Inventing Precision

Unit

3

Mathematical Concepts

- Different statistics emphasize different approaches toward measuring variability.
- Precision of measure is grounded in the consistency of measurement—the tendency of one measurement to be similar to another measurement. More variable measurements are less consistent, and less variable measurements are more consistent.
- Traditional measures of variability (e.g., range, interquartile range, average deviation, standard deviation and variance) emphasize different characteristics of the distribution of data.
- The interquartile range (IQR) and measures employing deviation, such as average deviation, coordinate center with spread to describe variability (see Mathematical Background)

Unit Overview

In Unit 3, students invent a measure of precision—the tendency of the measurements to agree. Focusing on precision, and not only spread, invites students to consider how to develop a quantity that measures "clumpiness" or proximity of the measurements. Focusing on clumps, especially the center clump, often spurs simultaneous consideration of center and spread as students focus on distances (deviations) from the center or consider the neighborhood of values around the center.

Day 1: Measuring Precision
Students begin by designing a measure for describing the precision of measurements (e.g., arm span measurements) that others can use and obtain the same result. They use the same displays developed in Unit 1 to guide their sense of precision. The algorithm designed by the students should produce a quantity—a numeric value—that indicates precision. It is often helpful to have two collections of measurements of the same attribute available, one obtained with a crude tool, the other with a more precise tool, so that students can see whether or not their invented measures capture differences in the variability of different distributions. (See Unit 4, Day 1 if you choose this option at this time.)

Day 2: Comparing Methods: Measure Review
Next, students who did not author the method try to use it. Different methods are compared and contrasted with an eye towards discerning what aspects of the data the author noticed. During this activity, you may find it

Contents

useful to introduce traditional measures or traditional names for student-invented measures (see Mathematical Background).

Note: Make sure students have had a chance to invent a measure of center before talking about this more difficult problem of measuring variability.

Day 3: Exploring Traditional Measures of Precision
Students explore inter-quartile range and average deviation. After inventing statistics, students are more likely to understand why we have more than one way of measuring precision.

Days 4 and 5: Formative Assessment
Students respond to a brief quiz. Student responses that represent different ways of thinking according to the Conceptions of Statistics construct map are deliberately compared and contrasted.

Read

- ☐ **Unit 3**
  Start by reading the unit to learn the content and become familiar with the activities. You can also scan Unit 4 for other kinds of samples of measurement or production data that students could generate or use for this lesson.
- ☐ **Mathematical Background**
  Reread the Mathematical Background carefully to help you think about the important characteristics of measures of variability (precision).
- ☐ **Sample student thinking**
  Reread Student Thinking, pp. 8-13, to anticipate the kinds of ideas that typically guide students' inventions of measures of variability.
- ☐ **Conceptions of Statistics construct map**
  Read the construct map and/or visit the website (modelingdata.org) to view a progression of student thinking about statistics, beginning with qualitative perceptions of data and knowing how to calculate statistics, progressing toward understanding statistics as measures of distribution. Pay particular attention to the video and text examples of student invented measures of precision.
- ☐ **Using TinkerPlots**
  Read the TinkerPlots supplement on the web site. Pay close attention to the sections that outline the use of the ruler tool and partition tools (i.e., dividers, hat plots and reference lines)—these are the tools your students will find helpful when inventing a measure for precision.
- ☐ **Read Discussion Guide and Measures of Precision Planner** on Modelingdata.org to guide a whole-class conversation where students compare their inventions and rationales for measuring the precision of measure.

Gather

For the class

- ☐ Student displays (from Unit 1)
- ☐ Plain white paper (or computers) for writing directions

Prepare

- ☐ TinkerPlots
  You will need to enter the arm span measurement data into a TinkerPlots file in order to use TinkerPlots with your students.

We explore briefly some conventional statistics of variability with the aim of describing their motivations—what each measure attends to in a distribution of data

What is range?

The range is the difference between the lowest and highest values in the collection. Because it is defined by extremes, the range will often vary substantially from sample to sample. Students often have intuitions about this. For example, they notice that "bad" measurements may not be as likely if we mesured again, even without memory of the previous measurements. Hence, the range may be misleading about the variability of a collection of measurements.

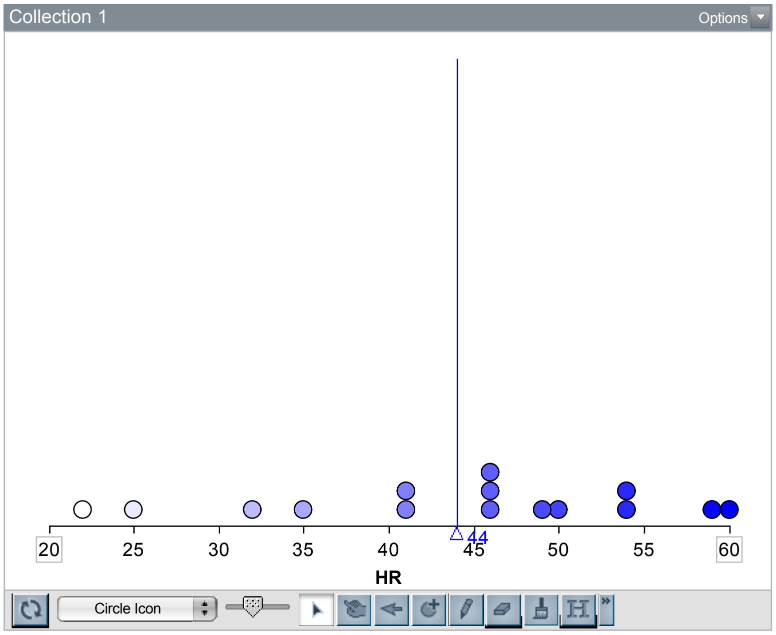
What is inter-quartile range?

The inter-quartile range designates the width of the central region of the distribution, the center clump of values about the sample median. It is the range (highest-lowest value) of the cases comprising the middle 50 percent, the second and third quartiles, of the distribution. Students often arrive at this statistic or something like it by reasoning about the center clump of a distribution.

