



Inquiry and Evidence:

What's the evidence about the evidence?

Mark LaCelle-Peterson, President
Teacher Education Accreditation Council

mark@teac.org
www.teac.org



Key Questions

Three questions drive the TEAC process:

1. What have candidates learned—do graduates' knowledge and performance match the faculty's claims?
- 2. What has the faculty learned from inquiry into the program's quality?**
3. Does institutional capacity support quality?



Arguments/rationale for the evidence

1. An argument for why the faculty believes its assessments are appropriate and meaningful measures of student performance,
2. A reasoned and empirical basis for the standards or criteria of success (cut scores).



Evidence about the evidence

Evidence of student learning in subject matter, pedagogical knowledge, and teaching skill, **must include evidence of the data's quality:**

- **How does the faculty know their interpretations are valid and reliable (or consistent and trustworthy)?**



What evidence is there?

Grades (content major, pedagogy, & clinical; course embedded assignments as assessments)

Scores on Standardized tests

(candidates' entrance, exit, and license scores; graduates' own students' test scores)

Surveys – students, alumni, employers

Ratings – portfolios, work samples, cases

Rates – hiring/tenure, certification, graduate study, awards, publications, NBPTS cert, etc.

How can the faculty assure reliable accuracy and valid interpretation of results?



Qualities of Evidence

- Evidence is reliable: chance is not a credible explanation for the findings
- Evidence is valid: rival explanations are not credible & the evidence is consistent with claims
- Evidence is of sufficient magnitude: when no other guidance is available, 75% guideline or heuristic is applied to the empirical maximum (the mean of the top ten percent observed)



TEAC's Five Standards of Evidence

Evidence must be:

1. **Representative**: sample must be appropriate
2. **Accurate**: at least 75% verified in the audit
3. **Consistent**: 75% of best observed ($r = .80$ [.68])
4. **Valid**: reliability & validity of evidence is known & adequate; rival explanations are ruled out
5. **Sufficient**: results meet established criteria or TEAC's 75% heuristic (.75 of mean of top 10%)



In other words...

Evidence must be:

1. **Fair**: a representative sample is required
2. **Trustworthy**: verified as accurate in the audit
3. **Reliable**: consistent w psychometric expectations
4. **True**: the validity of the evidence must be shown
5. **Sufficient**: meet established criteria or .75 of mean of top 10%



Example: Performance Rating

Imagine this: interns are evaluated by their mentor teacher and a faculty supervisor. How might we show that:

- The rating form gets at the right stuff? (Is it a valid measure of the construct or constructs?)
- Both raters understand the items and overall intent in the same way? (Do independent raters use the instrument consistently?)



Example: Performance Rating

Does our rating form get at the right stuff?

- Expert judgment: what do teachers say?
- Alignment with relevant standards
- Agreement with logically-related measures
- Is there sufficient variance in the evidence?

Is the instrument used consistently and does it produce consistent results reliably?

- Inter-rater agreement (**perhaps** correlation)
- Calibration exercises (let's watch a video...)



Evidence about the evidence: the technical stuff

The *Inquiry Brief* includes evidence about the reliability and validity of the reported evidence (QP 1.5).

All validity is local—i.e. validity is not a feature of an assessment instrument, but of its use (of the local meaning ascribed to scores).

Many tactics available—often, the simpler the better.



Handle data purposefully!

All measures for an individual candidate/completer should be linked.

Arguments for the validity of interpretations are enhanced by either convergent or divergent validity, as the case requires

Linking measures enables identification of 'pressure points' and learning from actions

Data from any single measure has limited 'reach'



Questions

Comments

Discussion

Thank you!