

The Answer is B: Better Multiple Choice Questions

Fall 2021 Professional Development Days
Jinger Doe, Jon Reddick-Lau, Melissa Deadmond

Workshop Learning Outcomes

- Distinguish between reliability and validity, and explain their importance to learning outcomes assessment.
- Develop and apply a table of specifications towards writing and reviewing multiple choice (MC) exams.
- Describe, interpret, and apply difficulty and discrimination indices in MC item analysis.
- Explain the relationship between item analysis and validity and reliability of MC exams.
- Apply best practices in writing MC items (questions) to help improve validity and reliability.

Reliability

Reliability - whether the results from your test measures something consistently

- Why is this important?
 - Ideally we want our tests to yield as close to a “true score” of what a student really knows, but sources of error contribute to unreliability
 - Method errors - source of error resides in the testing situation
- Improve reliability by:
 - Standardizing instructions across all settings when you give the test
 - Increasing number of items/observations
 - Moderating easiness and difficulty. Too easy/hard not accurate picture of performance
 - Deleting unclear items (questions) following analysis

Poor Reliability



Good Reliability, but . . .



Modified from http://media.acc.qcc.cuny.edu/faculty/volchok/Measurement_Volchok/Measurement_Volchok6.html

Validity

Validity - the extent to which conclusions drawn from your test are appropriate, meaningful, and useful, i.e. the test does what it is supposed to

- First establish reliability. A test cannot be valid before it's reliable.



Modified from http://media.acc.qcc.cuny.edu/faculty/volchok/Measurement_Volchok/Measurement_Volchok6.html

- Mostly concerned with **content validity** - your test items (a sample) truly reflects a universe (chapter, class time, assignments) of test items on a topic

Multiple Choice Achievement Tests

Writing good multiple choice (MC) items (questions) is not easy!

- Can take a long time
- MC items must have good distractors (well-written, focused) and one clear, correct answer

But there are advantages:

- Scoring is easy and reliable
- Can measure learning outcomes at almost any level of Bloom's taxonomy (if well-written)
- You can analyze the items to see if what they do what they're supposed to (validity) and refine further so they perform better (improve validity)



Item Analysis

To be reliable and valid, MC items (questions) should:

- Be clearly-written
- Not be too easy or too *difficult*
- *Discriminate* between those students who really know the answer and those who don't

Item analysis - Consists of two indices for each item (question) that suggest whether that item discriminates between those who know the material and those who do not

- Difficulty index
- Discrimination index

Item Analysis: Difficulty and Discrimination Indices

Difficulty Index - the number of students who got an item correct out of the total who answered the question

High and low groups - the top and bottom 27% of total exam scores (students)

$$D = (S_H + S_L)/T$$

Ideal D = 0.5. The more it differs, the more poorly the item discriminates.

D > 0.5 means the item may be too easy.

D < 0.5 means the item may be too hard.

Where:

D = difficulty index

S_H = number of students in high group who answered item correctly

S_L = number of students in low group who answered item correctly

T = total number of responses for the item

Discrimination Index

Discrimination Index - describes how effectively an item discriminates between the high and low groups

$$d = (S_H - S_L) / (0.5 * T)$$

Where:

- d = discrimination index
- S_H = number of students in high group who answered item correctly
- S_L = number of students in low group who answered item correctly
- T = total number of responses for the item

d ranges from -1.0 to 1.0
0.0 = no discrimination
1.0 = perfect discrimination
-1.0 = perfect discrimination,
but not how you want it

General Interpretations: Difficulty and Discrimination Indices

Difficulty Index

Range	Interpretation	Action
0 – 0.25	Difficulty item	May need to revise or discard item
0.26 – 0.75	Generally Appropriate	Retain item
0.76 – 1.0	Easy item	May need to revise or discard item

Discrimination Index

Range	Interpretation	Action
-1.0 to -0.5	Item discriminates but not in a desired way. The low group answered it correctly more often.	May need to discard the item
-0.51 to 0.45	Item does not discriminate well	May need to revise the item
0.46 to 1.0	Item discriminates between high and low groups	Include the item

Table of Specifications

Table of Specifications – A 2-dimensional table that describes topics/ SLOs to be covered by a test and the number of questions/ points that will be associated with each topic/SLO.

