Project Title: Developing Critical Thinking through 6th Grade Statistics

Name: Zach Mineroff

Draft Date: 12/17/17

Table of Contents:

Contents

Step A: Learners in Context

Context for Educational Design

Topic Context

In what domain have you or other educators noted learning challenges, and which of them will your project target?

A 2005 paper by Capraro et al., "Middle Grades: Misconceptions in Statistical Thinking," provides some good insights about learning challenges 6th grade students face in statistics. Even if students can construct various graphs and charts, they often have trouble choosing which one best represents the dataset given the problem at hand. They tend to have trouble with certain edge cases as well. For example, if they are constructing a bar chart and one of the categories has a value of 0, they might give that bar a height above the x-axis to show that it is "there." Furthermore, many students do not see the value of visual representations of data when they have computational methods (e.g., mean, median, and mode) at their disposal.

On a more informal level, I also browsed through the comments section of relevant Khan Academy videos. Quite a few students remarked on the lack of trust in their own ability to understand concepts in statistics ("oh my GOD idk how 2 do this"). Others indicated that they could follow certain procedures, but did not understand their purpose ("I still don't really get what the IQR really is, and why the concept is like that."). Many comments reflected the difficulties hinted at in the Capraro et al., paper regarding edge cases ("If two numbers are the most common in a set… what would be the mode?") and seeing value in the different visual representations of data:

> *"why does it matter which way you represent the data? if each representation holds the same data, why do we have to learn so many different kinds of data plots, and why can't we just use one of them? (like, to me a histogram makes a lot more sense than a dot plot) besides, if you get the same info out of so many different graphs and tables, isn't it just confusing to have to use so many? and which one is best?"*

Finally, the Math standards themselves note some common gaps in knowledge. In the introduction, the authors highlight that students who can explain where a rule comes from (rather than just using a mnemonic to remember it) will have a much better chance at succeeding in a less familiar task.

Based on the common learning challenges gathered from these sources, it will be important for my project to target conceptual understanding of different statistical methods and knowledge of when to apply one over the other. Dispositions about self-efficacy are also essential.

In a paragraph, explain what knowledge, skills, and dispositions are central to what you plan to teach.

Section 6.SP of the Math Standards lays out the core knowledge and skills that should be taught in a 6th grade statistics class. Students need to be able to recognize a statistical question, understand that datasets have distributions whose shapes can help us make conclusions, and distinguish between measures of center and spread. Given a dataset, students should be able to construct various visual representations of it and provide summary statistics that relate the data points to the context in which they were collected. I will make sure to address common misconceptions associated with these facts and procedures (including metacognitive strategies for students to notice when they themselves might be misinterpreting something). Furthermore, dispositions relating to self-efficacy and the value of

statistics are central to this project. At the metacognitive level, this includes self-monitoring of frustration and thinking through *why* they might believe they are less capable. In addition to knowing the various ways of representing the same data (graphs and summary statistics) and how to calculate them, students should know when to apply the appropriate method and what insights each one offers. This involves the dispositional goal of wanting to look at a problem from multiple angles in order to better understand it.

Which of the standards established by professionals in the field are related to your design focus?

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics: Grade 6 Statistics and Probability*. Retrieved from http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf

6.SP 1

Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.

6.SP 2

Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

6.SP 3

Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number. Summarize and describe distributions.

6.SP 4

Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

6.SP 5

Summarize numerical data sets in relation to their context, such as by:

a. Reporting the number of observations.
b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

This is the whole set of 6th grade statistics and probability standards. I think it is feasible to include all of them in one unit because the ideas are interconnected and build on each other. Additionally, I am including one more statement from the standards that is related to the idea of multiple representations:

STANDARDS FOR MATHEMATICAL PRACTICE PART 5

Proficient students are sufficiently familiar with tools appropriate for their grade or course to make sound decisions about when each of these tools might be helpful, recognizing both the insight to be gained and their limitations.

This last statement is not exactly a numbered "standard" from 6th grade statistics and probability, but is clearly important to the authors.

Why is it important for students to learn what you plan to teach?

Not every student will become a full-time statistician, but a community that encourages critical thinking will reap several benefits. People will have more sensible political discussions, interpret their peers' choices in a reasonable manner, and be less prone to taking over-glorified scientific studies from the media at face value. Groundless claims can spread around quickly, and can only be combated by people who stop and say, "Hmm, does this really make sense? Does this method of presenting the information mislead or outright lie in some way?" 6th graders might not quite be ready to run for office, join debate clubs, or have in-depth discussions about the economy, but they can begin to learn the basics of drawing appropriate conclusions from data. To use the cooking analogy from class, the students are not yet baking their own apple pies (analyzing scientific claims or political arguments from the media), but are chopping vegetables with help from their parents (knowing which forms of analysis to use for simple problems).

How can you maximize the continuity between your goals and the learners' goals (a la Bain's "promises") so that you can tap their natural motivation?

As Bain suggests, I can use the beginning of each instructional block to ask "big questions" regarding statistics in everyday life. The 6th graders I'm teaching will likely be very familiar with Google search, but perhaps they never thought about how it works. How does it come up with suggestions and results so quickly? In a similar vein, how does Netflix make recommendations about what to watch next? These sorts of questions demonstrate the power of statistics and may appeal to those who are curious about the way the world works.

Other students may be more interested in how statistics can help you make personal decisions. To take an example from Khan Academy, let's say you are a baker and are deciding between moving to City A and moving to City B. Both have the same average salary for bakers, but in City A, every individual baker has the same salary and in City B, one baker makes *a lot* more than the other bakers, who make very little. Which city would you rather move to? The benefit of this type of question is that, by the end of the course, students will actually be able to use the idea variance to make a more informed decision.

Throughout the intervention, I will periodically come back to these big questions to keep students interested in the material and prevent them from losing sight of the point of it all.

Learning Context

Who will you teach? Grade Level within K-12? Type of students? Type of community? Community values & priorities?

My project will target 6th graders in public school. Mainly motivated by familiarity, I will build the intervention around a community similar to the one I grew up in—a relatively affluent suburb on the east coast. Parents expect their kids to go to college right after high school and value performance on standardized tests. As work load starts to increase, the 6th graders are starting to learn how to balance

extra-curricular activities with school. Some students will have more interest in mathematics than others and may experience math anxiety perpetuated by role models (possibly parents or teachers) who have difficulty with math themselves.

Where will you teach them? In school or another learning environment? In what subject area within school?
I am envisioning teaching students mostly in the classroom during the math portion of the day. 6th grade is probably the first time students have separate classes/instructors for different subjects. I will also incorporate eLearning lessons that students can work through at their own pace at the end of the class and/or at home.

When will you teach them? Time of year? Anticipated length of instructional sequence? Estimated total time of lessons?
I will teach this chunk later in the school year, after students have had a good amount of practice with ratios and basic algebra. I will probably split the course up into 10 1-hour lessons. At one lesson per school day, the entire intervention will last two weeks.

How will the learning experiences you envision relate to other parts of the learners' program within the grade level or across grade levels?
According to the standards, 6th grade marks the beginning of lessons in statistical thinking. My intervention will draw on earlier experience with order of operations, ratios, and basic algebra. Having a good understanding of measures of variance and spread will prepare students for grade 7, where they will start making comparative inferences about two populations. In grade 8, the class will elaborate on possible relationships between variables. In high school, students will begin to use statistics to justify more-involved arguments. This requires a good understanding of which statistical technique to use based on the shape of the data distribution. A good foundation in this area in 6th grade will certainly help. Some students will study calculus in high school; experience with histograms should help them learn techniques for estimating the area under a curve.

Within the grade level, a solid understanding of statistics will benefit learning in other subjects. Most obviously, science requires analyzing, interpreting, and displaying data to test hypotheses. Furthermore, the standards for social studies involve collecting, using, and reporting data. For example, D2.Eco.11.6-8. (Use appropriate data to evaluate the state of employment, unemployment, inflation, total production, income, and economic growth in the economy) and D4.2.6-8. (Construct explanations using reasoning, correct sequence, examples, and details with relevant information and data, while acknowledging the strengths and weaknesses of the explanations) involve learning goals similar to those in the current project. Meeting the statistics standards will also help learners reach some of the English standards, such as Text, Types, and Purposes; Grade 6, 1b (Support claim(s) with logical reasoning and relevant, accurate data and evidence that demonstrate an understanding of the topic or text, using credible sources).

