Using Assessment to Improve and Evaluate Student Learning in Introductory Statistics

Bob delMas
University of Minnesota
Department of Educational Psychology
delma001@umn.edu
American Association for Higher Education (AAHE)
9 Principles of Good Practice for Assessing Student Learning


1. The assessment of student learning begins with educational values. Assessment is not an end in itself but a vehicle for educational improvement. Its effective practice, then, begins with and enacts a vision of the kinds of learning we most value for students and strive to help them achieve. Educational values should drive not only what we choose to assess but also how we do so. Where questions about educational mission and values are skipped over, assessment threatens to be an exercise in measuring what’s easy, rather than a process of improving what we really care about.

2. Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time. Learning is a complex process. It entails not only what students know but what they can do with what they know; it involves not only knowledge and abilities but values, attitudes, and habits of mind that affect both academic success and performance beyond the classroom. Assessment should reflect these understandings by employing a diverse array of methods, including those that call for actual performance, using them over time so as to reveal change, growth, and increasing degrees of integration. Such an approach aims for a more complete and accurate picture of learning, and therefore firmer bases for improving our students’ educational experience.

3. Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes. Assessment is a goal-oriented process. It entails comparing educational performance with educational purposes and expectations--these derived from the institution’s mission, from faculty intentions in program and course design, and from knowledge of students’ own goals. Where program purposes lack specificity or agreement, assessment as a process pushes a campus toward clarity about where to aim and what standards to apply; assessment also prompts attention to where and how program goals will be taught and learned. Clear, shared, implementable goals are the cornerstone for assessment that is focused and useful.

4. Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes. Information about outcomes is of high importance; where students “end up” matters greatly. But to improve outcomes, we need to know about student experience along the way--about the curricula, teaching, and kind of student effort that lead to particular outcomes. Assessment can help us understand which students learn best under what conditions; with such knowledge comes the capacity to improve the whole of their learning.

5. Assessment works best when it is ongoing, not episodic. Assessment is a process whose power is cumulative. Though isolated, “one-shot” assessment can be better than none, improvement over time is best fostered when assessment entails a linked series of cohorts of students; it may mean collecting the same examples of student performance or using the same instrument semester after semester. The point is to monitor progress toward intended goals in a spirit of continuous improvement. Along the way, the assessment process itself should be evaluated and refined in light of emerging insights.
6. **Assessment fosters wider improvement when representatives from across the educational community are involved.** Student learning is a campus-wide responsibility, and assessment is a way of enacting that responsibility. Thus, while assessment efforts may start small, the aim over time is to involve people from across the educational community. Faculty play an especially important role, but assessment’s questions can’t be fully addressed without participation by student-affairs educators, librarians, administrators, and students. Assessment may also involve individuals from beyond the campus (alumni/ae, trustees, employers) whose experience can enrich the sense of appropriate aims and standards for learning. Thus understood, assessment is not a task for small groups of experts but a collaborative activity; its aim is wider, better-informed attention to student learning by all parties with a stake in its improvement.

7. **Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.** Assessment recognizes the value of information in the process of improvement. But to be useful, information must be connected to issues or questions that people really care about. This implies assessment approaches that produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made. It means thinking in advance about how the information will be used, and by whom. The point of assessment is not to gather data and return “results”; it is a process that starts with the questions of decision-makers, that involves them in the gathering and interpreting of data, and that informs and helps guide continuous improvement.

8. **Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.** Assessment alone changes little. Its greatest contribution comes on campuses where the quality of teaching and learning is visibly valued and worked at. On such campuses, the push to improve educational performance is a visible and primary goal of leadership; improving the quality of undergraduate education is central to the institution’s planning, budgeting, and personnel decisions. On such campuses, information about learning outcomes is seen as an integral part of decision making, and avidly sought.

9. **Through assessment, educators meet responsibilities to students and to the public.** There is a compelling public stake in education. As educators, we have a responsibility to the publics that support or depend on us to provide information about the ways in which our students meet goals and expectations. But that responsibility goes beyond the reporting of such information; our deeper obligation--to ourselves, our students, and society--is to improve. Those to whom educators are accountable have a corresponding obligation to support such attempts at improvement.
Assessment Triangle


The three elements are interdependent.

