Evaluate Your Examinations: The Significance of High Stakes Examination Analysis

Dennis Edgerly, MEd, EMT-P

Paul L. Rosenberger, EdD, NRP

In school, you're taught a lesson and then given a test. In life, you're given a test that teaches you a lesson.

Tom Bodett (author, voice actor, radio personality, and spokesman for Motel 6)

Objectives

- After the presentation, the class participants will be able to:
  - Describe test reliability
    - Understand KR20 metrics
  - Describe test validity
  - Understand test item difficulty level
  - Understand item discrimination (Point bi-serial)
  - Design items that evaluate and enhance higher-order thinking

Types of Testing

▶Formative

▶Summative

Formative

- To monitor student learning and should be used to adjust instruction and learning
  - Individual students and the entire class
- Guides instruction and learning
- Students > Adjust learning; Teacher > Adjust curriculum
- To help "form" learning
- Informal (Classroom questions, observations, homework, quizzes)
- Low Stakes

Summative

- To assign grades to students after a period of time of instruction
- To "sum up" learning
- Formal (Finals, Projects, Term Papers)
- High Stakes
- High Technical Quality Desired

Test reliability

- Desired – A test that measures <u>consistently.</u>
  - Scale example
    - 200 pounds every time
  - Weight guesser example
    - ??? pounds
  - Paramedic class to paramedic class
    - Test average
    - Inter-rater reliability (-)
   - Essay (-)
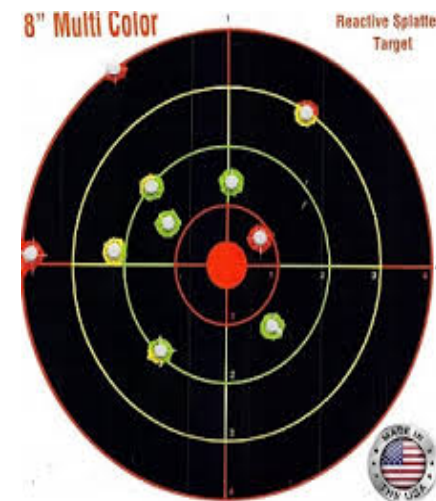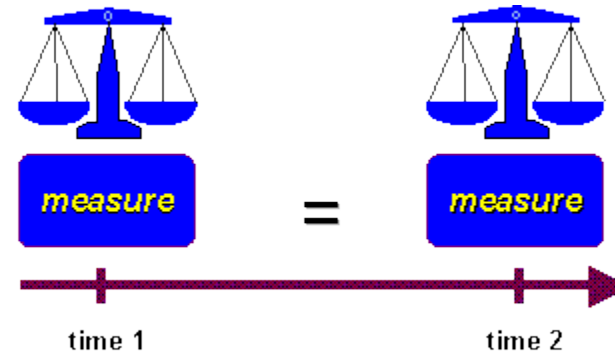
Not reliable or valid

Test reliability

- KR 20 (Kuder Richardson) Metrics
  - Post-test review
  - How well does the test measure a cognitive objective
  - Internal consistency
  - Split-half method
  - Score = 0.00 to 1.00
    - 0.6 or greater is desired
  - Score improvement
    - Longer tests
    - Heterogeneous candidates
    - Challenging items

Reliable, not valid

Test validity

- The test accurately measures what it was intended to measure (hopefully, based on course objectives)
  - Can conclusions be drawn from test results
- Scores correlate with student knowledge (or lack of)
- Bull's Eye Shooting – It hits the desired target!
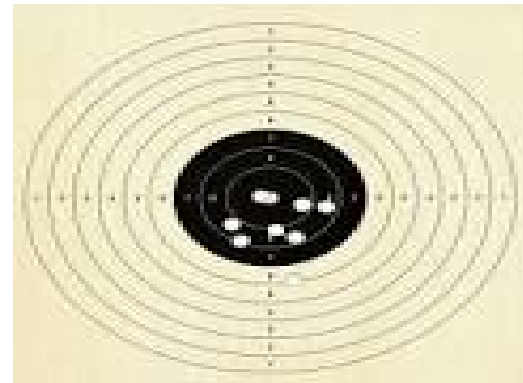
Reliable & Valid

Test validity

- Face validity
  - Common-sense
  - Least scientific
- Content/Construct validity
  - Items should representative of the wider body of knowledge
  - Blue printing
  - Low level objectives
  - High level objectives
  - Medical accuracy
- Criterion validity
  - Predictive value (clinical/field)
  - Piloting items

Item difficulty

- Post-test review
  - % of students who answer an item correctly
 - Ideally between 30% to 90%

Item discrimination (point bi-serial)

- Post-test review

- Compares the students who scored well with the students who did not score so well – It discriminates learners' ability

- Useful to identify well performing items and poor preforming items

  - Validation vs. revision

  - High difficulty level with solid bi-serial values

Item discrimination (point bi-serial)

- Possible scores
  - -1 to +1
  - 0 = No discrimination
- High discrimination
- Low discrimination
- Positive
- Negative

Item discrimination (point bi-serial)

- Reasons for negative discrimination
  - Item mis-keyed
  - Poorly constructed/Misleading
  - Content not adequately taught
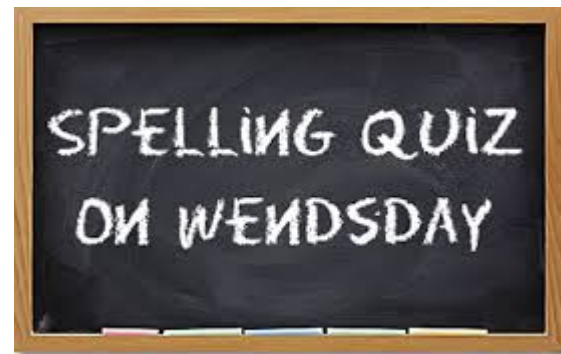
Poor Items

Attempts: 32 out of 32

When conducting an annual employee review for a paramedic that does not routinely meet quality management measures, you should include:

a labor union representative.	4 respondents	13 %	
his/her EMS partner.	1 respondents	3 %	
a human resources representative.	27 respondents	**84 %**	✓
his/her attorney.		0 %	

+0.02

Discrimination Index ⑦

84% answered correctly

Attempts: 34 out of 34

A paramedic inadvertently misplaced the endotracheal tube in the esophagus while performing an intubation. The patient became hypoxic and later died. Select the term that best describes this medical error.