Learner Profile as a Baseline

Developmental Level = learner characteristics based primarily on age / maturity
The developmental level of these learners is a double-edged sword. On one hand, 6th graders are able to think more critically and abstractly than they were in lower grade levels. On the other, they're facing a great deal of physical and social changes that might act as distractions from their learning. I examined

the developmental level of 11-12 year olds using three frameworks: Piaget's Cognitive Steps, the CDC's childhood development page, and Levine et al.'s Table of Neurodevelopmental Constructs.

Piaget's Cognitive Steps
6th graders are definitely able to handle concrete operational concepts. They can think logically about concrete objects and understand conversations. They are beginning to enter the formal operational stage, in which they can reason abstractly and think in hypothetical terms.

CDC
6th graders are growing out of middle childhood and into young teens. Increased attention spans and a higher ability for abstract thought are coupled with a growing independence regarding choices about friends, school, and other activities. They may feel less affection toward parents and might seem rude or short-tempered to teachers. A more-demanding workload and the onset of puberty may exacerbate stress and feelings of sadness/depression. Furthermore, an increase in peer pressure might distract students from learning. On the bright side, people entering this age group develop a better ability to express their own feelings through talking.

Levine et al.'s Table of Neurodevelopmental Constructs
This source is more about common demands/expectations that learners in this age group face. In school, 6th graders need to pay attention in contexts that may not interest them very much. An "explosion of decontextualized language" may require extended mental effort to understand. Classes begin to require more and more activation of prior knowledge. On the social level, students will likely face more distractions as they embark on a "quest for intimacy in friendship" and grapple with conforming and normality.

Knowledge Base = learner characteristics based primarily on experience
Pre-existing knowledge
6th graders will have a decent handle on order of operations. They will be able to list different ways that data might be collected as well as units of measurement. They should understand different forms that data can take (quantitative vs qualitative). Knowledge regarding data collection and analysis may be limited to certain contexts. For example, they will be familiar with using simple units of measurement (like inches) and familiar contexts (like heights of classmates), but not more complex, abstract domains (like m/s^2 to measure acceleration).

Pre-existing skills
By the time the instruction for this project begins, the students will have experience with multiplication, division, addition, and subtraction. A possible negative side effect is that students might be able to perform a lot of calculations without actually understanding how they work. Other procedures that students should have familiarity with include:

- Organizing information in rows and columns
- Observing basic patterns
- Ordering numbers
- Solving simple algebra problems

Pre-existing dispositions
Students will likely vary in the dispositions that they bring to the classroom. On the positive end, they may have an eagerness to reason through things and ask themselves "why" often. They may have developed a curiosity about how the world works and a willingness to check answers after solving

problems. On the harmful side, many students may have developed math anxiety (possibly worse for girls). They may feel that math has no real purpose or that they aren't meant to understand what a concept actually means beyond being able to perform the associated calculations.

Individual Differences = learner characteristics based primarily on stable, individual properties more than age or experience

Though my community might be relatively more affluent, there will still be students from lower-income families attending the school. These students may have less experience with math-related activities at home (e.g., board games). Parents will differ in how much they help their students with homework and place more or less value on mathematics in general. This will likely affect their children's math anxiety and interest in learning statistics. Students who are more visual learners will probably find constructing visual representations of data more natural than students who prefer to learn through verbal or kinesthetic instructional methods. Those who think more logically, as well as those who are careful, neat, and meticulous, will have an advantage.

Initial Resources Available

How much experience do you have in the project domain? As a student? As a teacher?

I have a good amount of experience in this domain as a student. I have a taken a few stats classes in college and on my own through online platforms. My last job involved using statistics at a higher level than I was initially prepared for, so I had to make sure I had the basics down pretty well. As far as being a teacher, I would sometimes go over statistical concepts with new lab members to make sure they were comfortable with the material.

Is there an educator studying this topic who can serve as a consultant for you?

Emily Keebler from class agreed to help me out. She has experience teaching statistics to this age group. I asked her a few basic questions about the direction of my project and I seem to be on the right track.

Start searching for articles and other web / print resources for your domain. Include your annotated bibliography here (i.e., reference & brief description of each).

Capraro, M.M., Kulm, G., & Capraro, R.M. (2005). Middle Grades: Misconceptions in Statistical Thinking. *School Science and Mathematics*, 105(4), 165-74.

This paper describes a study on 134 6th graders who were using the Connected Mathematics curriculum. Given a data table showing how many pets individual people have, the students were asked to construct a graph that shows how many people have no pets, one pet, two pets, and so on. They were also asked to determine the typical number of pets from the data/graph and justify their reasoning. After grading the papers on a rubric, the researchers also interviewed the students to gain further insights about their misconceptions. This is a good source for identifying potential learning challenges and understand why certain instructional methods can lead to misconceptions.

Beilock, S.L., Gunderson, E.A., Ramirez, G., & Levine, S.C. (2009). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*. 107(5), 1860-1863.

This study provides a good overview of the relationship between a female math teacher's own anxieties about math and her students' feelings about the stereotype that "boys are good at math and girls are

good at reading." It provides some suggestions for reducing instructors' math anxiety though training and education.

Boaler, Jo. (2015). Fluency Without Fear: Research Evidence on the Best Ways to Learn Math Facts. *YouCubed*. https://www.youcubed.org/evidence/fluency-without-fear/

While this source does not address statistics specifically, it's a good article that discusses the balance between achieving fluency and understanding in mathematics.

Are you aware of any educational materials (instruction and/or assessment) that have already been designed to teach this domain? List them here, along with your opinion of their quality.

Museum of Mathematics Lesson Plan for Hands-On Data Analysis

- Link
  - https://momath.org/wp-content/uploads/RosenthalPrize2012_Winning_Lesson_Plan.pdf
- Pros
  - Specifically addresses national standards
  - Ties lessons in with previous knowledge
  - Each day, asks students to recap what they accomplished
  - Includes instruction for metacognitive knowledge and skills
- Cons
  - Does not cover all of the goals for this project
  - Does not address dispositional goals

Khan Academy Lessons on 6th Grade Statistics

- Link
  - https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics
- Pros
  - Covers all of the national standards for this project
  - Set up in easy-to-use learning modules
  - Formative assessments included
  - Moderated comment section provides a platform for students to express what they're having trouble with and encourages students to help each other
- Cons
  - Instruction and assessment do not directly address dispositional goals
  - Does not specifically address common misconceptions
  - May encourage being able to follow procedures rather than really understanding mathematical concepts

Engage NY Grade 6 Mathematics: Statistics

- Link
  - https://www.engageny.org/resource/grade-6-mathematics-module-6
- Pros
  - Lesson plans directly relate to standards
  - Separation of student and teacher resources
- Cons

- There's *a lot* to comb through; not necessarily a bad thing in general, but this source might not be focused enough on the central chunk for my project

BetterLesson: 6th Grade Statistics

- Link
  - https://betterlesson.com/common_core/browse/374/ccss-math-content-6-sp-statistics-probability?from=breadcrumb_domain
- Pros
  - Lesson relate to standards
  - Encourages self-reflection
  - Lesson plans are iterative (they take instructor feedback)
  - Modules include Big Ideas in statistics
  - Lesson plan overviews include details about the environment they have been used in (e.g., urban/suburban, region of the US)
- Cons
  - Crowd-sourced nature of the lesson plans/voting system might mean some resources aren't properly vetted

Proposed Focus for a One-Week / 10-Hour Segment to Design in Detail

What central chunk of the educational program you envision do you propose to design for a one-week unit or 10-hour intervention spread over a period of time?

This educational intervention will focus on the 6th grade Statistics and Probability standards (6.SP). In addition to the knowledge and procedures outlined in the standards, this intervention will help develop metacognitive strategies for students to monitor their own understanding of measures of center, measures of variance, and ways of visualizing data. Furthermore, it will target dispositional goals related to self-efficacy, especially for those who are particularly prone to math anxiety.

Which of the educational standards you listed above are the most relevant to the central chunk you have selected?

6.SP 1

Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.

6.SP 3

Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number. Summarize and describe distributions.

6.SP 5

Summarize numerical data sets in relation to their context, such as by:

c. Giving quantitative measures of center and variability, as well as describing any overall pattern and any striking deviations from the overall pattern.
d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Step B: Goal Specification / Task Analysis

Conceptual Knowledge

C.C.1 Recognize a statistical question

- C.C.1.1 A statistical questions is a question that requires collecting multiple data points that (probably) vary in order to answer it.

