Objectives

- Apply current research in educational measurement specific to test item construction
- Identify the basic parts of a test item
- Distinguish good test items from ones that should be rewritten
- Apply Bloom’s “Revised” Taxonomy when writing or evaluating a test item for cognitive level
- Consider difficulty and item discrimination when reviewing a test item’s effectiveness
- Consider strategies for improving test items (and tests) over time
Anatomy of a Multiple-choice Question

Patients with congenital adrenal hyperplasia present with excessive circulating levels of ________________

a) ACTH
b) Aldosterone
c) BAM22
d) Cortisol
e) CXCR7

Key (correct answer)
Power Button

- Press once. Blue light indicates “on”.
- Automatically turns off in 5 minutes of non-use.

Response Buttons

- A – E
- Changed your mind? Press a different response button.
- Good Dog: A
- Bad Dog: B
Question 1
U.S. Grant was an ______________

a) president
b) man
c) alcoholic
d) general

Issues:
• Grammatical clue (a/an will fix that)
• Multiple defensible answers
Question 2

The free floating structures within the cell that synthesizes protein are called ______.

a) chromosomes 
b) lysosomes 
c) mitochondria 
d) free ribosomes

**Issues:**
• Stem clue
Question 3

The square root of 256 is _______

a) 14
b) 16
c) 4 X 4
d) both a and b
e) both b and c
f) all of the above

Issues:
• all/none of the above should be avoided
• can likely be figured out even if you can’t do the math!
Question 4

When 53 Americans were held hostage in Iran,

a) the US did nothing to try to free them
b) the US declared war on Iran
c) the US first attempted to free them by diplomatic means and later attempted a rescue
d) the US expelled all Iranian students

Issues:
- Put US in the stem to shorten the options
- Test writers tend to make the correct option longer than the distractors
Items to avoid
Type K (complex multiple-choice)

Which of the following behaviors suggests that you’re losing it?

A. You light a match to check a gas leak.

B. You pick apart your relationship with your significant other.

C. You advise your teenage son to use his own best judgment.

D. A and B

E. B and C

F. All of the above

Complex multiple-choice

- multiple combination choices of answers
  (1) A only; 2) both A and C; 3) both B and D; 4) A, B and C, 5) All of the Above)
- fewer can be answered in a given time period
- may be more dependent on test-taking skills than subject knowledge
- often have lower item discrimination scores

Items to avoid

Type K (complex true/false)

According to the laws of psychology, which of the following are true (A) and which are false (B)?

1. ___ Never ring a bell when a Pavlov’s dog is sitting on your lap

2. ___ Laws of behavior modification only apply to your neighbor’s children

3. ___ The right hand does know what the left hand is doing, it just doesn’t care.

4. ___ Adults get older faster than children and adults with children age the fastest

Which of the following are needed to calculate simple interest?

I. The amount of money borrowed
II. The interest rate
III. The length of the borrowing period

a) I only
b) I and II
c) I and III
d) I, II, and III
Type X: Research Shows …

**True/False**

- Difficult to write questions that avoid ambiguous statements without making the answer obvious.
- Writing true or false statements with no exceptions is difficult.
- Students have 50-50 chance of getting answer right.
- Students can make educated guesses increasing odds beyond 50-50 without knowing the answer outright.
Rules for *MCQ Test Items*

- Each item should focus on a *single* important concept
- Each item should assess *application of knowledge*, not recall of an isolated fact
- The stem of the item must pose a clear question
- All incorrect options should be homogenous and plausible
- Avoid technical flaws
And Remember Test-wiseness

*It’s real!*

- **Grammatical cues** (e.g., tense/case, singular/plural, nonparallel construction)
- **Logical cues** (e.g., some options illogical given the lead-in)
- **Absolute terms** (e.g., “never,” “always”)
- **Long correct answer** (e.g., the correct option is longer and more specific than the others)
- **Word repeats** (e.g., same/similar words in stem and correct option)
The pharmacological action of cortisol in the kidney is most similar to that of __________

a) Angiotensin II
b) Trimacinolone
c) Dexamethasone
d) Fludrocortisone
e) Betamethasone
An increase in the amplitude of cortisol secretion, with no change in the frequency or phase of cortisol secretion, in __________ is thought to result in ________________.

a) females, increased anxiety  
b) females, reduced anxiety  
c) males, cowardice  
d) males, reduced anxiety  
e) males, increased anxiety
Long-term therapy with prednisone (oral) in a female asthmatic patient would likely suppress levels of _________________ in that patient.