Burnout	2 respondents	6 %	
Iatrogenic	25 respondents	74 %	✓
Normalization of deviance	6 respondents	18 %	
Managed care	1 respondent	3 %	

-0.12

Discrimination Index ⑦

74% answered correctly

Attempts: 34 out of 34

Select the condition/event that can contribute to errors in pediatrics. ⑦

Kidney development is complete at 18 months		0 %	
Children easily communicate their clinical complaint		0 %	
Liver development is complete at 6 months		0 %	
Medications are calculated on weight/height	34 respondents	**100** %	✓

-0

Discrimination Index ⑦

100% answered correctly

Attempts: 31 out of 31

Select the characteristic that is **NOT** consistent with an effective quality improvement program.

Great quality management programs are managed and led.		0 %	
Great quality management programs are clinical-only initiatives which are managed by an individual.	24 respondents	**77** %	✓
Great quality management programs are functional and get results.	2 respondents	6 %	
Great quality management programs are based on a portfolio of projects.	5 respondents	16 %	

-0.06

Discrimination Index ⑦

77% answered correctly

Attempts: 31 out of 31

Select the item that is consistent with coaching (as opposed to management).

Directs people what to do		0 %	
Built on authority		0 %	
Immediate need focused	1 respondent	3 %	
More likely to facilitate	30 respondents	**97** %	✓

-0.15

Discrimination Index ⓘ

97% answered correctly

Good Items

Attempts: 32 out of 32

Expecting team members to understand what you are telling them without providing them with clear direction or without asking for feedback (to ensure that they understand the directive) are examples of which barriers?

Sender barrier	11 respondents	**34** %		✓
Receiver barrier	4 respondents	13 %		
Decoding barrier	17 respondents	53 %		
Medium barrier		0 %		

+0.32

Discrimination Index ⓘ

34% answered correctly

Attempts: 34 out of 34

Determine and select the proper answer for this statement. Normalization of deviance is a _____ phenomenon.

widespread	6 respondents	18 %	
short-term		0 %	
managed care	1 respondent	3 %	
long-term	27 respondents	**79** %	✓

+0.51

Discrimination Index ⑦

79% answered correctly

Attempts: 34 out of 34

Which physician group is not known for contributing to EMS quality management?

American College of Emergency Physicians	1 respondent	3 %	
Society for Academic Emergency Medicine	2 respondents	6 %	
American Academy of Surgeons	30 respondents	**88 %**	✓
National Association of EMS Physicians	1 respondent	3 %	

+0.43

Discrimination Index ⃝

88% answered correctly

Attempts: 33 out of 33

Which quality management tool is used to determine root cause?

The PDCA cycle	2 respondents	6 %	
Fishbone diagrams	28 respondents	**85** %	✓
DMAIC	2 respondents	6 %	
SWOT analysis	1 respondent	3 %	

+0.48

Discrimination Index ⑦

85% answered correctly

Attempts: 6 out of 6

+0.9

Determine if this statement is true or false. Is this statement consistent with the course readings and instruction?

Discrimination Index ⓘ

The key to subordinate survival is to recognize when your manager is acting in an administrative capacity and when they are acting in an authoritarian capacity.

True	4 respondents	67 %	
False	2 respondents	**33** %	✓

33% answered correctly

Attempts: 6 out of 6

Decide which of the following loans itself to good organizational decisions and communications. Which action is best for an organization?

Groupthink	1 respondent	17 %	
Conflict	1 respondent	17 %	
Disagreement	4 respondents	67 %	✓

+0.55

Discrimination Index ⑦

67% answered correctly

Continuous Improvement

Posttest Actions

Low discriminators (list items with Disc index of .15 or less or negative discrimination and with difficulty levels less than 85%)

Item #	Diff level (only those less than 85%)	Disc index of .15 or less	Comments/Disposition
1	84%	.08	Injury prevention - OK
2	71%	.12	Medical director - Keep
4	78%	.08	OK - legal - Keep
19	68%	-.33	Took C&D - Δ for next Exam
29	84%	.01	OK
32	78%	-.17	Took C&D scheduled Controlled Sbtance - Δ for Next
39	76%	-.06	Took A&C Δ for next Exam
41	84	.14	OK
42	84	.12	OK
53	68	.07	GCS - Keep
59	76	.08	Legal - Keep

Major/High Stakes Exam Analysis

Program Name:

Course Name: **Course Dates:**
Course Number: **Exam Date:**
Exam Name/Topic: **Lead Instructor:**

Number of students
completing exam: Is statistical analysis of
                                    questions conducted?   Y ☐   N ☐

Highest Score:

Lowest Score:

Cut Score:

Class Average:

Class Median:

Content areas receiving unexpected low scores

Recommendations from learning prescription if applicable

Changes made to exam if applicable

Strategies to address low performing areas

Program Director/Lead Instructor comments:

Exam results reviewed with:			
Faculty	Y ☐	N ☐	Date:
Medical Director	Y ☐	N ☐	Date:

Program Director: **Date:**

Questions

Thank you for your interest!