C.C.2 Understand the forms statistical data can take

- C.C.2.1 There can be many data points, or just a few
- C.C.2.2 The data points can be close together (similar in value) or stretched apart (different from each other)
- C.C.2.3 Data can be quantitative or qualitative
    - C.C.2.3.1 Quantitative data deals with numbers and can be measured
    - C.C.2.3.2 Qualitative data deals with descriptions and can be observed, but not measured.
- C.C.2.4 Data can be continuous or categorical
    - C.C.2.4.1 Continuous can take any value (within a range)
    - C.C.2.4.2 Categorical data are grouped into non-overlapping categories

C.C.3 Understand information we might want to extract from a dataset to answer a statistical question

- C.C.3.1 Center is a single number that indicates the typical value of a dataset
- C.C.3.2 Spread is single number that indicates the extent to which the data points are stretched out or squeezed together

C.C.4 Recognize numerical ways of finding center

- C.C.4.1 The mean is the sum of all values in a data set divided by the number of values in the data set
- C.C.4.2 The median is the middle value in a data set that has been arranged in order of magnitude
    - C.C.4.2.1 The first quartile is the median of the minimum and median of the entire data set
    - C.C.4.2.1 The third quartile is the median of the median of the entire data set and the maximum
- C.C.4.3 The mode is the value that appears most often in a dataset

C.C.5 Recognize numerical ways of finding spread

- C.C.5.1 The inter-quartile range (IQR) is the difference between the third quartile and the first quartile
- C.C.5.2 The mean absolute deviation is the mean difference from the mean of the entire dataset

C.C.6 Recognize visual methods of displaying data

- C.C.6.1 A dot plot includes all values from the data set, with one dot for each occurrence of a value
- C.C.6.2 A histogram divides data into intervals and uses a bar graph to show the number of observations in each interval
- C.C.6.3 A box plot shows the minimum, first quartile, median, third quartile, and maximum

C.C.7 Recognize the advantages and disadvantages of visual methods used to summarize statistical data

- C.C.7.1 From a dot plot, you can easily tell how many of a specific value is in a data set, but can become cluttered and tedious to construct if there are a lot of data points.
- C.C.7.2 Histograms can help you visualize the distribution of larger datasets, but you cannot see if a specific value is in the dataset.
- C.C.7.3 With a box blot, you can easily see the median, minimum, maximum, and IQR, but you cannot tell if specific values are in the dataset.

Meta-Level Conceptual

M.C.1 Self-awareness of deep understanding of these concepts (beyond ability to perform calculations or construct graph)

- Examples
    - Do I understand what conclusions I can draw from the IQR instead of just being able to calculate it?
    - Do I know why I should use a histogram to represent this dataset?

M.C.2 Shift learning focus based on difficulty mapping terms to definitions

- Examples
    - Am I getting mean, median, and mode mixed up? Should I make some sort of mnemonic to help me remember which is which?

Procedural Skills

C.P.1 Form a statistical question

- C.P.1.1 Given a general domain or something the student would like to know, formulate a question that requires collecting data to answer.
    - Step 1: Identify the property that you would like to know the typical value of.
    - Step 2: Identify the population you want to know the typical value for.
    - Step 3: Ask a question about the typical value of the property in the population that can be answered by collecting data. Here are some examples:
        - How [property] is the average [population]? (How *tall* is the average *6th grader*?)
        - What is the typical [property] that a [population] [action]? (What is the typical *number of pets* that a *family in Doylestown owns*?)

C.P.2 Describe the context surrounding a dataset

- C.P.2.1 Report the number of observations
    - If the number of observations is given, find the number in the problem statement.
    - If not, count the number of data points or rows in the table.
- C.P.2.2 Describe how the observations were collected, including units of measurement
    - Who collected the observations?
    - What strategy was used to collect observations (surveys, measuring, etc.)?
    - What measuring devices were used to collect the data?
    - What units of measurement were used (or what set of categories was used)?

C.P.3 Make and justify choices about the method of summarizing a dataset

- **C.P.3.1** Given a dataset and a statistical question, identify the graph that will best help answer the question
    - If the dataset is qualitative
        - If there are fewer than 20 data points, use a dot plot
        - If there are many data points, use a histogram
    - If the dataset is quantitative
        - If you need to know the number of values in a specific range
            - If there are fewer than 20 data points, use a dot plot
            - If there are many data points, use a histogram
        - If you need to highlight the median, minimum, maximum, and IQR, use a box plot
- **C.P.3.2** Identify the measure of center that is most appropriate
    - If the dataset is qualitative
        - Use the mode
    - If the dataset is quantitative
        - If there are clear outliers on one end, use the median
        - If there are no clear outliers data has a lot of variation
            - If the data has a lot of variation, use the mean
            - If there is very little variation, use the mode
- **C.P.3.3** Identify the measure of spread that is most appropriate
    - If the dataset is quantitative
        - If there are clear outliers on one end, use the IQR
        - If there are no clear outliers, use MAD

C.P.4 Calculate measures of center

- **C.P.4.1** Mean
    - Sum the individual values
    - Determine the number of observations (by counting or looking at the problem statement)
    - Divide the sum of values by the number of observations
- **C.P.4.2** Median
    - Order the terms from smallest value to largest
    - Find the middle value by crossing off the smallest and largest terms until there is only one term left (if there are an odd number of observations) or two terms left (if there are an even number of observations)
    - If there are two terms left, calculate the mean of these numbers
- **C.P.4.3** Mode
    - Count the number of times each value appears in the dataset
    - Identify the value that appears the most times
    - If two or more values appear the most (equal number of) times, then all of those values are the mode (there can be more than one mode)

C.P.5 Calculate measures of spread

- C.P.5.1 Inter-quartile range
  - Find Q2, the median of the entire dataset
  - Find Q1, the median of the first half of the data (values less than Q2)
  - Find Q3, the median of the second half of the data (values greater than Q2)
  - Subtract Q1 from Q3
- C.P.5.2 Mean absolute deviation
  - Calculate the mean of the dataset
  - Find the deviation of each data point from the mean
    - Subtract the value from the mean
    - Find the absolute value of that value
  - Calculate the mean of the deviations
    - Sum the value of each deviation
    - Divide this sum by the total number of observations

C.P.6 Construct relevant plots to display data

- C.P.6.1 Dot plot
  - List the values on a horizontal line (in increasing order if applicable)
  - For each time a value occurs in the dataset, draw a dot above that value on the line
    - Dots should be evenly spaced and drawn vertically, perpendicular to the number line
- C.P.6.2 Histogram
  - Determine the bin size you want to use (can be 1)
  - On the x-axis, list the ranges on a horizontal line (in increasing order if applicable)
  - On the y-axis, label the number of values on each tick mark
  - For each range, draw a box that has a height of the number of values in that range
  - Label the x and y axis, and write a title
- C.P.6.3 Box plot
  - On the x-axis, label the number values on each tick mark
  - Above the appropriate place on the x-axis, draw a vertical line at Q2 (median)
  - Above the appropriate place on the x-axis, draw a vertical line at Q1
  - Above the appropriate place on the x-axis, draw a vertical line at Q3
  - Draw a box around Q1, Q2, and Q3
  - Above the appropriate place on the x-axis, draw a vertical line at the minimum
  - Above the appropriate place on the x-axis, draw a vertical line at the maximum
  - Connect the minimum line to the Q1 line with a horizontal line, connected at the centers
  - Connect the maximum line to the Q3 line with a horizontal line, connected at the centers

Meta-Level Procedural

M.P.1 Self-monitoring of willingness to plan solution before starting calculations

- Examples
  - Am I taking time to think about which graph is best or am I just using the easiest one?

M.P.2 Monitor and evaluate process and change course if necessary

- Examples
    - Does my current procedure make sense? If not, am I willing to start over with a new method?
    - Am I taking too many unnecessary steps to solve this problem? Is there a way to cut down on the number of calculations I need to do or points I need to draw in order to make a conclusion?
    - Does my solution make sense given the problem statement and common sense? Is there a way for me to double-check my answer?

Dispositions

C.D.1 Self-efficacy

- Examples
    - I know that if I continue to practice, I will make sense of these concepts.
    - I understand that I am, in a way, learning a new language. It will take some time before I am familiar with all of the terms and when to apply each concept.
    - Everyone can be good at math!

C.D.2 Willingness to experiment

- Examples
    - The first graph I create might not actually be the best way to summarize the data. I am willing to try another way if I recognize this.

C.D.3 Value different forms of visualizations and measures

- Examples
    - It's good that I know how to represent data in different ways. For some problems, a histogram will help me reach a solution better than a box plot would, because it allows me to see how many values are in each bin.
    - I'm glad that I know how to find the mode, because in this dataset, the mean would be misleading because there are outliers that would make me think the typical value is much higher than it really is.

Meta-Level Dispositions

M.D.1 Monitor own anxiety and frustration

- Examples
    - Am I getting frustrated because I didn't get the correct answer right away?

