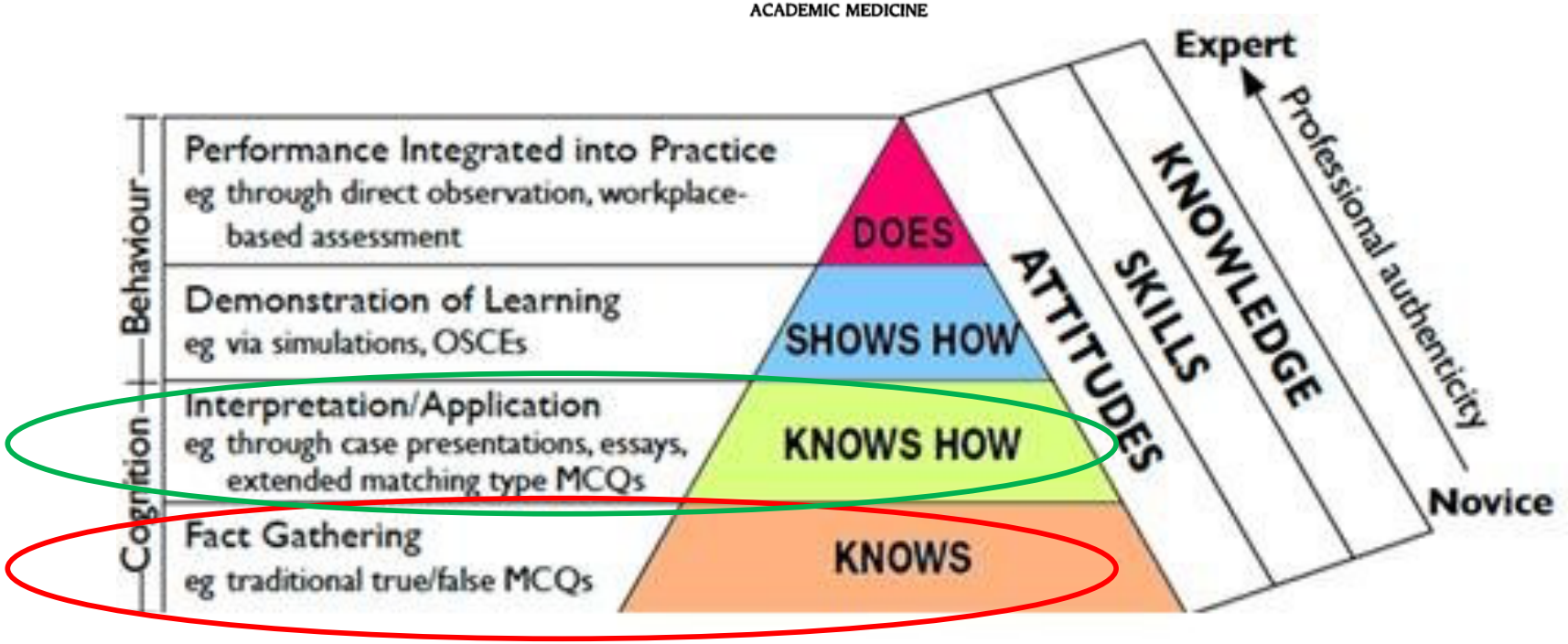
Invited Reviews

The Assessment of Clinical Skills/Competence/Performance

GEORGE E. MILLER, M.D.

Volume 65 • Number 9 • SEPTEMBER SUPPLEMENT 1990

ACADEMIC MEDICINE



"The merely well-informed man is the most useless bore on God's earth"
Alfred North Whitehead

"It is this quality of being functionally adequate, or of having sufficient knowledge, judgment, skill or strength for a particular duty that Webster defines as *competence*" George Miller, MD

Anatomy of a MC *Item*

What muscle is responsible for retracting the suprapatellar fat pad during terminal knee extension?

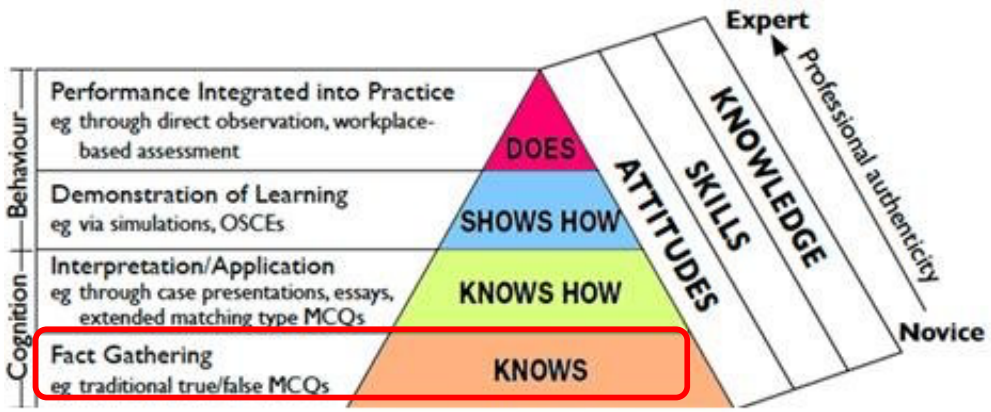
STEM

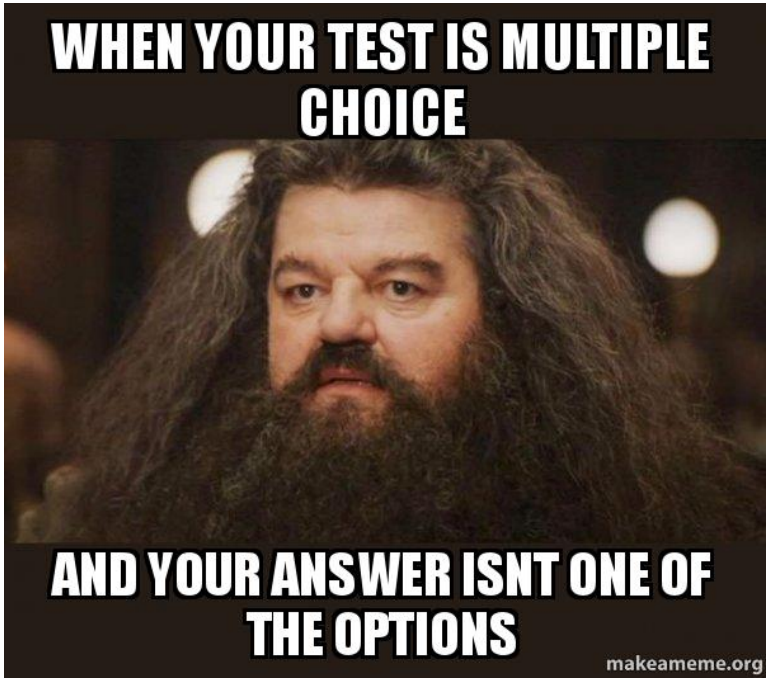
Options

a. VMO
b. Rectus Femoris
c. Genus Articularis

Distractors

KEY





Introduction to Item Writing Webinar, PR Geisler, August 4, 2020

Writing Clear Stems

1. STEM should be meaningful alone, BY ITSELF

2. STEM should NOT contain irrelevant material, generally
   - Unless, assessing ability to ascertain "relevant from irrelevant" in a case scenario, etc.*

