Objective Test: Blueprint and Item Analysis

Jaime O’Connor
Assessment Coordinator
Students stress over tests!

The American Test Anxieties Association (amttaa.org) reports:

• Up to 16-20% of students experience high levels of test anxiety
• 18% experience moderate test anxiety symptoms
Causes of test anxiety

The Georgia Southern counseling center website notes two primary contributors:

- Inadequate preparation
- Worrying about past experiences, comparing to other students, consequences of failure
Hidden causes of test anxiety

The test itself can contribute to test anxiety and continue the pattern of negative testing experiences:

• Poor test construction
• Irrelevant or obscure material coverage
• Unclear directions
Students (and faculty) stress over tests!

Sources of faculty test stress:

- Writing a good test is time consuming
- Too much material to cover well in a test
- Grades and grade disputes
Why do we give exams?

• Measure individual student’s learning
  • Grades
    • Accurately measure “desired” learning

• Evaluate teaching effectiveness
  • Assessment
    • Accurately evaluate teaching strategies
What makes an exam a *good* exam?

- **Validity**
  - Indicator of how well an exam is measuring what it is supposed to measure
    - Must accurately measure content and depth of knowledge
    - Must be consistent with instruction and assignments

- **Reliability**
  - Indicator of a test’s consistency
    - Test should produce consistent results (across sections and semesters)
Steps for exam construction

1. Plan for exam – construct test blueprint
2. Prepare exam – select/write test items and arrange
3. Analyze and revise exam – perform test analysis
1. Plan for exam – purpose of test blueprint

- A test blueprint defines the parameters of an exam before you begin the exam creation process:
  - Ensures alignment of exam to content and depth of knowledge
  - Ensures alignment of exam with instruction
  - Ensures adequate “stretch” for students – Including opportunities for both high and low achieving students
  - Allows you to consider appropriate weighting schemes
1. Plan for exam – steps for constructing test blueprint

1. List important course student learning outcomes or content/topics
2. Identify cognitive levels expected or desired
3. Determine number of items for entire test and each cell based on: emphasis, time, importance, testing time available, and ability to help students generalize.
4. Consider appropriate weighting for individual test items as well opportunity for “stretch”
<table>
<thead>
<tr>
<th>Content</th>
<th>Knowledge (Recall)</th>
<th>Comprehension (Interpretation)</th>
<th>Application (New Setting)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts of Motion</td>
<td>1C</td>
<td>2C</td>
<td>1Q</td>
<td>1Q</td>
</tr>
<tr>
<td>1D Kinematics</td>
<td></td>
<td>1C</td>
<td>2Q</td>
<td></td>
</tr>
<tr>
<td>Vectors</td>
<td></td>
<td>1C</td>
<td>2Q</td>
<td>1C</td>
</tr>
<tr>
<td>2D Kinematics – Projectile Motion</td>
<td>1C</td>
<td>1C</td>
<td>1Q</td>
<td>1C</td>
</tr>
<tr>
<td>2D Kinematics – Circular Motion</td>
<td>1C</td>
<td></td>
<td>1Q</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>3C</td>
<td>5C</td>
<td>7Q</td>
<td>2C</td>
</tr>
</tbody>
</table>
Bloom’s taxonomy of educational objectives

Knowledge
- Recognition, recall

Comprehension
- Paraphrase, explain in own words

Application
- Apply knowledge/skill to a new problem or situation

Analysis, Synthesis, Evaluation
- Extract, combine, judge, draw conclusions

Created by Deemak Daksina from Noun Project
Activity – create test blueprint

<table>
<thead>
<tr>
<th>Content</th>
<th></th>
<th></th>
<th></th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Steps for exam construction

1. Plan for exam – construct test blueprint
2. Prepare exam – select/write test items and arrange
3. Analyze and revise exam – perform test analysis
2. Prepare exam — overall considerations

• Time
  • For students to complete the exam
  • For you to grade the exam

• Cognitive Objectives
  • Task necessary for students to demonstrate abilities
    • Selected response
    • Constructed response
    • Performance assessment
  • Task allowing opportunities for “stretch”
2. Prepare exam – individual item types

<table>
<thead>
<tr>
<th>Selected response</th>
<th>Constructed response</th>
<th>Performance assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• True/false</td>
<td>• Fill in the blank</td>
<td>• Oral exams</td>
</tr>
<tr>
<td>• Multiple choice</td>
<td>• Short answer</td>
<td>• Oral presentations</td>
</tr>
<tr>
<td>• Matching</td>
<td>• Essay</td>
<td>• Writing</td>
</tr>
<tr>
<td></td>
<td>• Problems/computational</td>
<td>• Building/creating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Skill demonstration</td>
</tr>
</tbody>
</table>