A successful assessment synchronizes all three elements.

**Cognition**: the aspects of achievement or competencies that are to be assessed.

**Observation**: the tasks used to collect evidence about students’ achievement (i.e., the assessments).

**Interpretation**: the methods used to analyze the evidence resulting from the tasks.
Ideas from Beth Chance (Cal Poly, San Luis Obispo, CA).


An Assessment Cycle

- **Set goals**
  - What should students know, be able to do?
  - At what point in the course?
  - Identify assessable learning outcomes that match goals

- **Select methods**
  - Identify an assessment that matches the type of learning outcome
  - Consider minute papers, article reviews, newspaper assignments, projects, short answer items, multiple choice
  - Can the assessment be built into the activity?

- **Gather evidence** (i.e., administer the assessment)

- **Draw inference**
  - Don’t use results just to assign a grade
  - Consider what responses indicate about student understanding

- **Take action**
  - Provide feedback
  - What can be done to remedy a misunderstanding (an activity; extra reading; more experience with a procedure or a concept)
  - Re-examine goals and methods
Standard Normal Distribution Activity

Z-scores and the Normal Curve (textbook reference: pages 101 - 104)

To complete this activity, you may need to review several properties of z-scores.

1. Symbols: \( y = \) any value; \( \mu = \) mean of a population; \( \sigma = \) standard deviation of a population

2. The formula for a z-value is \( z = \frac{y - \mu}{\sigma} \)

3. A z-score represents a DISTANCE along the horizontal axis.

4. More precisely, a z-score represents how many standard deviations a value \( y \) falls above or below the population mean \( \mu \).
   a. When z is NEGATIVE, the value \( y \) falls BELOW the population mean \( \mu \).
   b. When z is POSITIVE, the value \( y \) falls ABOVE the population mean \( \mu \).

Finding the area BELOW a z-score

To complete the problems in this activity, you need to work with a table of standard normal values. You can use Table Z: Areas Under the Standard Normal Curve located in Appendix E at the back of your textbook, or you can use the Normal Density Tool that comes with ActivStats.

To use the Normal Density Tool:

1. Start ActivStats
2. Click on the CONTENTS tab
3. Scroll to the bottom of the list of Contents. Click the VII Appendix A
4. Scroll down and click 4. Additional Resources
5. In the ActivStats window, click the icon for the Normal Density Tool
6. From the SETTINGS menu, select Show Flag Values.

To use the Normal Density Tool, just point to one of the vertical lines that mark a z-score value, hold down the mouse button, and slide the vertical line to the left or the right. The z-score value at the top of the vertical line automatically changes as you move the mouse. The area in each tail of the normal distribution is marked in red, and the size of each area is reported in a box.

Use either Table Z or the Normal Density Tool to find the probability that a z-score is less than \(-1\).

Probability = \( P(z < -1) = \) 

Another way to write “the probability that \( z \) is less than \(-1\)” is to use the notation \( P(z < -1) \). This statement is just a formal way of writing the probability of getting a z-score less than \(-1\) in a normal distribution. Here, the “\( P \)” stand for “probability” and the statement inside the parentheses “\( z < -1 \)” simply means “\( z \) is less than \(-1\)”.

- 6 -
This probability indicates the proportion of the area under the normal curve that is below, or to the left, of the z-score of -1. Label and shade the graph on the right to indicate:

- a) The point where \( z = -1 \) falls on the number line
- b) The area under the curve that represents \( P(z \leq -1) \)
- c) The probability of \( P(z < -1) \)

**Finding the area ABOVE a z-score**

Next, calculate the probability that \( z \) is greater than -1 (in other words, \( P(z > -1) \)). Here’s a trick you can use to find \( P(z > -1) \). Because the normal distribution is perfectly symmetric, then the area above \( z = -1 \) is exactly the same size as the area below \( z = 1 \). In other words, to find \( P(z > -1) \), simply go to the table and find \( P(z < 1) \). This will give you the answer.

\[
\text{Probability} = P(z > -1) = P(z < 1) = \_\_\_\_\_
\]

You can also see this by using the **Normal Density Tool** in ActivStats.

