Assessment Tool Continuum

Design all to foster learning outcomes

**Formative**

<table>
<thead>
<tr>
<th>Assignments/Activities</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journals</td>
<td>Check lists</td>
</tr>
<tr>
<td>Reading Logs</td>
<td>Content Analyses</td>
</tr>
<tr>
<td>Reflections</td>
<td>Select correct response</td>
</tr>
<tr>
<td>Minute Papers</td>
<td></td>
</tr>
<tr>
<td>Blogs</td>
<td></td>
</tr>
<tr>
<td>In Class Questions-closed response</td>
<td></td>
</tr>
<tr>
<td>Brief Quizzes-open ended on objective</td>
<td></td>
</tr>
<tr>
<td>Muddiest Point</td>
<td></td>
</tr>
<tr>
<td>Clickers</td>
<td></td>
</tr>
</tbody>
</table>

**Indirect**

Surveys, Focus Groups, and Interviews

**Summative**

<table>
<thead>
<tr>
<th>Assignments/Activities</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Exams</td>
<td>Rubrics</td>
</tr>
<tr>
<td>Capstone Projects</td>
<td>Exam- Multiple Choice</td>
</tr>
<tr>
<td>Exam- Multiple Choice</td>
<td>Holistic Rubric</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assignments/Activities</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term Papers</td>
<td>Rubrics</td>
</tr>
<tr>
<td>Reports</td>
<td></td>
</tr>
<tr>
<td>Formal Essays</td>
<td></td>
</tr>
<tr>
<td>Documented Papers</td>
<td>Reviews</td>
</tr>
<tr>
<td>Holistic Rubric</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assignments/Activities</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentations</td>
<td></td>
</tr>
<tr>
<td>Posters</td>
<td></td>
</tr>
<tr>
<td>Lab Reports</td>
<td></td>
</tr>
</tbody>
</table>

*Will align with program outcomes

Adapted with permission from Chris Anson (2016) by Georgia Southern University
Presented at the 2016 Annual AALHE Conference, *Assessment 102: Measuring Learning*
Testing Your Multiple-Choice Test-Wiseness
(Adapted from Constructing Classroom Tests: Why and How, James Eison, Southeast Missouri State University, Cape Girardeau, MO, 1985)

DIRECTIONS: The seven multiple-choice questions below cover historical topics that you are not likely to know. See if you are able to determine the correct answer by carefully reading each item. Please circle the correct answer for each item.

1. The Locarno pact:
   a. is an international agreement for the maintenance of peace through the guarantee of national boundaries of France, Germany, Italy, and Belgium.
   b. allowed France to occupy the Ruhr Valley.
   c. provided for the dismemberment of Austria-Hungary.
   d. provided for the protection of Red Cross bases during war times.

2. The disputed Hayes-Tilden election of 1876 was settled by an:
   a. resolution of the House of Representatives.
   b. decision of the United States Supreme Court.
   c. Electoral Commission
   d. joint resolution of Congress.

3. The august character of the work of Pericles in Athens frequently causes his work to be likened to that in Rome of:
   a. Augustus.
   b. Sulla.
   c. Pompey.
   d. Claudius.

4. The Declaration of the Rights of Man was:
   a. adopted by the French National Assembly.
   b. adopted by every Western European legislature.
   c. immediately ratified by every nation in the world.
   d. hailed by every person in England.

5. The Locarno pact:
   a. was an agreement between Greece and Turkey.
   b. gave the Tyrol to Italy.
   c. was a conspiracy to blow up the League of Nations’ building at Locarno.
   d. guaranteed the boundary arrangements in Western Europe.
6. Horace in the 16th Epode emphasizes the:

a. despair of the average man confronted by sweeping social change.
b. elation of the average man confronted by sweeping social change.
c. optimism of the common man about sweeping social change.
d. all of the above.

7. About what fraction of the 1920 population of the United States was foreign-born?

a. less than five percent.
b. between fourteen and twenty-eight percent.
c. twenty-five percent.
d. between thirty and fifty percent.