I. ACTH
II. Cortisone
III. Aldosterone

a) I only
b) III only
c) I and II only
d) II and III only
e) I, II, and III
Analyze and Re-write

Anatomy of a Multiple-choice Question

STEM

Patients with congenital adrenal hyperplasia present with excessive circulating levels of _____________.

Options

- a) ACTH
- b) Aldosterone
- c) BAM22
- d) Cortisol
- e) CXCR7

Writing the Stem

- Include the central idea in the stem instead of the distractors.
- Ensure that the stem is clear.
- Avoid excessive verbiage.
- Avoid negatives such as NOT and EXCEPT. If used, use them sparingly (per exam) and ensure they are capitalized or in boldface.

Writing Distractors

- Develop as many effective distractors as you can but research suggests that three are adequate.
- Make all distractors plausible.
- Make sure that only one distractor is the right answer.
- Vary the location of the “key” (right answer) per question.
- List distractors in a logical or numeric order.
- Keep distractors homogenous in content and grammatical structure. One distractor should not be significantly longer or shorter than the others.
- **None of the above, All of the above, A and B only** etc. should not be used.
- Phrase choices positively, avoid negatives such as NOT.
- Use humor only if it is compatible with the teacher and the learning environment.

<table>
<thead>
<tr>
<th>Bloom’s Taxonomy</th>
<th>Student Activity</th>
<th>Words or Phrases to Use in the Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Design, construct, develop, formulate, imagine, create, change, write a poem or short story</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Appraise, evaluate, justify, critique, recommend, which would be better?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Compare/contrast, distinguish between...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
On to matching test items with instructional goals
In assessing Mr. Delgado, which behavior is the most reassuring sign that he has been following his treatment plan for his hypertension and diabetes?

A. He has a list of glucose readings for the past 10 days
B. He has a list of medications along with newly refilled meds.
C. He has kept a nutritional log for a 3-day period
D. He can verbalize the side effects of all his medications
Goal:
- Learn all the important content
- Learn how to think critically about the subject

Teaching Activities?
- Lecture - experts conduct hour-long lectures

Feedback/Assessment: Mid-term exam
Result: Students could not reason through to the right answer
Discussion: Should you assess what you haven’t taught?
The Cognitive Domain
Bloom’s Taxonomy


Before you can ...

- **understand** a concept, you have to **remember** it
- **apply** a concept, you must **understand** it
- **analyze** a concept, you must be able to **apply** it
- **evaluate** its impact, you must have **analyzed** it
- **create**, you must have **remembered**, **understood**, **applied**, **analyzed**, and **evaluated**.
<table>
<thead>
<tr>
<th>Taxonomy Level</th>
<th>Verbs to trigger thinking at this level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creating:</strong> can the student create new product or point of view?</td>
<td>assemble, construct, create, design, develop, formulate, write.</td>
</tr>
<tr>
<td><strong>Evaluating:</strong> can the student justify a stand or decision?</td>
<td>appraise, argue, defend, judge, select, support, value, evaluate</td>
</tr>
<tr>
<td><strong>Analyzing:</strong> can the student distinguish between the different parts?</td>
<td>appraise, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test.</td>
</tr>
<tr>
<td><strong>Applying:</strong> can the student use the information in a new way?</td>
<td>choose, demonstrate, dramatize, employ, illustrate, interpret, operate, schedule, sketch, solve, use, write.</td>
</tr>
<tr>
<td><strong>Understanding:</strong> can the student explain ideas or concepts?</td>
<td>classify, describe, discuss, explain, identify, locate, recognize, report, select, translate, paraphrase</td>
</tr>
<tr>
<td><strong>Remembering:</strong> can the student recall or remember the information?</td>
<td>define, duplicate, list, memorize, recall, repeat, reproduce state</td>
</tr>
</tbody>
</table>
What was the learning objective? And what level of the taxonomy was tapped?