M.D.2 Reflect on own beliefs about what it takes to learn statistics/math in general

- Examples
    - I can really understand these things if I try. I may not immediately get something when the teacher first introduces a concept, but I will understand it after I practice some problems.

M.D.3 Adjust how they approach statistical problems based on values

- Examples
    - If I really care about this, maybe I should think about which graph works best here.

Part C: Assessment Design

General Description

What is the focus of your assessment efforts in the context of your complete goal specification? What will be the primary focus (usually "core tasks") vs. secondary and tertiary coverage (more likely part of "other evidence")?

Two main assessment activities will provide measures for the degree of mastery of each goal listed in part B. The first major assessment is a unit-long statistics project designed to address the primary focus of the intervention. The second is a series of eLearning problems that are meant to assess secondary goals.[1] In addition to achieving sufficiency, the nature of these two activities allows a researcher to test the reliability and validity of each measure.[2]

Furthermore, the assessments help accommodate the different types of learners mentioned in part A. The measures, as well as the timing of the assessments within the unit, enable instructors (and the eLearning platform) to detect students who may be struggling at an early stage and offer them additional instruction.[3] Because the students receive feedback throughout the unit, they are also encouraged to monitor their own learning.[4]

Statistics Project

The unit-long project targets the primary goals of the course, which mainly involve justifying choices in statistics, seeing value in the course goals, and developing a growth mindset. They include:

- C.C.1
- C.C.7
- M.C.1
- C.P.1
- C.P.3
- M.P.1
- C.D.1
- C.D.2
- C.D.3
- M.D.2
- M.D.3

As described in more detail later, students will come up with a statistical question that they would be interested in answering and collect real-world data to draw a conclusion. As they learn about new topics in statistics, they will make choices about which methods best support their arguments and reflect on their own learning through guided questions. The project starts early in the unit and is segmented to enable continuous feedback. At each stage, the instructor uses a rubric to ensure that learners are completing each procedure effectively, using the dispositions we are trying to teach, and doing the metacognitive reflecting. This also helps students monitor their own learning and gives them multiple chances to reflect on and revise their work.

---

[1] Big Idea 4.1: Scale complexity of assessment with depth of learning goal

[2] Big Idea 4.3: Standards for the assessment triangle

[3] Big Idea 1.2: Structure for organizing knowledge

[4] Big Idea 3.4: Metacognitive skills for self-directed learners

eLearning Problems
For homework and end-of-class "sponge" assignments, students will complete problems on an eLearning platform that target the secondary goals of the unit. These goals mainly involve recognizing vocabulary, calculating summary statistics, and constructing graphs. They include:

- C.C.2
- C.C.3
- C.C.4
- C.C.5
- C.C.6
- M.C.2
- C.P.2
- C.P.4
- C.P.5
- C.P.6
- M.P.2
- M.D.1

In each module, students complete a series of formative assessments that provide them with instant feedback based on their performance. Data logs allow the platform to tailor scaffolding for individual students on the fly. After the formative assessments, students complete a summative quiz. The teacher can use the results to identify muddy points and adjust the next class accordingly. Furthermore, the summative quizzes include assessments of prior topics to make sure students are retaining what they learn. Finally, the module ends with a preview assessment to get students thinking about the next topic. This also allows the teacher to get a sense of the class's level of prior knowledge.

Tertiary Coverage
Some topics in this statistics unit are not essential, but good to know. These include:

- The extent of the domains for which statistics can be useful
- Other terms that can be used to describe the concepts taught in this unit (e.g., center=central tendency, spread=variation)

It's great if students are familiar with these ideas, but this unit will not assess mastery of them.

What is the purpose(s) of your assessments? Which are formative vs. summative? Student vs. Program focused?
The assessments are meant to provide genuine insight into the students' level of understanding and mastery of the learning goals, both to the instructors and the students themselves.[5] This helps teachers (and the eLearning platform) modify instruction for different types of learners and encourages students to monitor their own progress. Additionally, it enables researchers to test the efficacy of the program by providing valid, reliable, and equitable metrics.

It's worth noting that the assessments described in the table below are not all distinctly student-focused or program-focused. Some of the activities both provide evidence to support improvement of the program on the next iteration and help improve students' learning on the current iteration.

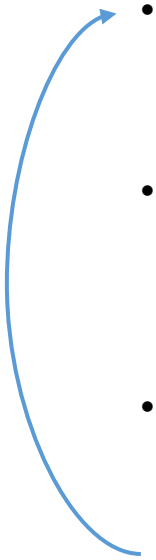
[5] Big Idea 4: Authentic assessments

Type	Description	Student Focus	Program Focus
Formative eLearning	"Learn by doing" problems that provide personally-tailored scaffolding and feedback, as well as prompts for self-explanation[6]	Helps students monitor their own learning and encourages use of metacognitive skills Enables students to catch errors early enough to avoid wasting time, but late enough for them to see the implications of their mistakes Provides spaced practice for knowledge and procedures	Analysis of log data ensures that students are interacting with the program in the intended ways (e.g., not abusing features that provide hints, spending too much/little time on instruction, etc.)
Formative project	Segmented milestones that ask the student to incorporate what they learned so far and reflect on previous sections[7]	Encourages students to continuously connect new information to prior knowledge Helps develop an appreciation for statistics by allowing them to apply it to a real-world context Gives students plenty of opportunities to reflect on and revise their work, promoting self-efficacy	
Summative eLearning	Pre and post quizzes (no scaffolding) for each topic; platform does not give feedback on questions other than correctness	Allow instructor to identify muddy points and provide more/alternative instruction to students as necessary Get students to think about how they might tackle problems before learning the ideal solution	Provides data that helps identify questions that are unclear or unreliable (e.g., are there certain items that all students are getting wrong?) Evaluate efficacy of instruction with learning curve analysis
Summative project	Final project submission that incorporates feedback and self-reflection	Gives students chance to reflect on entire unit and consolidate what they learned.	Test efficacy of instruction for primary goals by evaluating students' ability to justify decisions and tracking reflections throughout course

[6] Big Idea 3.2: Formative assessments that drive deep understanding
[7] Big Idea 3.4: Metacognitive skills for self-directed learners

What is the timing of your assessments? How frequently is each administered within and beyond the program you design?

The assessments are structured in a way that (i) allows students to grapple with new problems before getting instruction on the best solution[8] and (ii) gives instructors enough time to restructure lessons if necessary given performance on summative assessments. Here is what the assessment-instruction loop will look like:



- Before class
  - Instructor looks over eLearning summative assessments from homework and notes anything in particular that students struggled with
  - If students just submitted a project milestone, instructor assesses progress using a rubric
- In class (60 minutes)
  - Review of previous lessons, with more or less detail as dictated by summative assessments
  - Application of previous lessons to statistics project
  - Instruction on new topic
  - eLearning formative assessments on new topic
- Homework (15 minutes)
  - Go over formative assessments for that day
  - Summative assessments for that day, plus previous days
  - Quick introduction to next topic with a pre-quiz
  - Work on project if necessary

I assigned 15 minutes for homework based on a recommendation from TeacherVision, which states that a good estimate for the *total* minutes of homework is the student's grade level multiplied by 10 (in this case, 6 x 10 = 60).[9]

[POSSIBLY HERE: Of course different students will take more or less time; project acts as a sponge activity; encourage self-regulation to realize this; instructor can note time spent on eLearning, and balance that with expectations for project (help instructor give feedback and make recommendation to student)]

Overall Timeline

See "Detailed Schedule" in Part D to see exactly when the project milestones and eLearning assessments are due and how they fit in with the instructional material.

[8] Big Idea 3.1: Encourage grappling with ideas instead of memorization

[9] Big Idea 1.3: Engage students, but avoid cognitive overload

Specific Assessments

Assessment Type 1: Formative eLearning

Cognition

Targeted goals

- C.C.5.1 (definition of IQR)
- C.P.5.1 (calculating IQR)
- C.C.4.2 (definition of median)
- C.P.4.2 (calculating median)

Observation

The student sees a prompt and then a series of steps one at a time (steps remain on the screen once they are completed so they can be referenced).[10] The student must answer each step correctly to move on. If the student enters an incorrect response, the program gives relevant feedback based on the nature of the error. If appropriate, the student can also ask for hints (but the hint button does not appear until a response is given, as an indication that he or she at least made an attempt).[11] Student interactions with the eLearning platform (such as time spent on step, number of incorrect responses, number of hints requested, time between hint requests, etc.) are logged and available for interpretation by the instructor. As the student gains mastery of the relevant learning goals (see interpretation), the program displays fewer and fewer intermediate steps.[12]

Prompt

Here are the ages in years of some teachers at Kutz Elementary School: 36, 28, 42, 31, 68, 38, 41, 34, 45 Find the IQR of this dataset.