Some of these items may have been answered correctly even with little or no exposure to the content. All of these “clues” should be avoided when developing items. Many “test-wise” students are able to guess the correct answers to all of the items in the exercise and would make the same guesses if the same items were administered a second time. Thus, the test would be reliable, but the results would not be valid as a measure of achievement of course content. By following the test blueprint and the guidelines offered on the next pages, the chances of achieving valid and reliable test results are increased.

“Designing Multiple-Choice Tests That Foster and Reflect Learning Outcomes,” SACSCOC Annual Meeting 2012, Dr. Teresa L. Flateby, Associate Vice President, Institutional Effectiveness, Georgia Southern University, Statesboro, GA
Designing Multiple Choice Tests That Foster and Reflect Learning Outcomes

Teresa Flateby, Ph.D., Georgia Southern University, tflateby@georgiasouthern.edu
http://academics.georgiasouthern.edu/vpie

Test Construction Guidelines:

1. Begin with course or program outcomes – these should guide the development of tests. (See scenario below)
2. Develop a test blueprint to reflect content and cognitive levels desired and reflected in instructional strategies and activities and assignments. (See below)
3. Follow item writing guidelines and ensure items elicit appropriate cognitive levels. (Page 3)
4. Use item analysis to evaluate the test and test items (reliability coefficient, difficulty level, discrimination index, and distractor analysis – See item analysis table.).
5. Compare results of item analysis to the test blueprint to determine learning weaknesses.

Scenario:

The instructor of an undergraduate measurement course with 70 students developed a multiple-choice test to address several outcomes, one of which is “students will interpret and analyze multiple-choice test results and make accurate suggestions for improvement.” This outcome was derived from the institutional outcome “Students will demonstrate the critical thinking skills of analysis and evaluation in the general education curriculum and in the major.” The instructor created the following Table of Item Specifications to development of a 40 item test:

<table>
<thead>
<tr>
<th>Content</th>
<th>Knowledge</th>
<th>Comprehension</th>
<th>Application and Above</th>
<th>Total Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purposes of Testing</td>
<td>1</td>
<td>3</td>
<td></td>
<td>4 (10%)</td>
</tr>
<tr>
<td>Necessary Criteria for Tests</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3 (7.5%)</td>
</tr>
<tr>
<td>Reliability</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3 (7.5%)</td>
</tr>
<tr>
<td>Validity</td>
<td>1</td>
<td>3</td>
<td></td>
<td>4 (10%)</td>
</tr>
<tr>
<td>Test Development</td>
<td>1</td>
<td>2</td>
<td></td>
<td>6 (15%)</td>
</tr>
<tr>
<td>Table of Item Specifications</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Proper Item Construction</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Criteria for Evaluating Test</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Relevance, Variability, Difficulty</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Discrimination, Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Analysis</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>9 (22.5%)</td>
</tr>
<tr>
<td>Principles/Printout</td>
<td>10 (25%)</td>
<td>20 (50%)</td>
<td>10 (25%)</td>
<td>40 (100%)</td>
</tr>
</tbody>
</table>
Five Items from a 40-Item Test Developed for an Undergraduate Test and Measurement Course

Refer to the Item Analysis Table below when answering items 1-3 on the test:

