High School: Statistics & Probability
Interpreting Categorical & Quantitative Data

This learning progression will be applied in a 10th grade High school classroom and the common core standards that are aligned to this progression are HSS.ID.A.1**,** HSS.ID.A.2, and HSS.ID.A.3. The following standards for mathematical practice are also aligned to this learning progression: MP3: Construct viable arguments and critique the reasoning of others, MP6: Attend to precision, and MP7: Look for and make use of structure.

Students have been briefly introduced to statistics and probability in past years and will now be receiving a deeper understanding beginning with being able interpret quantitative data. Through this learning progression students will be working with the measures of central tendencies to describe the center of a set of data. As students begin to form an understanding of these measures of central tendencies and how they allow you to interpret data students will be introduced to the box-and-whisker plot and observe how this box plot also allows you to interpret the data even if you are not given the data directly. Students will also learn of outliers, how to determine if a set of data contains outliers, and how they affect the appearance of their box plots (skewedness) and the means of measuring the data's center.

In order to aid all students learning the teacher will incorporate cooperative learning throughout this learning progression. "Cooperative learning is the instructional use of small heterogeneous groups of students who work together to maximize their own and each other's learning" (Vaughan, 2002, p. 359). Studies have shows that group work aids student learning and retention. "The collaborative nature of cooperative learning gives students a chance to complete tasks and attain concepts they may not have been able to accomplish themselves" (The Multicultural Mathematics Classroom, 6). Cooperative learning especially benefits culturally diverse classrooms by allowing student to practice the new math language learned, through discussion with their peers.

Interpreting Categorical and Quantitative Data

Summarize, represent, and interpret data on a single count or measurement variable.

HSS.ID.A.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data.

In order for students to build on their knowledge they must have a clear understanding of the concepts they will be learning about. Students will first be introduced to the terms, mean, median, and mode. They will refer to these terms as measures of central tendency. Students by now are familiar with mean or the term average. The terms median and mode may be new to them, but as they are also used to measure or describe the center of a set of data students will be able to build off their previous understanding of the term mean. Students will learn the meaning of these terms, their uses, and

how to compute them through examples given in class and individual practice problems. This will benefit students as they are getting a better understanding of these terms through these examples and practice problems and increasing their procedural fluency. Giving students a strong foundation on the meanings of these terms will ease their transition as they go deeper into statistics and probability. It will also allow them attend to precision[M.P6], distinguish the terms from one to another, and know when it is appropriate to use term rather the other. Such as when outliers are in the data set and cause the distribution to be skewed, the median would be more appropriate and the mean is appropriate to use if the mean and the median are close in value, as that infers that the data is reasonably symmetric. Students will also be shown examples of when each term is appropriate to describe the center through different sets of data. These examples will show how outliers cause the mean to be skewed towards this extreme number and how the median is not affected by it. If I notice any struggling students through direct observation during my instruction, I will incorporate cooperative learning to aid their learning. Students will be paired into groups (struggling students with high achieving students) and given data sets for which they must chose the best measure of central tendency and explain their reasoning.

After students have an understanding of the measures of central tendencies students will also be introduced to Tukey's five number summary, which consists of the min, first quartile, median (second quartile), third quartile, and max. One of these terms students are already familiar with; the median. Their previous knowledge will help students fully understand what quartiles are and how to determine them. The first quartile will be explained as the "median" between the min and the median of the set of data as the third quartile will be the "median" of the data from the median to the max. By referring to the quartiles as medians, students will gain a better understanding of what they represent. It is important to help students understand that each quartile contains 25% of the data.

Once students have a better understanding of what the quartiles are and represent, they can be given the formula for the location of each quartile. When giving the students the

CCSS.MATH.PRACTICE.MP6 Attend to precision.

Benchmark for first lesson: The data below represents the wind velocity measurements at the summit of Mt. Rainer on May 15 for the last 28 years. Find the mean, median, and mode. Which would describe the center of this set of data? Explain.

21 25 34 18 17 3 21 24 32 20 17 25 27 26 23 30 9 24 19 16 17 26 23 17 29 25 12 10

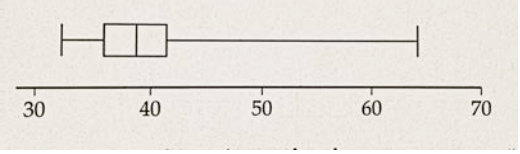
formulas, you can begin by giving them a set of data (with an obvious outlier for later use) and the formula to the median (second quartile) (n+1)/2. Show students how you would use the formula to find the second quartile/median and have students use their problem-solving skills to conjecture formulas to find the first and third quartile. This will be done in groups (cooperative learning) where they can brainstorm together and bounce ideas of each other until they conclude a formula which they will then get checked by me. I will then either confirm that it is correct or provide a hint to lead them in the right direction. Allowing students to conjecture the other formulas allows them to use their prior knowledge and problem solving skills and further assist their learning[MP7].

