



# Assessment *for* Learning

Andrew Zieffler, Ph. D.

---

*Quantitative Methods In Education*

*Department Of Educational Psychology*

# Assessment

---

Tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know.

*National Research Council*

# Assessment

---

Tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know.

*National Research Council*



# Assessment

---

Tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know.

*National Research Council*

# Assessment

---

Tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know.

*National Research Council*

Observation ————— Action



Interpretation

Observation



Action

Interpretation

Observation



Action



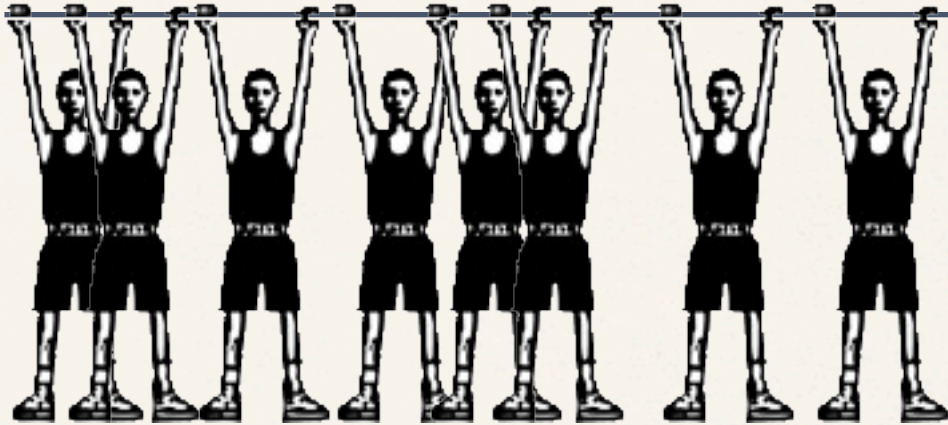
Validity Evidence



Interpretation

Observation

Action



Validity Evidence

# Purpose

---

- ❖ Clearly define your purpose
  - ❖ Progress Monitoring (formative assessment)
  - ❖ Objective / Instructional Feedback
  - ❖ Grading (summative assessment)
  - ❖ Placement

# Content Validity Evidence

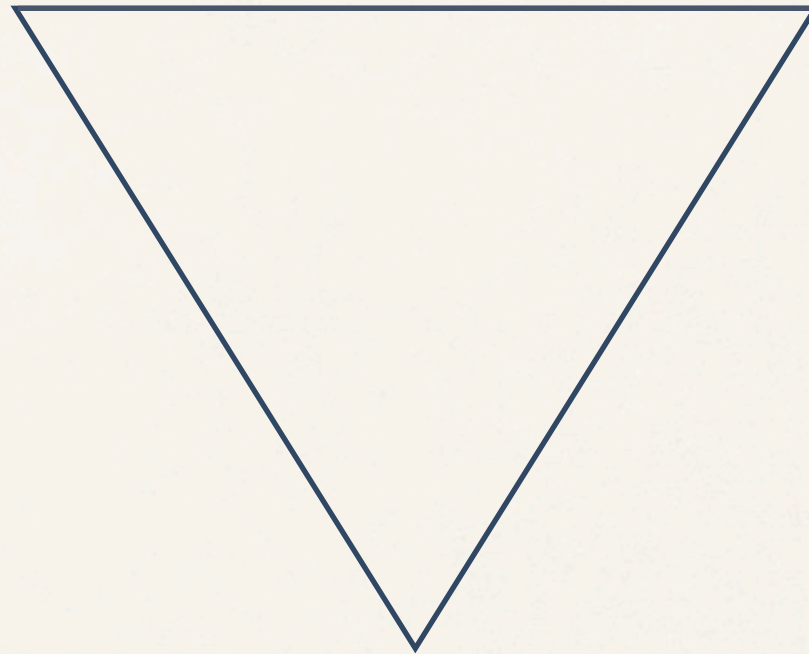
---

Content validity provides evidence of the relationship between test content and the construct to be measured.



Observation

Interpretation



Cognition

(National Research Council, 2001)

# Models of Learning

---

Assessment should be consistent with our understanding of learning in the subject matter—we need a *model of learning* to provide a guide for assessment design.

Allow us to describe the features of assessment tasks that illuminate aspects of knowledge, skills or abilities required for achieving content standards or learning objectives

# Models of Learning

---

- ❖ Ideas from research on teaching and learning statistics include:
  - ❖ Importance of context
  - ❖ Importance of sequencing tasks and knowledge structures
  - ❖ Importance of using multiple representations of ideas and concepts



# Item Writing

---

- ❖ Content
- ❖ Cognitive complexity
- ❖ Format

# Multiple-Choice (MC) Items

---

- ❖ Options should be plausible—common errors or misconceptions
- ❖ Use only the number of options you need *or* can develop (3 is sufficient)

Researchers surveyed 1,000 randomly selected adults in the U.S. A statistically significant, strong positive correlation was found between income level and the number of containers of recycling they typically collect in a week. Please select the *best interpretation* of this result.

