Turnkey Guidance for
The New York State Next Generation Mathematics Learning Standards
Statistics Progression Video (Pre-Kindergarten-Algebra II)

Goal: The following offers suggestive guidance on how districts can utilize the Statistics Progression Slideshow as a platform for collaborative conversations on the importance of statistical literacy. The slideshow takes an over-arching look at the progression of the statistics standards, Pre-Kindergarten through Algebra II, and provides an understanding of what the statistics standards require when it comes to student learning and instruction. What follows below are recommended stopping points, discussion points and guiding questions. Both the slideshow and guidance below are not intended to limit discussion and instruction.

Materials needed:

- Statistics Progression Slideshow for Pre-Kindergarten-Algebra II
- New York State Next Generation Mathematics Learning Standards

Optional Materials:

- Draft K-5 Progression on Measurement and Data (measurement part)
- Draft K-5 Progression on Measurement and Data (data part)
- Draft 6-8 Progression on Statistics and Probability
- Draft High School Progression on Statistics and Probability
- Guidelines for Assessment and Instruction in Statistics Education Pk-12 Report (GAISE)

Instructions:

- Suggested discussion group configurations:
  - Grade-level bands: Pre-K thru Grade, and MS/HS.
  - One large group with representatives from all grades/courses, Pre-K thru Algebra II.
- Encourage all participants to view the slideshow in its entirety and make any notes they deem valuable for discussion. All participants should also bring their hard/digital copies of the standards to the session.

STOP 1: IN A DATA DRIVEN WORLD (time stamp 0:44)

Slide 3: Several images/words are shown that highlight where data appears in daily life.

Guiding questions:

- What impact/role does statistics have in the adult daily routine/life? What daily experiences/decisions involve having statistical knowledge? **(e.g., weather prediction, when is it appropriate to wear a jacket)**
- What impact/role does statistics have in our students' daily routines/lives? What role can statistics play in supporting culturally responsive instruction?
- What does "*statistically literate*" look like in the grade strands (similarities/differences)?
  - Pre-K thru Grade 2 **(e.g., sort and classify, construct and interpret picture/bar graphs)**
  - Grades 3-5 **(e.g., construct and interpret scaled picture/bar graphs, line plots with measurement data)**
  - Grades 6-8 **(e.g., sampling populations; construct and interpret dot plots, histograms and box plots; graph bivariate data)**
  - Grades 9-12 **(e.g., regression line, correlation coefficient, correlation vs causation, normal distribution, sampling distributions, simulation, and inference)**
- How are statistical reasoning and algebraic reasoning different? **(e.g., statistical reasoning includes data collection, representation, interpretation, and prediction whereas algebraic reasoning**

involves representing, generalizing, and formalizing patterns and regularity in all aspects of mathematics[1])

STOP 2: CATEGORICAL DATA (time stamp 1:24)

Slide 6: Slide highlights how students are developing mathematical language when analyzing categorical data.

Guiding questions:

- What is the definition of categorical data? **(e.g., data that can be classified by type such as colors or breeds of dogs)**
- Where do students first encounter/experience categorical data? **(e.g., lunch choices, favorite color or favorite ice cream flavor)**
- What specific Pre-K thru Grade 3 standards connected to categorical data help solidify the foundations for students' work with statistics in grade 6 and beyond? **(e.g., NY-PK.MD.2, NY-K.MD.3, NY-1.MD.4, NY-2.MD.10, and NY-3.MD.3)**
- What aspects of other content areas support students' work with categorical data?

STOP 3: MEASUREMENT DATA (time stamp 2:27)

Slide 8: Slide demonstrates the movement of complexity through early grade levels regarding measurement data, fractional data, and precision.

Guiding questions:

- When is measurement data first introduced and utilized? At which grade-levels? (**e.g., NY-PK.MD.1, NY-1.MD.1, NY-1.MD.2, and NY-2.MD.1)**
- What additional standards, thru Grade 5, speak specifically to measurement data? **(NY-2.MD.9, NY-3.MD.4, NY-4.MD.4, and NY-5.MD.2)**
- Discuss the role of precision in the statistics standards. **(e.g., NY-2.MD.1 whole unit, NY-3.MD.4 quarters and halves, and NY-4.MD.4 eighths)**

STOP 4: THE SHAPE OF A DATA DISTRIBUTION & OUTLIERS

(time stamp 3:20)

Slide 10: Slide shows a skewed data distribution with a clearly marked outlier.

Guiding questions:

- What does a dot represent on the dot plot? **(e.g., an individual)**
- At what grade levels could the concept of "outliers" be introduced? **(e.g., student generated measurement data in grade 2)**
- At the middle level, the focus is on interpreting the effect of an outlier on the center (mean, median), spread (range, IQR), and shape (symmetry) for a given data set. How far away should the outlier be from the rest of the data to be considered an outlier, without formally establishing fences? **(e.g., isolated or extreme data values)**

STOP 5: SAMPLING (time stamp 3:57)

Slide 11: Slide speaks to the topic of sampling and when collecting all data values for a population is unreasonable.

