Appendix A

# ASE Meeting Primer 2026

A Practical Guide for Attendees at All Levels

Based on the abstracts to be presented at the 2026 ASE Annual Meeting, the following primer was prepared, using artificial intelligence, to help attendees prepare for the meeting and to discuss the presentations. All content has been reviewed for accuracy and tailored for this audience.

## **Part 1:** The Three Main Research Approaches

### Quantitative Research: Measuring and Comparing

- **What it is:** Studies using numerical data to identify patterns, compare groups, and test relationships.
- **Common types you'll see:**
  - **Cross-sectional studies** - Snapshots at one point in time (surveys, performance assessments)
  - **Retrospective cohort studies** - Looking back at existing data (ACGME case logs, NSQIP data, workplace assessments)
  - **Experimental/quasi-experimental** - Comparing interventions (AI feedback vs. faculty feedback)
- **When you'll see it:** Assessment validation, curriculum comparisons, any study asking "does X improve Y?"

### Qualitative Research: Understanding Experiences

- **What it is:** Research exploring the "why" and "how" through interviews, focus groups, or document analysis.
- **Common approaches:**

- **Thematic analysis** - Identifying recurring patterns in text
- **Grounded theory** - Allowing theory to emerge from data
- **Content or document analysis** - Examining OR dialogue, letters of recommendation, feedback narratives

- **When you'll see it:** Studies of lived experiences, decision-making processes, cultural factors in education. Look for terms like "thematic analysis," "phenomenology," "constant comparative method."

### Mixed-Methods: Combining Both Worlds

- **What it is:** Intentionally combining quantitative and qualitative data for comprehensive understanding.
- **When you'll see it:** Comprehensive curriculum evaluations, needs assessments, implementation studies tracking both outcomes and experiences.

## **Part 2:** Theoretical Frameworks Shaping the Field

### Competency-Based Medical Education (CBME) & EPAs

- **Core principle:** Progression based on demonstrated ability, not time or years in training
- **Key concepts:**
  - **EPAs (Entrustable Professional Activities)** - Discrete tasks trainees can perform independently
  - **Nested EPAs** - Smaller components within broader competencies
  - **Entrustment decisions** - Determining when learners can work with less supervision
- **Why it matters:** Supports individualized training and graduated autonomy. Many studies show proficiency scores outperform PGY in predicting performance.

### Deliberate Practice Theory

- **Core principle:** Expertise develops through focused practice with immediate, specific feedback.

- **Application:** Underlies most AI feedback tools, simulation curricula, and skills training programs. Timing and specificity of feedback, not just the tool itself, drive improvement.

### Cognitive Load Theory

- **Core principle:** Working memory has limited capacity. Minimize extraneous cognitive load to maximize learning.
- **Application:** Studies examining how imposter syndrome, multitasking in the OR, or poorly designed instruction create cognitive overload. Scaffolding curricula intentionally balance task complexity.

### Simulation-Based Learning Theory

- **Core principle:** Deliberate practice in controlled environments with structured feedback and defined benchmarks.
- **Application:** Box trainers, VR, cadaver labs, standardized patients. Look for proficiency-based training with clear performance standards.

### Situated Learning

- **Core principle:** Learning occurs in authentic contexts through participation in real practice.
- **Application:** Validates studying the OR as a learning environment. Intraoperative dialogue studies, real-time feedback capture.

## **Part 3:** Statistical Methods Decoded

### Describing Your Data (Descriptive Statistics)

- **Mean** = average (can be skewed by outliers)
- **Median** = middle value (better for skewed data)
- **SD** = how spread out the data is

## Comparing Groups

- **T-tests:** Compare averages between two groups (resident vs. faculty perceptions)
- **ANOVA:** Compare averages across three or more groups (PGY-1 vs. PGY-3 vs. PGY-5)
- **Chi-square ($\chi^2$):** Compare proportions between groups (male vs. female interest in subspecialty)
- **P-value:** Probability the difference happened by chance. $p < 0.05$ typically means "statistically significant" (likely real, not random)

## Measuring Relationships

- **Correlation (Pearson's r, Spearman's rho):** Do two variables move together?
  - **Range:** -1 to +1
  - Example: Word count in feedback positively correlated with quality (rho=0.81)
- **$R^2$ (R-squared):** How much variance does the model explain?
  - Example: $R^2$=0.52 means the model explains 52% of performance differences
- **Regression:** Predicts one variable from others while controlling for confounds
  - **Linear regression:** Predicting continuous outcomes (operative time improvement)
  - **Logistic regression:** Predicting categorical outcomes (pass/fail)
  - **Mixed-effects models:** Accounting for clustering (multiple ratings per trainee)
- **Odds Ratio (OR):** The odds of an outcome
  - Example: "Junior faculty less likely to provide appreciative feedback (OR=0.36)"

## Measuring Reliability & Agreement

- **Cronbach's alpha (α):** Internal consistency of a survey
  - 0.70 is generally acceptable
  - Example: Survey reliability α = 0.68–0.94
- Intraclass Correlation Coefficient (ICC): Agreement between raters

    - Example: ICC=0.93 (95% CI: 0.79-0.995) = excellent agreement
- **Cohen's Kappa (κ):** Agreement between two raters
    - 0.75 = substantial agreement
    - Used for AI vs. human coding comparisons

## Evaluating Test Performance

- **ROC curves and AUC (Area Under Curve):** How well does a model distinguish between categories?
    - AUC = 1.0 is perfect
    - AUC = 0.5 is no better than chance
    - Used for AI models distinguishing novice vs. expert
- **Sensitivity:** Ability to correctly identify positive cases
- **Specificity:** Ability to correctly identify negative cases

## Effect Size: How Big Is the Difference?

- **Cohen's d:** Standardized measure of difference magnitude
    - Small = 0.2, Medium = 0.5, Large = 0.8
    - Example: d=0.97 for AI vs. control feedback = large effect

## Assessing Qualitative Rigor

- **Saturation:** Interviewed enough participants that no new themes emerge
- **Multiple coders:** Independent analysis by 2+ researchers
- **Member checking:** Participants validate interpretations
- **Thick description:** Rich quotes supporting themes
- **Reflexivity:** Researchers acknowledge their own biases