- A. We can not conclude whether earning more money causes more recycling among U.S. adults because this type of design does not allow us to infer causation.
- B. This sample is too small to draw any conclusions about the relationship between income level and amount of recycling for adults in the U.S.
- C. This result indicates that earning more money influences people to recycle more than people who earn less money.



# Constructed Response (CR) Items

---

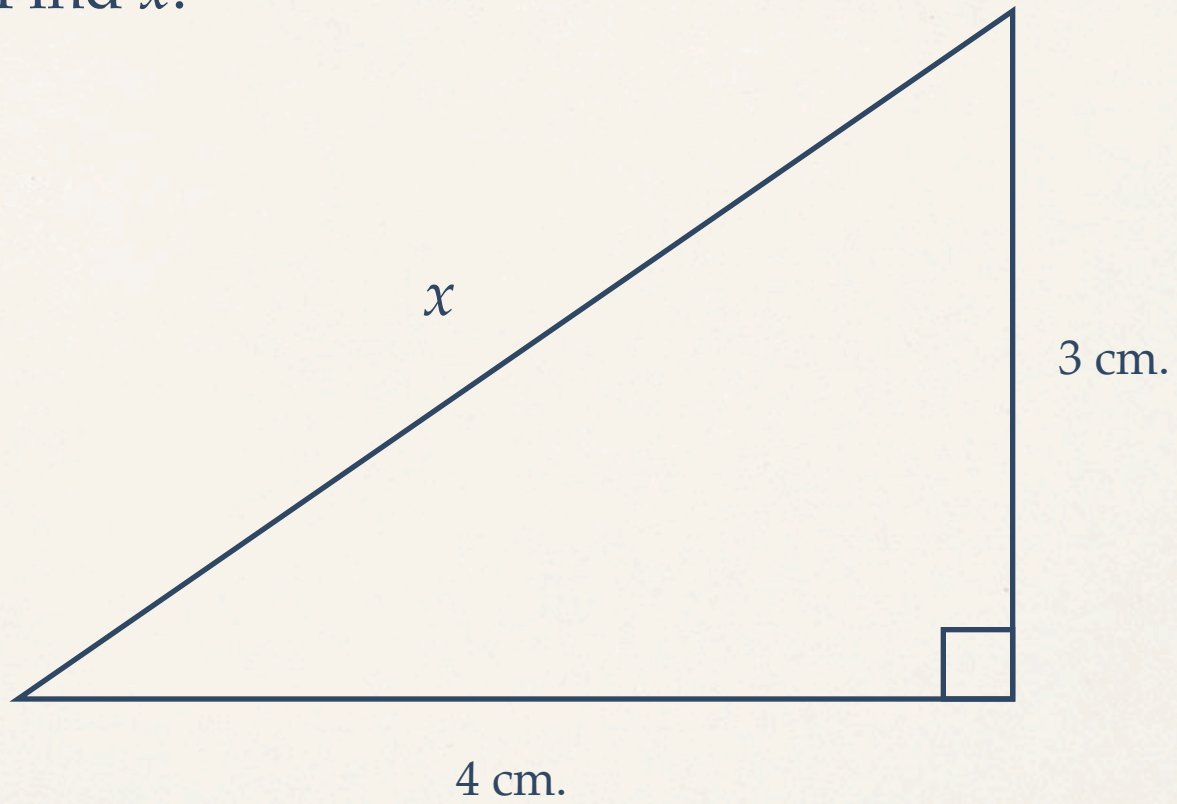
- ❖ Use CR tasks to assess thinking and skills that cannot easily be measured by MC items

