



a Webinar for Program Directors & Faculty:
Written Exam Item Analysis



Presenters

Deb Cason, RN, MS, EMT-P, Board of Directors, CoAEMSP
Associate Professor, UT Southwestern Medical Center, Dallas, TX

Gordon Kokx, PhD(c), NRP
Assistant Director of Accreditation Services, CoAEMSP, Rowlett, TX

Michael Miller, EdD, RN, NRP, Board of Directors, CoAEMSP
Director, EMS Education and Assistant Professor, Creighton
University, Omaha, NE

Patricia L. Tritt, MA, RN, Board of Directors, CoAEMSP
Dean of Instruction, NCTI, Englewood, CO

Objectives



- Identify CoAEMSP requirements for item analysis for major exams
- Describe the 2 components that are required on item analysis
- Discuss the use of difficulty level and discrimination index (or RPBI) to evaluate exam items

Why is Item Analysis Important?



- Student evaluation important in all domains
- Written exams need to accurately evaluate student competencies
- Item analysis assists in the development of fair and accurate student evaluation for student improvement

Why is Item Analysis Important?



- Valid testing prepares students for another valid exam: NREMT or state exam
- Evaluation of exams help faculty assess teaching effectiveness
- Accurate assessment = better determination of continuation in paramedic school or needed remediation

What does the CoAEMSP require for High Stakes Exams?



- What ***is*** a **high stakes** exam?
- Medical director must review/approve items/exams
- Continue to review the evaluation process
- Validity/reliability with item analysis

What does the CoAEMSP require for High Stakes Exams?



- **Validity** → **RELEVANCY**
 - Must accurately reflect the job Paramedics perform
 - Includes higher level thinking for decision making
- **Reliability** → **CONSISTENCY**

What does the CoAEMSP require for Item Analysis?



1. Difficulty level
2. Discrimination index (RPBI)
3. Use of these tools to improve exams

Difficulty Index



Percentage of students getting the item correct

- Difficulty Index = 100% → *easy* item
- Difficulty Index = 46% → *difficult* item

Discrimination Index (or RPBI)



- **Correlation coefficient:**
measure of *direction* and *strength* of relationship between 2 variables
 - examples:
 - relationship between **smoking** and **COPD**
 - high fat diet and heart disease
 - Both have a positive and close relationship
 - In this case, relationship between **high performers** and selecting **correct answer** on the item

Discrimination Index (or RPBI)



Very easy items and **very difficult items** do not discriminate well and consequently this index may not be useful in those cases

Frequent Questions



What if there are only a handful of students in the class?

Does EVERY program need to do computer item analysis?

- The fewer the numbers, the less accurate the data
- Computer item analysis is NOT REQUIRED
- Some version of evaluation of exam items are required

Difficulty Index



- Does not mean **easy** is good or bad
- Does not mean **hard** is good or bad

- Too many easy items might not accurately reflect the job Paramedics perform

- Too many hard items might not accurately reflect the job Paramedics perform

Discrimination Index (or RPBI)



- A positive number means a positive relationship
- A negative number means a negative relationship
- The closer to 1.0 the better
- “Teacher made tests” will likely be in the 0.3 - 0.5 range (some positive correlation)
- National certification exams will likely be in the 0.7-0.9 range (high positive correlation)

Discrimination Index (or RPBI)



Negative is bad—there should be a relationship between exam performance and item performance

Sample Statistical Analysis



- Response Frequency
- Key
- Percentage of each response
- Point Biserial Coefficient (Discrimination)

Sample #1

No Discrimination



Response	Frequency	Percent	Point Biserial
A	0	0.00	---
B	0	0.00	---
C**	29	100.00	---
D	0	0.00	---
	29	100.00	

Sample #2



Good discrimination between 2 answer options,
but no one drawn to C & D answers

Response	Frequency	Percent	Point Biserial
A	8	27.59	-0.60
B**	21	72.41	0.60
C	0	0.00	---
D	0	0.00	---
	29	100.00	

Sample #3



Solid item though no one drawn to answer A

Response	Frequency	Percent	Point Biserial
A	0	0.00	---
B**	21	72.41	0.69
C	1	3.45	-0.35
D	7	24.14	-0.57
	29	100.00	

Sample #4

All distractors working



Response	Frequency	Percent	Point Biserial
A	1	3.45	-0.35
B**	23	79.31	0.46
C	2	6.90	-0.20
D	3	10.34	-0.23
	29	100.00	

Sample #5

10% correct



Majority like the same answer, why?

