

DEVELOPING EFFECTIVE MULTIPLE CHOICE TESTS

Megan Welsh
Department of Educational Psychology
2/19/08

Goals for Today

By the end of today, you will be able to:

1. Create a table or specifications, or a plan, to help you construct a multiple choice test that addresses the skills you want to assess
2. Write multiple choice items that address lower-order and higher-order thinking
3. Review and rewrite multiple choice items to improve their quality

But first... A little theory



Like love, we can't ever truly ascertain or quantify how much someone knows.

Tests are imperfect!

But we do gauge their quality

Two key questions:

1. Would my clone get the same score, even if assessed by a different instructor or using a different test of the same skill? (Reliability)
2. Does the test measure what was intended? (Validity)

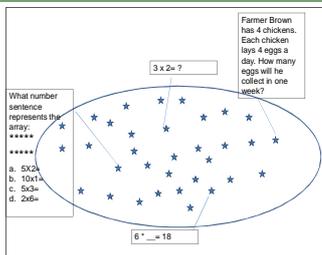


To improve reliability...

- Include many items
- Write items that can be objectively scored
- Minimize students' ability to guess the correct response

To improve validity...

- ❑ Make sure that your test addresses the skills you intended
- ❑ Write items that get at the skill in different ways
- ❑ Address key aspects of the skill



Traditional test development: Planning is key

1. Decide what you are going to assess
2. Write a test outline or table of specifications
3. Select the item types that best meet your needs
4. Write items
5. Edit and rewrite items
6. Administer
7. Analyze responses to see how your items worked

Traditional test development: Planning is key

1. Decide what you are going to assess
2. Write a test outline or table of specifications
3. Select the item types that best meet your needs
4. Write items
5. Edit and rewrite items
6. Administer
7. Analyze responses to see how your items worked

Today

Traditional test development: Planning is key

1. Decide what you are going to assess
2. Write a test outline or table of specifications
3. Select the item types that best meet your needs
4. Write items
5. Edit and rewrite items
6. Administer
7. Analyze responses to see how items worked

3/19

Test Planning

Decide if you need a traditional test.

ALWAYS use the assessment approach that will tell you what you need to know most efficiently, in terms of the time and effort it takes you to develop and score the assessment and in terms of the time it takes students to complete the assessment.

When don't you want a traditional test?



When don't you want a traditional test?

- When the skill you are assessing:
- Can not be gauged by answering questions or writing an essay
 - Involves very involved work that takes a great deal of time and thought to produce (synthesizing information from multiple sources, complex analyses, etc.)
 - Others?

For any assessment

- Start with the topics you want to assess, consider both the CONTENT and the SKILL you want to address.

CONTENT: The topic of instruction (noun and sometimes adjectives)

Write multiple choice items that address lower-order and higher-order thinking

For any assessment

- Start with the topics you want to assess, consider both the CONTENT and the SKILL you want to address.

SKILL: What the student must do to convey knowledge (verb)

Write multiple choice items that address lower-order and higher-order thinking

For any assessment

The way you taught it is how you should assess it:

CONTENT and SKILL in assessment should match content and skill in instruction.

If a traditional test is appropriate

Start with a table of specifications that outlines how many questions you will ask for each content area and the cognitive demand of the item (higher-order or lower order).

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33, Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (6%) Q23, Q24, Q36	5 (10%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33, Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Content Assessed

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Cognitive Demand

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Number (and percent) of items of each content/cognitive demand combination

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Items on test at each content/cognitive demand combination, inserted as test is developed

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Number (percent) of items addressing each content area

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Number (percent) of items addressing each level of cognitive demand

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33 Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (14%) Q23, Q24, Q36	5 (18%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Anatomy of a Table of Specifications

Total number of items

	Lower order	Higher-order	Total
Assessment Uses		2 (4%) Q11, Q12	2 (4%)
Objectives		8 (16%) Q13, Q14, Q15, Q31, Q32, Q33, Q34, Q35	8 (16%)
Reliability	2 (4%) Q1, Q2	3 (6%) Q16, Q17, Q18	5 (10%)
Validity	2 (4%) Q3, Q4	4 (8%) Q19, Q20, Q21, Q22	6 (12%)
Item Analysis	2 (4%) Q5, Q6	3 (6%) Q23, Q24, Q36	5 (10%)
Standardized Tests	4 (10%) Q7, Q8, Q25, Q37	3 (12%) Q26, Q27, Q28	7 (22%)
NCLB	3 (14%) Q9, Q10, Q38	2 (4%) Q29, Q30	5 (18%)
Total	13 (36%)	25 (64%)	38 (100%)

Purpose of a Table of Specifications

- Ensures you assess what you want to assess
- Helps you think through the weight each topic gets on a test (you want the most important topics to get the most weight)
- Prevents you from weighting items based on factors like the amount of time they take to answer
- Helps you think about the cognitive demand of the test

So what is cognitive demand???