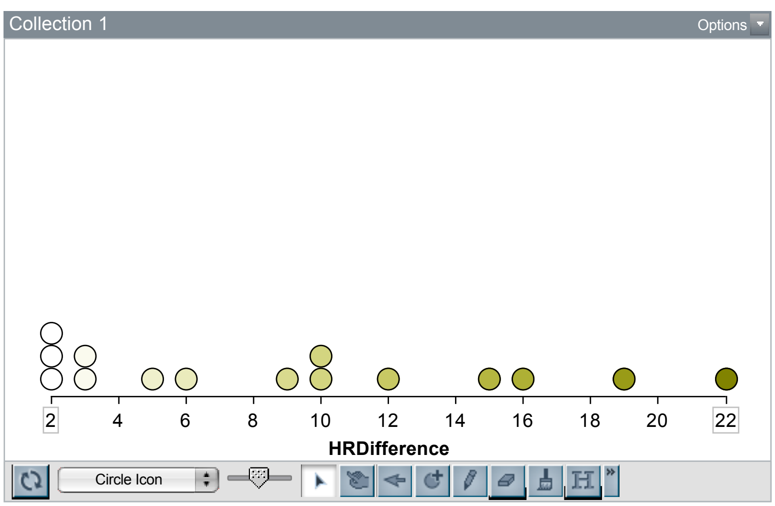
What is average deviation?

Average deviation is the mean (average) of the absolute values of differences (the deviations) between each measured value and the mean. To calculate the average deviation (mean absolute difference), find the sum of the absolute values of the deviations and then divide the sum by the number of deviations (the sample size). Average deviation attends to the magnitude of differences and ignores the direction of difference. Negative deviations correspond to under-estimates and positive deviations correspond to over-estimates in the context measuring a true value. In contexts of production, negative deviations correspond to under-shooting the target value and positive deviations to over-shooting the target value. But, for average deviation, these differences in direction are ignored and only the magnitudes of the deviations are considered.

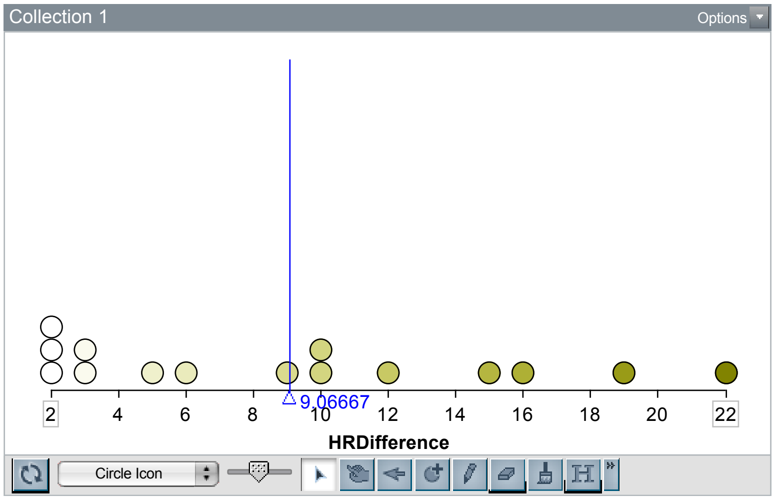
For example, the number of home runs hit by a baseball player during each of 15 seasons, with a mean of 44 home runs:



The absolute values of the differences (deviations) between the number of home runs hit each season and this mean:



The mean (average) deviation, which can be interpreted as a measure of consistency in home run production:



Students sometimes discover that the sum of signed deviations from the mean is zero. This is an excellent occasion to introduce the absolute value function, which focuses only on the magnitude of the deviations. Some students will invent a sensible counterpart to average deviation, **median deviation**: the median of the absolute values of the differences (deviations) between each measured value and the sample median.

What is variance?

Variance also builds upon differences (deviations) between each measured value and the mean, but each difference is first squared, and then the mean of the squared differences is computed. This mean value is called the variance. For example, if the values of a set of data are 2, 4, 6, the mean is 4. The sum of squared deviations is found as $(2\text{-}4)^2 + (4\text{-}4)^2 + (6\text{-}4)^2$ for a total of 8. The variance is then 8/3 or 2.66. It is very unlikely that students will invent this measure. It is not intuitive.

What is standard deviation?

The length of the side of the average (standard) square representing the variance is the standard deviation. It is obtained by finding the square root of the variance. In the previous example, the standard deviation is 1.63, the square root of 2.66.

Measuring Precision

To introduce the measure of variability, students are challenged to invent a measure of precision. Precision refers to the tendency of the measurements to agree. More tightly clustered measurements are more precise; less tightly clustered measurements are less precise. Students invent and write algorithms that other students can follow. The result of the algorithm should be a precision number—a number that tells how precise the measurements are. After students invent statistics, they are more likely to understand why statisticians have developed more than one way of measuring variability. In any group of students, some are likely to focus on the center clump (conventionally, the IQR), others the range, and still others, the distances among the measurements (conventionally, the average deviation). Students should not be introduced to conventional measures of variability until they have the opportunity to invent their own measures of it, because then students are in a better position to understand the rationales for conventions. We recommend that all students learn three conventional statistics: range, inter-quartile range (IQR), and either average deviation or median deviation.

Note: Precision of measure is not the same as accuracy. A measure can be biased yet precise, as in measures that consistently over- or under-estimate a true measure. For example, a grocer's scale that does not account for the weight of a container consistently overestimates the weight of the container's contents, and hence, customers pay more than they should.

Whole Group

1. **Introduce the activity: Inventing a method for measuring precision.**

   a. Make available the displays students created in Unit 1. You might consider posting them on the walls or making copies of a few and passing them out to individual students.

   b. Remind students of the finding that the measurements were not identical. Ask questions to emphasize this point, such as:
   Q: If our measures were all exact, all precisely the same, what would our display look like? (A stack as tall as the number of observations.)
   Q: If our measures had no relationship to one another at all, what might they look like? (A random scatter.)

Construct: CoS1(a)
This discussion often reveals what students notice about the visual quality of the data.

Measuring Precision
Comparing Methods
Traditional Measures of Precision
Formative Assessment

c. Provoke a discussion about why a precision measure might be important using questions like:

Q: Why is it important to know if our measurements tend to agree with one another? What does the tendency for the measurements to be alike tell us about the measurement method?
Q: What might we consider about the data to think about precision?

d. Tell students that their job is to invent a method that will measure how precise (how consistent, the tendency for measurements to agree) the measurements were. The result should be a precision number—a number that tells us how precise the measurements are. For example, the precision number, if all the measurements were exactly the same, would be different than if the measurements were all over the place.

Construct: CoS2(b), CoS3(a), and CoS3(b)
This activity engages students in inventing a measure of variability. In particular, this activity can focus on moving students from CoS3(a) to CoS3(b).