Step 1

Sort the data points from least to most _____ [check]

If correct

- Feedback: Great job!
- [Show next step]

If incorrect

- If sorted most to least
  - Feedback: Whoops! It looks like you sorted the numbers from most to least. Try sorting the numbers from **least to most.**
- Otherwise
  - Feedback: Not quite right! There's no need to rush. Are you asking yourself if your answer makes sense before submitting?

If incorrect again

- Let me help you out. The correct ordering is: 28, 31, 34, 36, 38, 41, 42, 45, 68

[10] Big Idea 1.2: Structure for organizing knowledge
[11] Big Idea 3.1: Encourage grappling with ideas instead of memorization
[12] Big Idea 3.1: Encourage grappling with ideas instead of memorization

Step 2

Find the median of the entire dataset _____ [check]

If correct

- Feedback: Great job!
- [Show next step]

If incorrect

- If response equals mean or mode
  - Feedback: Hmmm, that looks like the [mean|mode] of the dataset. Remember, the median is the middle value of a sorted dataset.
- Otherwise
  - Feedback: Not quite right! Try again. If you're stuck, you can ask for hints.
  - [Show hint button]

Hint 1

- Remember, the median is the middle value of a sorted dataset.

Hint 2

- You've already ordered the dataset from least to most. Now try crossing numbers off at each end until you reach the middle. Let me start this process for you.

  ~~28~~, ~~31~~, 34, 36, 38, 41, 42, ~~45~~, ~~68~~

  See if you can finish it up!

Hint 3

- ~~28~~, ~~31~~, ~~34~~, ~~36~~, 38, ~~41~~, ~~42~~, ~~45~~, ~~68~~

  After crossing off the numbers at each side, we see that the median of this dataset is 38

Step 3

Find Q1 _____ [check]

If correct

- Feedback: Great job!
- [Show next step]

If incorrect

- Feedback: Not quite right! Try again. If you're stuck, you can ask for hints.
- [Show hint button]

Hint 1

- Remember, Q1 is the median of the first half of the dataset

Hint 2

- The first half of the dataset is 28, 31, 34, 36. This has an even number of values. Do you remember how to find the median of an even number of values?

Hint 3

- The 2 middle numbers of the first half of the data are 31 and 34. To find the median, calculate the mean of these 2 numbers.

Hint 4

- The average of 31 and 34 is (31+34)/2 = 32.5, so Q1 is 32.5

Step 4

Find Q3 _____ [check]

If correct

- Feedback: Great job!
- [Show next step]

If incorrect

- Feedback: Not quite right! Try again. If you're stuck, you can ask for hints.
- [Show hint button]

Hint 1

- Remember, Q3 is the median of the first half of the dataset

[Hint sequence similar to Step 3]

Step 5

Calculate the IQR _____ years [check]

If correct

- Feedback: Great job!
- [Show next step]

If incorrect

- Feedback: Not quite right! Try again. If you're stuck, you can ask for hints. Are you asking yourself if your answer makes sense before submitting?
- [Show hint button]

Hint 1

- Remember, the IQR is the difference between Q3 and Q1

Hint 2

- Q3 - Q1 = 43.5 - 32.5 = 11

Interpretation

One criteria for evaluating performance is whether or not the student eventually answered individual steps correctly. If a student gets a step correct on the first try, the instructor has some evidence that she has a good grasp on the relevant learning goal. If a student answers incorrectly on the first try and then correctly on the second, we can infer that he didn't fully understand it at first, but used the feedback to alter his knowledge structure. Instructors can look at which incorrect answers were given to see what misconceptions a student might have. For example, a student who inputs the mean or mode in step 2 instead of the median might have trouble keeping terminology straight.

More attempts and hint requests may indicate persistence, but could also mean that the student was trying to game the system through abusing the hints feature. Other features, such as time spent on the problem and responses to self-explanation problems later in the module, can help instructors make a judgement call.

While the data regarding number of attempts, time spent on problem, and other features is useful for the teacher, I would not report this information to the students themselves. In this learning environment, I do not want students to think they are worse at math because they answered more slowly than their peers. Johnny should focus on his own improvement, not the fact that Billy is three times faster. That said, the instant feedback provides a tool for the students to monitor their own understanding of the problem.

Key Features

Criteria	Features
Internal validity	The steps in this assessment have direct links to at least one of the targeted goals. Steps 1-4 have to do with calculating the median. Steps 2-4 involve knowing the definition of the median. All steps are related to calculating the IQR and step 5 requires knowing the definition of the IQR. It may be hard to tell if a student went through the exercise on his own or got help from a parent. Clear discussions with parents about how the program is being used to help their children's learning could help with this issue (e.g., make sure parents know the student does not need a perfect score on the first try of formative assessments).
Ecological validity	When a student sees a formative assessment about the IQR for the first time, there is a lot of scaffolding and chances to request hints. This is not necessarily ecologically valid, as students probably won't have these steps explicitly laid out for them in the future. However, as the student becomes more and more comfortable with these topics (answers relevant questions quickly and accurately), these scaffolds can be gradually removed (e.g., showing fewer steps and offering fewer hints). The task becomes more ecologically valid as the student needs to complete more of the steps without guidance.
External validity	Familiarity with the concept presented in the question could be an issue. For example, people from other parts of the world might not know "Kutz Elementary School." This can be alleviated by substituting the problem context with one familiar to students.
Test-retest reliability	Problem presentation, scaffolding, and hint-giving are handled by the program, so those aspects should remain consistent across students. What may vary is the student's environment (completing the assessment at home vs in school vs in a library). Some places may be more distracting than others.
Inter-rater reliability	While the platform is designed to give consistent instantaneous feedback and track click logs, different instructors may interpret the students' data in different ways. The fact that goals are explicitly linked to individual steps should help with this issue. If necessary, training sessions for teachers and/or support from data scientists could help instructors interpret results.
Equity	There are no time constraints on these kind of formative assessments, so students who read or think at a slower pace are not put at a severe disadvantage. There is no punishment for needing more tries, and the scaffolding can be taken away at slower or faster rates depending on the student's individual progress.

ASSESSMENT TYPE 2: FORMATIVE PROJECT

Cognition

Targeted goals

- C.P.3 (make and justify choices about the method of summarizing a dataset)
- C.P.3.1 (identify graph that best helps answer statistical question)
- C.P.6 (construct relevant plots to display data)
- M.P.1 (self-monitoring of willingness to plan solution)
- C.D.2 (willingness to experiment)
- C.D.3 (value different forms of visualizations and measures)
- M.D.3 (adjust approach based on values)

Observation

This is an example of a project milestone that students will complete for homework after they have already received instruction on the various measures of center, spread, and the types of plots. Previously, they were asked to choose a measure of center to help answer their statistical question and justify that choice. At this point, the instructor has already provided feedback on their argument, so they can use that information in their current milestone.

Prompt 1

Now that you have collected some data, think about how you can represent it visually to best answer your statistical question. Remember, we learned about three kinds of graphs: dot plots, histograms, and box plots. Visualize your data using one of those three here:

Prompt 2

Which graph did you use and why?

Prompt 3

Did you use the teacher's feedback from the last project milestone on the current one? How?

Prompt 4

Which of the following sentences describe your graph-making process? Choose all that apply. a) I just made the first graph I could think of. b) I chose the graph that's easiest for me to make. c) I thought about which graph would help answer my statistical question using ideas we learned about in class. d) I made a rough sketch of each type of graph to help me choose which one is best for my question. e) Something else? _____

Interpretation

Instructors can use a rubric to guide scoring and help give feedback:

Relevant Goals	4	3	2	1
C.P.3 C.P.3.1	Argument in prompt 2 is well-supported using ideas from the class. Takes qualitative vs quantitative nature of data into account. Considers outliers and number of data points.	Argument in prompt 2 is somewhat supported using ideas from the class. May consider qualitative vs quantitative, but does not consider outliers or number of data points.	Attempt is made at an argument for prompt 2, but irrelevant ideas or are used to justify choice in prompt 1.	No attempt is made to make an argument for prompt 2.
C.P.6	Graph constructed in prompt 1 includes all key features, including title and axis labels.	Graph constructed in prompt 1 includes key features, except title and axis labels.	Graph constructed in prompt 1 missing some key features.	No or incomplete graph constructed in prompt 1.
M.P.1 M.D.3	Selects choices c and d in prompt 4 and/or writes a similar response for choice e.	Selects choice c or d in prompt 4 and/or writes a similar response for choice e.	Selects choice b (indicates some level of self-reflection) in prompt 1.	Selects choice a in prompt 1.
C.D.2 C.D.3 M.D.3	Incorporates instructor's feedback from previous milestones well, as indicated by argument in prompt 2 and self-explanation in prompt 3. Clearly values importance of using one method over the other.	Incorporates part of instructor's feedback into prompt 2 and demonstrates some awareness of own process in prompt 3. Sees some importance of using one method over the other.	Makes an attempt to incorporate instructor's feedback, but does so incorrectly/ Inappropriately. Sees some importance of using one method over the other.	Does not incorporate feedback or make an attempt to do so. Little to no evidence that student values using one method over the other.