In assessing Mr. Delgado, which behavior is the most reassuring sign that he has been following his treatment plan for his hypertension and diabetes?

A. He has a list of glucose readings for the past 10 days
B. He has a list of medications along with newly refilled meds.
C. He has kept a nutritional log for a 3-day period
D. He can verbalize the side effects of all his medications

<table>
<thead>
<tr>
<th>Taxonomy Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating: can the student create new product or point of view?</td>
</tr>
<tr>
<td>Evaluating: can the student justify a stand or decision?</td>
</tr>
<tr>
<td>Analyzing: can the student distinguish between the different parts?</td>
</tr>
<tr>
<td>Applying: can the student use the information in a new way?</td>
</tr>
<tr>
<td>Understanding: can the student explain ideas or concepts?</td>
</tr>
<tr>
<td>Remembering: can the student recall or remember the information?</td>
</tr>
</tbody>
</table>
Gotta love Iowa State

Retrieved from: http://www.celt.iastate.edu/teaching-resources/effective-practice/revised-blooms-taxonomy/
What’s the Bloomin’ Level?

Writing the Stem
- Include the central idea in the stem instead of the distractors.
- Ensure that the stem is clear.
- Avoid excessive verbiage.
- Avoid negatives such as NOT and EXCEPT. If used, use them sparingly (per exam) and ensure they are capitalized or in boldface.

Writing Distractors
- Develop as many effective distractors as you can but research suggests that three are adequate.
- Make all distractors plausible.
- Make sure that only one distractor is the right answer.
- Vary the location of the “key” (right answer) per question.
- List distractors in a logical or numeric order.
- Keep distractors homogenous in content and grammatical structure. One distractor should not be significantly longer or shorter than the others.
- **None of the above, All of the above, A and B only**, etc. should not be used.
- Phrase choices positively; avoid negatives such as NOT.
- Use humor only if it is compatible with the teacher and the learning environment.

Anatomy of a Multiple-choice Question

**STEM**

- Patients with congenital adrenal hyperplasia present with excessive circulating levels of ____________.
- Key (correct answer)
  - a) ACTH
  - b) Aldosterone
  - c) BAM22
  - d) Cortisol
  - e) CXCR7

**OPTIONS**

**Words or Phrases to Use in the Stem**

<table>
<thead>
<tr>
<th>Bloom’s Taxonomy</th>
<th>Student Activity (Anderson &amp; Krathwohl, 2001, pp. 67-68)</th>
<th>Words or Phrases to Use in the Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design, construct, develop, formulate, imagine, create, change, write a poem or short story</td>
<td><strong>Appraise, evaluate, justify, critique, recommend, which would be better?</strong></td>
<td></td>
</tr>
</tbody>
</table>
On to Psychometrics …
Nine out of Ten Psychometricians Say …

The best tests:

- Include questions from across the spectrum of the curriculum being tested
- Have a mix of item difficulty
- Do not include difficult items just for the sake of it
- Are analyzed after administration
- Use item discrimination to think about an item’s effectiveness
- **NOTE:** You can’t estimate item effectiveness in advance
Two measures of item effectiveness

**Difficulty and Discrimination**

- **Difficulty** (p-value)
  - The number of examinees who answer an item correctly

- **Discrimination** (iD and/or point biserial)
  - A comparison of top scorers with low scorers
Item Difficulty

$p$-value

42 students answered the item

$$\frac{8}{42} \approx 0.19$$
The higher the value, the easier the item.

- **Above 0.90** -- too easy; review for question’s purpose (warm up? fundamental?)

