

UTTyler™

THE UNIVERSITY OF TEXAS AT TYLER

UTTyler
THE UNIVERSITY OF TEXAS AT TYLER

Foundations of High-Quality Assessments: Purpose, Clarity, and Fairness

Brandon L. Bretl, PhD
Assistant Professor
School of Education
University of Texas at Tyler

Foundations of High-Quality Assessment: Purpose, Clarity, and Fairness

- Match assessment to learning objectives and validity needs
- Apply practical criteria to judge your existing tests/surveys
- Revise items to reduce ambiguity and bias

Introductions



Dr. Brandon Bretl

Assistant Professor, School of Education

PhD in Human Development and Learning

University of Kansas

- Math and science teacher
- Researcher on state standardized science tests
- Cognitive and social psych research



Introductions

Why attending?

What assessments/surveys are you using, creating, or plan on creating?

Question

What makes an assessment good?

Definition

An assessment is an evidence system to generate information that reduces uncertainty.

A ***good*** assessment provides high quality evidence and the greatest reduction in uncertainty.

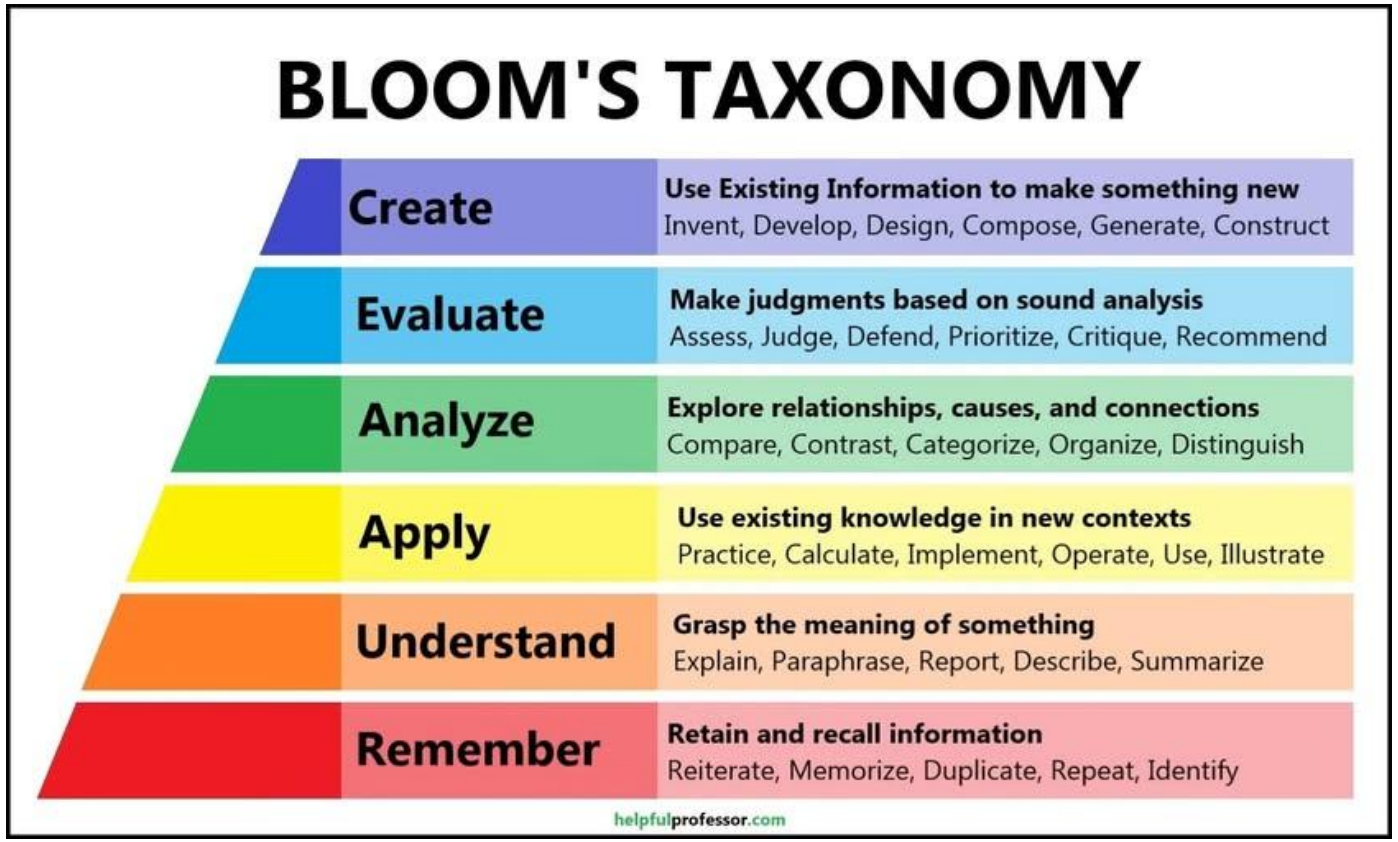
First Question to ask…

What do I want to reduce uncertainty about?

Purpose

- Epistemological
  - Knowledge
  - Conceptual understanding
- Latent Construct
  - Psychological constructs
- Fact gathering

Epistemological



Epistemological

- Starts with a standard, objective, or concept
- Pay attention to verbs
  - List, define, summarize, classify, use, solve, differentiate, justify, critique, create