1. Go to the **SETTINGS** menu and select **Synchronize 2-tails** (to turn this off).
2. Point to the vertical line for the left z-score and drag it off the graph to the left.
3. Point to the vertical line for the right z-score and drag it to the left to a z-score of -1. The value in the area box should be equal to the value you wrote above for \( P(z > -1) \).

Below are two graphs. Use the graph on the left below to label and shade in the area that corresponds to \( P(z > -1) \). Then use the graph on the right to label and shade in the area that corresponds to \( P(z < 1) \). When you look at the two graphs, you should see that the two shaded areas are just mirror images of each other.

Use Table Z or the Normal Density Tool to test this out for three other values of \( z \). For each \( z \)-score, find both the **area BELOW the z-score** and the **area ABOVE the NEGATIVE of the z-score**. Check that the two areas are equal. Label and shade in the areas that correspond to each probability.
Characteristics of the Normal Distribution

One of the interesting features of the normal distribution is its symmetry. If you draw a vertical line through any normal distribution at the point that corresponds to the population mean on the horizontal axis (i.e., where $z = 0$), you divide the normal distribution into two halves that are mirror images of each other. You can illustrate this point by using Table Z or the Normal Density Tool.

First, find that probability $P(z < -0.00) = \underline{\quad} \quad P(z > 0.00) = \underline{\quad}$

Then, find that probability $P(z < 0.00) = \underline{\quad}$

You will notice that the two probabilities are both equal to .5, or one-half. In other words, the mean splits the normal distribution into equal halves.

You can find this same symmetry for any $z$-score and its opposite. Select any NEGATIVE $z$-score that you want, but don’t select a value of $z = -1$, $z = -2$, or $z = -3$. For the first pair listed below, write the negative $z$-score into the space for the expression $P(z < -\quad)$. Now, find the area BELOW this $z$-
score and write the probability in to the blank space after the equal sign. Take the absolute value of the negative z-score and write it into the second expression for the first pair, P(z >  ).

For example, if you chose P(z < -1.57), then the second expression would be P(z > 1.57). Find the probability that z is GREATER than this second value and record this value below.

You can do this easily with the Normal Density Tool. Just go to the SETTINGS menu and select Synchronize 2-tails (to turn it on again). When you move one of the vertical lines for a z-score, both vertical lines will move, and the size of both areas are reported.

Finally, label and shade the graph to correspond to the two z-scores and their respective areas.

First pair: P(z < - ) = ________  P(z > ) ________

Use the same process for one more pair of z-scores. Choose a NEGATIVE z-score for the first expression and then the absolute value of this expression for the second expression. Label and shade the graph to correspond to the z-values and the probabilities.

Second pair: P(z < - ) = ________  P(z > ) ________


You can use the symmetry of the normal distribution to come up with what is referred to as the 68-95-99.7 rule.

First, find P(z < -1) = __________.
Next, find P(z > 1) = __________.
Label and shade both of these areas in the graph below.

These two areas represent the area under the normal curve that are NOT between \( z = -1 \) and \( z = 1 \). In other words, they represent the area that is NOT between one standard deviation below and above the mean. So, if you add the two probabilities together, you get the probability of NOT being between \( z = -1 \) and \( z = 1 \). Calculate this value: \( P(z < -1) + P(z > 1) = \) __________.

To find the area between the \( z = -1 \) and \( z = 1 \), you simply subtract the probability you just calculated from 1. The probability that \( z \) is between \(-1 \) and \( 1 \) can be written as \( P(-1 < z < 1) \). Therefore,

\[
P(-1 < z < 1) = 1 - \left[ P(z < -1) + P(z > 1) \right]
\]

You should find that this probability comes very close to .68, or 68%.

