

Student Evaluation

George W. Hatch, Jr., EdD, LP, EMT-P

June, 2015

DSHS Course Coordinator Course

Thanks to ***Deb Cason*** for some of the slides *

Objectives

- Describe the principles of student evaluation
- Discuss various types of student evaluation
- Discuss the use of affective evaluation
- Describe the use and improvement of written cognitive written exams
- Discuss the use of item analysis to improve exams

What's the Purpose of Student Evaluation?

- Measures student's progress on objectives (are they learning?)
- Feedback for student & instructor to Identify strengths and weaknesses (Motivate, How do we need to structure their study/class/free time?)
- Measures program & instructor's effectiveness (How effective am I as a teacher?)
- Reinforces learning

Davis, 2001 *Tools for Teaching*

Multiple Information

- Student evaluation is a two part message:
 - Assessing the Student
 - Assessing instructor effectiveness
- Both must be thoughtfully considered!



Is the teacher responsible to “teach”
everything that is on the test?

Definitions

Evaluate: to ascertain, judge, fix the value or worth

Test: a series of questions or problems designed to determine knowledge; a criterion, standard

Examine: inspect or analyze in detail

Assess: to estimate value for taxation

Different Tools

- Never rely on a single evaluation instrument
- **Formal methods** include: written exams, research projects, practical exams & observational reports
- **Informal methods** include: questions in class & homework assignments, observation

Formal Tools are used to

- Formulate a “grade”
- Provide sound judgment regarding mastery of objectives
- May NOT be useful in providing timely feedback for students and modifying the teaching strategy

Informal Tools may:

- NOT be used to formulate a “grade”
- Provide more timely feedback for modifying teaching strategy
- Lack the rigor for justifying student competency OR pass/fail
- How quickly do you get quizzes graded and back to students?



Formative vs. Summative

Formative- Ongoing throughout course

- Allows for “In-Flight Corrections”

Summative- Final & comprehensive

- Measures if competencies/goals met
- Useful for future enhancements

High-Stakes Evaluations

- Pass = continuation of the program
- Fail = no program continuation

- Usually summative
- Little possibility that feedback can be used to modify learning strategy

High-Stakes Evaluations

- Students HIGHLY stressed
- Exams require HIGH defensibility
- Example: NREMT exams must meet the highest levels of scrutiny



Low-Stakes Evaluations

- Little impact on continuation of the program
- Formative, timely feedback, minimal scrutiny
- Students maybe less prepared for low-stakes exams
- Example: Classroom quizzes & homework



So I can throw these together
quickly, right?

RELIABILITY

- Reliability is **CONSISTENCY**

Is the test reproducible?

Reliability

- The exam must **CONSISTENTLY** measure what it intends to measure
- If reliable, it will produce similar results in different students of the same level of mastery

RELIABILITY

Does the test measure true factors or does it measure chance?

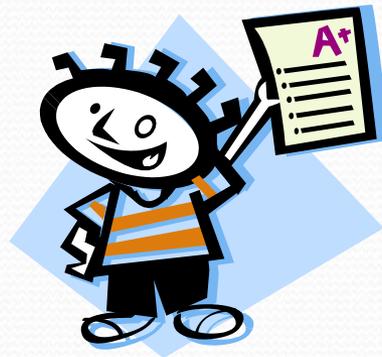
Examinee-fatigue, carelessness, discomfort, anxiety, distractions



RELIABILITY

Test- ambiguous items, trick questions, poorly worded directions

Scoring- carelessness, counting errors, computing errors



Inter-rater Reliability

The exam must CONSISTENTLY measure DESPITE the grader (ie. skills, essays, scenarios, fill in the blank)

If reliable, it will produce similar results in different students of the same level of mastery unrelated to WHO administered the exam

