**Writing High Quality Traditional Assessment Items**

by Jay Rasmussen, Bethel University

This document is designed to assist you in the preparation of high quality traditional assessment items. Please think fully about the following topics once you’ve determined that a particular learning outcome is best assessed through the use of traditional assessment:

* general guidelines for item writing,
* advantages and disadvantages of each item type,
* writing guideline considerations for each item type, and
* sample item(s) for each item type.

Based on work by….

Miller, M. D., Linn, R. L., & Gronlund, N. E (2013). *Measurement and assessment in teaching.* Upper Saddle River, NJ: Pearson.

**General Guidelines for Item Writing**

1. Relate items to instructional objectives.
2. Write items with academic and non-academic vocabulary at the appropriate level.
3. Write items as clearly as possible. Use a sentence structure that is clear and simple – keep it brief.
4. Avoid using words that are ambiguous. (e.g. hot, cold, few, many, large, small, high, low)
5. Don’t lift items word for word from the text. Because the item is being taken out of context, the meaning may be lost.
6. Don’t include items that are interrelated – i.e. students must be able to correctly answer one item before another item may be answered correctly.
7. Write questions with one best answer. If it’s an opinion question, state whose opinion you are looking for.

For example:

Poor – A major cause of heart attacks is cholesterol.

Better – According to Dr. White, a major cause of heart attacks is cholesterol.

1. Avoid negative questions when possible. Reason – the student needs to change normal thought processes and it will take more time and produce more errors.

For example:

Poor – T/F: A United States congressman is not elected for a 2-year term.

Better – T/F: A United States congressman is elected for a 2-year term.

1. Don’t give the answer away, either in that question or a previous one.
2. Avoid racial, ethnic, or gender bias.

**Short Answer Items**

Advantages

1. Require the student to produce rather than recognize an answer.
2. Are easy to administer.
3. Provide diagnostic information.

Disadvantages

1. Require more time to score than many other item types.
2. May not be effective at measuring higher level thinking skills.

Writing Considerations

Yes No

1. Can the items be answered with a number, symbol, word, or a brief phrase?
2. Has textbook language been avoided?
3. Have the items been stated so that only one response is correct?
4. Are the answer blanks equal in length?
5. Are the answer blanks at the end of the items?
6. Are the items free of clues (such as *a* or *an*)?
7. Has the degree of precision been indicated for numerical answers?
8. Have the units been indicated when numerical answers are expressed in units?
9. Have the items been phrased so as to minimize spelling errors?
10. If revised, are the items still relevant to the intended learning outcomes?

Sample Items

(Completion Variety)

Lines on a weather map that join points of the same barometric pressure are called \_\_\_\_\_\_\_\_\_\_

(Direct Variety)

If the temperature of a gas is held constant while the pressure applied to it is increased, what will happen to its volume? \_\_\_\_\_\_\_\_\_\_

------------------------------------------------------------------------------------------------------------------------------------------

**T-F Items**

Advantages

1. Can cover a large amount of subject matter in a short-period of time. (students can answer three T-F items for every two multiple choice items)
2. Can be scored quickly.
3. Can be written easily from a salvage pool of multiple choice items.

Disadvantages

1. Are most affected by guessing.
2. Are most susceptible to ambiguity.
3. Make it easier to cheat.
4. Allow students to enter a pattern of responding without thinking.

Writing Considerations

Yes No

1. Can each statement be clearly judged true or false?
2. Have specific determiners (e.g., usually, always) been avoided?
3. Have trivial statements been avoided?
4. Have negative statements (especially double negatives) been avoided?
5. Have the items been stated in simple, clear language?
6. Are opinion statements attributed to some source?
7. Are the true and false items approximately equal in length?
8. Is there an approximately equal number of true and false items?
9. Has a detectable pattern of answers (e.g. T, F, T, F) been avoided?

Sample Items

(True-False Variety)

T F A virus is the smallest known organism.

(Cluster Variety)

Mary Ann wanted her rose bush to grow faster, so she applied twice as much chemical fertilizer as was recommended and watered the bush every morning. About a month later she noticed that the rose bush was dying.

T F The following principles are necessary in explaining why the rose bush is dying.

1. A chemical compound is changed into other compounds by taking up the elements of water. (F)
2. Semipermeable membranes permit the passage of fluid. (T)
3. Water condenses when cooled. (F)
4. When two solutions of different concentration are separated by a porous partition, their concentration tends to equalize. (T)

(Correction Variety)

If False Option is Selected, Students Must Replace the Underlined Word with Word to Make the Statement True

T F The green coloring material in a plant leaf is called chlorophyll.

**------------------------------------------------------------------------------------------------------------------------------------------**

**Matching Items**

Advantages

1. Are especially effective at measuring learning of items, definitions, dates, locations, and events.
2. Require a limited amount reading so many questions may be asked.

Disadvantages

1. Require the student to produce rather than recognize an answer.
2. May overemphasize memorization.
3. Are difficult to write because items must be clustered.

Writing Consideration Yes No

1. Is the material in the two lists homogenous?
2. Is the list of responses longer or shorter than the list of premises?
3. Are the responses brief and on the right-hand side?
4. Have the responses been placed in alphabetical or numerical order?
5. Do the directions indicate the basis for matching?
6. Do the directions indicate that each response may be used more than once?
7. Is all of each matching item on the same page?
8. If revised, are the items still relevant to the intended learning outcomes?