After scoring, instructors can give this rubric directly to students to help them see where they could use improvement (though the goal numbers should be replaced with 6th-grader friendly descriptions). If necessary, the instructor could provide written feedback to students based on individual project needs. For example, say a student gives this response for prompt 2:

"My dataset is qualitative, so that rules out a box plot. I chose a histogram because it's easier to make than a dot plot."

The instructor could write:

"Good job considering whether your data is qualitative or quantitative! Instead of just using a dot plot because it's easier, try to think about the size of your dataset. Remember, histograms are good for datasets with at least 20 values, while dot plots are appropriate for smaller datasets."

On the next milestone, the instructor can again use a rubric to see how well the student incorporates the feedback.

Key Features

Criteria	Features
Internal validity	Relevant goals are linked to the scoring guidelines in the rubric. Though we'd want students to complete this assessment on their own, they may receive a lot of help from parents. If a teacher suspects that this is the case, he could compare the quality of work with directly-observed, in-class assignments or listen in during discussions with peers.
Ecological validity	This kind of assessment is closer to a real-world scenario than the formative eLearning example is. Before the instructor reviews the responses, the student is not given personalized scaffolding based on her individual project. The student must determine how the concepts in the course apply to her dataset on her own. In future statistics courses, she may not be explicitly asked to reflect on her own process, but she will likely do so if she sees the value in self-reflection.
External validity	Because students choose their own statistical question and collect data themselves, this assessment will likely generalize well to other populations. Students will likely be familiar with their own topic at this point in the unit.
Test-retest reliability	This is tricky to evaluate because the project formative assessments double as instructional events. If a student completes the same one again later, he will likely do better because he had a chance to review the feedback and revise his work. Administration could use classroom observations to make sure that certain teachers don't give *very* personalized instructions about choosing graphs (based on individual student datasets) before assigning this homework.
Inter-rater reliability	I suspect that the scoring guidelines in the rubric are clear enough that different raters would give the same ratings for the same set of responses. If two classes are running at the same time (with different instructors), the instructors could grade students from the other class in addition to their own. We could compare the scores to make sure one isn't being too strict or lenient (see part E).
Equity	Slower learners are not at a severe disadvantage because they will have the chance to reflect on and revise their work in future iterations of the project. Instructors can tailor their feedback based on their students' learning styles. If certain students have trouble expressing their thoughts in writing, teachers can listen to them during class discussions about this milestone and weigh that into their evaluations.

ASSESSMENT TYPE 3: SUMMATIVE ELEARNING

Cognition

Targeted goals

- C.C.4.1 (definition of mean)
- C.P.4.1 (calculating mean)

Observation

After the eLearning formative assessments, students take a summative quiz including questions about the current topic and possibly previous topics. The quiz does not provide instant feedback, but shows them which questions they got correct/incorrect after submitting all questions. Here are two example questions for the targeted goal.

Prompt 1

Here are the sizes of households for some houses in Doylestown: 4, 3, 6, 3, 4 What is the mean household size for this dataset? [After submission: correct answer is 4]

Prompt 2

Jane ran an average of 3 miles per day over a period of 5 days. How many total miles did she run in those 5 days? [After submission: correct answer is 15]

Interpretation

Because these summative assessments are due at a reasonable hour before the next class, the instructor has a chance to identify muddy points in the students' understanding of the mean. For example, if most people answer prompt 1 correctly, but struggle with prompt 2, students may have the formula memorized, but have a hard time grappling with its real meaning (e.g., working backwards from the mean to determine the sum of the original dataset). If certain students consistently answer questions about the current topic correctly, but struggle with questions about earlier topic, that may be a sign of bulimic learning. If students respond with the median or mode instead of the mean, they may have trouble keeping terminology straight. With this information, the instructor can add more or less review time to the next class as necessary.

Key Features

Criteria	Features
Internal validity	The questions in the quizzes are linked to specific goals. It may be hard to tell if a student went through the exercise on his own or got help from a parent. Clear discussions with parents about how the program is being used to help their children's learning could help with this issue (e.g., make sure parents know the summative assessments are used to help the instructor identify muddy points rather than rank their children's abilities).
Ecological validity	Though these assessment questions are closer to real-world problems than the formative eLearning assessments are (due to less scaffolding), students who actually need to calculate the mean in the future will almost certainly use software. Still, knowing how the calculation works can support reaching primary goals involving making choices about measures of center.
External validity	These types of assessments are relatively narrow and straightforward, so they likely generalize to other populations. General context surrounding the questions (such as city names and references to sports teams) can be adapted based on location.
Test-retest reliability	Problem presentation is handled by the program, so it should remain consistent across students. What may vary is the student's environment (completing the assessment at home vs in school vs in a library). Some places may be more distracting than others.
Inter-rater reliability	Different instructors may have different opinions about what the muddiest points are for certain students/classes. If necessary, teachers could have discussions with coaches and/or data scientists to help interpret the results.
Equity	There are no time constraints on the summative quizzes, so slower readers/thinkers will not be at a severe disadvantage. The quizzes are not very long, so those with shorter attention spans should still be able to complete them.

Assessment Type 4: Summative Project

Cognition

Targeted goals

- C.D.3 (value different forms of visualizations and measures)
- M.D.2 (reflect on own beliefs about what it takes to learn statistics/math in general)

Observation

The final project submission includes a self-reflection component that particularly target dispositions. Here is a sample of questions.

Prompt 1

In this unit, we learned that there are often several ways to represent similar ideas in statistics. For example, mean, median, and mode are all measures of center. Mean absolute deviation and inter-quartile range are both measures of spread. Visually, dot plots, histograms, and box plots can all represent the same dataset. Do you think it's important to know all of these methods? Why?

Prompt 2

Imagine you are in a statistics class next year and the teacher says you are going to learn about "standard deviation." How does this make you feel? Do you think you would become more comfortable with the idea of "standard deviation" as time went on? Why or why not?

Interpretation

Relevant Goals	4	3	2	1
C.D.3	Student clearly values the ability to represent data in multiple visual ways, use various measures of center to indicate typical value, and calculate different measures of spread. Aware that using one over the other could mislead other readers.	Student may see value in different forms of visualizations or measures of center/spread for personal use, but does not reflect on how other readers may be misled by certain choices.	Student values learning one way of visualizing the data, one way of calculating center, and one method for spread, but sees no point in learning more than one of each.	Student sees no point in learning any methods for summarizing datasets visually or numerically.
M.D.3	Clearly understands that he or she will probably not understand a brand new concept right away. Knows that he or she can improve with deliberate practice and reflective thinking.	Student may experience some frustration when learning about a new topic, but ultimately knows he or she can improve with practice and some self-reflection.	Student experiences a lot of frustration when introduced to a new topic, and engages in little self-reflection.	Gets frustrated when he or she does not understand something right away and sees no reason to try to improve.

In addition to using this rubric to evaluate the students' final reflections, the instructor can look at reflections from previous milestones to see how their dispositions and metacognitive abilities improved over time. These scores can be reported to administrators and researchers to help determine the overall efficacy of the instruction for primary goals.

Key Features

Criteria	Features
Internal validity	The prompts and scoring guides in the rubric are linked to relevant learning goals. The instructor likely has a good sense of the students' writing/reflecting ability by now, so she can probably tell if they did their own work. It is still possible they received help from a parent, though.
Ecological validity	Students may not be explicitly asked about their dispositions regarding statistics in the future, but they will need to deal with them nonetheless. This assessment probably provides at least a decent prediction about how they will approach statistics outside of this unit.
External validity	This assessment uses fairly simple, standard language for 6th grade statistics. Other typical populations at this level should be able understand what's being asked and craft a response. The questions can likely be translated to other languages if necessary.
Test-retest reliability	Certain teachers may bias the results by providing their students with heavily-scaffolded templates. Ensuring that instructors are clear on the purpose of the assessment could alleviate this potential issue.
Inter-rater reliability	I suspect that the scoring guidelines in the rubric are clear enough that different raters would give the same ratings for the same set of responses. If two classes are running at the same time (with different instructors), the instructors could grade students from the other class in addition to their own. We could compare the scores to make sure one isn't being too strict or lenient (see part E).
Equity	Because there is not strict time limit for these questions, slower thinkers/readers are not at a severe disadvantage. If certain students have trouble expressing their thoughts in writing, teachers can opt for a different form of evaluation, such as an in-class discussion.