Note: There is often great value in asking students to think about and propose definitions of precision as a whole class discussion. The diversity of student inventions helps all students develop more refined senses of the meaning of variability. However, some teachers have found more extensive support of the conversation is required. They suggest that at some point in the conversation, a teacher could mention:

- Another way of thinking about precision is to consider how consistent—how much alike—our measurements were.
- Another way of thinking about precision is to consider how much our measurements tend to agree.

Individual or Small Group

2. Let students invent a method that will measure how precise (how consistent) their measurements were.

a. Ask students to work together to invent a method by writing it out on a piece of paper.

b. Explain to students that when they have decided on a method, they should write it down so that someone else could follow it and get a precision number.

Measuring Precision
Comparing Methods
Traditional Measures of Precision
Formative Assessment

Note: Encourage students using TinkerPlots to make use of the ruler tool and the partition tools (dividers, hat plots, reference lines) to develop a measure of precision.

3. Circulate among groups and ask thought-revealing questions like:

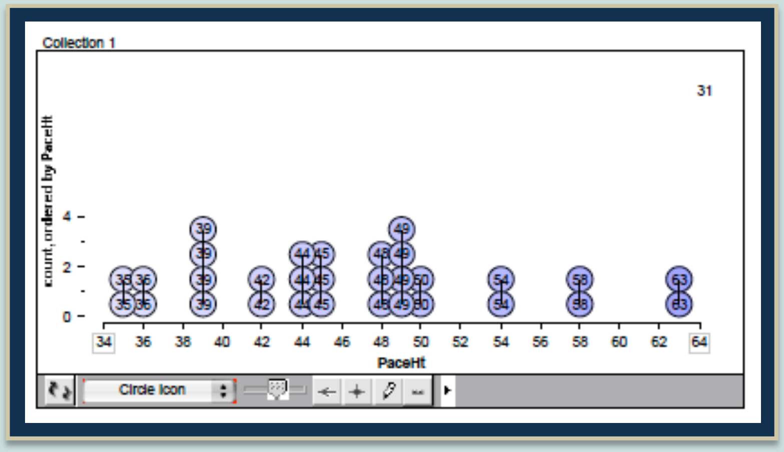
Q: Based on your method, how much do the measurements agree?
Q: What would be the number for precision if all the measurements were in agreement?
Q: What would happen to your precision number, if we were not very careful in our measurements?

Note: As in the measuring center activity in Unit 2, often students will not understand at first what they should be doing, or they have difficulty coming up with ideas. It is often helpful to let students work for about 10 minutes and then convene the whole class to quickly identify promising approaches and challenges. Remind groups that someone else should be able to follow their directions to come up with the same precision number. After a brief discussion, give the students plenty of time to invent their methods.

Students' Ways of Thinking about Variability

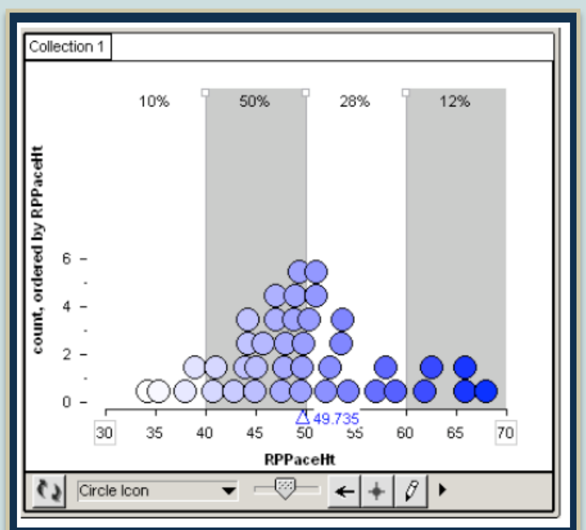
When students invent measures of precision, they tend to generate novel ways of capturing variability. We have noticed three general patterns: (a) approaches based on repeated values; (b) approaches based on thinking about the center clump of the data; and (c) approaches based on distance of each case value from a common reference point, usually the sample mean or sample median. Reasoning guided by (b) and (c) have counterparts in conventional statistics (IQR and Average Deviation, respectively), but the reasoning based on repeated measures provides an important avenue for considering the nature of variability and hence should not be discouraged. In the section that follows, we provide illustrations of each form of reasoning.

Repeated Values. Some students reason that repeated values suggest higher levels of agreement among measurers. If everyone agreed, everyone would obtain the same measurement. Such measurements would be highly precise and have no variability. On the other hand, if everyone obtained a different measurement, then variability would be high and precision of measurement would be low. For example, Jamir used TinkerPlots to create a frequency plot of 49 measurements made by the class of the height of the school's flagpole. He used the "connect cases with equal values" feature to highlight all the repeated measurements, and then he used the "hide cases" feature of TinkerPlots to suppress the values that were not repeated. Working from the display, Jamir counted the number of distinct repeated values, 31. At first, "31" was his measure of precision, but later he amended this to 63% by considering 31 of 49 cases.



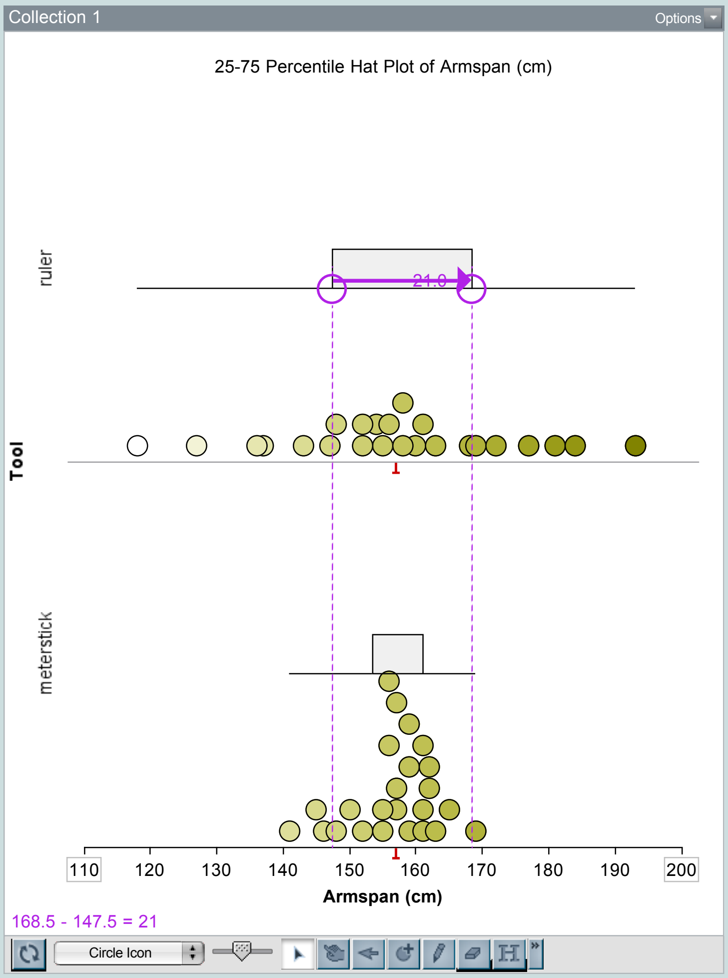
Using connect cases to invent a measure of precision.