# Caution

---

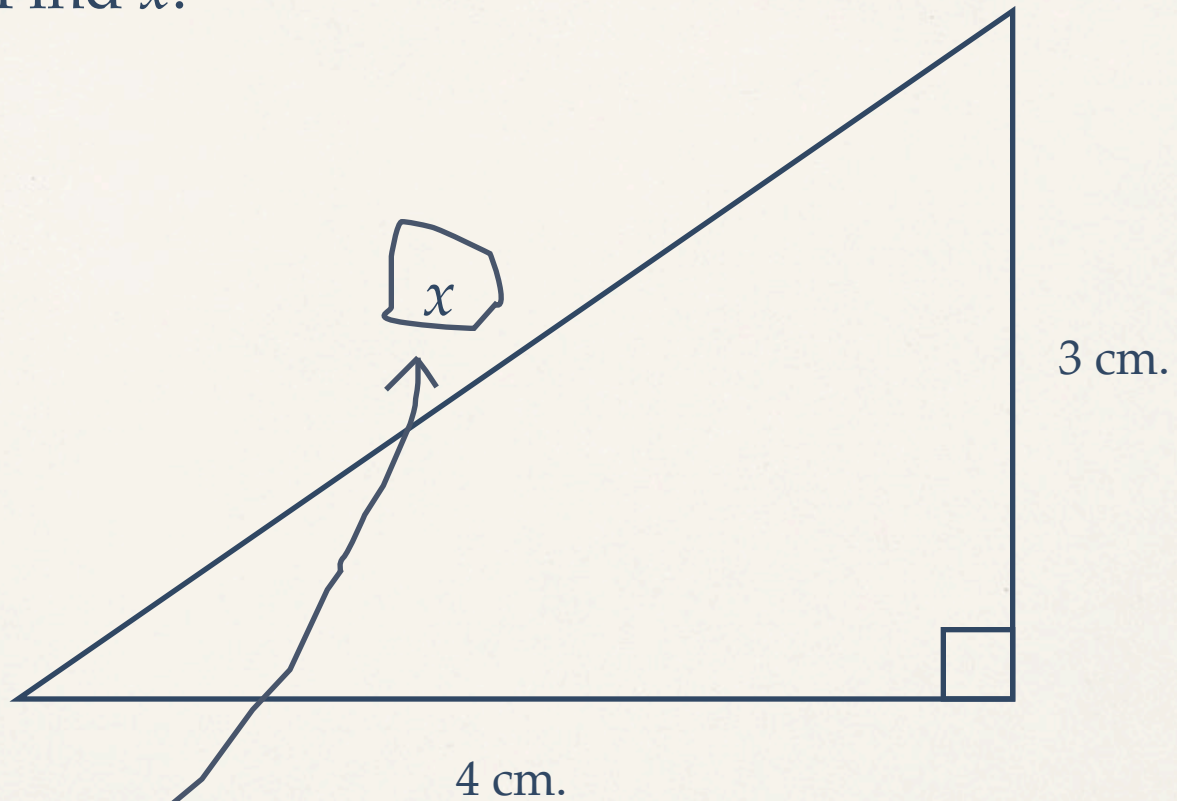
Avoid ambiguous task features

3. Find  $x$ .





3. Find  $x$ .



*Here it is*

# Caution

---

Assumptions necessary to respond correctly should be related to the content demands of the assessment.

Mr. Wilson and 3 friends dined at a popular restaurant. The bill was \$77 and they left a \$15 tip. Approximately what percentage of the total bill did they leave as a tip?

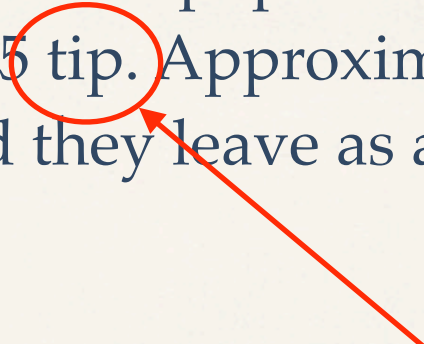
- A. 10%
- B. 13%
- C. 15%
- D. 20%
- E. 25%

Embretson (2007)



Mr. Wilson and 3 friends dined at a popular restaurant. The bill was \$77 and they left a \$15 tip. Approximately what percentage of the total bill did they leave as a tip?

- A. 10%
- B. 13%
- C. 15%
- D. 20%
- E. 25%



Some students might not know what this is

Embretson (2007)

Mr. Wilson and 3 friends dined at a popular restaurant. The bill was \$77 and they left a \$15 tip. Approximately what percentage of the total bill did they leave as a tip?

- A. 10%
- B. 13%
- C. 15%
- D. 20%
- E. 25%

May seem very unreasonable

Embretson (2007)

Mr. Wilson and 3 friends dined at a popular restaurant. The bill was \$77 and they left a \$15 tip. Approximately what percentage of the total bill did they leave as a tip?

- A. 10%
- B. 13%
- C. 15%
- D. 20%
- E. 25%

Are these the same?

Embretson (2007)



# Minimizing Irrelevant Task Performance

---

- \* Questions should require students to consider novel contexts
- \* Use reference materials (graphical displays) that are authentic

# Caution

---

Learning can (and will) take place during an assessment.

Researchers surveyed 1,000,000 randomly selected adults in the United States. A statistically significant, strong positive correlation was found between income level and the weekly number of containers of recycling collected. Please select the *best* interpretation of this result.

- A. The researchers *can* conclude that earning more money influences adults in the United States to recycle more.
- B. The researchers *cannot* conclude that earning more money influences adults in the United States to recycle more.

Explain your choice.



Researchers surveyed 1,000,000 randomly selected adults in the United States. A statistically significant, strong positive correlation was found between income level and the weekly number of containers of recycling collected.

Why is it NOT appropriate for the researchers to conclude that income level causes recycling behavior?

# Item Review

---

- ✦ Cognitive item-level analysis
  - ✦ Difficulty
  - ✦ Discrimination
- ✦ Verbal reports of examinees

# Some Advice

---

- ❖ Clarify expectations and standards
- ❖ Use accepted principles of item writing
- ❖ Peer review items
- ❖ Listen to your students



American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Embretson, S.E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Research*, 36(8), 449-455.

Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academy Press.

Rodriguez, M.C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3-13.