The depth of thinking required...

Lower-order

- Recall
- Identify
- Compute (when rote)
- Summarize

Higher-order

- Analyze
- Synthesize
- Evaluate
- Design

Relates to the skill emphasized in your instructional objectives.

Traditional tests can assess many levels of cognitive demand

But it is easier to assess some cognitive demand levels with certain item types.

Traditional test item types

- Multiple choice (the focus today)
- Alternate choice: True/False OR
The atmosphere near the equator flows to the (east/west)
Indicate if each underlined word is correctly spelled.
We lik to eat ice crem sundaes in the summr.
- Completion items: Fill in the blank OR
How many members are there in the House of Representatives?
- Essay items (Includes anything where more than a few words are required)

Traditional test item types

- Multiple choice (the focus today)
- Alternate choice: True/False OR
The atmosphere near the equator flows to the (east/west)
Indicate if each underlined word is correctly spelled.
We lik to eat ice crem sundaes in the summr.
- Completion items: Fill in the blank OR
How many members are there in the House of Representatives?
- Essay items (Includes anything where more than a few words are required)

Traditional test item types

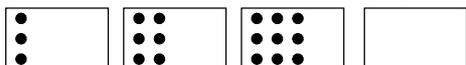
- Multiple choice (the focus today)
- Alternate choice: True/False OR
The atmosphere near the equator flows to the (east/west)
Indicate if each underlined word is correctly spelled.
We lik to eat ice crem sundaes in the summr.
- Constructed response { □ Completion items: Fill in the blank OR
How many members are there in the House of Representatives?
□ Essay items (Includes anything where more than a few words are required)

What have you heard about...

the benefits and drawbacks of multiple choice tests?

Anatomy of a multiple choice item

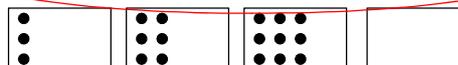
1. David drew this pattern of dots. How many dots should he draw on the last piece of paper?



- a. 10
- b. 12*
- c. 15
- d. 18

Anatomy of a multiple choice item

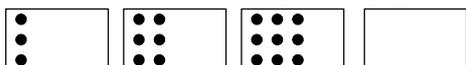
1. David drew this pattern of dots. How many dots should he draw on the last piece of paper?



- a. 10
- b. 12*
- c. 15
- d. 18

Anatomy of a multiple choice item

1. David drew this pattern of dots. How many dots should he draw on the last piece of paper?



- a. 10
- b. 12*
- c. 15
- d. 18

Distractors

Anatomy of a multiple choice item

1. David drew this pattern of dots. How many dots should he draw on the last piece of paper?



- a. 10
- b. 12*
- c. 15
- d. 18

Extra information

Interesting approaches to multiple choice

Directions: For each sentence, mark on your answer sheet the letter identifying the line that contains an adjective. Mark "E" if the sentence contains no adjective.

1. A. Krueger National
B. Park is a
C. game reserve
D. in South Africa.
E. (No adjective)

Interesting approaches to multiple choice

Which item format has each of the following qualities?

- A. Completion
 - B. Essay
 - C. Multiple choice
1. Answers can be quickly scored
 2. Test items can be quickly constructed
 3. Responses can be scored with minimal error
 4. Can assess students' abilities with written expression

Interesting approaches to multiple choice

January Weather Data for Selected Cities

	Low	High	Precipitation
A. Jacksonville	45	65	8
B. Key West	66	76	7
C. Miami	59	76	7
D. Orlando	50	71	6

Respond to the following questions using the table above. Write the letter associated with the correct city in the blank.

1. — Which city tends to have the lowest temperature in January?
2. — Which city tends to have the smallest range of temperatures in January?
3. — Which city is most likely to experience precipitation on a typical January day?
4. — Which city is least likely to experience precipitation on a typical January day?

Other items that count as multiple choice

- Matching. If using matching items, make sure that there are more options to select from than stems.

Match the politician to the position they hold

- | | |
|--------------------|-----------------------|
| 1. Barack Obama | a. Senator |
| 2. Joe Biden | b. Vice President |
| 3. Hillary Clinton | c. Secretary of State |
| 4. John McCain | d. President |

Other items that count as multiple choice

- Matching. If using matching items, make sure that there are more options to select from than stems.

Match the politician to the position they hold

- | | |
|--------------------|-----------------------|
| 1. Barack Obama | a. Senator |
| 2. Joe Biden | b. Vice President |
| 3. Hillary Clinton | c. Secretary of State |
| 4. John McCain | d. President |

Other items that count as multiple choice

- Matching. If using matching items, make sure that there are more options to select from than stems.