- **Below 0.20** -- too difficult; review for confusing language, remove from subsequent exams, and/or identify as area for re-instruction.
**Item Difficulty: Trivia**

*When guessing is taken into account*

\[
g = \frac{\text{# distractors}}{100}
\]

<table>
<thead>
<tr>
<th><strong>True/False</strong></th>
<th><strong>Multi-item MCQ</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• 2 items (g=.5)</td>
<td>• 4 items (g=.25)</td>
</tr>
<tr>
<td>• Optimal p = .75</td>
<td>• Optimal p = .63</td>
</tr>
<tr>
<td></td>
<td>• 5 items (g=.20)</td>
</tr>
<tr>
<td></td>
<td>• Optimal p = .60</td>
</tr>
</tbody>
</table>

\[
\text{Optimal p-value} = \frac{1.0 + g}{2}
\]
Item Discrimination

point-biserial correlation

Top 27%

Bottom 27%

\[
\frac{(# \text{ Upper Group Correct}) - (# \text{ Lower Group Correct})}{\text{Number of Students in the Upper Group}} = \frac{5 - 2}{6} = 0.50
\]

Image Sources:
## Item Discrimination

**point biserial range**

<table>
<thead>
<tr>
<th>Negative ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unacceptable</strong></td>
<td>Unacceptable – check for item error</td>
</tr>
<tr>
<td><strong>0% - 24%</strong></td>
<td>Usually unacceptable</td>
</tr>
<tr>
<td><strong>25% - 39%</strong></td>
<td>Good item</td>
</tr>
<tr>
<td><strong>40% - 100%</strong></td>
<td>Excellent item</td>
</tr>
</tbody>
</table>

Adapted from University of Wisconsin Oshkosh:
http://www.uwosh.edu/testing/facultyinfo/itemdiscrimone.php
### Oregon State University Test Scoring

#### Item Analysis

<table>
<thead>
<tr>
<th>QUESTION NBR ANSWER</th>
<th>ANSWER SELECTED</th>
<th>CORRECT</th>
<th>MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NONE</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>NUMBER OF QUESTIONS = 62</td>
<td>AVERAGE SCORE = 52.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER OF RESPONDENTS = 80</td>
<td>27% OF RESPONDENTS = 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAXIMUM SCORE POSS. = 62.0</td>
<td>MEDIAN SCORE = 53.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STANDARD DEVIATION = 3.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Item Analysis Details

<table>
<thead>
<tr>
<th>ITEM</th>
<th>TOTAL</th>
<th>DIFF INDEX</th>
<th>T-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74/92.5</td>
<td>0.925</td>
<td>0.631</td>
</tr>
<tr>
<td>2</td>
<td>74/92.5</td>
<td>0.925</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>78/97.5</td>
<td>0.975</td>
<td>2.212*</td>
</tr>
<tr>
<td>4</td>
<td>74/92.5</td>
<td>0.925</td>
<td>-0.433</td>
</tr>
<tr>
<td>5</td>
<td>79/98.7</td>
<td>0.987</td>
<td>1.794</td>
</tr>
</tbody>
</table>

*Denotes statistical significance.
T-values and Statistical Significance

• The score obtained when you perform a *T-Test*.
• Represents the difference between the mean or average scores of two groups while taking into account any variation in scores.
• The t-value measures the difference in scores between two groups.
  – Is the t-value is big enough for you to say that one group is significantly different from the other?
  – Was the result was something that could have just happened by chance?
## Condensed Test Report

- **Total Possible Points:** 60.00
- **Total Students:** 8.01
- **Standard Deviation:**