Use the method presented above to find the exact probabilities that go with the other two parts of the 68-95-99.7 rule. Be sure to label and shade the graph to represent the area.

\[
P(z < -2) = \quad P(z < 2) = \quad P(-2 < z < 2) =
\]

\[
P(z < -3) = \quad P(z < 3) = \quad P(-3 < z < 3) =
\]
The Situation: Joe, Talia, Cally, Desmond, and Nick are five clerks who work in the small claims office for the government of a large metropolitan city. Their supervisor periodically checks up on them to see how long it takes each of them to process claims. In order to deal with the large volume of small claims that come through the office each day, each clerk must take 6 minutes or less to process a claim, on the average. It would be too time consuming, if not impossible, for the supervisor to monitor every transaction performed by each clerk. The supervisor needs to decide on the best way to make a reliable estimate of the time it takes each clerk to process claims. In general, she has one of three options:

Small Sample: Pick a small random sample (somewhere between 2 and 4 claims).

Medium Sample: Pick a medium random sample (somewhere between 8 and 12 claims).

Large Sample: Pick a large random sample (somewhere between 15 and 20 claims).

The supervisor would choose only one sample. For any sample she chooses, she would calculate the average time it took the employee to process the claims in the sample. If a clerk takes longer than 6 minutes, on the average, to process small claims, the clerk may be moved to a less demanding position, or possibly fired. The supervisor does not want to make an incorrect decision and dismiss a clerk who is actually performing their job well. Your task is to find the method that will help the supervisor make the best decision.

Prediction: Which monitoring method should the supervisor choose?

<table>
<thead>
<tr>
<th>Small Sample</th>
<th>Medium Sample</th>
<th>Large Sample</th>
<th>Doesn’t Matter</th>
</tr>
</thead>
</table>

Your instructor will now lead you through several tasks that look at what might happen if the supervisor used each of the three methods with different employees.

First Employee: Joe

Joe’s claim processing times follow a Normal Distribution

Second Employee: Talia

Talia’s claim processing times do NOT follow a Normal Distribution

Third Employee: Cally

Cally’s claim processing times have a very Erratic (Irregular) pattern
Sampling Distributions Activity Scrapbook

Case 1: Joe
Normal Distribution
\( \mu = 5.00 \)
\( \sigma = 1.805 \)

Case 2: Talia
Negatively Skewed Population
\( \mu = 6.81 \)
\( \sigma = 2.063 \)

Case 3: Cally
Irregularly Shaped Population
\( \mu = 5.00 \)
\( \sigma = 3.410 \)

Distribution of Sample Means
\( n = 2 \)

Guess 1

A B C D E

Mean of \( \bar{X} = \) sd of \( \bar{X} = \)

Distribution of Sample Means
\( n = 9 \)

Guess 2

A B C D E

Mean of \( \bar{X} = \) sd of \( \bar{X} = \)

Distribution of Sample Means
\( n = 16 \)

Guess 3

A B C D E

Mean of \( \bar{X} = \) sd of \( \bar{X} = \)

Distribution of Sample Means
\( n = 16 \)

Guess 6

A B C D E

Mean of \( \bar{X} = \) sd of \( \bar{X} = \)

Distribution of Sample Means
\( n = 16 \)

Guess 9

A B C D E

Mean of \( \bar{X} = \) sd of \( \bar{X} = \)
Suggested Readings

**Book on Assessment in Statistics Education**


**Articles on Alternative Forms of Assessment (links available from ARTIST resource page)**

www.amstat.org/publications/jse/v5n3/chance.html

www.amstat.org/publications/jse/v2n2/fillebrown.html

www.amstat.org/publications/jse/v2n1/garfield.html

www.amstat.org/publications/jse/v5n1/hubbard.html

www.amstat.org/publications/jse/v3n1/konold.html

**Articles on Statistical Literacy, Reasoning, and Thinking**


www.amstat.org/publications/jse/v10n3/delmas_discussion.html


www.amstat.org/publications/jse/v10n3/rumsey2.html
Welcome to the ARTIST Web site!

Our goal is to help teachers assess statistical literacy, statistical reasoning, and statistical thinking in first courses of statistics.

This Web site provides a variety of assessment resources for teaching first courses in Statistics. Please use the navigation bar on your left to access these resources.