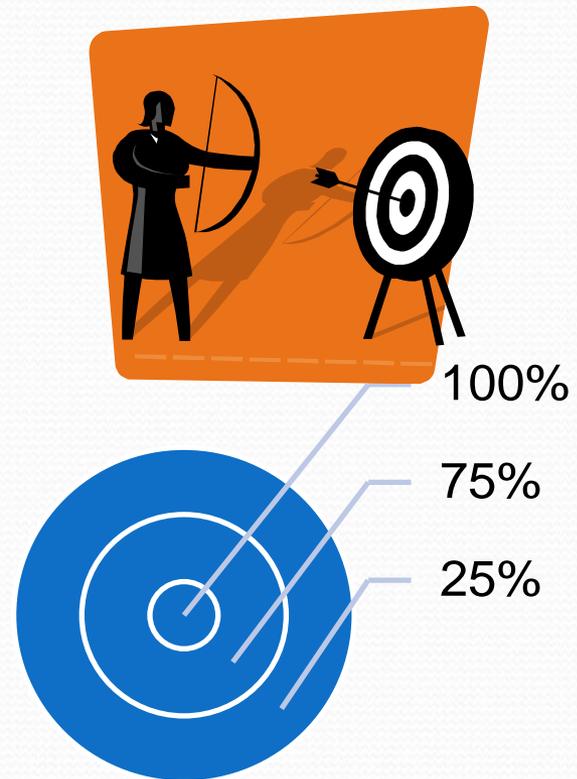
To Improve Inter-Rater Reliability

- Design subjective tool with specific behavioral anchors (see affective exams)
- Increase the number of evaluators (think Olympics)
- Average scores of evaluators or evaluations
- Eliminate high and low scores

Validity

Validity is **RELEVANCE**

Does the test measure what you want it to measure?



Validity

- The degree to which an exam measures the knowledge & skills it intends to measure
- If valid, knowledgeable students will perform better than those who are not

Types of Validity

- Face validity—commonsense validity
- Content validity—does the exam accurately represent the wider body of knowledge being tested?
- Construct validity—construction issues
- Criterion or Predictive Validity—does the exam predict future performance?

Written Exams

- True/False
- Matching
- Multiple choice
 - Tedious and time consuming to develop
 - Easy & quick to grade
 - Most objectives exams use this (NREMT)
 - Can test high level thinking
 - MORE TO COME!

Written Exams

- Completion/short answer
- Essay
 - Consider inter-rater reliability
 - One grader best
 - Skim all before to determine level & range of performance
 - Develop matrix before grading
 - Choose examples to serve as performance anchors (good, medium, poor)
 - Grade answer by answer if possible (all #1's)

Norm-Referenced Grading

- Compares student performance with each other to assign grades
- Commonly called “grading on the curve”
(Student performance based on group’s performance)
- Useful for formative exams—NOT summative exams

Criterion-Referenced Grading

- Based on MASTERY of objectives
- Use specific course objectives for evaluation tool
- Evaluation drawn from blueprint, grades set based on degree objectives are mastered
- Fairness requires valid/reliable items
- Generally the format in EMS classes

Other Evaluation Tools

- All domains must be evaluated—all domains must be acceptable to successfully complete course
- **Psychomotor:** formative needs detailed task analysis
 - PPCP on NAEMSE.org website
 - NREMT skills sheets are summative

Affective Evaluation

- See EMS.gov, go to Education
 - Bottom of the page is “Other”
 - 2002 Education Guidelines
 - Appendix V and VI are Affective Evaluation Tools
- CoA requires one comprehensive affective evaluation during class; each semester is desirable
- Suggest students self-evaluate

Rubrics

- Helpful when multiple facets are to be evaluated
- Converts a list of characteristics into observable criteria
- Rating scale: 0= unacceptable; 1 = acceptable
2 = above average
- Fewer than 3 or more than 9 = ↓ reliability

Professional Behavior Clinical Behavior & Judgment

Demonstrate professional behavior including: but not limited to, integrity, empathy, self-motivation, appearance/personal hygiene, self-confidence, communications, time management, teamwork/diplomacy, respect, patient advocacy, and careful delivery of service

- Education Standards, p. 53

Affective Evaluation Criteria

Integrity

Empathy

Self-motivation

Appearance/hygiene

Self-confidence

Communications

Time Management

Teamwork/diplomacy

Respect

Patient advocacy

Careful delivery of
service

* 1998 NSC & 2009 Ed
Standards

Sample Rubric Affective

Point Value

Criteria

1.
 - Inappropriate uniform or clothing worn to class or clinical; violates policies on clothing; poor hygiene or grooming
2.
 - Appropriate clothing most of the time; uniform unkempt (wrinkled or mildly soiled or in need of minor repairs). Occasional poor hygiene
3.
 - Clothing & uniform appropriate, neat, clean and well maintained
4.
 - Clothing & uniform above average, uniform pressed; grooming above average