Once students have found the formulas for the other quartiles, the box-and-whisker plot will be introduced. The set of data given to the students to find the quartiles will be used to create the box plot and to show the components of this box plot (Tukey's five number summary). Once the box plot is completed, point out how the outlier has caused the distribution to be skewed by comparing that box plot with a pre-made box plot of the set of data if the outlier was removed. Giving students a solid comparison on how outliers affect the box plot and distribution gives them a deeper understanding on what outliers are and how exactly they can affect data distribution. After this example students will be given another set of data and asked to create a box-and-whisker plot and interpret their results on their own. This way student can practice what they have just learned and transition this knowledge to their long term learning to retain in the future when needed. The box-and-whisker plot also allows the teacher to check students understanding and procedural fluency. Having them interpret their results also allows the teacher to assess their mathematical reasoning skills. This benchmark will be used to determine if an additional day is needed to further explain box-and-whisker plots and its components or if students are ready to move on. If common misconceptions are detected through analyzing this benchmark, the next lesson will address these misconceptions and be used to give alternative explanations/ activities to aid student learning.

In the third lesson students will be introduced to how to identify outliers using the interquartile range (IQR: third

CCSS.MATH.PRACTICE.MP7 Look for and make use of structure

Benchmark for the second lesson: The following data represents salaries of professors at a small university.
34 41 39 42 38 36 41 64 32 35 37 40 42 38
a) Present Tukey's five-number summary for these data.
b) Interpret the results.
c) Construct a box-and-whisker plot.



Seems the box plot is skewed to the right.

HSS.ID.A.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

quartile-first quartile), also a new term. The interquartile range will be referred to by the range of the center, as it gives you the range for the middle 50% of your data. By now students are already familiar with outliers, they just don't know what the exact criterion is for an outlier. Most outliers can be identified easily, but students must put them to the test in order to accurately identify them as outliers. In order to do so they will use the following criteria: values below (first quartile) - 1.5IQR and anything above (third quartile) + 1.5IQR and will be known as the lower and upper fence. Students will be given a few practice problems to address their understanding and procedural fluency. After students have gained a basic understanding of finding outliers they will be put into groups of 3 and given a task within their groups to come up with a quantitative topic to survey their classmates to create a set of data. Through the use of cooperative learning students will be provided with opportunities to express varied perspectives.

Students will have two sets of data. One will be of their female peers' responses and another of their male peers' responses. Students will then create a box-and-whisker plot for each males and females and compare the two data sets such as the one on the margin. They will also be asked to describe the shape, center, spread of their data, and if the median or mean would be more appropriate to describe the center of their data as well as to identify any outliers. In total they will be asked for the min, max, first quartile, median, third quartile, mean, and amount of male and female students surveyed.

The margin shows an example of a possible answer for this assessment/activity. It shows that both boxplots show distributions that are skewed to the right. It makes sense that most haircuts will not cost too much, but a few students will spend a large amount. Since the cost will always be a positive number, the minimum cannot be less than 0 and there is a long right tail. The centers and spreads are quite different. The median cost for females is about twice that of males, and there is much more variability in the haircut costs for women. The interquartile range (IQR) for women is $55, while for men it is $10.75. An appropriate response for #3 based on the box plot on the margin would be "It is obvious the mean is larger

HSS.ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Formative Assessment:
Seventy-five female college students and 24 male college students reported the cost (in dollars) of his or her most recent haircut. The resulting data are summarized in the following table.

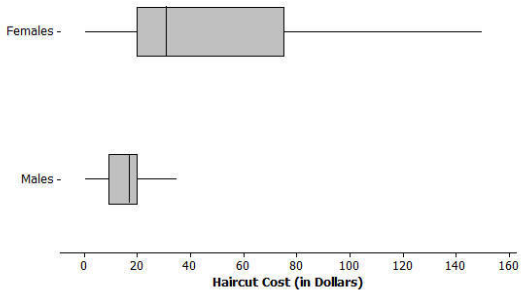
	Females	**Males**
No. of observations	75	24
Minimum	0	0
Maximum	150	35
1st Quartile	20	9.25
Median	31	17
3rd Quartile	75	20
Mean	52.53	20.13

1. Using the minimum, maximum, quartiles and median, sketch two side by side box plots to compare the hair cut costs between males and females in this student's school.
2. How would you describe the difference in haircut costs between males and females? Be sure you discuss differences/ similarities in shape, center, and spread.

than the median because the distribution appears to be skewed to the right. The mean averages all the values in the data, so is "pulled" toward the high ones. "Since the median gives a better description of the center, or a "typical" haircut cost, it is more appropriate. Note that the mean for males is about equal to the 3rd quartile, indicating that 75% of the males paid less than the mean haircut cost for males. For women, the median is $31, indicating that half of women spent $31 or less, but the mean haircut cost for women is $52.53. The mean doesn't give us a good idea of what we could expect for a typical student's haircut cost. It is best to only use the mean when the data distribution is reasonably symmetric.

This activity will be assessing students' understanding of the concepts and vocabulary we have been learning and their procedural fluency as they conduct their surveys and create their box plot. Through the prompts their reasoning skills will also be assessed. Throughout this learning progression students have been building on previous knowledge to meet the mentioned common core standards and this last assessment will inform me if students have successfully met them.

3. Is the mean greater or less than the median? Explain why.
4. Is the median or mean a more appropriate choice for describing the "centers" of these two distributions?
5. Are their any outliers? Explain.



CCSS.MATH.PRACTICE.MP3 Construct viable arguments and critique the reasoning of others.