Response	Frequency	Percent	Point Biserial
A	33	89.19	-0.13
B**	4	10.81	0.13
C	0	0.00	---
D	0	0.00	---
	37	100.00	

Sample #6

Excellent item



Response	Frequency	Percent	Point Biserial
A	13	35.14	-0.46
B**	20	54.05	0.58
C	3	8.11	-0.23
D	1	2.70	0.05
	37	100.00	

Sample #7 - *Poor* item Negative discrimination



Response	Frequency	Percent	Point Biserial
A	4	13.79	-0.26
B**	7	24.14	-0.30
C	0	0.00	---
D	18	62.07	0.45
	29	100.00	

Sample #8

Great item



Response	Frequency	Percent	Point Biserial
A	1	3.45	0.04
B	6	20.69	-0.35
C	3	10.34	-0.31
D**	19	65.52	0.48
	29	100.00	

What about commercial exams?



If a program chooses to use a commercial or vendor test bank, it needs to measure student performance on test items to determine if:

1. The topic was understood and covered by the instructor
2. The students did not understand the material
3. The question(s) had bias, were poorly written, or miskeyed

The biggest challenge the CoAEMSP often sees is that programs do not use the products for self-analysis at a program-level!

What about commercial exams?



What evidence does CoAEMSP/CAAHEP require if a program is using a commercial or vendor product?

1. The program is measuring performance at the program level
 - Why did so many students miss this question?
2. Changes made to poorly performing questions
 - New lesson plans
 - Change of instructors
 - Student remediation
3. During a site visit, be prepared to demonstrate “*how* the program used the information provided by the vendor to make changes to enhance student learning.”



Remember:

CoAEMSP/CAAHEP does not recommend or endorse a specific vendor or product.

Is defense of setting the cut score a CoAEMSP requirement?



- No, but it is a good idea for you to evaluate!

Does CoAEMSP require reliability studies?



- No, but the information may be helpful for you!
- KR20 & KR21 examples of test reliability statistics
- Similar to item analysis but considers all items of the exam, student performance on each item, and variance
- Index ranges from 0.00 to 1.00
 - 0 means you are measuring many unknown factors
 - 0.60 or above means the exam is producing reliable scores

Other Considerations to Improve Exam Development



- Provide a good mix of knowledge, application and problem solving questions (“you suspect...” and “you first...”)
- Remember: Certification exams like the Registry predict a student’s ability to practice, not just recite knowledge
- Choose the best questions from a variety of test banks from different publishers and vendors
- Don’t be afraid to modify questions to meet the program’s needs
- Include questions that address critical patient conditions (i.e., things that kill people)

Resources: Exam Development and Item Analysis



- Consider the NAEMSE textbook, *Foundations of Education: An EMS Approach*
- Volunteer for test item writing meetings at NREMT
- Attend the Evaluating Student Competency workshop co-sponsored by CoAEMSP & NAEMSE
- CECBEMS exam construction
- If available, utilize testing resources at your school

If you have additional suggestions, email them to jennifer@coaemsp.org; she will distribute them to the group

Question & Answer



- Have a question?
Post it in the **Questions** section of the GoToWebinar control panel.
- All questions will be collated and a FAQs created.

Thank You!



Questions?

www.coaemsp.org

- Gordy Kokx
 - gordy@coaemsp.org
- Debra Cason
 - debra.cason@UTSouthwestern.edu
- Mike Miller
 - mikemiller@creighton.edu
- Patricia L. Tritt
 - patricia.tritt@comcast.net



a Webinar for Program Directors & Faculty:
Written Exam Item Analysis