Analytical rubric
<table>
<thead>
<tr>
<th>Pros and cons of item types</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>Essay</strong></th>
<th><strong>Multiple Choice</strong></th>
<th><strong>Matching</strong></th>
<th><strong>True False</strong></th>
<th><strong>Completion</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended Uses</strong></td>
<td>When measuring the highest cognitive levels (synthesis and evaluation levels of Bloom’s Taxonomy of Educational Objectives – Cognitive Domain). When a response needs to be created. When evaluating writing ability.</td>
<td>When measuring content achievement at the knowledge, comprehension, application, analysis, and evaluation levels.</td>
<td>More appropriate for knowledge and comprehension learning levels. Reading time may be less, greater content coverage. When length of the test is an issue.</td>
<td>When a large amount of content needs to be tested in a short amount of time.</td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td>Relatively short amount of time required to construct the items. Allows for creativity, originality, and composition.</td>
<td>Objective scoring. (Once “correct” answers are decided). Efficient to score. Guessing factor is reduced. Evaluation of validity is possible by comparing the test and items to the Test Blueprint. Evaluation of reliability is possible. Thorough sampling of course content is possible. Item analysis resulting from test scores can reveal particular problems in the exam and/or in the instruction or learning.</td>
<td>Reading time may be less. Greater content coverage (may require less time to answer than multiple choice items). Efficient to score.</td>
<td>Relatively easy to construct and score. Efficient to score.</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Objective scoring is questionable, and more difficult. No generally acceptable criteria for demonstrating the validity and reliability of test. Course sampling is limited. Time consuming to evaluate responses.</td>
<td>Time consuming to construct. Difficult to construct items at the highest cognitive levels. Faculty must have some training or knowledge in test construction and item analysis techniques to write effective items that contribute to reliability.</td>
<td>Mostly reflects lower levels of understanding. Higher levels are more difficult to construct.</td>
<td>50% chance of guessing correctly. Mostly reflects lower levels of understanding.</td>
</tr>
<tr>
<td><strong>Item Writing Advice</strong></td>
<td>Limit choice of items. Keep each question short. Do not include sub questions. Use thesis statement students will support, modify, or refute.*</td>
<td>Stems should be correct and distractors should be similar in vocabulary, length, and are homogeneous. Distractors should be plausible to those without content mastery. Word items succinctly.</td>
<td>Develop homogeneous lists. Arrange numerically or alphabetically. May base the second list on examples or interpretations students have not previously seen. List on right should be longer than the stems.</td>
<td>Ask students to correct a false answer. Avoid negative wording. The number of true and false items should be approximately the same.</td>
</tr>
</tbody>
</table>

2. Prepare exam – maximize preparation

- Build a test bank

- Organize questions following your test blueprint
- Archive and rotate questions
- Retire or repurpose questions that are ineffective
## Activity

- develop items for a row from your test blueprint

<table>
<thead>
<tr>
<th>Content</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Steps for exam construction

1. Plan for exam – construct test blueprint
2. Prepare exam – select/write test items and arrange
3. Analyze and revise exam – perform test analysis
3. Analyze and revise exam – check for reliability

- Not a characteristic of the test itself - based upon the specific sample of examinees
- Typically based upon the correlation of two sets of scores - students’ performance remains consistent over two forms of the same exam
- Internal consistency is the typical method used to calculate reliability in the classroom
3. Analyze and revise exam – perform test analysis

1. KR-20: an outcome of .70 or higher
2. Point Bi-Serial Correlation (-1.0 – +1.0): should be positive
3. P-value (0% to 100% or 0 – 1.0): generally, >.20, <.90
4. Analysis of distracters
# Condensed Test Report