Center Clump-Based Solutions. A group of students claimed, "where the precision was where most people had their numbers." Then they found out that 50% of all measurements were in the 40s, and 28% of all measurements were in the 50s. So they decided to use the percentage of measures in the decade-interval containing the mean as their measure of precision. Their display is illustrated next.

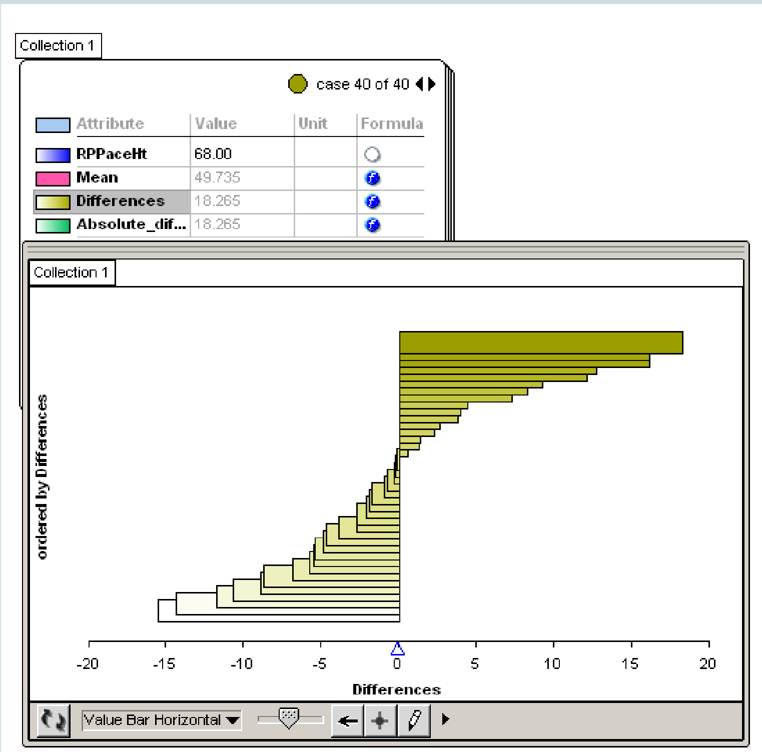


Center clump solution to measuring precision (variability)

Other students who focus on the center clump invent the IQR. For example, to compare the precision obtained with a ruler and a meter stick, a student first found the middle 50 percent of each distribution with TinkerPlots and then used the TinkerPlots Ruler tool to find each IQR, as shown next.

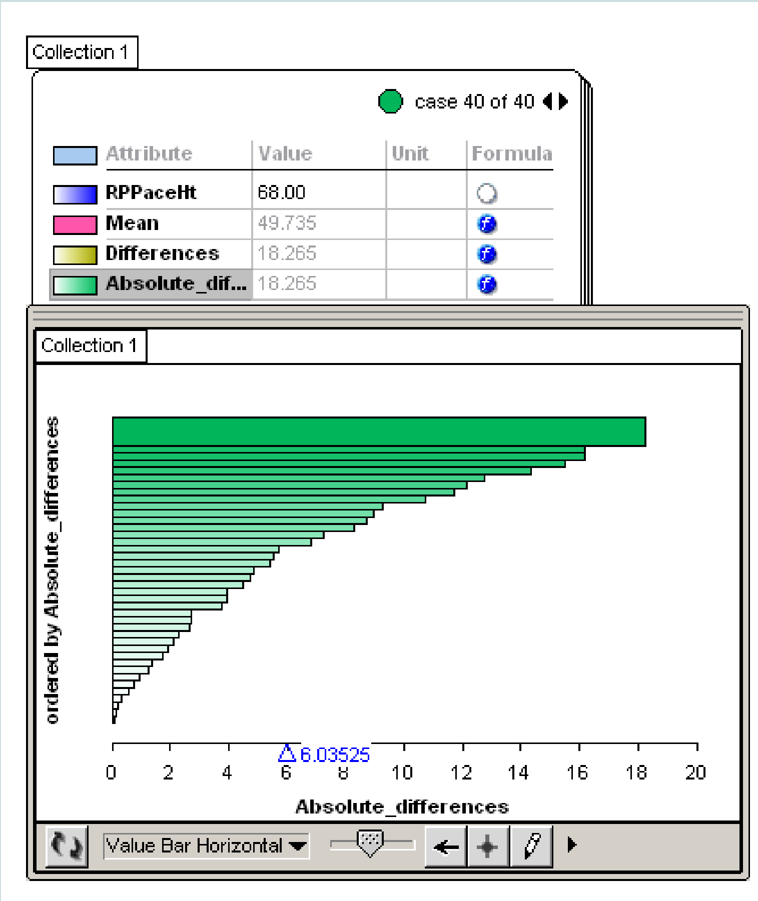


Distance-Based Solutions—Distance from the mean. A sixth-grader, Robert, started out trying to find out how far each value at the tails of the distribution was from the middle of the distribution. He used the mean to represent the middle. His teacher capitalized on this idea: "Good idea, but how would you describe the precision of the group as a whole?" After thinking hard about this, Robert suggested that he would average the differences between the mean and each measurement. Distances corresponding to over-estimates were positive and those corresponding to under-estimates were negative. Robert proposed to find their sum and then to divide by the number in the sample. Robert thought that this method would be "like the average," except that it would indicate how close the measures were, "on average." When he attempted to find the mean of the differences, he was surprised that the sum was zero. (This is a property of the sum of differences between each observation and the mean. It is a consequence of the definition of the mean.) Robert was puzzled but he reiterated that he thought his method was good for finding the distances between each score and the mean. He plotted each difference with TinkerPlots, and wondered what might have gone wrong.