**Item Analysis Table**

<table>
<thead>
<tr>
<th>Item</th>
<th>N</th>
<th>P</th>
<th>N</th>
<th>P</th>
<th>N</th>
<th>P</th>
<th>N</th>
<th>P</th>
<th>N</th>
<th>P</th>
<th>Prop. Passing</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>.08</td>
<td>0</td>
<td>.00</td>
<td>8</td>
<td>.10</td>
<td>6</td>
<td>.08</td>
<td>60*</td>
<td>.75</td>
<td>.75</td>
<td>.175</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>.05</td>
<td>3</td>
<td>.04</td>
<td>67*</td>
<td>.84</td>
<td>3</td>
<td>.04</td>
<td>3</td>
<td>.04</td>
<td>.84</td>
<td>.347</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>.22</td>
<td>21*</td>
<td>.26</td>
<td>26</td>
<td>.32</td>
<td>15</td>
<td>.19</td>
<td>0</td>
<td>.00</td>
<td>.26</td>
<td>-.015</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>.08</td>
<td>73*</td>
<td>.91</td>
<td>1</td>
<td>.01</td>
<td>0</td>
<td>.00</td>
<td>0</td>
<td>.00</td>
<td>.93</td>
<td>.095</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>.03</td>
<td>7</td>
<td>.09</td>
<td>2</td>
<td>.03</td>
<td>68*</td>
<td>.85</td>
<td>1</td>
<td>.01</td>
<td>.85</td>
<td>.107</td>
</tr>
<tr>
<td>6</td>
<td>43*</td>
<td>.54</td>
<td>1</td>
<td>.01</td>
<td>9</td>
<td>.11</td>
<td>26</td>
<td>.32</td>
<td>1</td>
<td>.01</td>
<td>.49</td>
<td>.235</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>.01</td>
<td>25</td>
<td>.31</td>
<td>1</td>
<td>.01</td>
<td>0</td>
<td>.00</td>
<td>52*</td>
<td>.65</td>
<td>.65</td>
<td>-.042</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>.05</td>
<td>9</td>
<td>.11</td>
<td>20*</td>
<td>.25</td>
<td>47</td>
<td>.59</td>
<td>0</td>
<td>.00</td>
<td>.25</td>
<td>.180</td>
</tr>
</tbody>
</table>

* Correct response

1. If an item from a multiple-choice test were judged strictly upon its item discrimination index and difficulty level, which of the following is the best performing item?
   a. #1
   b. #8
   c. #6
   d. #4

2. What is the most plausible interpretation of the results from item #3 in the Item Analysis?
   a. It is a difficult item, indicating Application and above on the Table of Item Specifications.
   b. It is an ambiguous item.
   c. It is a well-performing item because some students selected all distracters.

3. What is the best interpretation of the results from item #4 in the Item Analysis?
   a. It is a poorly performing item.
   b. It measures the lower cognitive levels.
   c. The better performing students answered this item incorrectly.
   d. It is an easy item.
Refer to the following options when answering items 4-5.

1. The consistency of the test
2. The item difficulty
3. The content and cognitive level sampling
4. The degree to which the better performing students on the test answer the item correctly

4. Which of the options above should be affected if an instructor increases the number of well-constructed items on an achievement test?
   a. 1 and 2
   b. 1 and 3
   c. 1 and 4
   d. 3 and 4
   e. 2 and 4

5. The item analysis of a classroom test indicates a faulty item. If the instructor decides to replace it, which options must be considered when creating the new item?
   a. 1 and 3
   b. 2 and 3
   c. 2 and 4
   d. 3 and 4

General Item Writing Guidelines

An item or question contains three parts:

- The stem, in which the question is asked or the problem is stated
- The correct option
- The incorrect options, also called foils or distracters

The item should have only one correct answer and should be based on significant information or concepts, not trivia. The item should be clearly defined and be worded precisely without ambiguities.

Write the stem:

1. as a complete sentence or question, or an incomplete statement which is completed by selecting one of the responses. It is easier to write complete statements or questions without ambiguity. Measurement experts recommend complete sentences for those new to test construction.
2. in a positive form. Negative items are easier to write and easier for students to answer.
3. with a single correct answer. The stem may ask for the best answer, which elicits finer discriminations.
4. as precisely as possible. However, given the choice between a longer stem or longer options, lengthen the stem.

All options, both incorrect and correct, should:
- be brief,
- be grammatically consistent with the stem,
• be approximately the same length,
• be equally complex,
• cover the same type of content,
• be independent of each other, and
• follow the rules of grammar,

The distracters (or foils) should:
• be written for students who have a partial understanding or misunderstanding of the content,
• be plausible, and
• be similar to the correct answer.

The correct options should NOT:
• provide clues or
• be longer, more technical, or repeat any important words from the stem.

Generally speaking, refrain from using “all of the above” because a student is able to select the correct response from partial information. Also refrain from using “none of the above” because the correct response merely indicates that a student is able to identify what is incorrect but does not provide evidence that the accurate information (correct answer) is known.

References:


Assessment: multiple-choice tests. Office of Faculty and Organizational Development, Michigan State University, Lansing, MI. Retrieved from http://oir.fod.msu.edu/oir/Assessment/multiple-choice.asp