Affective Evaluation

- Rubrics can be easily converted to grades
- Still must use as criterion referenced (do NOT average so that poor areas of performance are made up by areas of good performance)
- The more “behavioral anchors” the more inter-rater reliability

Oral Exams/Scenario

- Q & A are given verbally
- Scenario adds reality & requires “patients”
- Patients can either be manikins or “standardized patients” or “programmed patient” (actor)
- Moulage to better simulate reality
- Used to evaluate higher level thinking
- Can also evaluate affective domain

Oral Exams/Scenario

- Similar to essay
 - Can closely approximate entry level performance
 - Requires minimal or no distractions
-
- Requires as close to reality as possible
 - Very instructor intensive & time consuming



Oral Exams/Scenario

- Must be scripted with clear instructions
- Can not deviate from script
- Should include DOING vs. SAYING
- Examiner clues must be avoided—verbal and non-verbal
- Requires strong inter- rater reliability
- Video taping preferable

Tools for Simulation

Preparation

Objectives

Tasks

Conditions

Scenario

Timeframe

Participants

Equip/Set up

Endpoints &

Evaluation

Debriefing

- Foundations for Education:
An EMS Approach, 2nd ed

Field Internship Evaluation Tool

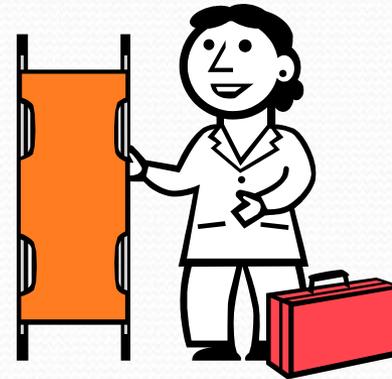
- PPCP materials has an excellent tool
- Access on NAEMSE.org website under “Instructor Resources”
- Also in *Foundations of Education: An EMS Approach*, 2nd ed book
- Includes definition of Team Leads (no prompts)

Improving Written Exams

- When should you use written exams?
- When should you use multiple choice exams?
- Who should write these exams?
- Who should understand exam writing?

Principles of Exam Improvement

1. Be clear of the purpose of each exam



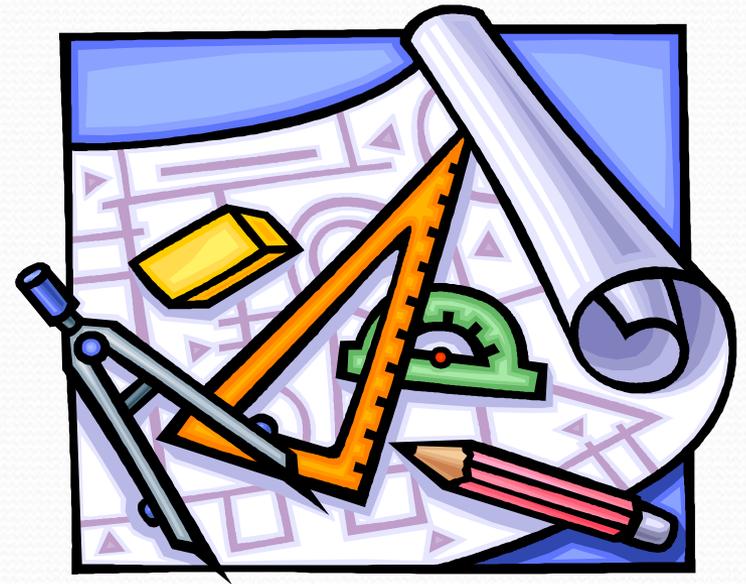
Principles of Exam Improvement

2. Base each item on an objective



Principles of Exam Improvement

3. Blueprint the exam to appropriately cover and weight topics



See handout Sample
Blueprint

Principles of Exam Improvement

4. Construct items appropriately



See Exam Item Construction
handout

Construction issues

- **Principles relating to the item as a whole**

Best answer

Immediately following the delivery of a newborn, the EMT should focus attention on