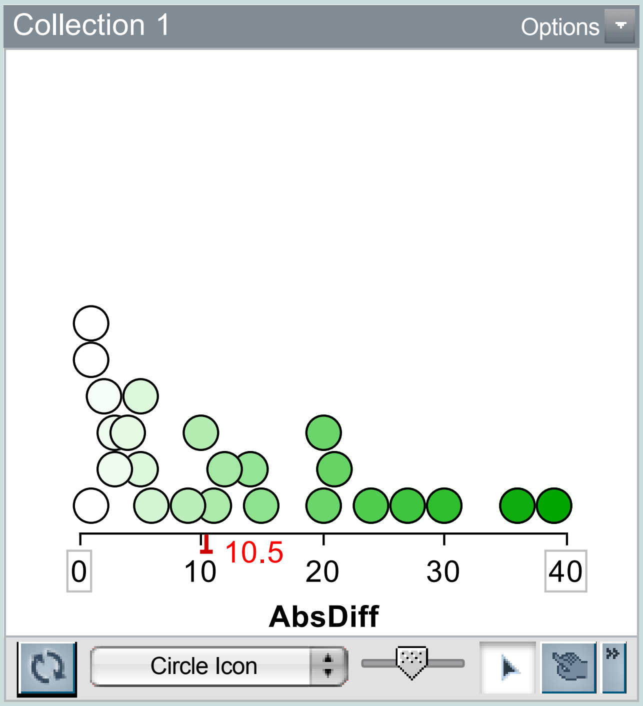
TinkerPlots plot of signed differences.

In light of class discussions about some estimates being over and some under the real height of the school's flagpole, the teacher asked if Robert were more concerned about the direction, or the magnitude, of each difference. Robert mentioned that the direction of the difference was not that important-some measures must be greater than the mean and others less. Hence, what mattered was how far each measure was from the mean. The teacher built on Robert's insight to introduce the absolute value function. [f (x) = x, if x > or equal to 0; f (-x) = x]. Robert used the absolute value function in the TinkerPlots formula menu to generate the average deviation. He then plotted the absolute values of the differences, and located their average value-the average deviation. Robert's solution is shown in the next display.



TinkerPlots plot of absolute values of differences

Distance-based solutions—Distance from the median. A pair of fifth-grade students found the differences between each observed value and the median. One of the two students said that the spread would be zero if all measurements perfectly agreed. However, the student did not know how they could get "zero" mathematically by using actual measurements. The teacher asked students that if the true measurement was 157 cm (median) and each measurement would be the same, what would be the differences between the median and each measurement? This teacher question helped the two students think about precision in terms of distances. Also, the teacher asked the students whether positive and negative differences would be important in the situation, and the students responded that how far each measurement was away from the median was more important than the direction. The students' initial plan was to sum all absolute differences. To extend this way of thinking, the teacher asked the students what would happen if 100 students measured the arm span with a more precise tool. The students answered that the spread number would go up, because the sum would increase. But then the sum might exceed that of fewer students with a crude tool. To account for differences in sample size, the students found the median of the differences (below).



TinkerPlots plot of median of absolute values of differences

Comparing Methods: Measure Review

The purpose of this activity is to support students to better understand both student-invented and conventional methods as ways to measure the precision of a distribution. At the beginning of the activity, each group shares their precision algorithm with another group. After trying out someone else's method, groups report back to the class. The students should answer three questions: (a) Is the method clear (i.e., does the method generate a reliable outcome no matter who follows the method)? (b) Which parts of the distribution are focused on by each method? and (c) Does the method result in a good estimate of the precision?

Whole Group

1. **Introduce the activity: comparing methods.**

   Ask students to pass their method for calculating a measure of precision along to another student group.

Small Groups

2. **Let students try out the algorithms other students have invented.**

3. **Listen and watch for algorithms to highlight in the whole-group discussion.**

   Circulate around the room and push students to think carefully about the comparisons by asking questions like:

   Q: What is the main idea behind the other group's method?
   Q: How is the other group's method a measure of precision? Is it good for finding the precision of our measurements? Why?
   Q: Using this method, how consistent were our measurements? How can this method describe the patterns and trends we found in the displays?
   Q: In what way are these methods the same? In what way are they different?

   Note: Watch for student invented methods that are similar (or identical) to traditional measures of variability (i.e., range, inter-quartile range, and average deviation). Each pays attention to a

Construct: CoS2(b), CoS3(a), CoS3(b), and CoS3(c)
This activity engages students in considering how measures of variability work.

different characteristic of the distribution to obtain an estimate of precision (variability).

Whole Group

4. Students report back to the class about the results of the comparison.

a. Ask one pair of students to explain their method to the class.

Note: You should select pairs deliberately. Make sure students have opportunities to compare different mathematical approaches to measuring variability, such as a method that uses the range and one that focuses on the center clump. Often, as few as two carefully selected measures can lead to productive discussions.

b. Use questions like the following to get students thinking and talking about the mathematical elements of the method after it is shared:

Q: Is this a good method for determining the precision of our measurements?
Q: What parts of the data does this measure use?
Q: Who can think of a situation where this method might not give us a very good value of precision? What would happen if (e.g., create extreme cases, change the shape of the data)?

c. Select additional groups to report back. Use questions like the ones below to get students talking about the differences in method. Work to make sure that students notice and talk about the mathematical differences in the methods and how these differences affect the measure in different scenarios.

Q: Which method is easiest to understand? Which is the hardest? Why?
Q: Which method would almost always be a good measure of precision, no matter what and how we measured? Why?
Q: What does it mean that we “invent a measure”? Can we do that?
Q: How does precision relate to how spread out the measurements are?
Q: What makes a measurement good or useful? What do you look for?

Note: Teachers should be alert to student thinking that seems to be guided by the "center clump"--perhaps asking students why they think the largest clump is at the center (most data sets collected in this manner will have a bell-shaped distribution) and what that implies about precision. Teachers should also be alert to student thinking that employs distance as a way of determining agreement --how close the measurements tend to be. It is also helpful for teachers to occasionally provoke consideration of the effect of sample size on the measure proposed. For example, some students invent sums of differences as measures of precision. But, what happens if the sample size grows or shrinks? (Conventional statistics resolve this problem by finding average or median values of differences.)

The Measure Review can be fruitfully extended to consideration of another sample—see Unit 4—so that students can determine whether or not the invented methods can in fact be employed with new samples.