- **Purpose** – To identify the areas of achievement to be measured and ensure that a fair and representative sample of questions appear on the test.
- Allows instructors to focus on the key areas and weight those areas based on their importance and time spent on them in class.
- Helps insure that an assessment measures what the assessment intends to measure - **CONTENT VALIDITY!!**

Topic	Amount of Time Spent in Class	Number of Items/ Questions	Percent of Items/ Questions on Test (Should be approximately equal to Amount of Time Spent in Class)

Example 1

Table of Specifications

Steps to Creating a Table of Specifications

1. Identify the domains/topics/SLOs to be measured on the next exam. These will vary in breadth from large domains to more specific SLOs.
2. Break the domains/topics/SLOs into key or fairly independent parts. SLOs can even serve as these key or independent parts.
3. Document the time spent in class on these topics/SLOs.
4. Construct the table.

SLO	Amount of Time Spent in Class	Level of <u>Blooms Taxonomy</u> (with <u>key words to help identify which level</u>)	Number of Items/ Questions	Percent of Items/ Questions on Test (Should be approximately equal to Amount of Time Spent in Class)

Example 2

Difficulty and Discrimination Indices in Canvas

Examples:

- A question with low difficulty and discrimination indices.
 - What does this mean?
 - Why might this be intentional?
- Two questions
 - Identical difficulty indices (0.42)
 - One with a positive discrimination index
 - One with a negative discrimination index

Writing good multiple choice questions: Anatomy

Susie Science has discovered a mutant form of insulin that lacks a signal peptide. What will be the final cellular destination of the mutant insulin?

Stem

Answer

Distractor

Distractor

Distractor

Distractor

Distractor

A. Cytosol

B. Endoplasmic reticulum

C. Extracellular space

D. Golgi apparatus

E. Peroxisome

F. Plasma membrane

Alternatives

Item

Writing good multiple choice questions: Guidelines

● Guidelines: Stem

- The stem should be meaningful by itself
- The stem should not contain irrelevant material
- The stem should be negatively stated only when significant learning outcomes require it
- The stem should be a question or a partial sentence.

STEM IS NOT MEANINGFUL

Which of the following is a true statement?

- A. Mitochondrial genomes are relatively constant in content (i.e., types of genes present).
- B. Mitochondrial genomes are relatively constant in organization.
- C. Mitochondrial genomes are relatively constant in size.

BETTER STEM

What characteristic is relatively constant in mitochondrial genomes across species?

- A. Content (i.e., types of genes)
- B. Organization
- C. Size

Writing good multiple choice questions: Guidelines

● Guidelines: Alternatives

- All alternatives should be plausible.
- Alternatives should be stated clearly and concisely.
- Alternatives should be mutually exclusive.
- Alternatives should be homogenous in content.
- Alternatives should be free from clues about which response is correct.
- The alternatives “all of the above” and “none of the above” should not be used.
- The alternatives should be presented in a logical order
- The number of alternatives can vary among items as long as all alternatives are plausible.

OVERLAPPING ALTERNATIVES

How many chromosomes are found in a typical human cell?

- A. 12
- B. 18
- C. 32
- D. 46
- E. 54

IMPLAUSIBLE ALTERNATIVES

Who gathered the data that helped reveal the structure of DNA?