3. "STEMLESS Items" (most common flaw)
   - No clear problem to solve, No question to answer

Writing Clear Stems

4. Avoid "Negative Stems"
   - "All of the following, *except*…"
   - "Which of the following is *not* true…"
   - "Which of the following is *least* likely…"

- Allows "rationalization" of distractors
- Reduces discrimination factor
- Mostly "recall" nature
- Knowing "one wrong answer" (the "exception" doesn't demonstrate knowledge of "why" other choices are correct*

*CAVEAT—OK for questions dealing with severe/fatal conditions and professional relevance/context (e.g., drug contraindications, emergent care procedures)

Writing Clear Stems

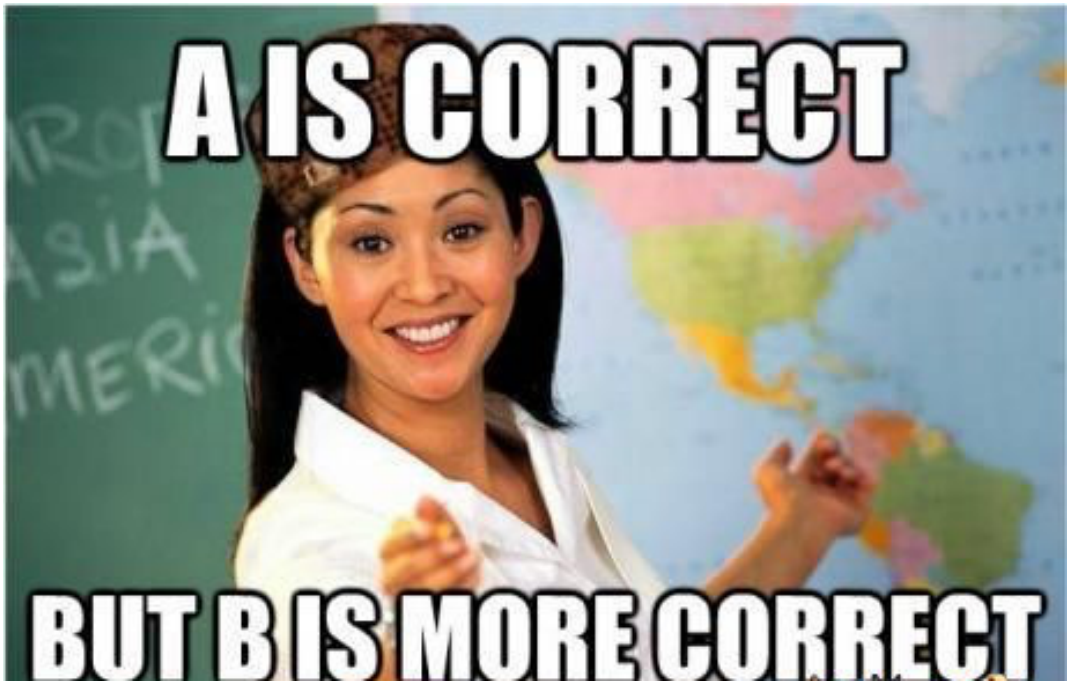
5. STEM should be a *question* or *partial sentence*

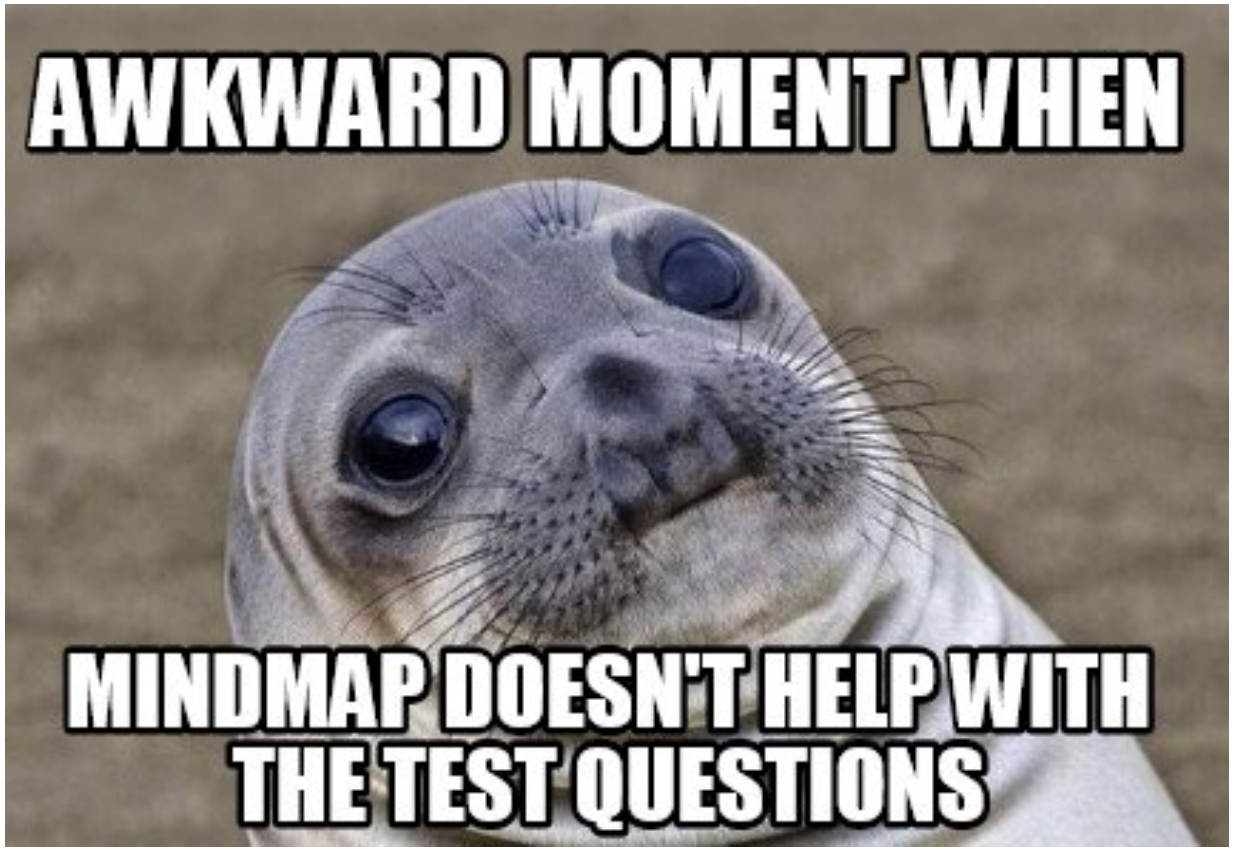
- Allows focus on "answering a question", not holding partial sentence in working memory and sequentially completing the phrase with each possible answer.
- Decreases cognitive load for student.
- Avoid "fill in blanks"