Sample Items

Column A (Premise) Column B (Response)

1. Air pressure A. Anemometer
2. Air temperature B. Barometer
3. Humidity C. Hygrometer
4. Wind velocity D. Rain gauge

E. Thermometer

F. Wind vane

**Multiple-Choice Items**

Advantages

1. Effectively measure various levels of thinking.
2. Avoid ambiguity because the thought is completed by the options.
3. Can be used to provide diagnostic information.
4. Are easily scored.
5. Are the most popular item type among students.
6. Can be controlled for difficulty by making the options more or less homogenous.
7. Can be used effectively with drawings, maps, graphs, and visuals.

Disadvantages

1. Are very time consuming to construct.
2. Require more student time to complete than some other item types.

Writing Considerations

Yes No

1. Does each item stem present a meaningful problem?
2. Are the item stems free of irrelevant material?
3. Are the item stems stated in positive terms (if possible)?
4. If used, has the negative wording been given special emphasis (e.g. capitalized)?
5. Are the alternatives grammatically consistent with the item stem?
6. Are the alternative answers brief and free of unnecessary words?
7. Are the alternatives similar in length and form?
8. Is there only one correct or clearly best answer?
9. Are the distractors plausible to low achievers?
10. Are the items free of verbal clues to the answer?
11. Are verbal alternatives in alphabetical order?
12. Are numerical alternatives in numerical order?
13. Have *none of the above* and *all of the above* been avoided (or used sparingly and appropriately)?
14. If revised, are the items still relevant to the intended learning outcomes?

Sample Items

(Direct Question Variety)

What is the function of the kidneys?

1. Eliminate waste products.
2. Improve the circulation of blood.
3. Maintain respiration.
4. Stimulate digestion.

(Incomplete Statement Variety)

Alternating electric current is changed to direct current by means of a

1. condenser.
2. generator.
3. rectifier.
4. transformer.

(Best-Answer Variety)

Which one of the following best illustrates the principle of capillarity?

1. Fluid is carried through the stems of plants.
2. Food is manufactured in the leaves of plants.
3. The leaves of deciduous plants lose their green color in winter.
4. Plants give off moisture through their stomata.

------------------------------------------------------------------------------------------------------------------------------------------

**Interpretive Items**

Advantages

1. Measures ability to interpret various forms of written information (charts, graphs, maps, etc.)
2. Possible to measure more complex learning outcomes than a single assessment item
3. Greater depth and breadth in measurement of intellectual skills when multiple items are based on a single set of data
4. Measurement of complex learning outcomes is not likely to be influenced by irrelevant information
5. More structured than performance assessment tasks
6. Ability to measure separate aspects of problem-solving

Disadvantages

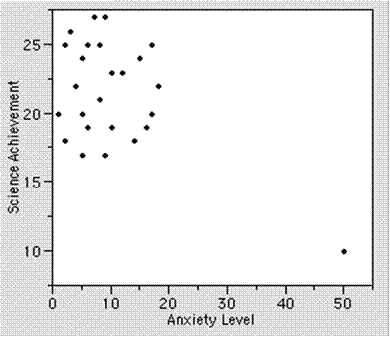
1. Difficult to construct (need to find new yet relevant and fair information; need to create questions that measure desired learning outcomes)
2. Highly dependent on reading skills – poor readers are at a serious disadvantage
3. Provides diagnostic rather than holistic information about problem-solving abilities
4. Limited to measuring recognition-level learning outcomes

Writing Considerations Yes No

1. Is the material to be interpreted relevant to the intended learning outcomes?
2. Is the material to be interpreted appropriate to the students’ curricular experience and reading level?
3. Have pictorial materials been used whenever appropriate?
4. Does the material to be interpreted contain some novelty (to require interpretation)?
5. Is the material to be interpreted brief, clear, and meaningful?
6. Are the test items based directly on the introductory material (cannot be answered without it) and do they call for interpretation (not just recall or simple reading skills)?
7. Has a reasonable number of test items been used in each interpretive exercise?
8. Do the test items meet the relevant criteria of effective item writing?
9. When key-type items are used, are the categories homogenous and mutually exclusive?
10. If revised, are the interpretive exercises still relevant to the intended learning outcomes?

Sample Item

21. The following scatterplot shows the relationship between scores on an anxiety scale and an achievement test for science. Choose the best interpretation of the relationship between anxiety level and science achievement based on the scatterplot (Developed by the Web ARTIST Project <https://app.gen.umn.edu/artist/>).



a. This graph shows a strong negative linear relationship between anxiety and achievement in science.

b. This graph shows a moderate linear relationship between anxiety and achievement in science.

c. This graph shows very little, if any, linear relationship between anxiety and achievement in science.

**Essay Items**

Advantages

1. Effectively measure the ability to organize ideas.
2. Are easily constructed.

Disadvantages

1. Require an expert to grade.
2. Have low content validity because sampling is limited.
3. Are often scored unfairly because of writing ability, length of paper, sequence of reading, and the “Halo Effect”.

Writing Considerations

Yes No

1. Are the questions designed to measure higher-level learning outcomes?
2. Are the questions relevant to the intended learning outcomes?
3. Does each question clearly indicate the response expected?
4. Are the students told the bases on which their answers will be evaluated?
5. Are generous time limits provided for responding to the questions?
6. Are students told the time limits and/or point values for each question?
7. Are all students required to respond to the same questions?
8. If revised, are the questions still relevant to the intended learning outcomes?

Scoring Guidelines

1. Prepare a short outline of the correct answers for each item.
2. Before any grading is done, read through a sample of papers to get an idea of student responses.
3. Grade without looking at names.
4. Grade the same question on all papers before moving to the next question.

Sample Item

*Restricted Response –* 20 minute time limit, must demonstrate: thorough understanding of relationship (5pts.) and writing clarity/grammar (5pts.).

1. The percentage of the workforce that is unemployed increased following an increase in interest rates. Briefly explain the relationship.