Learn more about Statistical Literacy, Statistical Reasoning, and Statistical Thinking:
- Definitions of Statistical Literacy, Reasoning, and Thinking
- Examples of Assessment Items coded as Statistical Literacy, Reasoning, and Thinking
- How Statistical Literacy, Reasoning, and Thinking are related
- How Statistical Literacy, Reasoning, and Thinking relate to Bloom’s and other taxonomies
- Words that characterize assessment items for Statistical Literacy, Reasoning, and Thinking

This project was funded by a grant from the National Science Foundation (NSF CCLI –ASA- 0206571) and supported by the following institutions:

[NSF Logo] [University of Minnesota Logo] [Cal Poly Logo]

For more information about this project, or to contribute materials, please contact Dr. Joan Garfield by e-mail: jbg@umn.edu.
ARTIST Assessment Builder

https://app.gen.umn.edu/artist/user/login.asp

Searching the ARTIST Item Database

<table>
<thead>
<tr>
<th>ARTIST Start Search Page</th>
<th>HELP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAUTION</strong></td>
<td>DO NOT use the BACK and FORWARD buttons to move from page to page</td>
</tr>
<tr>
<td>ARTIST Home Page</td>
<td>MANAGE ASSESSMENTS</td>
</tr>
</tbody>
</table>

Assessment: **Interpreting Graphs**  COURSE: **Introduction to Statistics**  Total Items in Assessment = 0

<table>
<thead>
<tr>
<th>ITEM FORMAT</th>
<th>Forced-Choice, Check List</th>
<th>Open-Ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERFORMANCE ITEMS:</td>
<td>Include</td>
<td>Exclude</td>
</tr>
<tr>
<td>ITEM SETS:</td>
<td>Only items that meet search criteria</td>
<td>ALL items in an Item Set</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOPICS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHECK ALL</td>
<td>CLEAR ALL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LEARNING OUTCOMES</th>
<th>CHECK ALL</th>
<th>CLEAR ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Literacy</td>
<td>□ Reasoning</td>
<td>□ Thinking</td>
</tr>
</tbody>
</table>

- Data Types
- Producing and Collecting Data
- Representing Data
- Measures of Center
- Measures of Spread
- Comparing Groups
- Measures of Position
- Normal Distribution
- Bivariate Data, Quantitative
- Bivariate Data, Linear Regression
- Bivariate Data, Categorical (includes Chi-Square)
- Probability
1. A class of students recorded the number of years their families had lived in their town. Here are two graphs that students drew to summarize the data. Which graph gives a more accurate representation of the data? Why?

**Graph 1**

*Diagram of a bar graph showing years in town with some years having multiple marks.*

**Graph 2**

*Diagram of a line graph showing years in town with some years having multiple marks.*
Assessment Builder: Downloading a Test

REGISTERED USER: Bob delMas

ARTIST Assessment Manager

NOTE: This session will time out if you do not use the Assessment Builder for 60 minutes.

<table>
<thead>
<tr>
<th>VIEW</th>
<th>DOWNLOAD</th>
<th>UPDATE</th>
<th>Assessment Name</th>
<th>Course Identifier</th>
<th>Course Name</th>
<th>Items</th>
<th>DELETE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST</td>
<td>GET</td>
<td>RENAME</td>
<td>Interpreting Graphs</td>
<td>GC 1454</td>
<td>Introduction to Statistics</td>
<td>7</td>
<td>DELETE</td>
</tr>
</tbody>
</table>

ARTIST Assessment Download

Assessment: Interpreting Graphs     COURSE: Introduction to Statistics

RTF File Created

Download Interpreting_Graphs.rtf
Data Representation Items

1. A baseball fan likes to keep track of statistics for the local high school baseball team. One of the statistics she recorded is the proportion of hits obtained by each player based on the number of times at bat as shown in the table below. Which of the following graphs gives the best display of the distribution of proportion of hits in that it allows the baseball fan to describe the shape, center and spread of the variable, proportion of hits?