Match the politician to the position they hold

- | | |
|--------------------|-----------------------|
| 1. Barack Obama | a. Senator |
| 2. Joe Biden | b. Vice President |
| 3. Hillary Clinton | c. Secretary of State |
| 4. John McCain | d. President |
| 1. | e. Governor |

Other items that count as multiple choice

Ranking. Putting a series of statements in the correct order.

What is the correct order of events involved in washing your hands?

1. Turn off the water
2. Dry your hands
3. Turn on the water
4. Rub your hands with soap
5. Rinse your hands

Now you try...

- Write a couple of multiple choice items for a class you are currently teaching. Discuss with a neighbor.

Multiple choice item review checklist

1. Does the item measure the specified skill?
2. Is the level of reading skill required by this item below the students' ability?
3. Does the stem clearly present the problem to be addressed?
4. Are the options parallel in type of content and is the grammar consistent with the stem?
5. Do the options avoid repetitive words?
6. Is extraneous content excluded from the stem?
7. Are adjectives and adverbs emphasized when they reverse or significantly alter the meaning of a stem or option? Is the word *not* excluded?
8. Is each distractor plausible?
9. Does the item exclude options equivalent to "all of the above" and "none of the above"?
10. Unless another order is more logical, are options arranged alphabetically?

Does the item measure the specified skill?

Seems basic, but easy to overlook...

5th grade objective: Capitalize brand names.

Which sentence does not have a capitalization error?

- A. I buy oatmeal from montana.
- B. I eat eggs.
- C. I love grits, johnny.
- D. all of the above

Is the level of reading skill required by this item below the students' ability?

Shouldn't be as much of a problem in post-secondary education, but

1. Don't use academic lingo unless you are testing knowledge of academic lingo
2. The simpler the wording, the more questions you can ask
 - more reliable test
 - more content assessed

Does the stem clearly present the problem to be addressed?

TIP: Write the stem as a question instead of an incomplete sentence

The artificial satellite with the highest orbital speed is

- A. the most recently launched satellite
- B. the satellite in higher orbit
- C. the satellite in lower orbit

Does the stem clearly present the problem to be addressed?

TIP: Write the stem as a question instead of an incomplete sentence

The artificial satellite with the highest orbital speed is

- A. the most recently launched satellite
- B. the satellite in higher orbit
- C. the satellite in lower orbit

Does the stem clearly present the problem to be addressed?

TIP: Write the stem as a question instead of an incomplete sentence

Does the artificial satellite in higher or lower orbit have the higher orbital speed:

- A. the satellite in higher orbit
- B. the satellite in lower orbit
- C. the satellites have the same orbital speed

Are the options parallel in type of content?

TIP: Attend both to grammar and to response type.

Which of the following represents the warmest temperature?

- A. 100 degrees Celsius
- B. 100 degrees Fahrenheit
- C. An oven set at medium

Do the options avoid repetitive words?

Okay....

The artificial satellite with the highest orbital speed is

- A. the most recently launched satellite
- B. the satellite in higher orbit
- C. the satellite in lower orbit

Do the options avoid repetitive words?

Better....

The artificial satellite with the highest orbital speed is the:

- A. most recently launched satellite
- B. satellite in higher orbit
- C. satellite in lower orbit

Is extraneous content excluded from the stem?

We use punctuation when we write. It is very important. Which sentence needs an exclamation point?

- A. You are five
- B. How old are you
- C. Kate, do you like ice cream
- D. The girl screamed, "Whee, whee"

Emphasize adjectives and adverbs and avoid not, especially in the stem

1. Which of the following conditions LEAST affects the speed at which the wind blows?
2. Which of the following is not an item format?
 - a. Not recall
 - b. Not short answer
 - c. Not reverse video

Is each distractor plausible?

- Who was U.S. President at the start of the Gulf War?
- A. George Bush
 - B. Dick Cheney
 - C. Mickey Mouse

Does the item exclude "all of the above" and "none of the above"?

Why???

Because they make it easier to guess.

Which of the following instruments uses reeds to make sound?

- A. Oboe
- B. French Horn
- C. None of the above

Are distractors arranged alphabetically?

Other logical orders (like greatest to least) are also fine.

Why?? The correct answer is most often B or C.

Which sentence needs an exclamation point?

- A. You are five
- B. How old are you
- C. Kate, do you like ice cream
- D. The girl screamed, "Whee, whee"

Are distractors arranged alphabetically?

Other logical orders (like greatest to least) are also fine.

Why?? The correct answer is most often B or C.

Which sentence needs an exclamation point?

- A. How old are you
- B. Kate, do you like ice cream
- C. The girl screamed, "Whee, whee"
- D. You are five

Now practice...

Questions?

Next session (3/19) will address how to analyze multiple choice responses to check item quality

Contact info:

megan.welsh@uconn.edu