6. Can it pass the COVER Test?

Common Item Writing Flaws
(Reduce Validity & Discrimination)

- Negative Stems
- Stemless Items
- Answer Key errors
- Outdated information
- Blueprinting errors (curriculum, domain, etc.)
- Technical writing flaws
  - Grammar, spelling, clues, syntax



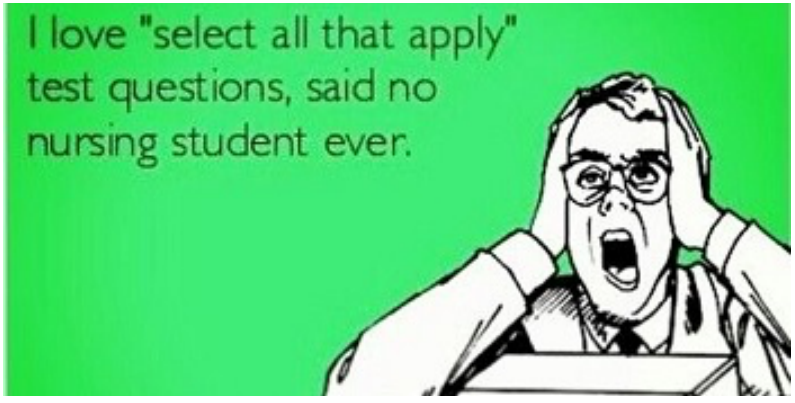


Simple Item Writing Tips to Increase V & D

- Avoid Adverbs (frequently, typically, etc.)
- Avoid absolutes ("never", "always", etc.)
- Avoid overlapping questions (part of 1 question contained in another)
- Avoid overly complicated verbiage
- Avoid "repeated elements" within questions
  - Allows "convergence strategy"

- All options ***HOMOGENOUS*** with stem
  - Same Style & Length
  - Same Language/Verbiage
  - Same/Consistent Grammar/Punctuation
  - Does stem and each answer make a complete sentence?
- All possibilities are **PARALLEL** (e.g., all anatomy choices, not mixture of anatomy & physiology)
- Avoid **OVERLAPPING** answers (esp. with numbers, time frames)
- All should be **PLAUSIBLE** options, presented in **LOGICAL** order
- Avoid "all the above", "none of the above" and "aggregated answers"
- VERIFY that all distractors ARE undoubtedly **WRONG** ☺

Writing Distractors and Keys



How many options to use for MCQ Items?

Michael C. Rodriguez, *University of Minnesota*

Multiple-choice items are a mainstay of achievement testing. The need to adequately cover the content domain to certify achievement proficiency by producing meaningful precise scores requires many high-quality items. More 3-option items can be administered than 4- or 5-option items per testing time while improving content coverage, without detrimental effects on psychometric quality of test scores. Researchers have endorsed 3-option items for over 80 years with empirical evidence—the results of which have been synthesized in an effort to unify this endorsement and encourage its adoption.

THREE OPTIONS

ptions Are Optimal for iple-Choice Items: A Meta-Analysis of 80 Years of Research

Based on the evidence synthesized in this meta-analysis, the item-writing rule can be revised: Three options are optimal for MC items in most settings.

Rodriquez M. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Ed Measure: Issues Practice*. 2005;Summer:3-13.

How many options to use for MCQ items?

Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting

Stephen D Schneid,[1] Chris Armour,[2] Yoon Soo Park,[3] Rachel Yudkowsky[3] & Georges Bordage[3]

OBJECTIVES Despite significant evidence sup
porting the use of three-option multiple-ch
questions (MCQs), these are rarely us
ten examinations for health profe
dents. The purpose of this st
the effects of reducing f
MCQs to three-optio
psychometric ch
dard setting
examin
st

two versions of a
containing 98 MCQs
students and 39 Year 3
nts. Four- and five-option MCQs
ted into three-option MCQs to cre-
versions of the examination. Differences
response time, item difficulty and discrimina-
tion, and reliability were evaluated. Medical and
pharmacy faculty judges provided three-level
Angoff (TLA) ratings for all MCQs for both ver-
sions of the examination to allow the assessment
of differences in cut scores.

ULTS Students answered three-option MCQs
rage of 5 seconds faster than they answered
d five-option MCQs (36 seconds versus
ds; p = 0.008). There were no significant
in item difficulty and discrimination,
reliability. Overall, the cut scores generated
three-option MCQs using the TLA ratings
were 8 percentage points higher (p = 0.04)

CONCLUSIONS The use of three-option MCQs in a health professions examination resulted in a time saving equivalent to the completion of 16% more MCQs per 1-hour testing period, which may increase content validity and test score reliability, and minimise construct under-representation. The higher cut scores may result in higher failure rates if an absolute standard setting method, such as the TLA method, is used. The results from this study provide a cautious indication to health professions educators that using three-option MCQs does not threaten validity and may strengthen it by allowing additional MCQs to be tested in a fixed amount of testing time with no deleterious effect on the reliability of the test scores.

THREE OPTIONS

Medical Education 2014: 48: 1020–1027
doi: 10.1111/medu.12525

Discuss ideas arising from the article at
www.mededuc.com 'discuss'

Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting

Stephen D Schneid,[1] Chris Armour,[2] Yoon Soo Park,[3] Rachel Yudkowsky[3] & Georges Bordage[3]

Although there is nearly a century of research supporting the use of three-option MCQs, it has not had much impact on the strong orthodoxies that exist regarding the number of options used in health professions' MCQ examinations.