<table>
<thead>
<tr>
<th>Player</th>
<th>Proportion of Hits</th>
<th>Player</th>
<th>Proportion of Hits</th>
<th>Player</th>
<th>Proportion of Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH</td>
<td>0.305</td>
<td>SU</td>
<td>0.270</td>
<td>BC</td>
<td>0.301</td>
</tr>
<tr>
<td>HA</td>
<td>0.259</td>
<td>DH</td>
<td>0.136</td>
<td>AA</td>
<td>0.143</td>
</tr>
<tr>
<td>JS</td>
<td>0.281</td>
<td>TO</td>
<td>0.218</td>
<td>HK</td>
<td>0.341</td>
</tr>
<tr>
<td>TC</td>
<td>0.087</td>
<td>RL</td>
<td>0.267</td>
<td>RS</td>
<td>0.261</td>
</tr>
<tr>
<td>MM</td>
<td>0.167</td>
<td>JB</td>
<td>0.270</td>
<td>CR</td>
<td>0.115</td>
</tr>
<tr>
<td>GV</td>
<td>0.333</td>
<td>WG</td>
<td>0.054</td>
<td>MD</td>
<td>0.125</td>
</tr>
<tr>
<td>RC</td>
<td>0.085</td>
<td>MH</td>
<td>0.108</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A

B

C

D

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>GUESS PERCENT</th>
<th>PERCENT (N = 1643)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. A local running club has its own track and keeps accurate records of each member's individual best lap time around the track, so members can make comparisons with their peers. Here are graphs of these data.

Which of the above graphs allows you to most easily see the shape of the distribution of running times?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>GUESS</th>
<th>PERCENT (N = 1345)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Graph A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Graph B.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Graph C.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. All of the above.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Topic: Data Representation

Names of Group Members

Why do you think students are selecting the incorrect responses for each item?
Outline an instructional activity to help students develop the correct understanding.
Sampling Variability Items

1. A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 1608)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Sam, because there are more candies, so his bag can have more brown candies.</td>
<td>5.3</td>
</tr>
<tr>
<td>b. Sam, because there is more variability in the proportion of browns among larger samples.</td>
<td>11.6</td>
</tr>
<tr>
<td><strong>C. Kerry, because there is more variability in the proportion of browns among smaller samples.</strong></td>
<td><strong>32.4</strong></td>
</tr>
<tr>
<td>d. Kerry, because most small bags will have more than 50% brown candies.</td>
<td>1.7</td>
</tr>
<tr>
<td>e. Both have the same chance because they are both random samples.</td>
<td>48.9</td>
</tr>
</tbody>
</table>

2. Consider the distribution of average number of hours that college students spend sleeping each weeknight. This distribution is very skewed to the right, with a mean of 5 and a standard deviation of 1. A researcher plans to take a simple random sample of 18 college students. If we were to imagine that we could take all possible random samples of size 18 from the population of college students, the sampling distribution of average number of hours spent sleeping will have a shape that is

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 872)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Exactly normal.</td>
<td>18.8</td>
</tr>
<tr>
<td><strong>B. Less skewed than the population.</strong></td>
<td><strong>34.4</strong></td>
</tr>
<tr>
<td>c. Just like the population (i.e., very skewed to the right).</td>
<td>34.7</td>
</tr>
<tr>
<td>d. It's impossible to predict the shape of the sampling distribution.</td>
<td>12.0</td>
</tr>
</tbody>
</table>
Confidence Intervals Items

1. Suppose two researchers want to estimate the proportion of American college students who favor abolishing the penny. They both want to have about the same margin of error to estimate this proportion. However, Researcher 1 wants to estimate with 99% confidence and Researcher 2 wants to estimate with 95% confidence. Which researcher would need more students for her study in order to obtain the desired margin of error?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 1296)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Researcher 1.</td>
<td>51.9</td>
</tr>
<tr>
<td>b. Researcher 2.</td>
<td>25.9</td>
</tr>
<tr>
<td>c. Both researchers would need the same number of subjects.</td>
<td>9.1</td>
</tr>
<tr>
<td>d. It is impossible to obtain the same margin of error with the two different confidence levels.</td>
<td>13.1</td>
</tr>
</tbody>
</table>

2. A high school statistics class wants to estimate the average number of chocolate chips per cookie in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a confidence interval for the average number of chips per cookie (18.6 to 21.3). Indicate if the following interpretations are valid or invalid.

