Council for the Accreditation of Educator Preparation

Establishing Content Validity

Dr. Stevie Chepko, Sr. VP for Accreditation
Stevie.chepko@caepnet.org

Content Validity Defined

- The extent to which a measure represents all facets of a given construct
  - Extent to which an indicator measures what it was designed to measure
  - Constructs include the concept, attribute, or variable that is the target of measurement
  - Estimate of how much a measure represents every single element of a construct
  - Used to assess constructs or domains
    - Based upon an analysis of the body of knowledge surveyed
    - Refers to the degree to which the content of the indicator reflects the content domain of interest

Determining the Body of Knowledge for the construct to be measured

- Level of subjectivity exists in determining content validity
- Qualitative in nature
- Requires a degree of agreement among "experts"
  - Requires the use of recognized subject matter experts
  - Based on the judgment of subject matter experts
  - Relies on individuals who are familiar with the construct such as –
    - Faculty members
    - EPP based clinical educators
    - P-12 based clinical educators
  - Ask the fundamental question – "Do the indicators really assess the construct to be measured?"

Aligning Indicators to Construct

- Indicators must assess some aspects or segment of the construct
- Indicators must align with the construct
- Example:
  - In an online business, an important construct could be "Customer Service"
    - Survey is developed to measure customer satisfaction with the service
    - Questions must measure/assess some aspect of customer service to successfully determine the quality of service
      - Alignment is key
      - Direct measure of some aspects

Example for Customer Service

- Could you please take a moment and rate your experience with our company?
- Question 1:

  Instruction: On a scale of 1-5 with 5 being excellent, rate the timeliness of delivery once your order was placed.

  Question 2:

  Instruction: On a scale of 1-5 with 5 being excellent, rate the affordability of the product that your ordered?

Which of the two questions is aligned with the construct to be measured?

Using Lawshe's Content Validity Ratio

- Indicators on assessments attempt to operationalize the construct to be measured
- Content validation approach requires judgment as to the correspondence of abilities (indicators) tapped by the assessment with abilities requisite for job success
  - Demonstrating the indicators on the assessment appropriately sample the content domain
  - **Question: How well do the indicators align with the construct to be measured?**

Lawshe (cont.)

- Performance domains:
  - Behaviors that are directly observable
  - Can be a simple proficiencies
  - Can be higher mental process (inductive/deductive reasoning)
  - Operational definition – Extent to which overlap exists between (a) performance on assessment under investigation and (b) ability to function in the defined job
  - Attempts to identify the extent of the overlap

Content Evaluation Panel

- Composed of persons knowledgeable about the job
  - Most successful when it is a combination of P-12 based clinical educators, EPP based clinical educators, and faculty
  - Each panel member is given the list of indicators or items independently
    - Ask to do the following
      - Rate the item as “essential”, “useful but not essential”, or “not necessary”
      - Items/indicators must be aligned with the construct being measured (think of the customer satisfaction survey)

Quantifying Consensus

- Quantifying consensus:
  - Any item/indicator which is perceived as "essential" by more than half of the panelists, has some degree of content validity
  - The more panelist (beyond 50%) who perceive the indicator as "essential," the greater the extent or degree of its content validity
  - Calculating the content validity ratio (CVR)

$$CVR = (n_e - n/2)/(n/2)$$

Quantifying Consensus (cont.)

$$CVR = (n_e - n/2)/(n/2)$$

n_e = number of panelists indicating "essential"

N = total number of panelists

If you have 20 panelists total and 12 indicated it was essential, what is the CVR?

Compare answer with CVR chart to determine CVR value based on the number of panelists

Quantifying Consensus (cont.)

- CVR is calculated for each indicator
- A minimum value of the CVR is based on the number of panelists and is on a CVR Table
  - CVR values range from -1.0 to + 1.0
  - The more panelists the lower the CVR value
    - For example –
      - 5 panelists requires minimum CVR value of .99
      - 15 panelists requires minimum CVR value of .49
      - 40 panelists requires minimum CVR value of .29
  - Allows for the retention or rejection of individual items

Defining the construct

- Need to define the construct to be measured
  - Agreement on the construct
  - Behaviors or strategies related to the construct
- For measuring candidates' effectiveness in teaching to college-and-career readiness - which of the following would not be essential?
  - Engaging students in learning experiences requiring critical thinking
  - Being prepared to teach each day
  - Creating learning experiences that require students to apply content knowledge across disciplines
  - Engaging students in learning experiences that require the summary and analysis of a written text

Worksheet on Indicators

- For the indicators identified on the worksheet, rank them as “essential”; “useful, but not essential”; and “not useful” for classroom management –
- Remember the indicators must align with classroom management skills

Council for the Accreditation of Educator Preparation

Inter-rater Reliability

Dr. Maria del Carmen Salazar

University of Denver Morgridge College of Education

Associate Professor, Teaching & Learning Sciences

Reliability & Inter-rater Reliability

- Reliability
  - The degree to which scores are consistent over repeated applications of a measurement procedure and hence are inferred to be **dependable** and **repeatable**. A measure is said to have a high reliability if it produces **consistent** results under consistent conditions (CAEP Accreditation Manual, 2015)

- Inter-rater Reliability
  - Degree of **agreement** among multiple raters (Gwet, 2012).

Challenge to EPPs

Developing and assessing equitable and effective teaching using observation instruments that maximize construct validity and inter-rater reliability

Objective of University of Denver Study

- Analyze measures of **reliability** across 4 facets
  - ✓ **Supervisor** (e.g., **inter-rater reliability**, internal consistency, bias)
  - ✓ Apprentice (e.g., distribution of ratings)
  - ✓ Item (e.g., variability of items)
  - ✓ Time (e.g., variability of apprentice performance across time)
- Analyze measures of **validity** (convergent validity)
- Identify **implications** for revising the FEET evaluation model and training for supervisors.