The findings from this study challenge yet again the traditional approach to standardising the number of options for an entire examination to four or five.

Thus, this study provides a cautious indication professions educators that using three-o threaten validity and may strengthen MCQs to be tested in a fixed a no deleterious effect on the re

From a practical perspective, can continue to write as many plausible option asible, but, most importantly, they should not discard three-option MCQs during the test development phase. The practice of eliminating poor distractors, which may allow the inclusion of more items per unit of testing time, may, depending on the content of the additional test material, provide for more valid and reliable test scores.

THREE OPTIONS

Educational Me ISSUES A NCME national council on measurement in education

Th Multiple-Choice Items: A Meta- rch

5 | https://doi.org/10.1111/j.1745-3992.2005.00006.x | Citations: 139

2 is Assistant Professor of Quantitative Methods in Education, College of Education and opment, University of Minnesota, 206 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN, 55455; umn.edu. His areas of specialization include item writing, test design and evaluation, meta-analysis, hierarchical linear modeling.

Read the full text > PDF TOOLS SHARE

Abstract

Multiple-choice items are a mainstay of achievement testing. The need to adequately cover the content domain to certify achievement proficiency by producing meaningful precise scores requires many high-quality items. More 3-option items can be administered than 4- or 5-option items per testing time while improving content coverage, without detrimental effects on psychometric quality of test scores. Researchers have endorsed 3-option items for over 80 years with empirical evidence—the results of which have been synthesized in an effort to unify this endorsement and encourage its adoption.

International Journal of Selection and Assessment Volume 20 Number 1 March 2012

The Three-option Format for Knowledge and Ability Multiple-choice Tests: A case for why it should be more commonly used in personnel testing

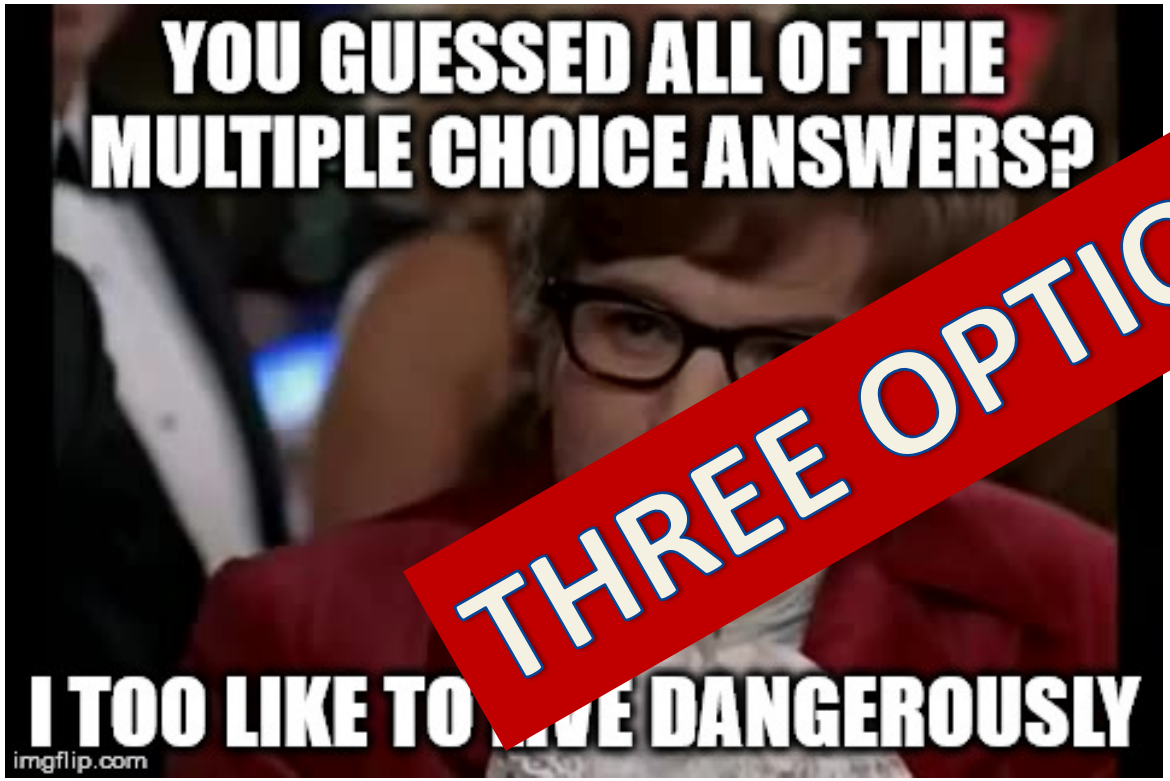
Bryan D. Edwards*, Winfred Arthur Jr and Leonardis L. Bruce*****

*Department of Management, Oklahoma State University, Stillwater, OK
bryan.edwards@okstate.edu
**Department of Psychology, Texas A&M University, College St
***Department of Psychology, Auburn University, Auburn

Multiple-choice (MC) tests are … esting format in appli… settings. In the psychometri… …ch on the optimal number of options for knowledge and ab… …hat three-option tests are psychometrically equivalent and, in so… …five-option tests. In addition, there are a number of practical, economic… …ative advantages associated with the use of three-option MC tests. Yet, despite… …ntages, the three-option format is underutilized in personnel selection. Across two st…dies, we compared test-taker perceptions, cri…ion-related validity, and …-based subgroup differences, and in Study 1 we …mpared race-based subgroup differences on three- and five-option tests. Participants in the two studies completed a three- or five-option version of ACT. Test perceptions, criterion-related validity, and race- and sex-based subgroup differences were similar across test formats. The implications for the expanded use of three-option tests in applied settings and future directions for research are discussed.

THREE OPTIONS

How many options to use for MCQ Items?