- A. Delivery of the placenta
- B. Massaging the uterus
- C. The baby's breathing**
- D. Controlling the mother's bleeding

Opinions

- **According to the American Heart Assn 2015 guidelines...**
- **According to the protocols of ACME Springs EMS...**
- **The Texas Dept of State Health Resources EMS Bureau states...**

Item Construction

- **Principles relating to the stem**

Negative statements

At the scene of an auto crash, a charged power line has fallen on a car. As an EMT, your actions should include all of the following EXCEPT:

- A. Order all bystanders to move back
- B. Call for the power company immediately
- C.** Attempt to cut the power line at the pole
- D. Tell the occupants to stay in the car

Item Construction

- **Principles relating to options**

Principles of Exam Improvement

5. Make adaptations from published item banks.



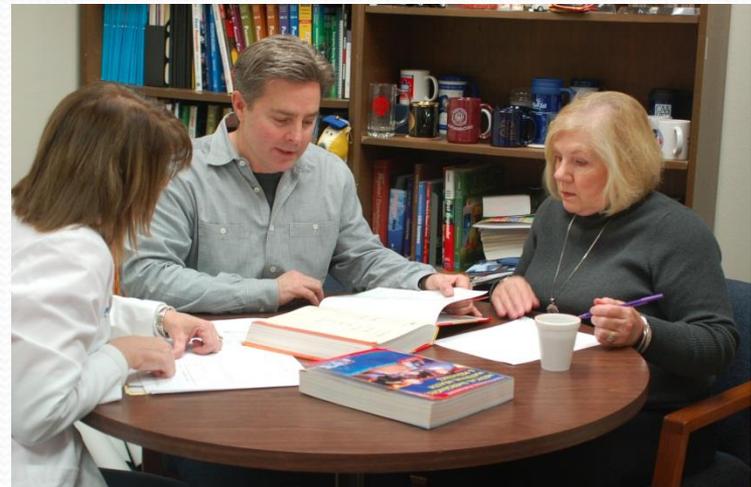
Fisdap exams (secure)

Publisher exam banks

See “Strategies...High Performing...”
handout

Principles of Exam Improvement

6. Obtain multiple sources of “expert” item review (include medical director).



Principles of Exam Improvement

7. Use “higher cognitive level” items rather than “recall”



Cognitive Domain

Knowledge Level:
recall, comprehend

Application Level: apply, use

Problem solving Level:
analyze, synthesize,
evaluate

RECALL

Which of the following parameters is included in initial patient assessment?

- A. Blood pressure
- B. Level of consciousness**
- C. Movement of extremities
- D. Bowel sounds

RECALL

The heart is enclosed in a membrane called the

- A. Pleura
- B. Myocardium
- C. Peritoneum
- D. Pericardium**

Application

Which of the following assessment findings is most helpful for determining the adequacy of ventilation?

- A. Skin color**
- B. Heart rate**
- C. Blood pressure**
- D. Respiratory rate**

Problem Solving

A patient from an auto crash presents with decreased LOC, b/p 170/100, p 60, and a resp rate 8. His skin is pale, cool and moist. Which of the following would be the most appropriate means of administering oxygen?

- A. Bag-valve-mask**
- B. Nasal cannula**
- C. Non rebreather face mask**
- D. Venturi mask**

Principles of Exam Improvement

- 8. Write clear, complete directions for each test and provide verbal instructions**

Directions...

Use cover sheet

Time limits

Bathroom rules

Answer sheet only will be graded

No calculators, etc

Only process questions

When/how grades posted

Principles of Exam Improvement

9. “Prepare”
examinee for
the test



Preparing Students

- **The “sleep and no meds” lecture**
- **Similar type of exam items previously**
- **Clear objectives**
- **Subjects to be covered**
- **Discuss rules**

Principles of Exam Improvement

10. Pay attention to exam security – before, during and after exam



Prevalence of Cheating

TYPE of CHEATING	students in 1961	students in 1991
Copied from another exam	26%	52%
Helped friend cheat on exam	23%	37%
Use crib notes to cheat on exam	16%	27%
McCabe, DL. What We Know About Cheating in College 1996 Change 28(1): 31		