Methods & Analysis

- Empirical study to establish **reliability** of the FEET
  - ✓ Design and implement protocols and supervisor training
  - ✓ Develop procedures to estimate inter-rater reliability and internal consistency
  - ✓ Analyze results using FACETS software four-faceted Rasch model
  - ✓ Identify implications to revise supervisor training and FEET items

Results: Inter-rater Reliability

- Except for severity of ratings, supervisors showed good understanding and application of both the items and the rating scale.
  - ✓ **rater separation reliability (e.g., severity in ratings)**
  - ✓ consistency in ratings (e.g., central tendency, halo effect)
  - ✓ bias in ratings (e.g., items, subjects, rating categories)

Results

Supervisor	Measure (logit)	Interpretation
1	-0.59	Lenient
2	-0.35	Slightly Lenient
5	-0.12	Slightly Lenient
9	-0.09	Slightly Lenient
3	0	Target
8	0.12	Slightly Severe
4	0.16	Slightly Severe
7	0.26	Slightly Severe

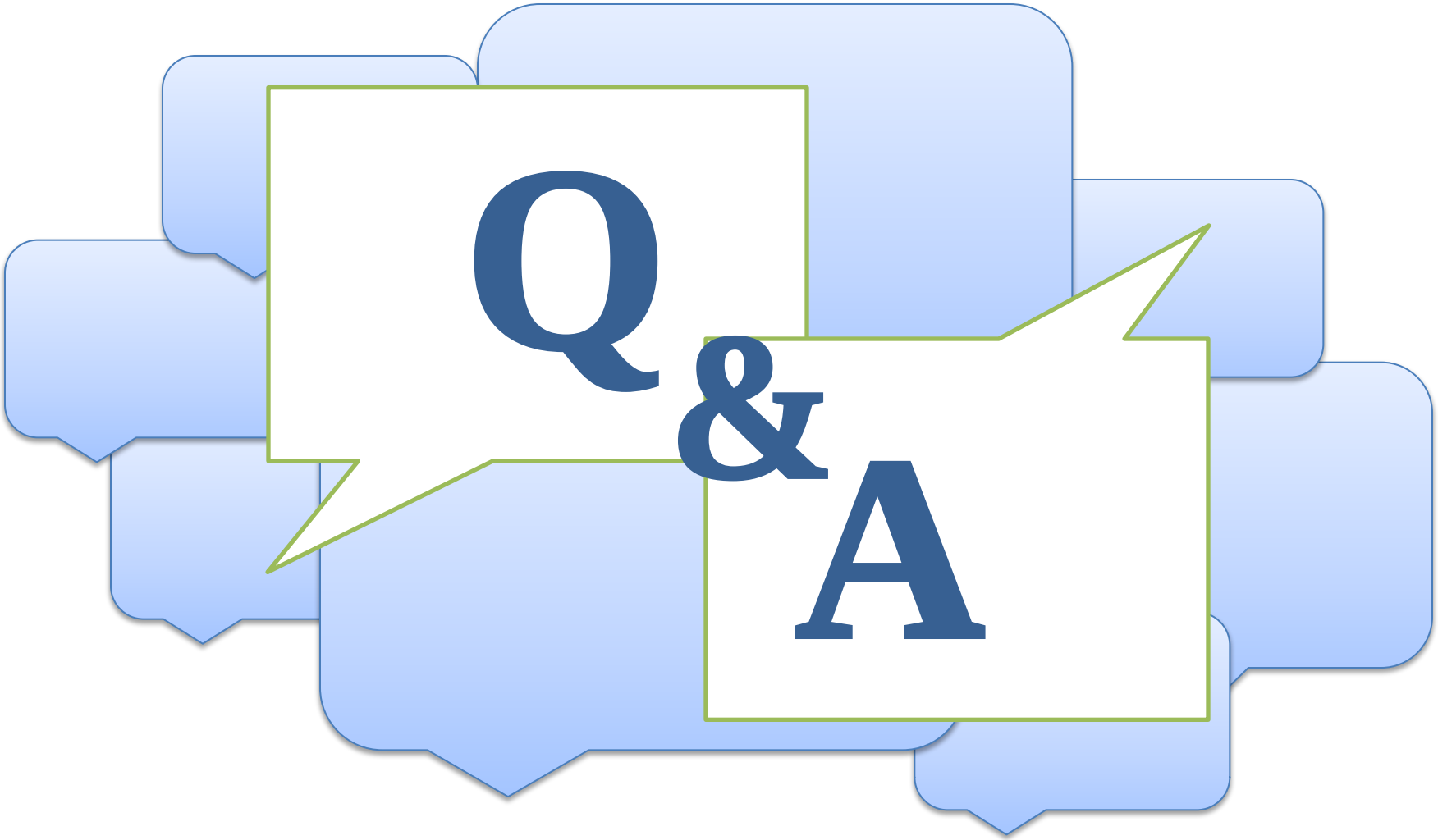
Continuous Improvement

- Inter-rater reliability needs to be an on-going focus of continuous improvement
  - ✓ Supervisor calibration, goal-setting, progress monitoring
- Revise FEET and supervisor training
- Replicate study
- Submit federal grant

CAEP – Establishing Inter-rater Reliability

- Need for EPPs to establish inter-rater agreement among raters. It gives a score of how much homogeneity or consensus, there is in the ratings given by judges.

- CAEP will take absolute percentage of agreement by raters using the same instrument and watching the same teaching performance.


Q&A