- A. Francis Crick
- B. George Washington
- C. James Watson
- D. Rosalind Franklin
- E. Snoopy

Writing good multiple choice questions

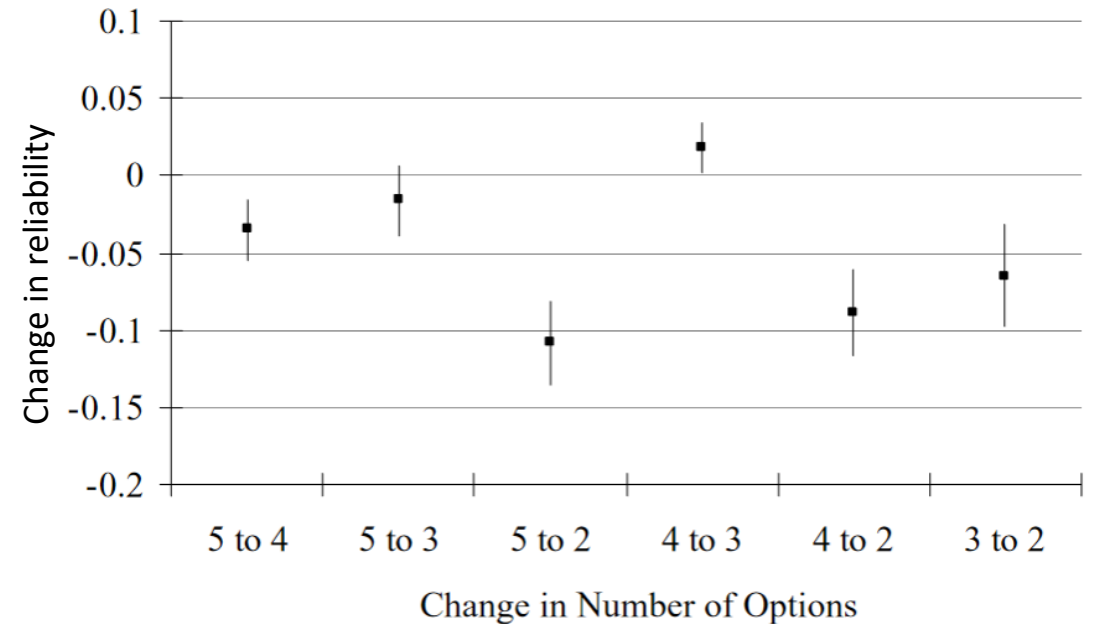
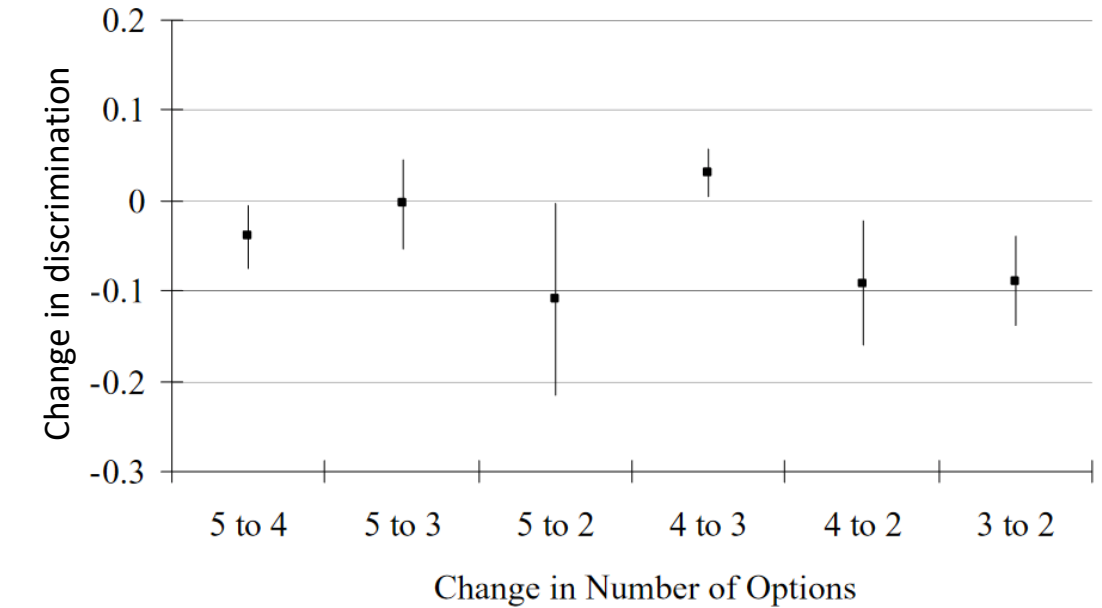
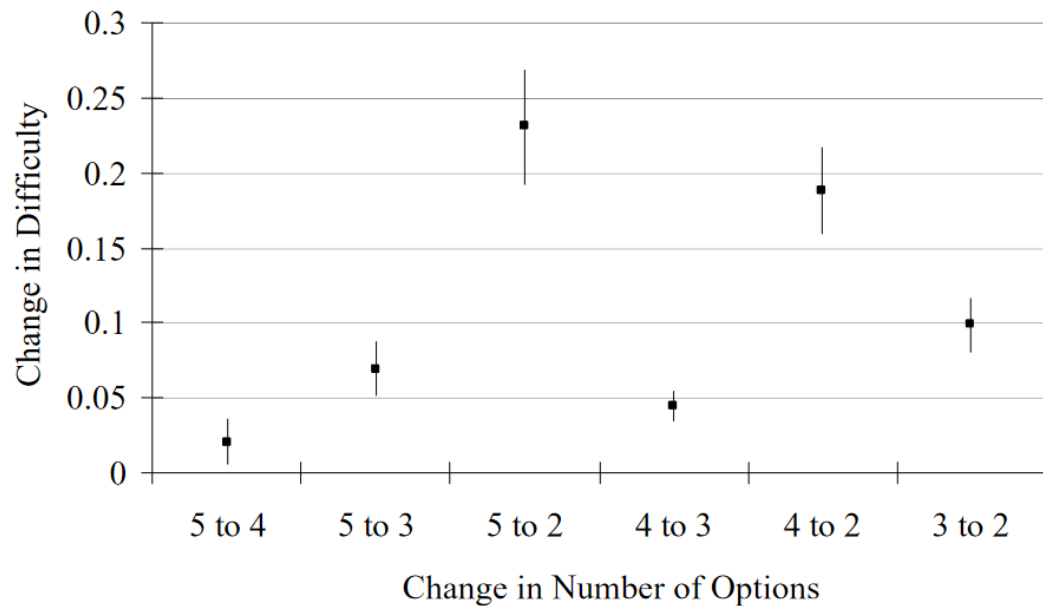
- **Table of Specifications**
 - Adds validity to our assessments
 - Can be used on the “front” or “back” end
- **3-item MC questions**
 - Alternative considered functional if 5% of respondents select it
 - 1-8% of distractors beyond one are functional
 - Benefits of 3-item MCQs:
 - More questions per assessment
 - Harder to guess
 - Easier to write
 - Higher discrimination scores
 - Unchanged difficulty/ reliability

IMPLAUSIBLE ALTERNATIVES

Who gathered the data that helped reveal the structure of DNA?

- A. Francis Crick
- B. George Washington
- C. James Watson
- D. Rosalind Franklin
- E. Snoopy

Discrimination, Difficulty, Reliability, in 3-option MC Items



Activity - Examine a MC question that you have used on a previous exam:

- Find the indices if you have the questions in Canvas
- What guidelines for writing a MC question did you use? (intentionally or unintentionally)
- What guidelines might you use to improve the question?

Questions?

References

Brame, C. (2013) Writing good multiple choice test questions. Retrieved [August 16, 2021] from <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>.

Fives, Helenrose & DiDonato-Barnes, Nicole (2013). Classroom Test Construction: The Power of a Table of Specifications. *Practical Assessment, Research & Evaluation*, 18(3). Available online: <http://pareonline.net/getvn.asp?v=18&n=3>

Haladyna, Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5

Rodriguez. (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. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>

Salkind, N. J. (2017). *Tests & measurement for people who (think they) hate tests & measurement*. 3rd ed. Thousand Oaks, Calif: SAGE Publications. Chapters 3-4.