Part D: Instructional Design

General Description

What are the key features of the learning environment you envision?

In this learning environment, students will recognize that it is safe to fail. The established norms will encourage experimentation and trying to understand the implications of mistakes. [13] These attributes are essential to help students construct knew knowledge about how the world works through statistics.[14] A major part of the instruction is having students grapple with problems before being provided with the ideal solution.[15] If students are uncomfortable with this uncertainty, they will likely experience frustration. The lessons are tied with what they learned on previous days in order to keep everything in context and fix misconceptions as they arise.[16]

Though it's not the most important aspect of the unit, the students will be learning a good amount of new terminology. The instruction will not assume that the students will be able to keep all of the vocabulary straight after presenting it one time. It will keep referring to terms and prompting for definitions as a spacing technique.[17] Posters in the classroom could also support memory.

In order to benefit different types of learners, students will also be provided with multiple forms of instructional activities.[18] For example, group activities give them a chance to discuss ideas with each other and clear up possible misconceptions. For those who are shy and/or cannot keep up with the pace of these discussions, eLearning instruction allows them to grapple with the concepts at their own pace.

Finally, the instruction is designed to be adaptable. Students may have more trouble with a given concept than anticipated, and thus require more instruction before moving on. The lesson plans are set up in a way that allows the instructor to go over the students' responses to the homework before the next class starts. This way, he or she can add more or less time to the review sessions as needed.[19]

What overall routines will you establish to guide the flow of each session, each week, etc.?

Each day will have a clear structure, which follows the assessment-instruction loop described in Part C. The first day of class will not have a pre-assessment and review session. Instead, we will kick things off with the "promises" of statistics, as recommended by Bain and described in part A.[20] The first day is also a good chance to introduce students to the eLearning platform we will use for in-class assessments and homework. Going over the nature of the pre-assessments provides a nice opportunity to establish the norm of encouraging experimentation and feeling safe to fail.

---

[13] Big Idea 2.3: Create a culture where it's safe to fail
[14] Big Idea 1: Constructing models of reality
[15] Big Idea 3.1: Encourage grappling with ideas instead of memorization
[16] Big Idea 1.1: Connecting prior knowledge
[17] Clark & Mayer, 2016; Ch13, Principle 4: Distribute and mix practice among learning events
[18] Big Idea 2.2: Take students' context into account when designing instruction and assessments
[19] Big Idea 4: Authentic assessments (see "Why it's important")
[20] Big Idea 1.1: Connecting prior knowledge

What range of teaching approaches are you planning to use?

Personalized eLearning

This is particularly useful for dealing with individual differences in speed and prior knowledge. Those who have trouble can receive more feedback and practice, while those who are faster learners can go on to more challenging problems.

Lectures

These will mainly be used to review the previous lesson's homework assignments, discuss the students' ideas about the new problems from the pre-assessments, and go over the new lessons.

Group Activities

These in-class activities give the students a chance to work together on problems, discuss their project milestones with peers, and clarify misconceptions. It also gives the instructor a chance to informally observe students' learning. Discussions will be guided by handouts or prompts projected on the wall.

Overview of Curriculum

This curriculum will introduce the concept of statistical questions and what it takes to answer them. We will go over visual and numerical ways of summarizing datasets and discuss the advantages and disadvantages of each one. This course involves the use of an eLearning platform for in-class "sponge" activities, as well as for homework. During the homework, students will be asked to think about problems that may be unfamiliar. These topics will be discussed in the next day's lecture. Students will be asked to complete the homework before the start of school the next day, so that the instructor has time to refine the review session. Furthermore, students will complete a unit-long statistics project in which they will continuously incorporate new knowledge. The project is segmented into milestones and the instructor will provide personalized feedback on each one, giving the students multiple opportunities to revise and reflect on their work.

The instructor may add more or less time to each day's activity based on performance on assessments, but in general the timing of each day will be as follows:

- Lecture
    - Review of previous lessons: 15
- Group activity
    - Application of previous lessons to project: 10
- Lecture
    - Discussion of pre-assessment: 5
    - Lecture on new topic: 15
- In-class eLearning: 10
- Homework (~15 minutes per day)

The extra 5 minutes of in-class time acts as a buffer and can be distributed to any activity and/or transitions between activities.[21]

[21] Big Idea 4.2: Keep assessments practical by considering context

Detailed Schedule

Day 1

- Main topic: Statistical questions
- Lecture
    - Establishing environment: promises of statistics [C.D.3]
    - Segway to statistical questions [C.C.1, C.P.1]
- Intro to the learning platform
    - Describe mechanics, set up logins, clear up any technical issues
- In-class eLearning
    - Formative assessments on statistical questions
- Establish norms
    - Tonight you will have HW on what you learned today, but also on what you will learn tomorrow. I'm not concerned if you get these questions right, I just want to see that you are thinking about them and are trying. [C.D.2, M.D.1, M.D.2]
- Homework
    - Formative assessments on statistical questions
    - Summative assessments on statistical questions
    - Pre-assessment on forms that statistical data can take

Day 2

- Main topic: Datasets in context/Forms that statistical data can take
- Review Lecture
    - Review of statistical questions
- Group Activity
    - Discuss in groups the kinds of statistical questions you personally want to answer
    - Get class together, ask some people to share and others to give feedback
- New Topic Lecture
    - Discussion of pre-assessment on forms of data
    - Lecture on datasets in context [C.C.4, C.P.2]
- In-class eLearning
    - Formative assessments on datasets in context
- Homework
    - Formative eLearning assessments on datasets in context
    - Summative eLearning assessments on datasets in context and statistical questions
    - eLearning pre-assessment on dot plots (e.g., how might we visualize this dataset?)
    - Project milestone: Define a statistical question you want to answer, and that you can practically collect data for

Day 3

- Main topic: Dot Plots
- Review Lecture
    - Review of datasets in context
- Group Activity
    - Discuss in groups what kind of data you might collect for your project.
- Lecture
    - Discussion of pre-assessment on dot plots
    - Lecture on dot plots [C.C.6.1, C.P.6.1]

- In-class eLearning
    - Formative assessments on dot plots
- Homework
    - Formative eLearning assessments on dot plots
    - Summative eLearning assessments on dot plots and previous lessons as needed
    - eLearning pre-assessment on histograms (e.g., how to visualize if you have a lot of data points?)
    - Project milestone: Reflect on feedback about statistical question. Identify what kinds of data you will need to collect for your project (qualitative vs quantitative, etc.)

Day 4

- Main topic: Histograms
- Review Lecture
    - Review of dot plots
- Group Activity
    - Given worksheet with example dataset, work together to make dot plot
- Lecture
    - Discussion of pre-assessment on histograms [C.C.7, C.P.3.1, C.D.3]
    - Lecture on histograms [C.C.6.2, C.P.6.2]
- In-class eLearning
    - Formative assessments on histograms
- Homework
    - Formative eLearning assessments on histograms
    - Summative eLearning assessments on histograms and previous lessons as needed
    - eLearning pre-assessment on center and mean (e.g., how can we describe the typical number?)
    - Project milestone: Reflect on feedback about previous milestones. Start a plan for collecting real-world data.

Day 5

- Main topic: Center and Mean
- Review Lecture
    - Review of histograms
- Group Activity
    - Given worksheet with example dataset, work together to make histogram
    - Discuss similarities and differences with dot plots; possible advantages of one over the other
- Lecture
    - Discussion of pre-assessment on center and mean
    - Lecture on center and mean [C.C.3.1, C.C.4.1, C.P.4.1]
- In-class eLearning
    - Formative assessments on mean
- Homework
    - Formative eLearning assessments on mean
    - Summative eLearning assessments on mean and previous lessons as needed

    - eLearning pre-assessment on median and mode (e.g., Consider this dataset (which has an outlier)? Does the mean really represent the typical value here? Why/why not? What would be a better way?)
    - Project milestone: Reflect on feedback about previous milestones (Is your plan for collecting data practical?). Start collecting a few data points.