Translation?

option MCQs are
as valid as 4 or 5
tion MCQs, and…

- Can save time making the test
- Increase content coverage by allowing more questions/test
- Increase test validity & discrimination

Big Picture-For Using 3 Option MCQs

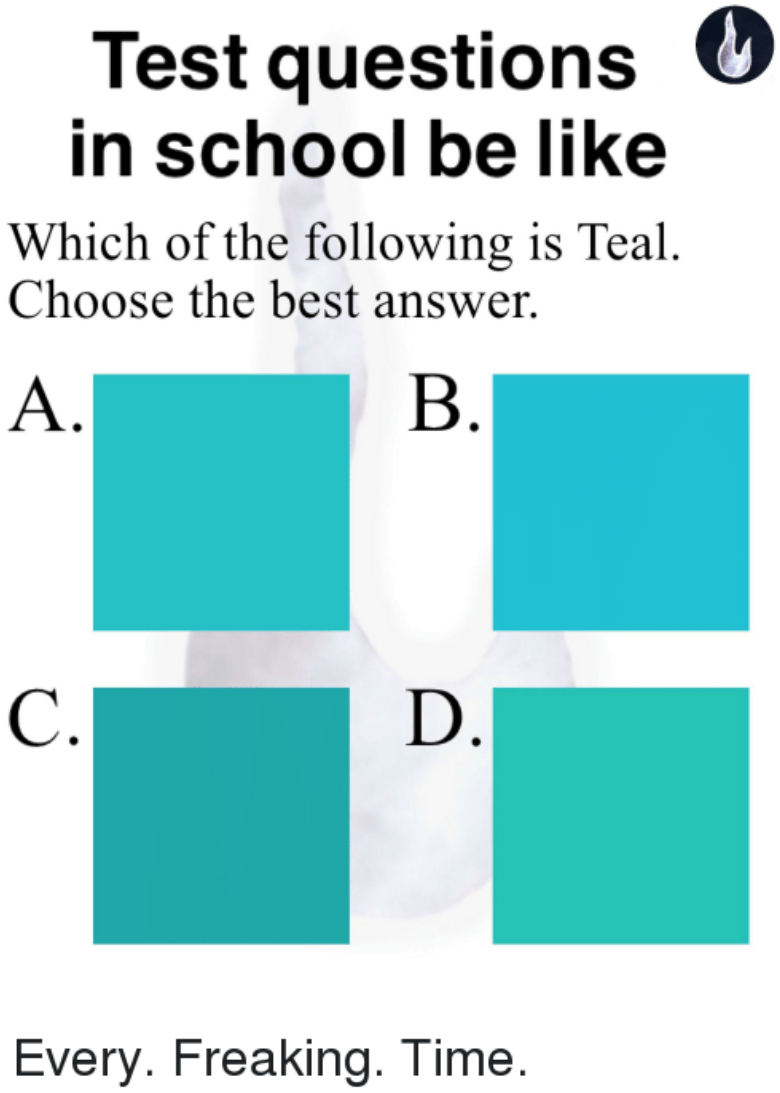
PHOP

1. **P**LAUSIBLE
2. **H**OMOGENOUS
3. **O**NE (correct option)
4. **P**ARALLEL

- Greater validity
- Greater distribution
- Greater discrimination
- Less guessing
- Save time

- OK to USE 4 or 5 when necessary/appropriate and follow other "rules"

Introduction to Item Writing **Webinar, PR Geisler, August 4, 2020**



What do you think?

- Which order of events is correct for an acute inflammatory response following tissue damage?

  I. cells release chemicals

  II. vasodilation

  III. tissue ischemia

  IV. fluids and proteins move into interstitial space

a. I, II, III, IV

b. I, III, IV, II

c. I, II, IV, III

d. II, IV, I, III

Overlapping, need 4 to get 1 correct; cognitive overload; time sucker

What do you think?

- Which of the following symptoms would be effectively treated by an antihistamine?

  a. headache

  b. cough

  c. runny nose

  d. nasal congestion

> 1 Correct Answer; heterogenous options; combo of S/S

What do you think?

- After a significant infection, strength may take _____ to fully recover.
  a. 2 to 4 days
  b. 2 to 4 months
  c. 2 to 4 weeks
  d. strength is not affected

Not Homogenous—Eliminate Option "D"; make game of 1/3s. Overlapping options

What do you think?

- Which of the following is NOT suggestive of potential hypertrophic cardiomyopathy?

  a. family history of sudden death under age 50

  b. headache every morning upon waking

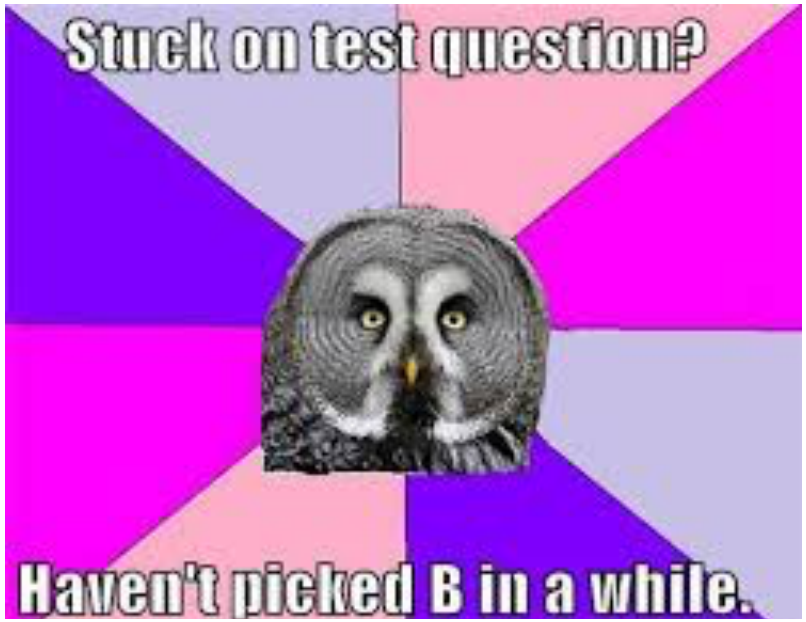
  c. personal history of syncope and dyspnea

  d. auscultated murmur with Valsalva maneuver

Negative Stem ("NOT")

Psychometrics 101

- Point Biserial Index (PBI) (-1 to +1)
  - Correlation between score on item & score on exam; differentiates between those who have high- or low-test scores
  - Positive PBI = those who scored well on exam, answered item correctly
  - < 2.0 = Poor (revise)
  - ≥ 0.2-0.29 = Fair
  - 0.3-0.39 = Good
  - 0.4-0.7 = Very Good
  - ≤ 0.10 raise suspicions of incorrect key
  - ≤ 0.05 PBI questions need to be discarded
  - NEGATIVE PBIs = BIG PROBLEM
- Item Difficulty (p-value)
  - % of correct responses
  - Want 0.3-0.8 (30-80%)
- Kuder-Richardson Formula 20 (KR-20)
  - Likelihood of obtaining similar results with another group of similar students (0-1)
  - > 0.5 = Good
  - > 0.65 = Very Good
- Exam Breakdown:
  - 5% very hard, 5% very easy, 20% difficult, 20% easy and 50% medium difficulty questions
  - Look at ALL PBIs, not just correct
  - Too many *very hard* and *very easy* items decrease exam reliability