<table>
<thead>
<tr>
<th>No.</th>
<th>Question</th>
<th>Correct Answer</th>
<th>Response Frequencies</th>
<th>Correct Group Responses</th>
<th>Non-Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A 65.00 B 20.00 C 20.00 D 0.00 E 10.00</td>
<td>Total % Upper 27% Lower 27% Point Biserial</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Q1</td>
<td>A</td>
<td>15.00 80.00 75.00 5.00 0.00 0.00 DE</td>
<td>60.00 60.00 60.00 60.00 0.34 0.23</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Q2</td>
<td>B</td>
<td>0.00 0.00 20.00 90.00 5.00 5.00 ABE</td>
<td>75.00 75.00 75.00 75.00 0.34 0.23</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q3</td>
<td>C</td>
<td>0.00 0.00 0.00 0.00 10.00 10.00 ABC</td>
<td>85.00 85.00 85.00 85.00 0.34 0.23</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Q4</td>
<td>D</td>
<td>0.00 0.00 0.00 0.00 0.00 0.00 ABC</td>
<td>90.00 90.00 90.00 90.00 0.34 0.23</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Q5</td>
<td>E</td>
<td>0.00 0.00 0.00 0.00 0.00 0.00 ABC</td>
<td>95.00 95.00 95.00 95.00 0.34 0.23</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Q6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Q10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Q11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Q12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Q13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Q14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Q15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Q16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Q17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Q18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Q19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Q20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Q21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Q22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Q23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Q24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Summary
- **Median Score:** 46.55
- **Mean Score:** 46.56
- **Reliability Coefficient (KR20):** 0.58
- **Maximum Score:** 66.00
- **Minimum Score:** 32.00
- **Range of Scores:** 34.00

---

A Kinder, Gentler Scantron Report
Reliability

Kuder-Richardson Formula 20 (KR-20)

• The measure obtained by administering the same test twice over a period of time to the same individuals.
• Scores from time 1 and time 2 are correlated to evaluate the test for stability over time.
• Acceptable reliability coefficients?
  – 0.60 is an acceptable lower value
**From 30,000 Feet**

### Oregon State University Test Scoring

#### Statistical Summary

<table>
<thead>
<tr>
<th>COURSE:</th>
<th>RANK</th>
<th>RAW SCORE</th>
<th>%</th>
<th>% TOTAL</th>
<th>CUM FREQ</th>
<th>FREQ</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>62</td>
<td>100.0</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>61</td>
<td>98.3</td>
<td>2.5</td>
<td>3</td>
<td>2</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>60</td>
<td>96.7</td>
<td>0.0</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>59</td>
<td>95.1</td>
<td>2.5</td>
<td>5</td>
<td>2</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>58</td>
<td>93.5</td>
<td>5.0</td>
<td>9</td>
<td>4</td>
<td>****</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>57</td>
<td>91.9</td>
<td>3.7</td>
<td>12</td>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>56</td>
<td>90.3</td>
<td>7.5</td>
<td>18</td>
<td>6</td>
<td>******</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>55</td>
<td>88.7</td>
<td>17.5</td>
<td>32</td>
<td>14</td>
<td>******************</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>54</td>
<td>87.1</td>
<td>10.0</td>
<td>40</td>
<td>8</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>53</td>
<td>85.4</td>
<td>11.2</td>
<td>49</td>
<td>9</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>52</td>
<td>83.8</td>
<td>11.2</td>
<td>58</td>
<td>9</td>
<td>********</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>51</td>
<td>82.2</td>
<td>2.5</td>
<td>60</td>
<td>2</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>50</td>
<td>80.6</td>
<td>3.7</td>
<td>63</td>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>49</td>
<td>79.0</td>
<td>3.7</td>
<td>66</td>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>48</td>
<td>77.4</td>
<td>6.2</td>
<td>71</td>
<td>5</td>
<td>*****</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>47</td>
<td>75.8</td>
<td>3.7</td>
<td>74</td>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>46</td>
<td>74.1</td>
<td>3.7</td>
<td>77</td>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>45</td>
<td>72.5</td>
<td>1.2</td>
<td>78</td>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>44</td>
<td>70.9</td>
<td>2.5</td>
<td>80</td>
<td>2</td>
<td>**</td>
</tr>
</tbody>
</table>

---

**Exam Statistics**

- **High Score** = 62.0
- **Maximum Score Poss.** = 62.0
- **Number of Questions** = 62
- **Average Score** = 52.9
- **Number of Respondents** = 80
- **Low Score** = 44.0
- **Median Score** = 53.5
- **Standard Deviation** = 3.94
Oregon State University Test Scoring