<table>
<thead>
<tr>
<th>Statement (N = 1609)</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td>We are 95% certain that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.</td>
<td>51.2</td>
<td>48.8</td>
</tr>
<tr>
<td>We expect 95% of the cookies to have between 18.6 and 21.3 chocolate chips.</td>
<td>34.1</td>
<td>65.9</td>
</tr>
<tr>
<td>We would expect about 95% of all possible sample means from this population to be between 18.6 and 21.3 chocolate chips.</td>
<td>53.1</td>
<td>46.9</td>
</tr>
<tr>
<td>We are 95% certain that the confidence interval of 18.6 to 21.3 includes the true average number of chocolate chips per cookie.</td>
<td>75.7</td>
<td>24.3</td>
</tr>
</tbody>
</table>
Test of Significance Items

1. A newspaper article claims that the average age for people who receive food stamps is 40 years. You believe that the average age is less than that. You take a random sample of 100 people who receive food stamps, and find their average age to be 39.2 years. You find that this is significantly lower than the age of 40 stated in the article ($p < .05$). What would be an appropriate interpretation of this result?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 1101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The statistically significant result indicates that the majority of people who receive food stamps is younger than 40.</td>
<td>33.8</td>
</tr>
<tr>
<td>B. Although the result is statistically significant, the difference in age is not of practical importance.</td>
<td>50.5</td>
</tr>
<tr>
<td>c. An error must have been made. This difference is too small to be statistically significant.</td>
<td>15.7</td>
</tr>
</tbody>
</table>

2. A researcher compares men and women on 100 different variables using a two-sample t-test. He sets the level of significance to .05 and then carries out 100 independent t-tests (one for each variable) on data from the same sample. If, in each case, the null hypothesis actually is true for every test, about how many "statistically significant" findings will this researcher report?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 1160)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 0</td>
<td>30.2</td>
</tr>
<tr>
<td>B. 5</td>
<td>45.7</td>
</tr>
<tr>
<td>c. 10</td>
<td>7.1</td>
</tr>
<tr>
<td>d. none of the above</td>
<td>17.1</td>
</tr>
</tbody>
</table>
Bivariate Quantitative Data Items

1. The number of people living on American farms has declined steadily during the last century. Data gathered on the U.S. farm population (millions of people) from 1910 to 2000 were used to generate the following regression equation: Predicted Farm Population = 1167 - .59 (YEAR). What method would you use to predict the number of people living on farms in 2050.

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 1591)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Substitute the value of 2050 for YEAR in the regression equation, and compute the predicted farm population.</td>
<td>19.8</td>
</tr>
<tr>
<td>b. Plot the regression line on a scatterplot, locate 2050 on the horizontal axis, and read off the corresponding value of population on the vertical axis.</td>
<td>15.6</td>
</tr>
<tr>
<td>C. Neither method is appropriate for making a prediction for the year 2050 based on these data.</td>
<td>28.4</td>
</tr>
<tr>
<td>d. Both methods are appropriate for making a prediction for the year 2050 based on these data.</td>
<td>36.2</td>
</tr>
</tbody>
</table>

2. A statistics instructor wants to use the number of hours studied to predict exam scores in his class. He wants to use a linear regression model. Data from previous years shows that the average number of hours studying for a final exam in statistics is 8.5, with a standard deviation of 1.5, and the average exam score is 75, with a standard deviation of 15. The correlation is .76. Should the instructor use linear regression to predict exam scores from hours studied?

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>PERCENT (N = 850)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Yes, there is a high correlation, so it is alright to use linear regression.</td>
<td>21.2</td>
</tr>
<tr>
<td>b. Yes, because linear regression is the statistical method used to make predictions when you have bivariate quantitative data.</td>
<td>27.1</td>
</tr>
<tr>
<td>C. Linear regression could be appropriate if the scatterplot shows a clear linear relationship.</td>
<td>46.2</td>
</tr>
<tr>
<td>d. No, because there is no way to prove that more hours of study causes higher exam scores.</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Topic: Names of Group Members

Why do you think students are selecting the incorrect responses for each item?
Outline an instructional activity to help students develop the correct understanding.