"There are three kinds of lies: lies, damned lies, and statistics."
Mark Twain

Hingorjo & Jaleel, 2012

"*Properly constructed MCQs* assess higher-order cognitive processing like interpretation, synthesis and application of knowledge, instead of just testing recall of isolated facts"

Carneson J, Delpierre G, Masters K. Designing and managing MCQs: Vol. 62, No. 2, February 2012 146. Appendix C: MCQs and Bloom's taxonomy. (Online) 2011 (Cited 2011 Feb 2). Available from URL: http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html

Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences, National Board of Medical Examiners 3rd ed. (Online) 2010 (Cited 2011 Feb 5). Available from URL: http://www.nbme.org/publications/item-writing-manual.html

Morrison S, Free KW. Writing multiple-choice test items that promote and measure critical thinking. *J Nurs Ed.* 2001;40:17-24.

Going from "Knows" to "Knows How"

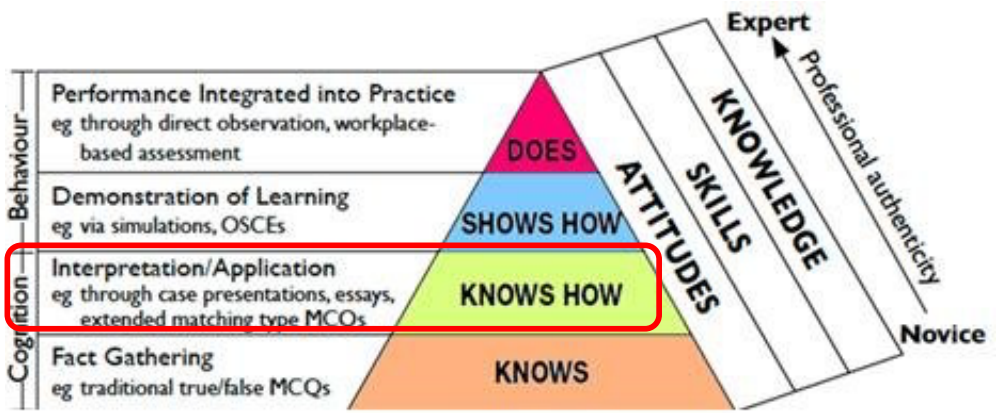
Your 6 mos. s/p ACL-R patient has full PROM in EXT but can't achieve last 10 degrees of active knee extension due to inhibition of this muscle. **STEM**

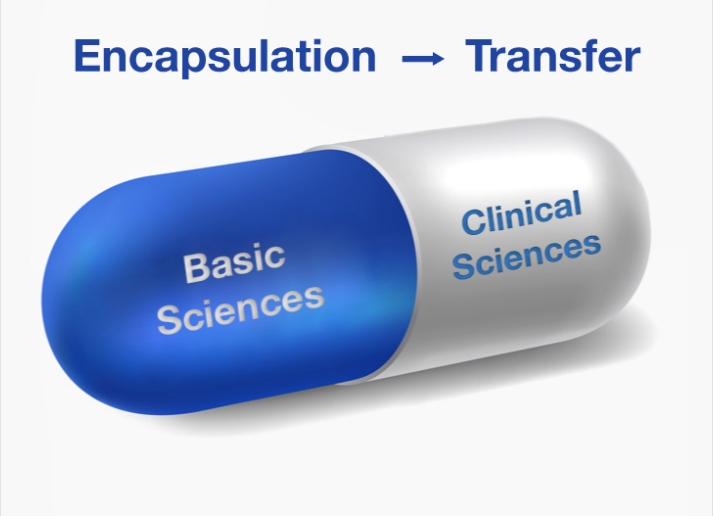
Options

a. VMO
b. Rectus Femoris
c. Genus Articularis

Distractors

KEY





Increased Authenticity and Validity with Media

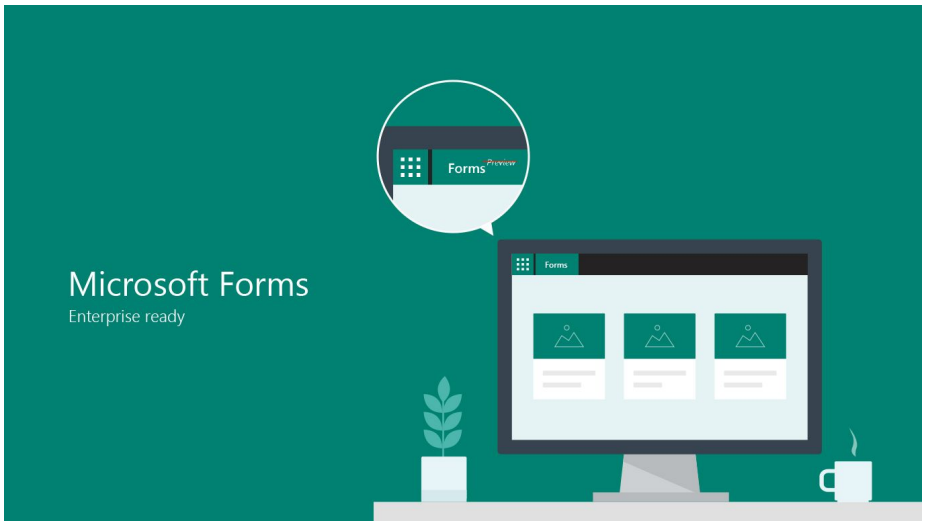
(pics, videos, etc.)

Liu M, Papathanasiou E & Hao Y (2001). Exploring the use of multimedia examination formats in undergraduate teaching: Results from the fielding testing. *Computers in Human Behavior*, 17: 225-248.