Exploring Traditional Measures of Precision

1. Explore inter-quartile range (IQR) and average deviation.

a. Distribute the worksheet, *Home Run Hitter*. Home Run Hitter presents the number of home runs hit by a baseball player during 15 seasons, arranged from least to greatest. The IQR and average deviation both measure the hitter's consistency from season-to-season. After finding the value of each measure, students are asked to choose which is the better measure of consistency for this purpose.

Note: The task provides students with the total number of home runs hit, so they should recognize that this total, coupled with the number of seasons, is sufficient to find the value of the mean. The hitter (Babe Ruth, with the exception of a small adjustment to one season's total) has a single season with a low number of home runs. This affects the average deviation more than the IQR.

b. If TinkerPlots is available, distribute the worksheet, *Speed Zone.* The file, Speedlimit.tp contains the speeds of 30 cars tracked by radar in a speed zone. Students use statistics to estimate the speed limit (the target value of the production process) and to estimate the tendency of drivers to adhere to the speed limit (the variability of this production process).

Formative Assessment

1. **Administer the quiz.**

   Leah and Mark's Method asks students to use and evaluate two different measures of variability so one can differentiate between levels 2 and 3 (and sub levels of 3) on CoS.

   Consistency of Water Treatments asks students to invent a measure of variability that indicates the consistencies of two different methods of water treatment. This can differentiate between levels 1 and 3 on CoS and among a few sublevels of 3 as well.

2. **Use the scoring guides to score student responses.**

3. **Use Leah and Mark's Method responses to generate a discussion of tradeoffs among different methods for indicating variability.**

   a. **Select student responses to compare and contrast.**

      While scoring, select four different responses to use in the conversation. These responses should include one that correctly calculates the range and draws the appropriate conclusion that according to this statistic, class B is more precise, one that correctly calculates the IQR and draws the appropriate conclusion that class A is more precise. If possible, include other responses that do not demonstrate correct calculation, perhaps due to a small mistake or perhaps due to not understanding how to find the range or the IQR.

   b. **Prepare questions to support and guide student thinking.**

      Q: Let's compare Leah and Mark's methods. How were their methods similar?
      Q: How were their methods different?
      Q: Which method was most helpful for figuring out which class was more precise?
      Q: Which method was easiest to understand? Which is the hardest? Why?
      Q: Which method would almost always be a good measure of precision, no matter what and how we measured? Why?

c. **Use an Assessment Conversation to help students consider the tradeoffs of methods.**

Compare and contrast as students present their responses. For parts A and B, highlight how each statistic makes use of different characteristics of the distribution to measure variability. For example, the IQR makes use of the tendency of the measurements to cluster around a central value, while the range relies on only two measurements. For part C, highlight generalization. If the classes measured again, which measurements or types of measurements are more likely to be repeated? Which might change? What would that do the range or to the IQR?

4. **Use Consistency of Water Treatments responses to generate a discussion of invented measures.**

a. **Select student responses to compare and contrast.**

While scoring, select four different responses to use in the conversation. These responses should include one from each level on the scoring guide. For example, some students may use the mean to represent consistency, confusing center with variability/consistency. Others may look at the data and conclude that Method 2 is less spread out--more consistent--but be unable to generate a statistic—a quantity—to represent the spread. Yet others may decide to use the range, and if so, they should conclude that there is no difference in consistency. Be sure to include student responses that approach the problem by finding the difference between each bacteria count and the mean count.

b. **Prepare questions to support and guide student thinking.**

Q: What makes X's method a good method for determining precision?
Q: What parts of the data does X's measure use?
Q: Can anyone think of a situation where X's method might not give us a very good value of precision? What would happen if (e.g., create extreme cases, change the shape of the data)?
Q: Which method was easiest to understand? Which is the hardest? Why?
Q: Which method would almost always be a good measure of precision, no matter what and how we measured? Why?

c. **Use an Assessment Conversation to help students consider the tradeoffs of students' invented measures.**

Invite students to present their responses. Guide the conversation with questions that direct the students to important elements of measures of precision. For example, for part C, when comparing a CoS(3f) response with CoS(3f-) and NL(ii) responses, it is important that students see how to use qualities of a distribution to choose a statistic. You may also wish to help students see that if 50 samples using Method 2 were compared to 10 samples of Method 1, if differences between each sample count and the mean were only added, then Method 2 would seem to be much less consistent than Method 1 (the sum of its differences would be very large compared to those of Method 1). One way to adjust for differences in sample sizes is to "fair share" the sum of the differences among the samples. This is an average deviation. The average deviation for Method 2 is lower than that of Method 1 and will remain so no matter how many samples are tested in each Method (assuming the same shape of the data).

Home Run Hitter

Investigating IQR and Average Deviation

When nominating players for the Baseball Hall of Fame, some sportswriters argue that the average number of home runs hit by a player and how consistent they are from year to year can help determine who is worthy of admission.

During 15 years, a player hit 660 home runs. The number of home runs per year, arranged from least to greatest, were:

22, 25, 32, 35, 41, 41, 46, 46, 46, 49, 50, 54, 54, 59, 60

1. What was the median number of home runs hit?

2. What was the IQR? What does the IQR mean for this situation? (What does it measure?)

3. What was the mean number of home runs hit?

4. What was the average deviation? What does it mean for this situation?

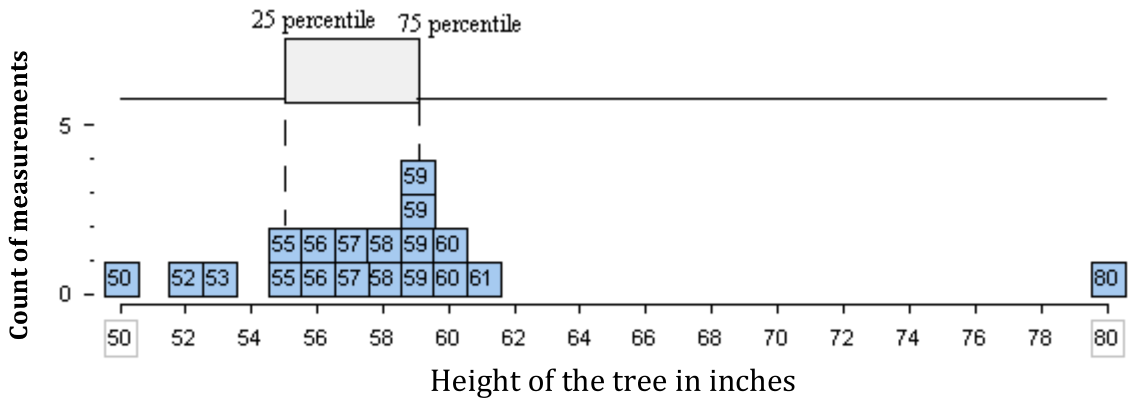
5. Which do you think is a better measure of consistency, the average deviation or the IQR? Why do you think so?