Guiding questions:

- What is the difference between Census and Sampling? **(e.g., every participant vs a representative subset)**
- Why use a sample? **(e.g., it is physically impossible or cost prohibitive to conduct a census)**
- Why should a sample be randomly selected? **(e.g., random sampling removes potential bias in the selection process while providing a representative subset)**

- When do students begin comparing two or more samples? **(e.g., NY-7. SP.3)**
- What should students be comparing when looking at two or more samples? **(e.g., measures of center and spread (NY-7. SP.3/4))**
- What inferences can be made about the mean heights of students from both Middle Schools? **(e.g., since the representative sample from Lincoln MS had a greater mean height, one could infer that Lincoln MS students are taller than Washington MS)**
- Bias is explicitly mentioned in the Algebra II standards, however, when is the concept of bias first introduced? **(e.g., NY-6. SP.1b, NY-6. SP.1c)**

STOP 6: REPRESENTING UNIVARIATE VS BIVARIATE DATA

(time stamp 5:28)

Slides 12-14: Slide 12 displays two univariate data distributions represented on a single number line (univariate measurement data represented on a boxplot). Slides 13 and 14 shows a bivariate data scatterplot on the coordinate plane.

Guiding questions:

- What is the difference between the boxplots and the scatterplot? **(e.g., the data values in the boxplots are univariate (one data measure per individual from the sample), whereas data values in the scatterplot are bivariate (two data measures per individual))**
- Why can't the two Middle School samples represented in the boxplots be plotted on a scatterplot? **(e.g., the data is univariate and unrelated to an individual)**
- A boxplot is a visual representation of the five-number summary. What are the five numbers in the five-number summary and what do those numbers tell us about the data set? **(e.g., the grade 6 five-number summary includes the minimum value (32), first quartile (50 (25% of the scores are at or below 50)), median (60), third quartile (70), and the maximum value (78))**
- What do the x and y represent in the scatterplot? **(e.g., height and arm span in inches)**
- Why construct a regression line? **(e.g., best approximates the relationship between height and arm span and allows for making predictions)**
- If predicting arm span for two students, one 60" tall and the other 68" tall, which prediction is more reliable and why? **(e.g., since we are extrapolating in both cases, 60" is more reliable because it is closer to the domain of the original data set)**

STOP 7: IDENTIFYING OUTLIERS (time stamp 5:54)

Slide 15: This slide introduces the formula for outliers used to determine when a data point is considered extreme (too far from the center values) for a univariate data set.

Guiding questions:

- If the data in the boxplot represents the number of chocolate candies in various 3.5 oz. packages, interpret the outlier in the context of this problem. **(e.g., a bag with 102 candies is extreme as compared to the IQR)**
- What is the shape of the distribution? **(e.g., strongly skewed left)**
- What is the appropriate measure of center, mean or median? **(e.g., the outlier has a greater impact on the mean so the median would best represent the center of the data (AI-S.ID.3))**

STOP 8: NORMAL PROBABILITY DISTRIBUTION (time stamp: 6:35)

Slides 16-17: Slide introduces standard deviation as a measure of spread about the mean.

Guiding questions:

- Compare and contrast the two measures of spread, the standard deviation and the IQR. **(e.g., standard deviation is the average of distances of each data point to the mean, whereas the IQR is the distance between the first and third quartile of the data set)**
- Is it appropriate to use standard deviation for a data set that contains outliers? **(e.g., no, because the standard deviation is not resistant to outliers; meaning that the outlier distances from the mean will cause to the standard deviation to be larger and grossly inflate the variability.)**
- Interpret the meaning of the 0.849 in the context of the problem presented on slide 17. **(e.g., 84.9% chance that Amy's swim time will fall between 37 and 44 seconds)**
- What interval would represent Amy's swim times that fall within 1 SD of the mean? What percent of her times fall within that interval? **(e.g., 37.4 to 42.0 and normalcdf(37.4, 42, 39.7, 2.3) = 0.683 = 68.3%)**

STOP 9: THE PREDICTIVE NATURE OF STATISTICS (time stamp 7:12)

Slides 18-19: Both slides highlight how a simulation can be utilized to determine if a value for a sample proportion or sample mean is likely to occur. Have participants engage with the following applet: http://onlinestatbook.com/stat_sim/sampling_dist/ Click [Begin] on the top left and then click the [Animate] button on the right several times.

Guiding questions:

- What do the three graphs represent? **(e.g., top: population; middle: sample; bottom: distribution of sample means)**
- Which of the three graphs would be used to simulate sample means? **(e.g., the bottom graph would be used to simulate multiple sample means from the population)**
- What is the purpose of calculating the interval containing the middle 95% of the data centered on the mean? **(e.g., identify when sample means are usual or unusual)**
- When is a sample mean statistically significant? **(e.g., when the sample mean falls outside the interval containing the middle 95% of the data centered on the mean (AII-S.IC.4))**
- What does a dot represent on the dot plot seen on slide 19? **(e.g., in Algebra II, dots no longer represent individuals, and in this case, the dots represent sample means based on 100 different simulated samples of size 200)**
- Why is a sample mean (or sample proportion) that falls outside the 95% interval considered unusual? **(e.g., statisticians have deemed that chance variation can be ruled out if a sample mean/proportion falls outside the two SD interval)**

NYSED would like to thank members of the Content Advisory Panel and the S/CDN Mathematics Framework for assisting in the development of this guidance document.

1. Van de Walle, J. A. (2004). *Elementary and middle school mathematics: teaching developmentally (5th ed.)*. New York: Longman, 417.