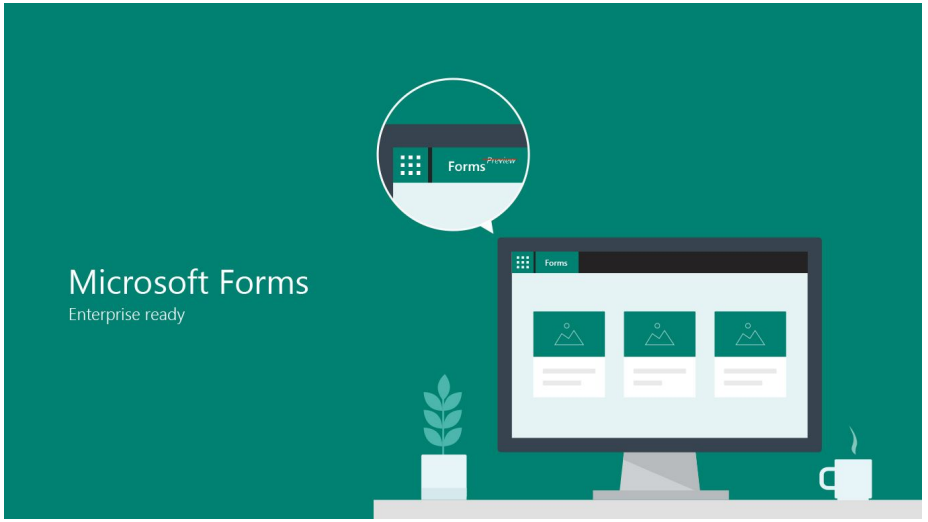
Using multimedia online examinations:

- Assessment more closely matched material taught
- Use of # media presentations aided student recall
- Questions better reflected real-world situations; more authentic
- Students learned more in these assessments, helping their learning move forward (*Assessment FOR Learning*)

