In this module we will cover assembling, administering and analyzing the test.
Previously, we addressed how to write good objective items. In this module, we will discuss how to put those objective items together in a test form. We'll also consider some good policies for administering tests.

In addition, we'll spend time on test analysis, particularly item analysis. Item analysis can help you evaluate how well your objective items are actually working. While item analysis is a tool that can help the classroom teacher, it is also an approach that is routinely used in the development of standardized tests.

This material marks the transition between the first half of the semester, where we concentrated on classroom testing, and the second half of the semester, where we will concentrate on standardized testing.
Assembling the test

- Item arrangement
- Test format
- Test check

For assembling the test, there are several things that need to be noticed: Item arrangement, test format, and test check.
Item arrangement

- Group together items of the same format
  - All true-false items or all completion items together
- Order the test from easy to hard; at least the first few items easier for each type of items
  - Builds confidence and reduce test anxiety

Regarding item arrangement, group together items of the same format.

For example, keep all the true-false items or completion items together.

Be sure to order the test from easy to hard, or at least order it so the first few items are easier for each type of item.

This is done with the hope to build student confidence and reduce test anxiety.
Test format

• Space the items for readability, if not, it will interfere with a student’s true ability

• Keep an item’s stem and response options on the same page

• Place supporting material immediately above any related items

Regarding test format: Space the items for readability, if not, it will interfere with a student’s true ability. Keep an item’s stem and response options on the same page. Place supporting material immediately above any related items. For instance, a map is placed immediately above the items to which they refer.
Test check

- Proofread the test
  - Check any typos or grammatical errors
  - Use test assembling checklist

- Review the test directions
  - To check if it is clear or it includes all elements

- Review the answer key
  - No correct answer pattern
  - Correct answers are distributed equally

Test check is an important process when assembling the test.

Proofread the test. Check for any typos or grammatical errors. Use a test assembling checklist.

Review the test directions to check if its clear or it includes all elements.

Review the answer key. Make sure that there is no correct answer pattern and that the correct answers are distributed evenly.
Once the test has been created and assembled it is ready to be administered. There are some guidelines to follow when administering the test in terms of attitude efforts and directions for students.
You can help the students prepare for the test by: encouraging a positive test-taking attitude, maximizing achievement motivation, and minimizing distractions during the test.

**Attitude efforts**

- Try to encourage a positive test-taking attitude in the students
- Try to maximize achievement motivation
  - Serious but not anxious test attitude
- Try to minimize distractions during the test
  - Avoid a test taking place with noise
Here are some directions that can be used when administering the test:

Provide reasonable test-taking strategy suggestions, such as “Don’t spend too much time on difficult items.”

Specify any administrative rules, in terms of time limit, restroom policy, or the answer sheet.

Remind students to check their test forms, for example check page, item number, or put personal information on their answer sheet.

Give students time warnings when there are 10 or 15 minutes left.
As for the principles of scoring the test, prepare an answer key, and then check your answer key before the test.

After the test administration, score the test without the students’ names, check any machine-scored answer sheets, and finally record scores before handing back the test.
Debriefing the students

**Before** handing back the tests:
- Discuss any items you found to be problematic
- Listen to general student reactions
- Avoid on-the-spot decisions
- Be fair with any changes

**After** handing back the tests:
- Ask students to double-check their math
- Ask students to identify any items they found problematic

When reviewing the test with students, you should debrief the students both before and after handing back the tests.

Before handing out the tests, discuss any items you found to be problematic (for example, an item that everyone or almost everyone got wrong).
- Listen to general student reactions
- Avoid on the spot decisions
- Be fair with any changes

After Handing back the tests:
- Ask students to double check their scores.
- Ask students to identify any items they found problematic.
After assembling and administering the test, the next step is item analysis.

Item analysis can help you evaluate how well your objective items are actually working. These problems can be corrected, resulting in a better test, and better measurement.
Item analysis is an approach that is routinely used in the development of standardized tests.

It is most useful when you are developing a bank, or pool, of items that you will continue to use.

It is also a useful tool, anytime students have complained about an item.

It can be used to identify mis-keyed items as well as when evaluating standardized tests.
Qualitative item analysis

- Careful scrutiny of items to identify
  - flaws
  - mis-matches with instructional objectives
- Appropriate to edit and re-write items, to assess their content validity, after administering a test
- Qualitative is the primary approach for performance-based tasks
- Quantitative analysis can also be used to examine the distribution of students earning each possible score

Item analysis can be “qualitative” or “quantitative.”

