

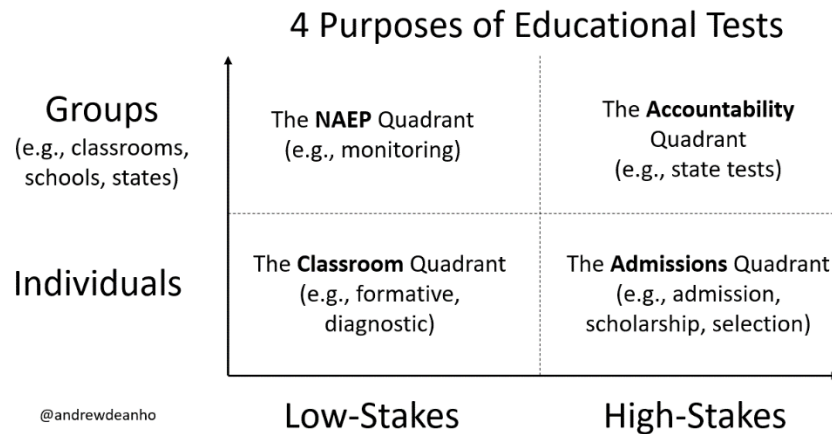
Questions to ask about educational tests

Andrew Ho, *Harvard Graduate School of Education*

(Last updated: 6/3/21)

Purpose: What are you trying to accomplish with test scores?

A 2x2 framework (see [Ho, 2014](#) for more about “purpose drift”):



Evidence: What validity evidence do you have that supports this use? (In my teaching, I call these, “The 5 Cs”) Five sources of validity evidence (AERA/APA/NCME Standards for Testing, [pp. 13-22](#))

- 1) **Content:** “Evidence based on test content”
 - a. What does the test measure? Can you give me some examples of test questions?
 - b. What are the standards, strands, or domains that are part of this content?
- 2) **Cognition:** “Evidence based on response processes”
 - a. Have you talked to kids while they take this test and asked them how they are reasoning through their responses?
 - b. How do you know that kids aren’t getting high scores using irrelevant tricks?
- 3) **Coherence:** “Evidence based on internal structure”
 - a. If you were to repeat the test, how much would you expect scores to change?
 - b. What evidence do you have that scores are stable enough for your proposed use?
- 4) **Correlation:** “Evidence based on relations to other variables”
 - a. What do your scores correlate with or predict? How strong is the relationship?
 - b. How do your scores differ from other similar variables or indices (e.g., SES)?
- 5) **Consequence:** “Evidence for validity and consequences of testing”
 - a. What theory suggests that your proposed use of tests is better than not testing?
 - b. What evidence do you have that testing has helped you to accomplish your goals?

Eight questions for additional context:

Population: How does the sample you have differ from the population you wish to measure?

Conditions: Under what conditions was the test administered? When? How? Where? Why? By whom?

Comparability: What evidence do you have that scores are comparable over time, grades, populations?

Stakes/Threat: What incentives or threats did test respondents perceive during administration?

Multiple Measures: What other measures will you use to inform the decision or proposed use, and how?

History: What is the historical legacy of this testing program or approach, including historical misuses?

Conflict of Interest: What testing agencies or organizations compensate you for advice or governance?

Transparency: Is your technical documentation public? Under what conditions do you allow access?

Inspired by John Willett and Judy Singer’s classic, *Questions to ask when reporting education research*, [here](#).