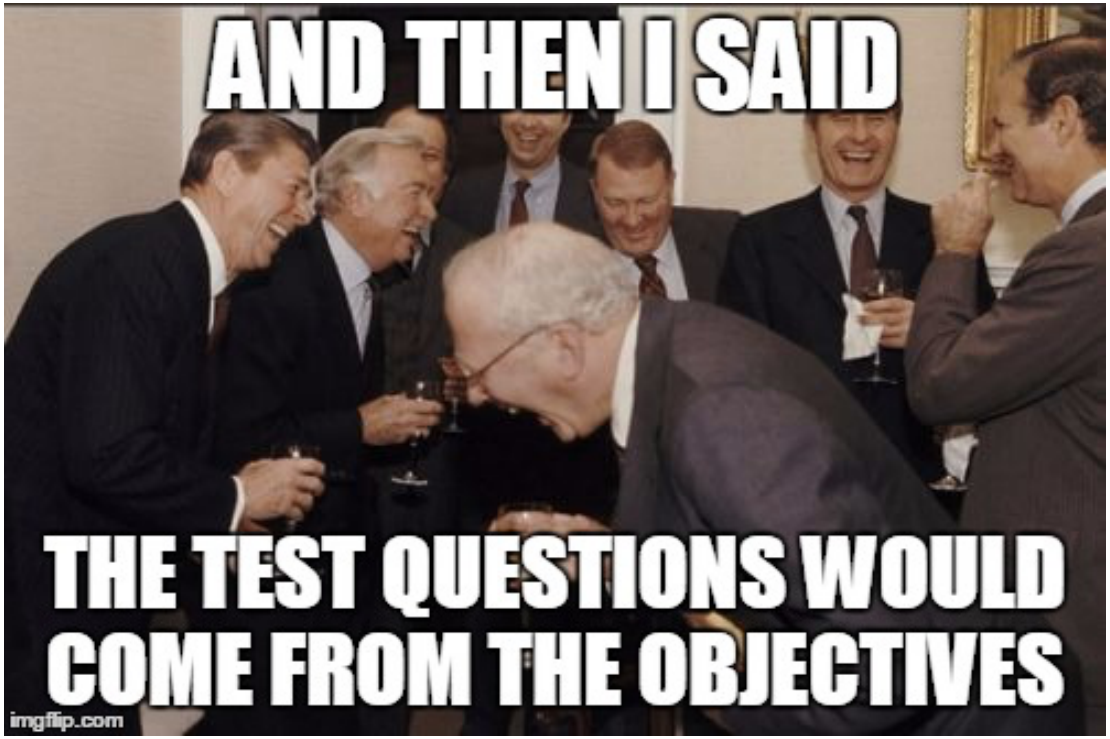
Using MS Forms



Using MS Forms

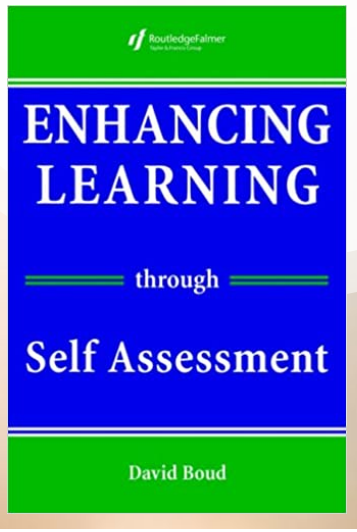


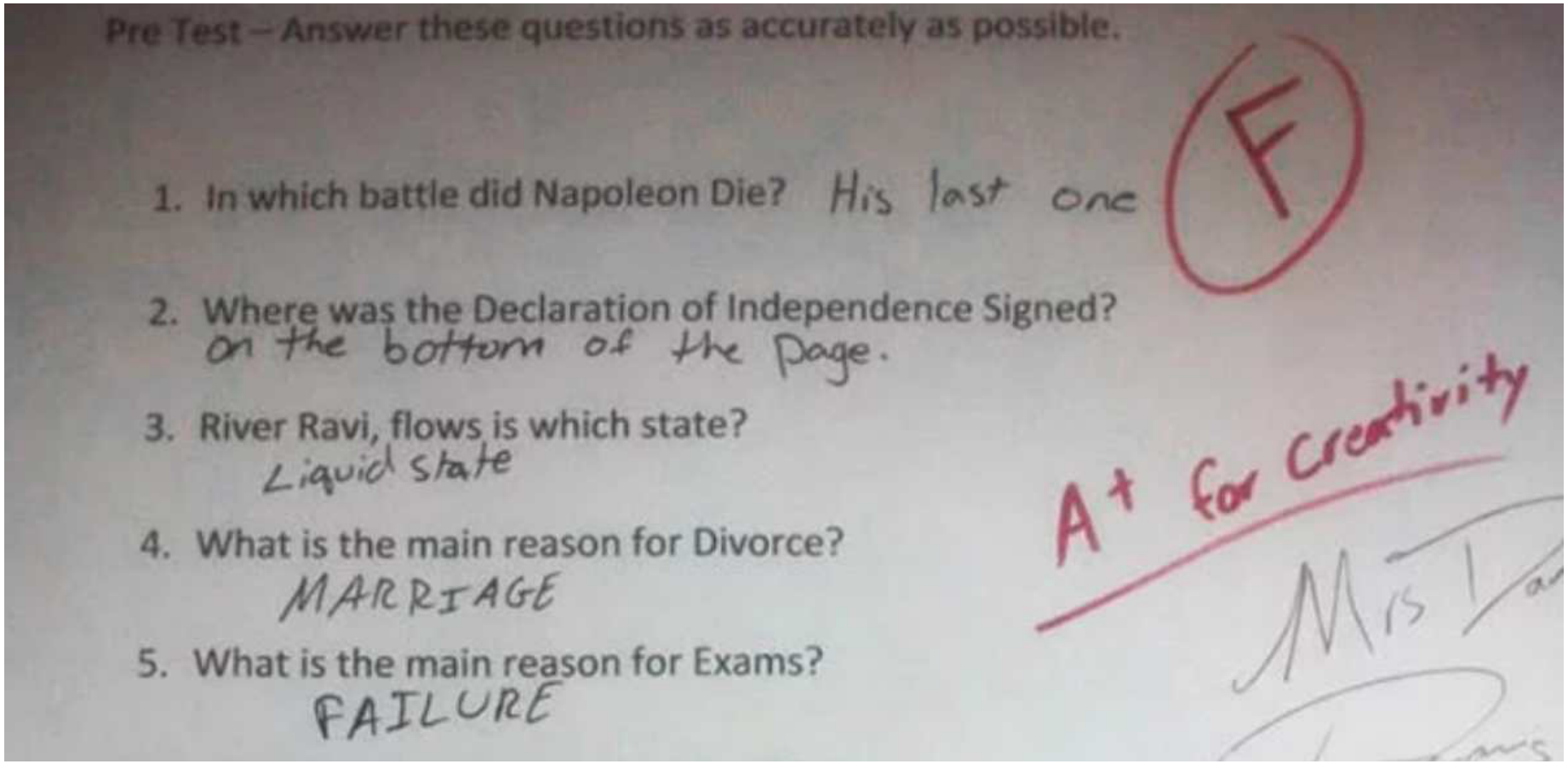
Quick Strike Poll #3 (2 ?s)



David Boud, 1995

"Students can, with difficulty escape from the effects of poor teaching, they cannot…escape the effects of poor assessment"




Pre Test – Answer these questions as accurately as possible.

1. In which battle did Napoleon Die? His last one

F

2. Where was the Declaration of Independence Signed?
on the bottom of the page.

3. River Ravi, flows is which state?
Liquid state

4. What is the main reason for Divorce?
MARRIAGE

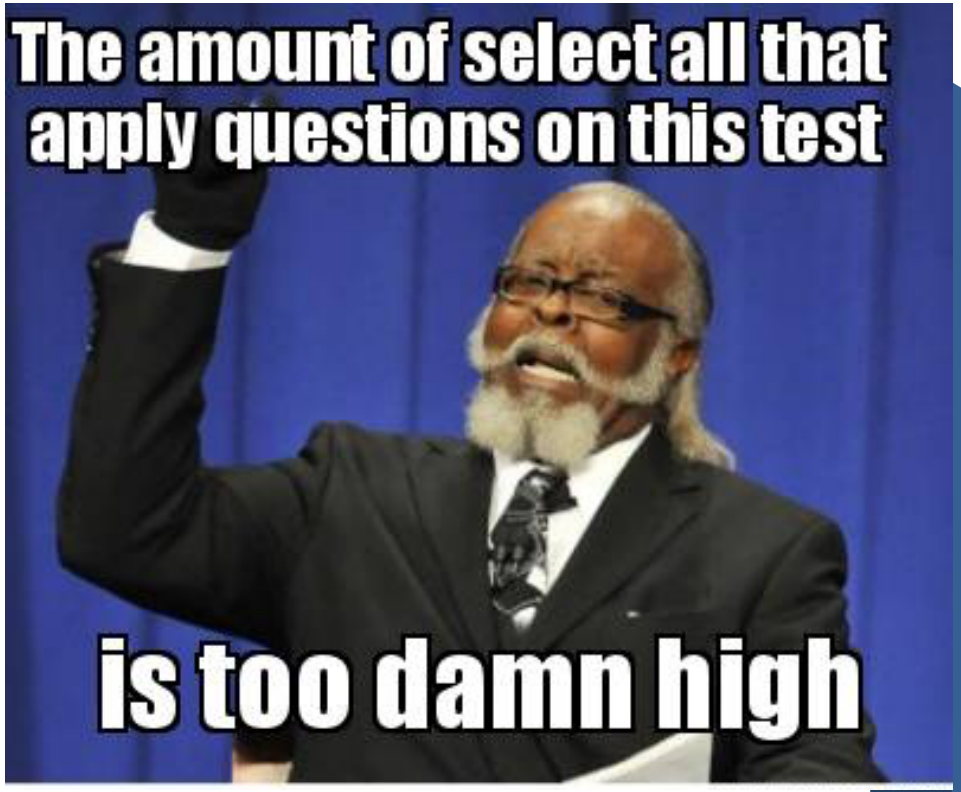
5. What is the main reason for Exams?
FAILURE

A+ for Creativity

Mrs

Thank You, Host & Sponsor

Glen Bergeron, WFATT & Oliver Coburn, BASRAT



Open Discussion

Thoughts, questions, input?

References

Ebel, RL. Writing the test item. In EF Linquist (Ed.), Educational measurement (pp. 185–249). 1951. Washington, DC: American Council on Education.

Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990; 65(9S):S63-67.

Raymond, MR, Stevens, C & Bucak, SD. The optimal number of options for multiple-choice questions on high-stakes tests: Application of a revised index for detecting nonfunctional distractors. *Adv in Health Sci Ed. 2019;***24:**141–150

Liu M, Papathanasiou E & Hao Y (2001). Exploring the use of multimedia examination formats in undergraduate teaching: Results from the fielding testing. *Computers in Human Behavior*, 17: 225-248.

Rodriquez M. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Ed Measure: Issues Practice.* 2005;Summer:3-13.

Morrison S, Free KW. Writing multiple-choice test items that promote and measure critical thinking. *J Nurs Ed.* 2001;40:17-24.

Edwards, BD, Arthur, W, Jr and Bruce, LL Three-option Test Format. *Int J Select Assess*. 2012;20:65-81.

Schneid SD, Armour C, Park YS, Yudkowsky R, Bordage G. Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting. *Med Educ*. 2014;48(10):1020-1027.