



When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments

Reference:

Cook DA¹, Kuper A², Hatala R³, Ginsburg S². *When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments*. *Academic Medicine*. 2016 Apr 5. [Epub ahead of print]

Author Institutions:

¹Mayo Clinic College of Medicine, ²University of Toronto, ³University of British Columbia

Tags

Clinical domain

General

Educational domain

Education research

Background

Which comment is more formative for a learner? A program director?

- "You are a 3.4 out of 5."
- "You used technical terms without explanation; however, your nonverbal skills engaged the patient to make them feel comfortable."

Both comments are assessments about patient–physician communication.

Quantitative data has long reigned supreme in medical education. The translation of a complex judgment into a standardized representation of that judgment allows for ease of aggregation of many judgments and statistical manipulation of the data to determine trends about the learner and the raters. While statistics can seem magical, there are lots of issues around the "truth" behind these numbers. (See KeyLIME episodes [49](#), [59](#), [78](#), [86](#)... and [63](#) for a counterpoint.)

As qualitative research methodologies from the social sciences have influenced medical education, the adoption of narrative within assessment programs has also increased. (See KeyLIME 91 for a partial description of the field notes instrument used by family medicine training programs in Canada.) So, how can the rigor of qualitative research inform the increasingly complex needs of programmatic assessment in this new era of medical education? That's the issue this paper tackles.

Purpose

“The purpose of this article is to:

- *articulate the role of qualitative assessment as part of a comprehensive program of assessment,*
 - *translate the concept and language of validity*
- ..., and*
- *elaborate on principles ... that relate the validity argument to both quantitative and qualitative assessment”*

Type of paper

Commentary

Key Points on the Methods

Best practices from qualitative research are aligned with both Messick’s (source of evidence to support the construct being assessed) and Kane’s (process and assumptions in collecting evidence, i.e. “the argument”) validity frameworks. This is a theory paper (appropriately) without a methods section.

Key Outcomes

Qualitative research is considered rigorous if it demonstrates:

- a theoretical frame
- an explicit question
- reflexivity (influence of assessors’ /analysts’ background and their relationship with learners)
- responsiveness in data collection
- purposive sampling
- thick description
- triangulation of data sources
- transparent, defensible analysis
- transferability
- relevance

Kane’s Inferences Framework & Qualitative Assessment

DOMAIN	(SELECTED) EVIDENCE OF RIGOR
Scoring (How observations become a narrative)	<ul style="list-style-type: none">• Questions / prompts stimulate rich response• Observer credibility• Varied reflexivity of assessors
Generalization (Aggregated data accurately reflects performance when observed)	<ul style="list-style-type: none">• Analysts are credible• “Auditable” analysis• Iterative/responsive/meaningful analysis (e.g. seeks counter examples)• Triangulation of data• Sufficiency of data
Extrapolation (Generalization of judgment extends to new “real-life” contexts)	<ul style="list-style-type: none">• Authentic / real-life data (and process for data collection)• Member check (stakeholders agree with final interpretation)• Analysis consistent with other external data
Implication	<ul style="list-style-type: none">• Interpretation leads to appropriate advancement /

(Acting on the judgment leads to meaningful decisions and minimal negative downstream effects)	remediation <ul style="list-style-type: none"> • Unintended consequence of assessment is favourable
--	--

For a superior review of evidence of rigor in qualitative assessments check out Tables 1 to 3 in the manuscript.

The authors acknowledge that their theory paper is limited by hypothetical examples of evidence of rigor not supported by a systematic search of the literature. However, this is the first description of a framework to defend the rigor of qualitative assessments, thus the limitation is over stated.

Finally, while narrative provides rich data, operational issues preclude it from appropriate use in “all” scenarios. Programmatic assessment is complex and benefits from an integration of quantitative and qualitative data, an argument made by the authors.

Key Conclusions

The authors conclude...

“We vigorously oppose the segregation of quantitative and qualitative assessment methods. Rather, we advocate a “methods-neutral” approach, in which a clearly stated purpose determines the nature of and approach to data collection and analysis ... we urge the use of a contemporary validity framework when evaluating any assessment, quantitative or qualitative... What matters most in validation is that evidence is strategically sought to inform a coherent argument that evaluates the defensibility of intended decisions.”

Spare Keys – other take home points for clinician educators

Language is complex and nuanced. A shout out to the authors for their constructivist approach that suggests that “reality” is a social phenomenon, interpreted through the shared meaning of words, words that evolve in definition and hold different interpretations between people.