  - For elementary: show, sort, tell, make, build, act out, draw, retell

Epistemological

Construct a scientific explanation based on evidence for how environmental and genetic factors influence the growth of organisms.

Can the student **construct** a scientific explanation?

Can test that *or* something that they can only do if they construct a scientific explanation…

Epistemological

“Given the table/graph on light hours and height for two varieties of radish, write a **claim** about which factor(s)—environment (light) and/or genetics (variety)—influenced growth. Cite **evidence** from the data (specific numbers), and give a **reasoning** sentence linking evidence to the claim using a correct principle (e.g., keeping other variables constant allows attribution).”

- **Scoring (0–2 each; total 6)**
- **Claim:** 0 = off-target; 1 = partial (names a factor); 2 = correctly attributes environment and/or genetics based on dataset.
- **Evidence:** 0 = none; 1 = general trend; 2 = precise, comparative values (e.g., “Variety B 24 cm > Variety A 18 cm under same light”).
- **Reasoning:** 0 = none/incorrect; 1 = vague; 2 = correct causal logic (controls/comparison).

Latent Constructs

Real measurements



Latent Constructs



Latent Factor Analysis



PATIENT HEALTH QUESTIONNAIRE (PHQ-9)

NAME:_____ **DATE:**_____

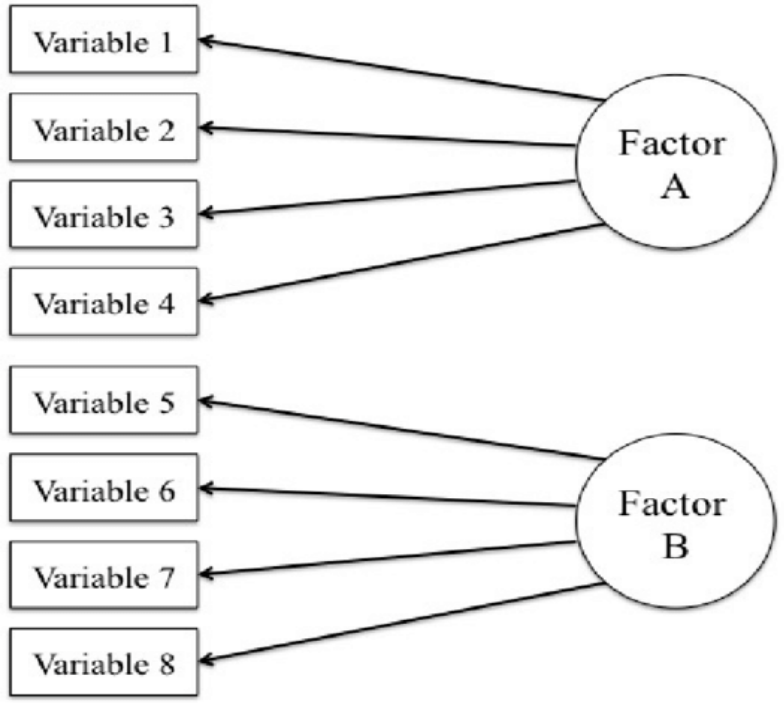
Over the last *2 weeks*, how often have you been bothered by any of the following problems?
(use "✓" to indicate your answer)