Unit Quiz

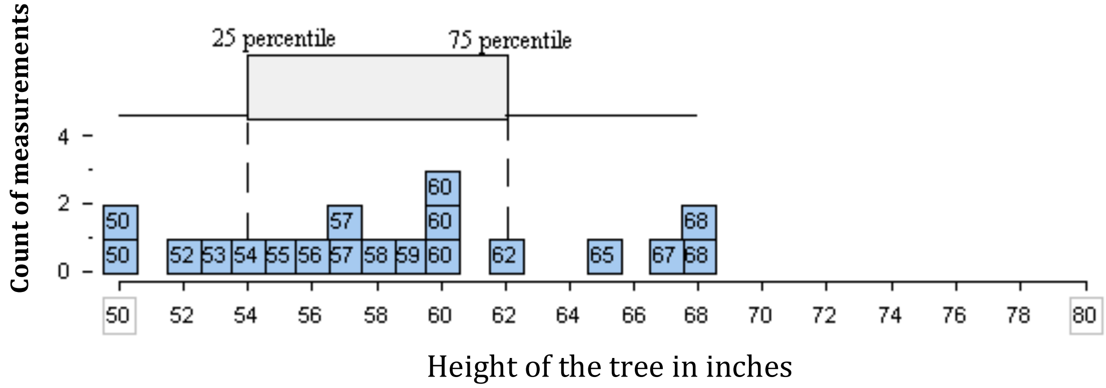
Leah and Mark's Method

Two classes of students measured the height of a young tree. Here are their results:

Class A Measurements



Class B Measurements



Leah and Mark wanted to know which class was more precise (less spread out in their measurements). They each came up with a different way to show how precise each class's measurements were.

Unit Quiz

Leah's Method:
She subtracted the lowest measurement from the highest measurement to get a measure of precision.

Mark's Method:
He subtracted the value at the 25^{th} percentile from the value at the 75^{th} percentile.

A. If you use **Leah's method**:

1. What is the measure of precision for **class A?** _________
   Show how you got the result.

2. What is the measure of precision for **class B?** _________
   Show how you got the result.

3. Based on Leah's method, which class is more precise, A or B?

Unit Quiz

B. If you use **Mark's method**:

4. What is the measure of precision for **class A?** _________
   Show how you got the result.

5. What is the measure of precision for **class B?** _________
   Show how you got the result.

6. Based on Mark's method, which class is more precise, **A or B?**

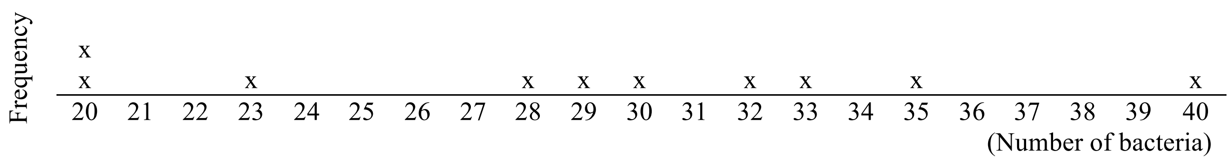
C. **Compare** Leah and Mark's methods. Which is a better method?
   Why do you think so?

Unit Quiz

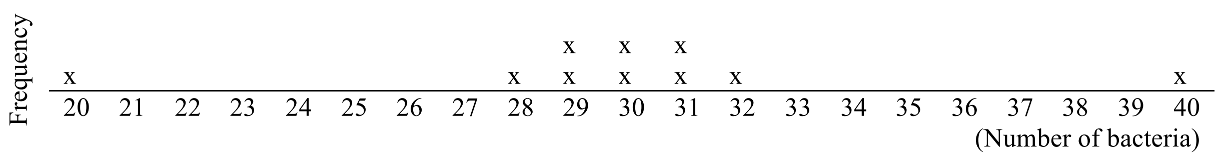
Consistency of Water Treatment Methods

A manager of a water treatment plant has two different methods for making water safe for animals to drink. The manager wants to use the method that is most consistent—the one that produces results that are most nearly alike. The chart shows the number of bacteria left after treatment in 100 ml samples of water, using each treatment method.

Method 1 (Mean = 29)



Method 2 (Mean = 30)



Show a way to calculate a measure of consistency that helps the manager make a decision on which treatment to use. Your way should work for other samples too.

1. Describe your way of calculating consistency (you do not need to do the actual calculations).

Unit Quiz

2. Explain why it is a good method for this situation.

3. Using the measure you chose, which method of water treatment is more consistent?

Unit Quiz Scoring Guide

Leah and Mark's Method

Part A: (1) Measure of Precision for Class A, Leah's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(2b)	**Calculates statistics indicating variability.** Student correctly applies Leah's method, shows work, and arrives at correct result.	• Class A: 30. 80-50=30
CoS(2b-)	**Calculates statistics indicating variability.** Student correctly applies Leah's method, however, the result is incorrect because of errors such as computational mistakes. OR student gives the correct result without showing work.	• Class A: 30. • Class A: 80-50=20*
NL(ii)	**Student incorrectly applies Leah's method and gets incorrect results.**	• Class A: 58. I subtracted the smallest measurement which was 50-61 which gave me 11 so I counted 11 and it gave me 58 which was the median and mode. • Class A: -30. 50-80=-30.
NL(i)	**Attempts item but answers are irrelevant, unclear, implausible, unreasonable, or demonstrate that student did not understand the item.**	• I don't know.*
M	**Missing response.**	

Part A: (2) Measure of Precision for Class B, Leah's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(2b)	**Calculates statistics indicating variability.** Student correctly applies Leah's method, shows work, and arrives at correct result.	• Class B: 18. 68-50=18
CoS(2b-)	**Calculates statistics indicating variability.** Student correctly applies Leah's method, however, the result is incorrect because of errors such as computational mistakes. OR student gives the correct result without showing work.	• Class B: 16. • Class B: 68-50=8 *
NL(ii)	**Student incorrectly applies Leah's method and gets incorrect results.**	• Class B: 50. I subtracted 75-25=50 which was the lowest number on the chart.
NL(i)	**Attempts item but answers are irrelevant, unclear, implausible, unreasonable, or demonstrate that student did not understand the item.**	• I don't know.*
M	**Missing response.**	

