METHODS FOR DOCUMENTING THE QUALITY OF YOUR ASSESSMENTS

M. David Miller

2nd Annual UF Assessment Conference

March 30, 2018

ARE WE USING QUALITY ASSESSMENTS?

ARE WE DOCUMENTING THE QUALITY?

HOW DO WE KNOW OUR ASSESSMENT IS A QUALITY ASSESSMENT?

You are the expert.

There are methods for documenting quality.

But you and your opinions are the clear foundation of quality assessment in your discipline!!

HOW DO WE (ASSESSMENT EXPERTS) DEFINE QUALITY?

- Validity
- Reliability
- Fairness

VALIDITY

- Strength of the interpretations of test scores for proposed uses of tests
- Most important part of test quality

WHAT TO DOCUMENT FOR VALIDITY (EVIDENCE)

- Content and skills required
- Response process
- Internal structure
- Relationship with other information
- Consequences of use

VALIDATION OF CONTENT BASED ON ALIGNMENT (DOCUMENTING)

Planning
- Objectives
- Student Learning Outcomes

Teaching
- Curriculum
- Instruction

Learning
- Assessment

EXAMPLE

- Ask peers to help by reviewing tests for content or response process

- Ask Simple questions (document results)
  - Do these items or scoring procedure align with expected uses or interpretations?
  - Do these items or scoring procedures align with the SLOs?
  - Do these items or scoring procedures ask students to respond in a way that is consistent with SLOs?

RELIABILITY

- Consistency of scores across replications
- Minimizing random errors

TYPES OF RELIABILITY

- Consistency in scoring if open-ended – written assignments or presentations (inter-rater)
- Consistency based on pool of items (internal consistency)

HOW TO HAVE HIGH INTER-RATER AGREEMENT

- Use Rubrics (see VALUE rubrics)
- Discuss how to use
- Provide examples at score points
- Work with multiple scorers to reach agreement

ints possible * 3 for grade)

	Outstanding (A)	Very Good (B)	Satisfactory (C)	Unsatisfactory (D/E)
Topic and framing	RQ evokes compelling grand challenge through multivariate perspective of a social scientist (10-9 pts)	RQ evokes compelling grand challenge while addressing much of its social complexity (8 pts)	RQ draws from a grand challenge while addressing some of its social complexity (7 pts)	RQ unclear, or evokes a peripheral social challenge, or fails to address its social complexity (6-0 pts)
Annotated bibliography	5 timely empirical research articles used and summarized accurately (10-9 pts)	4 timely empirical research articles used and summarized accurately (8 pts)	3 timely empirical research articles used and summarized accurately (7 pts)	Fewer than 3 empirical timely articles used or are summarized inaccurately (6-0 pts)
Lit review analysis and synthesis	Findings synthesized to offer significant insight (10-9 pts)	Findings synthesized to offer some insight (8 pts)	Findings summarized with little synthesis (7 pts)	Findings merely summarized (6-0 pts)
Data critique	Data expertly critiqued for: 1) sampling, 2) measurements, 3) data collection design and 4) analysis (10-9 pts)	Data critiqued for: 1) sampling, 2) measurements, 3) data collection design and 4) analysis (8 pts)	Data critiqued for 3 of 4 criteria (7 pts)	Data critiqued for less than 3 of 4 criteria (6-0 pts)
Grand Challenge solution	Findings used to offer a fresh, compelling, and realistic solution to RQ (10-9 pts)	Findings used to offer a compelling and realistic solution to RQ (8 pts)	Findings used to offer a realistic solution to the RQ (7 pts)	No realistic solution to the RQ offered (6-0 pts)
Future research	Study proposed to better answer RQ with data collection design and limitations described (10-9 pts)	Study proposed to better answer RQ with data collection design described (8 pts)	Study proposed to better answer RQ but data collection design inadequate (7 pts)	Any suggested study would not better answer RQ (6-0 pts)
Data charts	Findings conveyed through 3 clear, fair and interesting data charts (10-9 pts)	Findings conveyed through 2 clear, fair and interesting data charts (8 pts)	Findings conveyed through 1 clear, fair and interesting data chart (7 pts)	Chart missing, or is unclear, unfair or uninteresting (6-0 pts)
Mechanics	All sources cited per APA or MLA style; all charts labeled: axes, title, legend (10-9 pts)	All sources cited; all charts labeled: axes, title, legend (8 pts)	No more than one error or omission in attributing sources or labeling charts (7 pts)	More than one error or omission in attributing sources or labeling charts (6-0 pts)
Visual clarity	Meaning of research enhanced through compelling visual design (10-9 pts)	Meaning of research conveyed through competent visual design (8 pts)	Design conveys meaning of research if a little cluttered or boring (7 pts)	Design clashes with meaning of research or impairs readership (6-0 pts)
Oral presentation	Subject mastery shown in concise presentation and in answering questions (10-9 pts)	Subject familiarity shown in presentation and in answering questions (8 pts)	Subject familiarity shown in presentation (7 pts)	Little subject knowledge shown beyond written presentation (6-0 pts)

TRAINING TO SCORE CONSISTENTLY

1. Discuss scoring rubrics or other criteria in a group
2. Individually score assignments
3. Look at percent agreement
4. Meet and discuss where there is disagreement and why
5. Repeat until you have sufficient agreement

RELIABILITY FOR OBJECTIVE ASSESSMENTS

- Item analysis – which items work well – Canvas
- Reliability (Cronbach's Alpha)

RELIABILITY ON CANVAS: ITEM ANALYSIS

1. In course navigation, click Quizzes link
2. Click the title of the quiz you want to open
3. Open Quiz Statistics
4. View Statistics for Quiz
5. View Question Breakdown (Item Analysis):
   - Percent correct
   - Discrimination Index (above .2?)

FAIRNESS

- Many broad definitions including tests having the same meaning for different groups and empirical effects
- I think it means that scores have the same interpretations and usefulness for everyone
- Must first consider who are groups of interest – frequently demographics gender or ethnicity

SENSITIVITY OR OFFENSIVENESS REVIEW

- Ask peers with either content expertise or group representation to help review items, scoring rubrics,...
- Identify relevant groups
- Ask simple questions:
  - Do these items or scoring procedures contain any biased or offensive materials for group X?
  - Do these items or scoring procedures contain any stereotypes for group X?

FAIRNESS AND DATA

- Look at mean differences between groups on assessment…but often difficult to interpret

- Look at mean differences between groups on items…can provide helpful information about specific content and skills

SUMMARY

- You are probably already using quality assessments!
- Best to set up procedures to document the quality of your assessments!
- Two ways to think about this:
    - Expertise counts and collaborating with other experts helps. Review for each other.
    - Data can be informative so learn to use all of the features of Canvas and other programs when needed

SUMMARY

- Remember you want tests that lead to
  - Interpretable and useful scores (validity)
  - Consistency in measurement rather than random scores (reliability)
  - The same interpretability and usefulness for all students (fairness)