	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3

Latent Factor Analysis

- The latent factor is **causing** some variation in ratings of items (measurements)

Latent Factor Analysis



Latent Constructs

- Cognitive
  - General intelligence
  - Working memory
  - Processing speed
  - Executive function
- Personality
  - Openness
  - Neuroticism
- Self-efficacy
- Social Psych
  - Empathy
  - Right-wing Authoritarianism
- Clinical
  - Depression
  - Schizotypy
  - Impulsivity
- Developmental
  - Attachment style

Practical Considerations

- Clearly define the construct to be assessed
- Is it unidimensional or multidimensional?
  - One or several related ideas?
- When possible, use existing, valid and reliable instruments
- If creating your own…
  - Research
  - Draft
  - Get broad expert feedback
  - Pilot
  - Revise
- Pilot 6 items per construct, keep at least 4
- Again, the idea being that the construct will cause variation in ratings

Fact gathering

- BMI and VO2 max of student athletes
- Preferences
- Subjective experiences
- Direct views or opinions
- Demographics

Validity

Is an argument for how your assessment is measuring what you are claiming it is measuring.

Never valid or not valid…

Only better and worse validity arguments.

Does Johnny know the meaning of absolute value bars?

Does Johnny know the meaning of absolute value bars?

Complete the following:

1. $|-\pi r^2| =$
2. $|\sqrt{9} - 100| =$
3. $|\frac{-6}{12}| =$

Does Johnny know the meaning of absolute value bars?

Complete the following:

1. $|-6|$ =
2. $|6|$ =
3. $|-320|$ =
4. $|400|$ =
5. $|-12|$ =

Choosing the Right Format

- Multiple Choice Questions (MCQs)
  - Efficient
  - Ability to assess nuanced understanding and misconceptions
  - Risk of cueing
- Short answer
  - Authentic reasoning
  - Difficult to score reliably
- Performance task/project
  - Authentic
  - Requires rubric
- Surveys
  - Construct validity

Clarity and Reducing Bias

A bias in assessment is a feature of the assessment that causes a decrease in the reliability and/or validity of the information obtained.

- Gender
- Race/ethnicity
- Socioeconomic status
- Religion
- Etc.

Example, elementary reading

My uncle's field is computer programming.

Which sentence uses the word *field* in the same way?

A. The softball pitcher knew how to *field* her position.
B. They prepared the *field* by spraying and plowing it.
C. I know the *field* I plan to enter when I finish college.
D. The doctor used a wall chart to examine my *field* of vision.

Avoiding Bias

- Content Bias
  - Don’t assume cultural or life experiences
  - E.g., not everyone has been sailing
- Language Bias
  - Idioms and double negatives
  - Highlight or make **bold** words likely to be missed
- Avoid irrelevant difficulty
- Stereotype threat
  - Avoid asking for demographic information prior to assessment

Anatomy of an MCQ

In which region of the vertebrate digestive system does the majority of nutrient absorption into the blood stream occur?

Stem

a) Stomach
b) Colon
c) Small intestine
d) Large intestine
e) Esophagus

Distractors

Key

MCQs and Information

- Assessing multiple things in one MCQ provides less information on student knowledge
  - We don’t know “why” they got it wrong

Best Practices for MCQs

- Should be complete in information needed to answer
- Distractors should be…
  - Plausible
  - Grammatically consistent
  - Similar in length and structure
  - Distractors address common misconceptions
- Don’t use “all of the above” or “none of the above”
- Use 3-5 distractors

Your Turn: Practical Criteria

- Purpose Fit
  - Does the stated purpose match how scores will be used?
- Knowledge, Objective Alignment
  - Do items require the cognitive work named in the objective?
- Stakes
  - Does it require a high degree of accuracy, reliability?
- Fairness/Bias
  - No idioms, insider info, unnecessary reading?
- Actionability
  - You get exactly what you were looking for… Does it actually result in something actionable or conclusive?

Red Flag!

- Double negatives or double-barreled items.
- Cultural/insider references not taught in the course.
- MCQ options of noticeably different length or specificity.
- Distractors are not plausible misconceptions.
- “All/none of the above.”
- Using a perception survey as evidence of *achievement*.
- No plan for rater calibration on performance tasks.

Need additional help?

Thank you!

BBRETL@uttyler.edu