Unit Quiz Scoring Guide

Inventing Precision Unit 3

Part A: (3) Which class is more precise, Leah's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(3c)	Generalizes the use of a statistic beyond its original context of application or invention. Student who calculate the precision scores correctly for both a(1) and a(2) arrives at conclusion consistent to previous calculations.	• Class A: 30, Class B: 18, Class B is more precise according to Leah's method.
CoS(3c-)	Generalizes the use of a statistic beyond its original context of application or invention. Student, despite not calculating the precision scores correctly for both a(1) and a(2), arrives at conclusion consistent to previous calculations (i.e., the one with smaller absolute value is more precise).	• Class A: 28, Class B: 18, Class B is more precise according to Leah's method. • Class A: -30, 50-80=-30, Class B: -18, 50-68=-18, Class B is more precise according to Leah's method.
NL(ii)	Student arrives at a conclusion that is inconsistent with previous calculation. Indicates student probably does not understand the concept of precision.	• Class A: 30, Class B: 18, Class A because they get closest to the measurement. • Class A: 30, Class B: 18, Class A because your value is 30, and it is bigger.
NL(i)	Attempts item but answers are irrelevant or unclear.	• I would draw a graph of 18 going to 30.
M	Missing response.	

Part B: (1) Measure for Class A, Mark's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(2b)	Calculates statistics indicating variability. Student correctly applies Mark's method, shows work, and arrives at correct result.	• Class A: 4. 59-55=4.
CoS(2b-)	Calculates statistics indicating variability. Student correctly applies Mark's method, however, the result is incorrect because of errors such as computational mistakes. OR student gives the correct result without showing work.	• Class A: 4 • Class A: 5. 59-55=5*
NL(ii)	Student incorrectly applies Mark's method and gets incorrect result.	• Class A: -4
NL(i)	Attempts item but answers are irrelevant or unclear.	• I don't know.*
M	Missing response.	

Part B: (2) Measure for Class B, Mark's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(2b)	Calculates statistics indicating variability. Student correctly applies Mark's method, shows work, and arrives at correct result.	• Class B: 8. 62-54=8
CoS(2b-)	Calculates statistics indicating variability. Student correctly applies Mark's method, however, the result is incorrect because of errors such as computational mistakes. OR student gives the correct result without showing work.	• Class B: 8 • Class B: 4. 62-54=12*
NL(ii)	Student incorrectly applies Mark's method and gets incorrect result.	• Class B: 12. 59-62=12
NL(i)	Attempts item but answers are irrelevant or unclear.	• I don't know.*
M	Missing response.	

Part B: (3) Which class is more precise, Mark's Method **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(3c)	Generalizes the use of a statistic beyond its original context of application or invention. Students who calculate the precision scores correctly for both b(1) and b(2) arrives at conclusion consistent to previous calculations.	• Class A: 4, Class B: 8. Class A is more precise according to Mark's method.
CoS(3c-)	Generalizes the use of a statistic beyond its original context of application or invention. Student, despite not calculating the precision scores correctly for both b(1) and b(2), arrives at conclusion consistent to previous calculations (i.e., the one with smaller absolute value is more precise).	• Class A: 1. 55-54=1. Class B: 3. 62-59=3. Class A is more precise according to Mark's method. • Class A: 1. 59-55=4. Class B: 3. 62-54=2. Class B is more precise according to Mark's method.
NL(ii)	Student arrives at a conclusion that is inconsistent with previous calculation. Indicates student probably does not understand the concept of precision.	• Class A: 18. Class B: 18. Class B is more precise if you take it look and look class A you know there is a big gap and that's why the hat line aren't equal.
NL(i)	Attempts item but answers are irrelevant or unclear.	• You already know what you are subtracting when you get your measurement.
M	Missing response.	

Part C, Comparing Leah and Mark's Methods **Leah and Mark's Method and Conceptions of Statistics (CoS)**		
Level	**Performance**	**Example**
CoS(3f)	Choose statistics by considering qualities of a distribution.	• "Leah's method makes you think class B is more precise. Mark's method makes you think class A is more precise. I think Mark's method is better, because look at class A. You know there is a big gap but otherwise they are clumped together. Class B is more evenly spread out."*
CoS(3f-)	Considers qualities of a distribution, but incorrectly applies those qualities to choose a statistic.	• "I think Leah's method is better for Class A because there is an outlier."* • "Leah's is better because Mark's leaves out the big differences, and we are trying to measure how much we agree."
CoS (2b)	Using ease of calculation to justify choice of method, without regard to what the statistic is measuring.	• "Leah's method is easier. You just subtract."
NL(ii)	Student chooses a statistic without taking into consideration of the sample qualities. OR student considers both lead to the same conclusion either because they previously solve the prior two problems incorrectly, or because misunderstanding of the question.	• "Leah's method makes you think class B is more precise. Mark's method makes you think class A is more precise. I think I will choose the lowest number 4 to be most precise. And that's Mark's method. "* • "They do lead you to the same conclusion because class a is more precise than class b in both cases." [However, she previously had answered class b was more precise according to Leah's method.]
NL(i)	Attempts item but answers are irrelevant or unclear.	• I don't know.
M	Missing response.	

**Mock student responses*

Consistency of Water Treatments

Consistency of Water Treatments and Conceptions of Statistics (CoS)		
Level	**Performance**	**Example**
CoS(3b)	**Invent a sharable (replicable) measurement process to quantify a quality of the sample.**	• "I would subtract the number from the mean and add up the differences. Method 2 would be more consistent because it is less spread out than Method 1." *
CoS(2b)	**Calculate statistic indicating variability.** Student calculates a statistic without further considering it in relation to characteristics of the distribution.	• "They are about the same because the range is the same. Both are 40-20=20."*
CoS(1a)	**Use visual qualities of the data to summarize the distribution and provides a method that relies on only eyeballing the data.**	• "You can just look and tell. Method 2 is more consistent because it is less spread out." *
NL(ii)	Student makes a claim but does not describe ways to determine who is more consistent. Students use measures of center to compare consistency.	• "Method 2 is more consistent because its mean is 30 and method 1 is 29, which is less."* • "Method 2 is more consistent." *
NL(i)	Student does not understand what the question asks for.	• "I don't know what consistent means."*
M	Missing explanation.	

**Mock student responses*

