Principles of Test Building

Well-constructed tests can test a range of knowledge and skills involving higher levels of thinking. They need not be limited to recall items testing student knowledge of concrete information. As well, multiple-choice and numeric response examination items can provide useful diagnostic information about individual achievement and group achievement.

Writing a "Stem"

An effective stem

- has clear and grammatically correct language
- is appropriate in reading level as well as content
- contains enough detail to provide a clear basis for selecting/providing an answer

Effective questions

- arise naturally through exploration of a context
- are based upon real or true data
- are able to be completed in 2 minutes or less
- are restricted to one or two steps in order to obtain the answer
- have instructional validity
- vary in focus, nature and type
- have curricular validity and test something worth asking

Numerical Response Questions

Numerical response (NR) questions need not be limited to straight calculation questions.
NR type questions could be used in the following situations:

- asking for a calculation where the response is not negative and/or where the response fits in the NR boxes
- asking for a listing of events to be ordered (eg: order of operations)
- asking a matching type question
- asking for a selection of procedures from a list (not necessarily order dependent)

Use of Context Boxes/Contexts

If the stem of a question is longer than 3 lines, put the information in a context box and the main question as the stem. All graphs or artwork go in a context box.

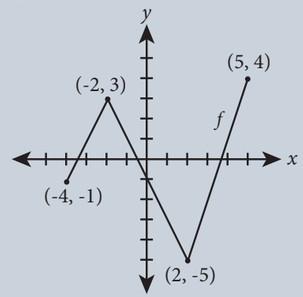
Pitfalls to Avoid When Writing MC Alternatives

1. Cueing – from stem to alternatives or within a question set
2. 3-1 Split – one alternative is unlike the other three
3. Length Imbalance – alternatives are of different lengths, or keyed response stands out because it is too detailed, too wordy or too brief
4. Grammatical Imbalance – alternatives are not grammatically matched to others and/or to the stem
5. Double Key – the key and another alternative are the same in value or the same in meaning
6. Implausible/Irrational Alternatives – alternatives that are meaningless or have no rational basis
7. Repetition – words or phrases that are repeated in each alternative

Anatomy of a Multiple-Choice Item

Use the following information to answer the first question.

The graph of $y = f(x)$ is shown below.



stem → 1. The range of $y + 5 = f(x - 3)$ is

distractors
A. $0 \le y \le 9$
B. $-1 \le y \le 8$
C. $-7 \le y \le 2$
key → D. $-10 \le y \le -1$

alternatives

Questions to Consider

- Does the test have blueprint balance? Blueprint balance includes varying levels of difficulty, representation from the three cognitive levels, both acceptable and excellence level questions, and attention to the mathematical processes.
- Do the questions in the test cover the full range of the program?
- Does the test have instructional validity?
- Do any questions cue the answer to other questions in the test?
- Are instructions clear for each section of the test and for the test overall?

Multiple Choice

Advantages: easy to mark; potential to assess multiple learning targets at once; can assess various question types; including higher level 'type 3'

Disadvantages: difficult to construct effective items; test wiseness can skew results; inappropriate for measuring outcomes that require most skilled performance; can lead the instructor to favour simple recall of facts; guess can be a factor (although only 1 in 4); sometimes tests reading comprehension rather than content or other skill; cannot provide partial marks

Elements of a well-designed item: clear instructions – especially on source-based questions; grammatically correct - clearly and concisely written; conscious of response patterns – balance correct responses (A, B, C, D); avoid absolutes ("all of the above"); avoid using universal descriptors such as "never", "always", and "all"; distractors should be organized by length or numerically

MC Question Data Analysis

Difficulty: how easy or difficult this question was for the class; in item analysis, the lower the number, the more difficult the item was; can also be written as % of students who got question

Discrimination: how well the question can distinguish between the high group and the low group on this test; how well this item contributed to test results; a discrimination above 0.2 is good; a negative discrimination means the item is doing opposite of what it is supposed to be doing

Distractors: each one should have students drawn to it; if more students are drawn to a distractor than the correct answer – indicates a flaw in the question or a common misconception among students

Source-Based Multiple Choice Quality Checklist

An item is **technically sound** when...