Statistical Summary

<table>
<thead>
<tr>
<th>RAW SCORE</th>
<th>%TILE</th>
<th>Z-SCORE</th>
<th>T-SCORE</th>
<th>STANINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.0</td>
<td>98.75</td>
<td>+2.30</td>
<td>73.09</td>
<td>10</td>
</tr>
<tr>
<td>61.0</td>
<td>96.25</td>
<td>+2.05</td>
<td>70.55</td>
<td>10</td>
</tr>
<tr>
<td>59.0</td>
<td>93.75</td>
<td>+1.54</td>
<td>65.48</td>
<td>10</td>
</tr>
<tr>
<td>58.0</td>
<td>88.75</td>
<td>+1.29</td>
<td>62.94</td>
<td>9</td>
</tr>
<tr>
<td>57.0</td>
<td>85.00</td>
<td>+1.04</td>
<td>60.40</td>
<td>9</td>
</tr>
<tr>
<td>56.0</td>
<td>77.50</td>
<td>+0.78</td>
<td>57.86</td>
<td>8</td>
</tr>
<tr>
<td>55.0</td>
<td>60.00</td>
<td>+0.53</td>
<td>55.32</td>
<td>7</td>
</tr>
<tr>
<td>54.0</td>
<td>50.00</td>
<td>+0.27</td>
<td>52.79</td>
<td>6</td>
</tr>
<tr>
<td>53.0</td>
<td>38.75</td>
<td>+0.02</td>
<td>50.25</td>
<td>4</td>
</tr>
<tr>
<td>52.0</td>
<td>27.50</td>
<td>-0.22</td>
<td>47.72</td>
<td>3</td>
</tr>
<tr>
<td>51.0</td>
<td>25.00</td>
<td>-0.48</td>
<td>45.18</td>
<td>3</td>
</tr>
<tr>
<td>50.0</td>
<td>21.25</td>
<td>-0.73</td>
<td>42.64</td>
<td>3</td>
</tr>
<tr>
<td>49.0</td>
<td>17.50</td>
<td>-0.98</td>
<td>40.11</td>
<td>2</td>
</tr>
<tr>
<td>48.0</td>
<td>11.25</td>
<td>-1.24</td>
<td>37.57</td>
<td>2</td>
</tr>
<tr>
<td>47.0</td>
<td>7.50</td>
<td>-1.49</td>
<td>35.03</td>
<td>1</td>
</tr>
<tr>
<td>46.0</td>
<td>3.75</td>
<td>-1.75</td>
<td>32.49</td>
<td>1</td>
</tr>
<tr>
<td>45.0</td>
<td>2.50</td>
<td>-2.00</td>
<td>29.95</td>
<td>1</td>
</tr>
<tr>
<td>44.0</td>
<td>0.00</td>
<td>-2.25</td>
<td>27.42</td>
<td>1</td>
</tr>
</tbody>
</table>

---FORM ALL ---SECTION ALL

EXAM STATISTICS

<table>
<thead>
<tr>
<th></th>
<th>HIGH SCORE</th>
<th>MAXIMUM SCORE POSS.</th>
<th>AVERAGE SCORE</th>
<th>MEDIAN SCORE</th>
<th>LOW SCORE</th>
<th>STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF QUESTIONS</td>
<td>62</td>
<td></td>
<td>52.9</td>
<td>53.5</td>
<td>44.0</td>
<td>3.94</td>
</tr>
<tr>
<td>NUMBER OF RESPONDENTS</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PAGE 2
Finding Good Dogs and Bad Dogs

- Which items had the best difficulty scores?
- discrimination scores?

- Which items were good foundational questions?

- Comparing difficulty AND discrimination, which items had the best balance of the two?

- What is your overall “take” about this exam?
Objectives review

• Apply current research in educational measurement specific to test item construction
• Identify the basic parts of a test item
• Distinguish good test items from ones that should be rewritten
• Apply Bloom’s “Revised” Taxonomy when writing or evaluating a test item for cognitive level
• Consider difficulty and item discrimination when reviewing a test item’s effectiveness
• Consider strategies for improving test items (and tests) over time