MEDEDPREP

NREMT ITEM WRITING & TEST STRATEGY

Professional Reference for EMS Educators

7 Essential Rules for Item Quality

💡 **What This Solves:** Stops your team from writing questions with unclear wording, missing scope, or random wrong answers that don't teach anything.

1	Independent Lead-in	Stem answerable before seeing options. Tests clinical knowledge, not reading comprehension.
2	Single Defensible Key	Only one answer works given the constraints. If two could work, add vitals or scope.
3	Homogeneous Options	Match length, grammar, and units. Inconsistent options telegraph the correct answer.
4	Diagnostic Distractors	Wrong answers reveal real misconceptions. Shows what students misunderstand, not random guessing.
5	Scope in the Stem	Provider level (BLS/AEMT/Paramedic) must be explicit. Don't make students guess what's available.
6	Word Count Discipline	Keep stems under 120–140 words. Long stems test reading speed, not knowledge.
7	Track Performance	Monitor difficulty & discrimination after each exam. See stats card for interpreting these metrics.

🚩 Red Flags to Catch in Peer Review

- Two answers both work → Add vitals or scope constraint
- Wrong answers are random → Replace with common student mistakes
- Can't answer without seeing options → Rewrite the stem
- Scope isn't stated → Add "BLS unit" or "Paramedic" to stem

✨ Quick Pearls

- ✨ If two answers work, you didn't fence scope tight enough
- ✨ Random distractors don't teach—use real student errors
- ✨ Long stems test reading speed, not clinical knowledge

Process of Elimination (POE) Speed Strategy

💡 **What This Solves:** Gives students a repeatable elimination sequence that works under exam pressure and prevents paralysis.

1. **Predict** — Read lead-in first; cover options and answer silently
2. **Disconfirm** — Re-read stem for details that contradict your prediction
3. **Kill by Rule** — Priority → Contraindication → Scope (in that order)
4. **Compare Survivors** — Least risk / highest benefit given presentation

💡 How to Teach Step 1

Ask students to distinguish between these questions:

"What should be done first?"
→ Tests priority sequence (oxygen first—keeps airway open)

"What will benefit the patient most?"
→ Tests understanding of definitive treatment (epinephrine—treats anaphylaxis)

Same scenario. Same answer choices. Different correct answer.
Priority isn't just "what saves them"—it's "what kills them FIRST."

Real Example: Anaphylaxis with severe respiratory distress
- "What should be done first?" → **Oxygen** (airway failure kills in minutes)
- "What will benefit most?" → **Epinephrine** (reverses allergic cascade)

Why? Oxygen keeps them alive long enough FOR the epi to work. Epi takes 3–5 minutes to kick in.

⚠ Common Student Error

Asking "What would YOU do?" gets different answers based on who's answering.
Train them to ask: **"What should the [EMT/AEMT/Paramedic] do?"**
→ Forces scope awareness and removes personal bias.

✨ Quick Pearls

- ✨ Priority error kills faster than wrong medication choice
- ✨ Contraindications = instant elimination (no exceptions)
- ✨ When in doubt, protect the airway first

Understanding Item Statistics

💡 **What This Solves:** Takes the mystery out of post-exam numbers. You don't need a PhD—just these two metrics to make smart decisions about your item bank.

p-value (Item Difficulty)
This is just the percentage of students who got the question right. Written as a decimal (0.00 to 1.00) or percent (0% to 100%).

What the numbers mean:
- p = 0.90 (90% correct) → Too easy, everyone's getting it
- p = 0.60 (60% correct) → Good difficulty for most items
- p = 0.30 (30% correct) → Hard question, only strong students pass
- p = 0.10 (10% correct) → Way too hard or broken

What you want:
- Most items: p = 0.50–0.70 (moderate difficulty)
- Some stretch items: p = 0.30–0.49 (challenging)
- Few anchors: p = 0.71–0.85 (easier, builds confidence)

Point-biserial (Discrimination)
This tells you if the item separates your strong students from weak ones. It's a correlation number from -1.00 to +1.00.

What the numbers mean:
- ≥0.30 = Excellent (strong students get it, weak students don't)
- 0.20–0.29 = Good (acceptable for most uses)
- 0.10–0.19 = Weak (barely separating students—needs work)
- <0.10 = Problem (not working—fix or toss it)
- Negative = Red flag (weak students passing, strong students failing—miskey or trick question)

Plain English: If your best students are missing it at the same rate as your weakest students, the question isn't testing knowledge—it's testing luck or reading tricks.

✨ Important Note About Terminology

✨ **Point-biserial** and **discrimination index** are NOT the same thing. Point-biserial is a correlation coefficient (what we use here). Discrimination index is calculated differently and uses different cutoff values. Most modern testing software reports point-biserial—that's what these guidelines are for.

Item Triage Decision Grid

💡 **What This Solves:** Post-exam bank review in 30 seconds per item. Look at the grid, find your item's location, do what it says.

	Low Point-Biserial (<0.10)	High Point-Biserial (≥0.20)
Low p (0.20–0.40)	**RETIRE / REWRITE** Hard question that confuses everyone equally. Likely has ambiguous wording or a defensible distractor.	**KEEP (Advanced)** Difficult but clean. Strong students get it, weak don't. Good for stretching top performers.
Mid p (0.50–0.70)	**REVISE** Moderate difficulty but not discriminating. Clarify stem, rebuild distractors to target real student errors.	**KEEP (Core Zone)** Perfect. Healthy difficulty with strong discrimination. Most of your bank should be here.
High p (0.75+)	**WATCH / TUTORIAL** Too easy and not discriminating. Good for confidence-building early in course. Monitor for answer leakage.	**KEEP (Anchor)** Easy but still separates students. Good for preventing zero-scores and test calibration.

📊 Quick Decision Rules

High p + High biser → Keep it (easy but still shows who knows their stuff)
Low p + Low biser → Fix or toss (hard AND confusing = broken)
Negative biser → Suspect miskey (your best students are failing while weak students pass)
Mid p + Low biser → Your distractors aren't attracting the right mistakes

🚨 How to Fix Low-Discrimination Items

- Shorten the stem—remove double negatives and extra words
- Add ONE deciding piece of info (vital sign, history detail, physical finding)
- Make sure scope is crystal clear in the stem
- Replace at least one random distractor with a common student error
- If you can't fix it in 10 minutes, retire it and write a new one

✨ Bank Management Pearls

✨ Aim for 60-70% of your bank in the "mid-p + high-biser" core zone
✨ Low-discrimination items waste everyone's time—they don't measure anything
✨ Don't defend a broken item because "it tests important content"—write a better one
✨ Track items over time—if p-value jumps suddenly, suspect answer leakage