<table>
<thead>
<tr>
<th>No.</th>
<th>Question</th>
<th>Correct Answer</th>
<th>Response Frequencies</th>
<th>Non-Distractor</th>
<th>Correct Group Responses</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A  B  C  D  E</td>
<td></td>
<td>Total %  Upper 27%  Lower 27%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Q1</td>
<td>A</td>
<td>65.00 20.00 5.00 0.00 10.00</td>
<td>D</td>
<td>65.00 100.00 20.00</td>
<td>0.52</td>
</tr>
<tr>
<td>2</td>
<td>Q2</td>
<td>B</td>
<td>15.00 80.00 5.00 0.00 0.00</td>
<td>DE</td>
<td>80.00 100.00 80.00</td>
<td>0.34</td>
</tr>
<tr>
<td>3</td>
<td>Q3</td>
<td>C</td>
<td>0.00 20.00 75.00 5.00 0.00</td>
<td>AE</td>
<td>75.00 100.00 40.00</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>Q4</td>
<td>D</td>
<td>0.00 0.00 10.00 85.00 0.00</td>
<td>ABE</td>
<td>85.00 80.00 80.00</td>
<td>-0.10</td>
</tr>
<tr>
<td>5</td>
<td>Q5</td>
<td>E</td>
<td>0.00 0.00 0.00 10.00 90.00</td>
<td>ABC</td>
<td>90.00 100.00 80.00</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>Q6</td>
<td>D</td>
<td>0.00 5.00 0.00 55.00 40.00</td>
<td>AC</td>
<td>55.00 100.00 60.00</td>
<td>0.42</td>
</tr>
<tr>
<td>7</td>
<td>Q7</td>
<td>C</td>
<td>0.00 0.00 85.00 15.00 0.00</td>
<td>ABE</td>
<td>85.00 80.00 80.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>8</td>
<td>Q8</td>
<td>B</td>
<td>0.00 90.00 10.00 0.00 0.00</td>
<td>ADE</td>
<td>90.00 100.00 100.00</td>
<td>0.14</td>
</tr>
<tr>
<td>9</td>
<td>Q9</td>
<td>A</td>
<td>65.00 35.00 0.00 0.00 0.00</td>
<td>CDE</td>
<td>65.00 60.00 40.00</td>
<td>0.17</td>
</tr>
<tr>
<td>10</td>
<td>Q10</td>
<td>B</td>
<td>30.00 65.00 5.00 0.00 0.00</td>
<td>DE</td>
<td>65.00 40.00 60.00</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>Q11</td>
<td>C</td>
<td>0.00 10.00 90.00 0.00 0.00</td>
<td>ADE</td>
<td>90.00 100.00 80.00</td>
<td>0.18</td>
</tr>
<tr>
<td>12</td>
<td>Q12</td>
<td>D</td>
<td>0.00 0.00 25.00 75.00 0.00</td>
<td>ABE</td>
<td>75.00 80.00 80.00</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Example 1

Interpretation:

- Correct answer is A
- p value = 83.51
  - question answered correctly by 83.51% of class
- point biserial of .40
  - ≥.20 – high scoring students were more likely to choose the correct answer
- all distractors chosen

✓GOOD QUESTION
Example 2

<table>
<thead>
<tr>
<th>Response</th>
<th>Frequency</th>
<th>Percent</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>1.14</td>
<td>-0.27</td>
</tr>
<tr>
<td>°C</td>
<td>45</td>
<td>25.57</td>
<td>-0.14</td>
</tr>
<tr>
<td>D</td>
<td>129</td>
<td>73.30</td>
<td>0.20</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

Interpretation:

- Correct answer is C
- p value = 25.57
  - question answered correctly by only 25.57% of class
- point biserial = -0.14
  - <.20 – low scoring students were more likely to choose the correct answer

❌ POOR QUESTION
Review these test item statistics:

What can you conclude about this test item?

A. The distractors were implausible and should be replaced.

B. Low scoring students got this item correct more often than high-scoring students.

C. More than 10% of the class answered this question incorrectly.

D. This test item showed high discriminative ability and should be retained.
A. Incorrect. The distractors were all chosen by a small number of students and were plausible choices.

B. Incorrect. The point biserial was .57. This indicates that high scoring students were more likely to answer this question correctly.

C. Incorrect 90.29% of the class answered this item Correctly, which means less than 10% answered incorrectly.

D: Correct. 90.29% of the class answered this item correctly. The point biserial was 0.57 which indicates high discriminative ability (> .20 cutoff criterion). High-scoring students were more likely to answer this item correctly.
3. Analyze and revise exam – improve reliability

• Develop longer tests with well-constructed items
• Make sure items are positive discriminators; students who perform well on tests generally answer individual questions correctly
• Develop items of moderate difficulty; extremely easy or difficult questions do not add to reliability estimations
3. Analyze and revise exam –
use results for assessment purposes

- Analyze performance on each item according to the outcome evaluated
- Determine reasons for poor test performance
  - Faulty item
  - Irrelevant material
  - Unclear directions
- Make adjustments to remedy these problems
- Update test question bank
Steps for exam construction

1. Plan for exam – construct test blueprint
2. Prepare exam – select/write test items and arrange
3. Analyze and revise exam – perform test analysis
Thank you

For additional resources or information:

joconnor@georgiasouthern.edu
912-478-8442

Created by Adrien Coquet from Noun Project