Detecting Cheating

- Observation
- Statistical evidence: most useful on multiple choice examinations

BOTH usually needed to prove an incident



Exam Security

- **Appropriate spacing**
- **Use cover sheets**
- **Stroll the classroom**



Exam Security

- **Bathroom breaks?**
- **Books, back packs, cell phones, etc NOT at desks**
- **Enforce silence during the testing period**
- **Remove caps, hats and sunglasses**

Exam Security



- **Students can't keep exams or take notes from exam**
- **Avoid teaching the test**
- **Continually revise high stakes exams**

Principles of Exam Improvement

11. Evaluate your exam and items after it is administered



Test Evaluation

**ALL Exams are
IMPERFECT Measuring
Devices!**

**Student
feedback —very important**



Difficulty Index

% of students selecting correct answer

Ideal –above 40%

**Below 40%--possible problem,
may just be difficult concept**



Discrimination Index

How well the item discriminates between high and low scorers

- **Ranges from +1.0 to – 1.0**
- **The higher the score, the more discriminating**

Discrimination Index

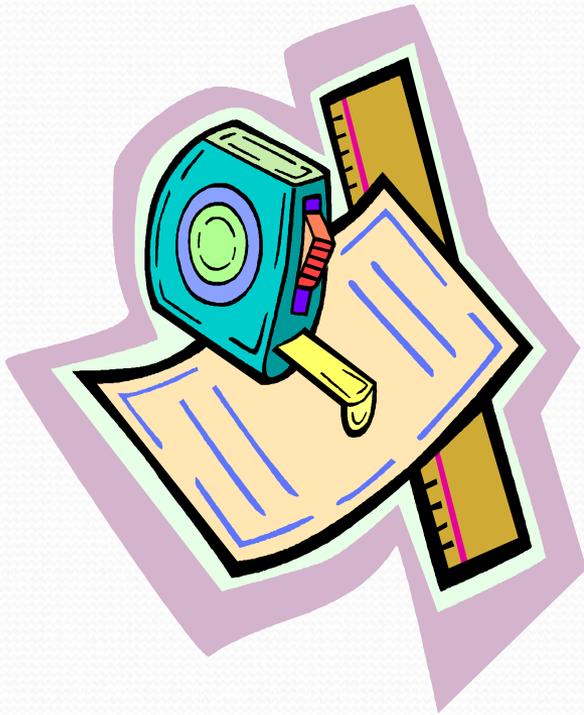
- **Above .40 is good**
- **.20 - .40 is probably okay**
- **Negative discrimination (-.1 or less) usually means tricky or mis-keyed**

Distracter Analysis

- **How many students selected each distracter**



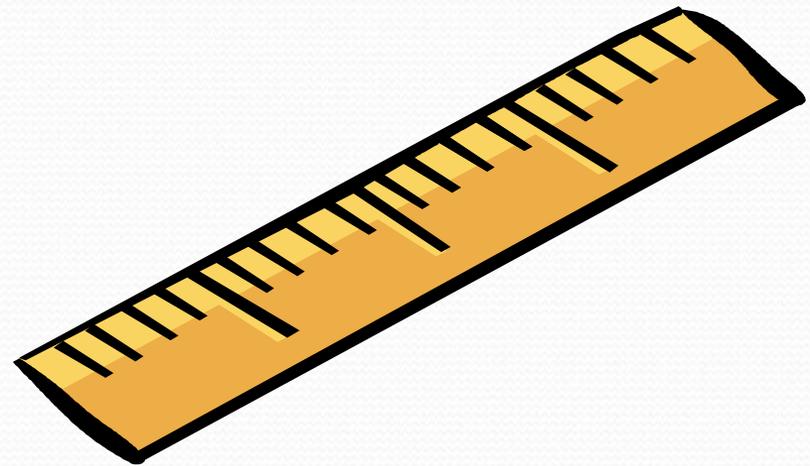
Reliability Coefficient



- An indirect measure of consistency
- Usually “KR 20”
- Goal for teacher made tests: .60-.80

Principles of Exam Improvement

- **12. Set an appropriate cut score**



NOT THE SAME:

- **St. George's EMT school is harder! They require an 80% on their final exam!**



- **Go to ACME's EMT school—they only require a 70%!**

St. George's EMT Class

- **Most items are in the 85-100% difficulty level range (NOT a BAD thing necessarily)**
- **Most items are recall level**
(This IS a BAD thing if they are to function in the field!)