Qualitative item analysis includes the careful scrutiny of items to identify flaws such as overlooking possible grammatical cues, specific determiners, double negatives as well as multiple defensible answers. It is also important to scrutinize the items for any mis-matches with instructional objectives.

It’s appropriate to edit and re-write items, and to assess their content validity, after administering a test, too.

For performance-based tasks, qualitative is the primary approach to item analysis. Quantitative analysis such as the examination of the distribution of students earning each possible score can be useful.
There are three primary quantitative indices or approaches for conducting item analyses on objective items.

Item difficulty measures the proportion correct and is represented with a lowercase “p.” Item discrimination measures extent to which a test item discriminates or differentiates between students who do well on the overall test and those who do not. A capital “D” is used to refer to the discrimination index.

A distractor analysis is conducted to measure the proportion of examinees selecting each response option.
To compute item difficulty, which tells us how easy or difficult the item is, we calculate the proportion of examinees tested that answered the item correctly. Item difficulty statistic is referred to as lowercase “p.”
Item difficulty can range from 0 to 1.

An item with a difficulty level of .75 to 1 indicates a moderately easy to very easy item, where an item with a difficulty level of 0 to .25 indicates a very difficult to moderately difficult item.

An item with a difficulty level of .25 to .75 indicates a moderate item.
Item discrimination describes the extent to which a test item discriminates or differentiates between students who do well on the overall test and those who do not.

This measure requires the class to be split into two groups, an upper group and a lower group. These groups are established on the scores received on the assessment.

To compute the discrimination for an particular item, subtract the number of students in the lower group who answer that item correctly from the number of students in the upper group who answer that item correctly, then divide by the number of students in either one group.

If the number of group members is not equal, use the number of the group with more students.

The higher the discrimination the better the item is able to discriminate, or distinguish between the two groups of students.
For large samples, item discrimination can be computed using correlation showing the relationship between a single item and the total test. This is referred to as a point-biserial correlation.
Item discrimination can range from -1 to 1.

A discrimination of .30 or higher indicates a moderate to high discrimination.

A discrimination of between .30 and 0 indicates little to no discrimination.

A discrimination of under 0, is a negative discrimination and is not desired.
Item difficulty and discrimination are related and together give information about how the items function.

The item can be little discrimination, which means that the item is not able to distinguish well between high and low performing groups if nearly everyone gets the item correct or nearly everyone gets the item wrong. In other words, when the item is too easy or too hard, there will be little discrimination of that item.

In contrast, there can be maximum discrimination if about half of the people got the item correct, and about half got the item wrong.
This graph depicts the relationship between discrimination and difficulty.

Here it can be seen that when achieving maximum discrimination the item difficulty will be about .5 and that when the item difficulty is an extreme measure, the discrimination will be low.
Distractor analysis is used to determine which of distractors or “wrong response options” students find attractive.

Distractor analysis can be done by comparing the number of examinees selecting each option in high and low groups, or consider the proportion of (total) students choosing each option.

<table>
<thead>
<tr>
<th></th>
<th>a*</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example:
Proportion of total examinees selecting each option

Distractor analysis can also consider the proportion of all the students choosing each option.

This example indicates that 78% of the students chose Response “a”, where 11% chose option b, 3% option c and 8% option d. The asterisk refers to which answer is the correct answer to the item.
Upper and lower groups are needed to compute discrimination values as well as for distractor analysis when comparing numbers of examinees.

To place examinees in upper and lower groups, the first step is to arrange the examinees by the total score. Then divide them into each group: Top half who has a higher score becomes upper group and bottom half who has a lower score becomes lower group.
Here is an example of an item with its corresponding distractor analysis.

The correct answer to the question is “A. Bern.” 13 students in the upper group and 1 student in the lower group chose this option. However, option “d” was an attractive response for students in the lower group as 9 students chose it on the test.

Each distractor was chosen at least once by students so they are all functioning properly as well.
For this same item, use this data table to compute the p (item difficulty), and D (item discrimination), and consider the distractors.

<table>
<thead>
<tr>
<th></th>
<th>a*</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lower</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

Total class size = 30

\[
p = \frac{N_{\text{correct}}}{N_{\text{total}}}
\]

\[
D = \frac{N_{\text{correct upper group}} - N_{\text{correct lower group}}}{N_{\text{either (one) group}}}
\]

For this same item, use this data table to compute the p (item difficulty), and D (item discrimination), and consider the distractors.