- the item uses source material that is familiar but new to the student
- the source material is an appropriate length and at an appropriate reading level
- the source material is properly cited
- the items cannot be answered without the source material
- the items are at the applied taxonomy level or above
- the item has correct spelling and grammar
- the items meet the technically sound checklist for mulitple choice items

An item is **equitable** when the item...
is free from cultural and se xual bias
is free of irrelevant material
is stated in appropriate and clear language
is free fron cultural references that would not be familiair to all students

Multiple-Choice Item Quality Checklist

An item is **technically sound** when...

- the item is free of cues to the answer
- the stem is focused on a single, meaningful problem
- key words in the stem are emphasized as needed
- distractors are all plausible
- alternatives are in an order that is logical and easily understood
- the item has correct spelling and grammar

An item is **equitable** when the item...

- is free from cultural and sexual bias
- is free of irrelevant material
- is stated in appropriate and clear language
- is free from cultural references that would not be familiar to all students

Designed by Richelle Marynowski September 2016 Adapted from Alberta Education 2010

True/False

Advantages: easy to mark and easy to design; can cover a lot of knowledge content; space-efficient

Disadvantages: prone to guessing (50/50) – questionable validity; limited to lower level questioning

Elements of a well-designed item: clear instructions; grammatically correct - clearly and concisely written; avoid trick questions – statements must be entirely true or entirely false; avoid using universal descriptors such as "never", "none", "always", and "all"; avoid negative words, as they are often overlooked; do not include two ideas in one statement unless you are evaluating student's understanding of cause and effect relationships; provide a "T" and "F" beside each statement and ask students to circle correct answer; consider including more false than true statements and vary the number of false statements from test to test.

Fill-in-the-blank

Advantages: easy to mark and easy to design; can cover a lot of knowledge content; space-efficient

Disadvantages: limited to lower level questioning; possibility of getting more than one question wrong if there is only one response for each question

Elements of a well-designed item: questions must be carefully worded so that all students understand the specific nature of the question asked and the answer required; instructions and teacher's expectations about filling in blanks should be made clear - indicate whether each blank of equal length represents one word or several words, whether long blanks require sentences or phrases, and whether synonymous terms are accepted; when an answer is to be expressed in numerical units, the unit should be stated; one blank per statement; indicate whether the same response can be used more than once (if there is a list of possible answers).

Points to consider: offer a list of possible responses (more responses than needed as distractors); be clear about whether spelling is important and if synonyms are acceptable where then there is no list of possible answers.

Constructed Response

Advantages: relatively easy to design; teacher can be creative in how question is asked – connected to a variety of sources; higher level thinking can be assessed; students not limited by information provided – can craft unique ways to demonstrate understanding

Disadvantages: more time consuming to mark; subjectivity may be a factor in grading (teachers must guard against personal bias); can be more stressful for students if time is a factor

Elements of a well-designed item: clear instructions; question should be specific (avoid broad and/or vague questions); grammatically correct - clearly and concisely written; clear instructions on how to use source information (if needed); provide marking criteria for how response will be assessed - perhaps a rubric or clear instructions from teacher about what is expected; provide enough space for students to write (or offer extra paper)

Matching Questions

Advantages: easy to mark and easy to design; can cover a lot of knowledge content; space-efficient

Disadvantages: possibility of getting more than one question wrong if there is only one response for each question; prone to guessing – process of elimination

Elements of a well-designed item: clear instructions; grammatically correct - clearly and concisely written; indicate whether the same response can be used more than once; maintain grammatical consistency within and between columns; ensure that any matching question appears entirely on one page; provide an unequal number of premises and responses; provide space for letter or number answers; make sure lists are homogeneous; wording of the premises longer than the wording of the responses; identify the items in one list with numbers and those in the second list with letters.