ACME EMT Class

Some items are in the 40-60% or 60-80% difficulty range

- **Problem solving items included**



MY EMT CLASS

- **Fall class final exam average is 89%**
- **Spring class final exam average is 82%**

WHY ARE THEY DIFFERENT?

-students -instructors -exam

Other resources

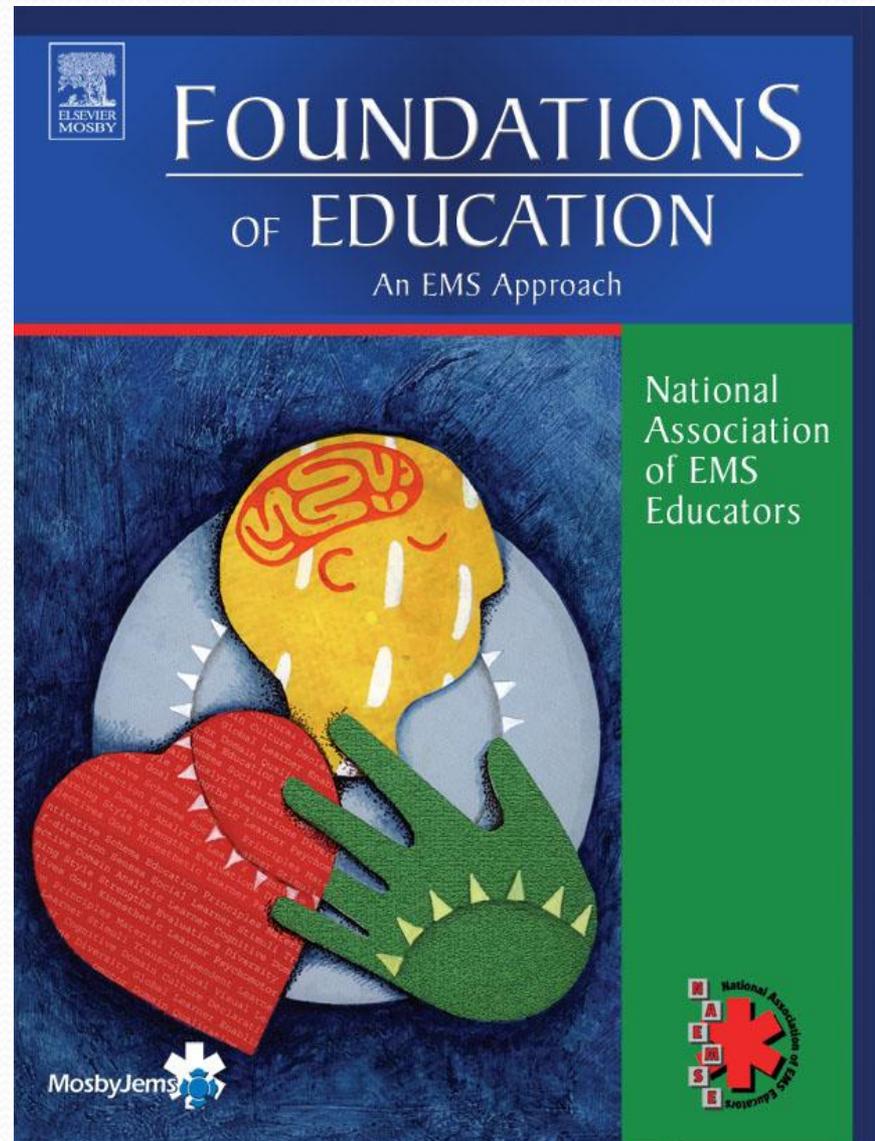
- National Association of EMS Educators
- www.Naemse.org
- NAEMSE meeting
Sept, 2014, Reno
- CECBEMS exam guidelines
- CoA/NAEMSE workshop
Evaluating Student
Competency



NAEMSE Text

Foundations of Education: An EMS Approach

2nd ed.



The road is LONG and Bumpy!!





QUESTIONS?