To compute item difficulty is to use the number of students who answer the item correctly divided by the total number of students who answer the item, which is 14 (students who answer the item right) divided by 30 (students who answer the item). The item difficulty is .47.

To compute the discrimination, the information about the numbers of upper and lower groups should be ready.

Recall that to compute the discrimination for an item, the number of students who answer that item correctly in the lower group is subtracted from the number of students who answer that item correctly in the upper group, divided by the number of students in either one group. In this case, 13 students in the upper group and 1 student in the lower group answered the item correctly. The total number of students for each group is 15 so the item discrimination is (13 minus 1) divided by 15, which is .80.

This suggests that the item is moderately difficult and has high discrimination.

Looking at the distractors, it is clear that option d was an attractive response for students in the lower group as 9 students chose it on the test.
Example 2: Item and analysis

2. The most important part of test planning is creating:
   a. sound instruction.
   b. a test blueprint.
   c. an item analysis plan.
   d. the grading curve.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b*</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lower</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Numbers in the upper and lower groups who selected each option

Here is another example.

The same number of students in the upper group got this answer correct as in the lower group.
Example 2: Compute $p$ and $D$

- Using this data table, compute $p$ and $D$, and consider the distractors.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b*</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lower</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total class size = 20

\[
p = \frac{N_{\text{correct}}}{N_{\text{total}}}
\]
\[
D = \frac{N_{\text{correct upper group}} - N_{\text{correct lower group}}}{N_{\text{either (one) group}}}
\]

Now using this example, compute the $p$ (item difficulty), and $D$ (item discrimination), and consider the distractors.

Item difficulty for this item is that 16 (students) divided by 20 (students) equals 0.80, indicating that this item is very easy.

Item discrimination is that 8 minus 8 divided by 10 equals 0.00, indicating that this item has no power to distinguish the students between upper and lower groups.

When examining the distractors, no one in the lower group chose option “c” and no one in the class chose option “d.” This would suggest that maybe these distractors need to be revisited.
Example 3: Item and analysis

3. Which type of essay item contains the most explicit instructions to students?
   a. extended response
   b. fixed response
   c. explicit response
   d. restricted response

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c*</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Lower</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Numbers in the upper and lower groups who selected each option

Now, examine this item and its corresponding distractor analysis table.
Example 3: Compute $p$ and $D$

- Using this data table, compute $p$ and $D$, and consider the distractors.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c*</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Lower</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Total class size = 40

\[ p = \frac{N_{\text{correct}}}{N_{\text{total}}} \]

\[ D = \frac{N_{\text{correct, upper group}} - N_{\text{correct, lower group}}}{N_{\text{either (one) group}}} \]

Compute the $p$, $D$, and consider the distractors.

The total class size is 40.

Item difficulty is 9 divided by 40, which is 0.23. This item is hard item.

Item discrimination is that 2 minus 7 divided by 20 equals -0.25. The negative value of item discrimination means that this item found the students in the lower group perform better than the students in the upper group. There may be something wrong in this item.

Distractor analysis suggests that majority of students in the upper group chose option “d” and only few students in the upper group chose option “c,” which is indicated to be a key for this item. Based on this information from distraction analysis, the item may be mis-keyed.
Let’s look at the item and its options more carefully. Yes, this item was mis-keyed. Option “d” was actually the correct answer. Restricted response essay items not explicit response essay items actually contains the most explicit instructions to students.

Now recalculate the $p$ and $D$, and consider distractor analysis with the correct key.

The $p$ is 22 divided by 40, which is 0.55 and the $D$ is (14 minus 8) divided by 20, which is 0.30, suggesting the item is moderately difficult and discriminates somewhat the students between the upper and lower groups.

With respect to the distractors, all of the options were selected at least once, suggesting they function well.
Distractor analysis

- Mis-keyed items
- Guessing
- Ambiguity

There are three issues regarding distractors that may be found after conducting distractor analysis.

First, when most students in the upper group fail to select the keyed option the item may be considered mis-keyed.

Second, guessing may occur when students in the upper group respond to an item in random fashion. In other words, there is not a majority of the students in the upper group choosing a particular option. This item maybe have not been taught yet.

Third, when a wrong answer is selected about the same amount of time as the correct answer this suggests ambiguity in the response options.