Day 6

- Main topic: Median, Mode
- Review Lecture
    - Review of mean
- Group Activity
    - Given worksheet with example dataset, work together to calculate mean
    - Discuss expected mean of data from project based on what you have so far
- Lecture
    - Discussion of pre-assessment on median/mode [C.P.3.2, C.D.3]
    - Lecture on median/mode [C.C.4.2, C.C.4.3, C.P.4.2, C.P.4.3]
- In-class eLearning
    - Formative assessments on median/mode
- Homework
    - Formative eLearning assessments on median/mode
    - Summative eLearning assessments on median/mode and previous lessons as needed
    - eLearning pre-assessment on MAD (e.g., what about spread?)
    - Project milestone: Reflect on previous milestones. Continue collecting data.

Day 7

- Main topic: Spread and MAD
- Review Lecture
    - Review of measures of center
- Groups activity
    - Discuss which measures of center would be the most useful for individual projects
- Lecture
    - Discussion of pre-assessment on MAD
    - Lecture on concept of spread and MAD [C.C.3.2, C.C.5.2, C.P.5.2]
- In-class eLearning
    - Formative assessments on MAD
- Homework
    - Formative eLearning assessments on MAD
    - Summative eLearning assessments on MAD and previous lessons as needed
    - eLearning pre-assessment on IQR (e.g., other ways of measuring spread (think about median)?)
    - Project milestone: Reflect on previous milestones. Make and justify choice about which measure of center to use for your statistical question and dataset.

Day 8

- Main topic: IQR, Box Plots
- Review Lecture
    - Review of MAD

- Group Activity
    - Given worksheet with example dataset, work together to calculate MAD
    - Discuss expected MAD of data from project based on what you have so far
- Lecture
    - Discussion of pre-assessment on IQR and Box Plots [C.C.7, C.P.3, C.D.3]
    - Lecture on IQR and Box Plots [C.C.5.1, C.C.6.3, C.P.5.1, C.P.6.3]
- In-class eLearning
    - Formative assessments on IQR and Box Plots
- Homework
    - Formative eLearning assessments on IQR and Box Plots
    - Summative eLearning assessments on IQR and Box Plots and previous lessons as needed
    - eLearning pre-assessment on advantages/disadvantages of all the different methods of visualization and measuring center/spread
        - This is not an entirely new concept, as students have been grappling with this through the pre-assessments and in-class discussions
    - Project milestone: Reflect on previous milestones. Make and justify choice about which graph to use to summarize dataset.

Day 9

- Main topic: How to choose which method to use
- Review Lecture
    - Review of IQR and Box plots
- Group activity
    - Review each other's graphs from the last project milestone
    - Discuss which measure of center and spread would work well for each other's projects
- Lecture
    - Discussion of pre-assessment on how to choose a method
    - Lecture on choosing methods (point out that they already have a lot of experience with this through grappling with the pre-assessments and questions in the formative assessments) [C.C.7, C.P.3]
    - Tying back to day 1 to show how much students know now [C.D.1. C.D.3, M.D.2]
- In-class eLearning
    - Formative assessments on choosing the right method
- Homework
    - Formative eLearning assessments on choosing the right method
    - Summative eLearning assessments on all topics
    - Project milestone: Reflect on previous milestones. Make and justify choice about which measure of spread to use for dataset.

Day 10

- Lecture
    - Reflection on the course and how students might use statistics in the future [M.D.2, C.D.3]
    - Demonstrate how class improved by tracking quiz scores [C.D.1]
- Group activity
    - In class workshop; receive feedback on milestones from peer and instructor
- In-class project

    - Start working on final project milestone
- Homework
    - Final project milestone: Reflect on previous milestones. Self-reflection on project and course

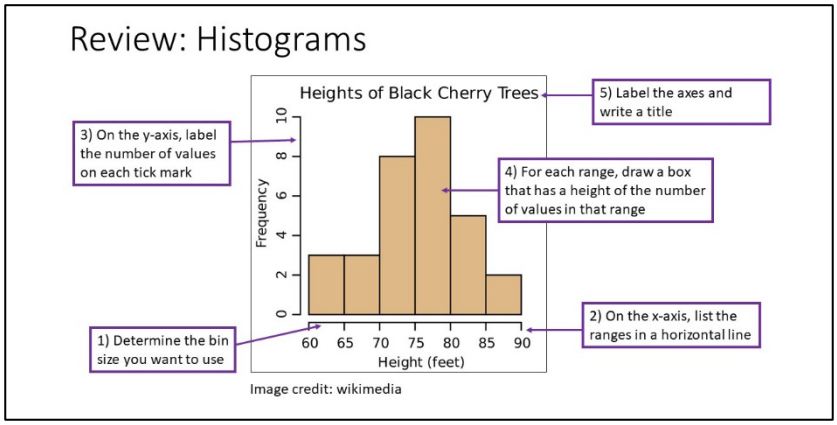
Specific Activities

Instruction Type 1: Review Lecture

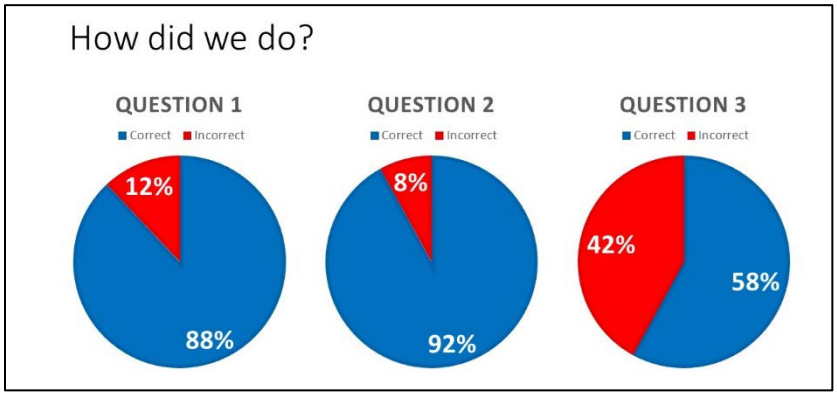
Context

- Focus: Histograms [C.C.6.2, C.P.6.2]
- Given on day 5
    - Students already had previous lecture introducing this topic
    - Students have already completed relevant formative and summative eLearning assessments
- Type of content: facts, basic explanations, procedures
- Key approach: didactic

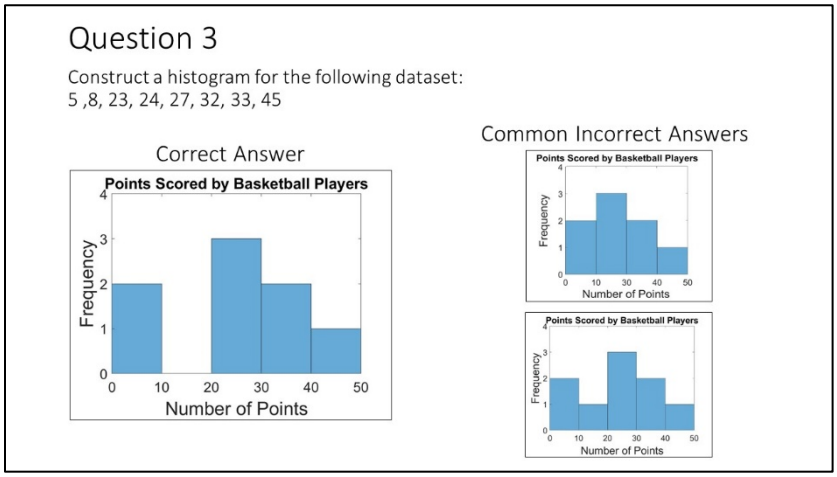
Representations, Guidance, and Interactions



Instructor Notes
Remind students of the definition of a histogram and give a brief overview of the procedure. Steps appear on screen one at a time.



Instructor Notes
Offer encouragement about doing well on questions 1 and 2 for histogram quiz, but don't worry if you slipped up; there will be other chances to try again. Note that a lot of us had trouble with question 3, so we should take a closer look.



Instructor Notes
Discuss muddy point: constructing histograms when there are 0 values in a certain range. Ask students if they can point out difference between correct answer and incorrect answers.

Accommodations for different learners
People who do not learn well from lectures will have chances to discuss this muddy point during the group activity and do practice problems in future eLearning formative assessments.

INSTRUCTION TYPE 2: GROUP ACTIVITY

Context

- Focus: Histograms [C.C.6.2, C.P.6.2], Making & justifying choices about visualizations [C.P.3.1]
- Given on day 5
    - Students already had lectures and eLearning formative assessments on histograms
- Type of content: examining and extending meaning, choosing and applying strategies
- Key approach: constructivist

Directions and Products
Have students get together in groups of 3 and go through this worksheet:

Here are the number of minutes Jess spent practicing ballet each day last week:
22, 35, 15, 8, 0, 38, 35
Using a bin width of 10, create a histogram representing this dataset.

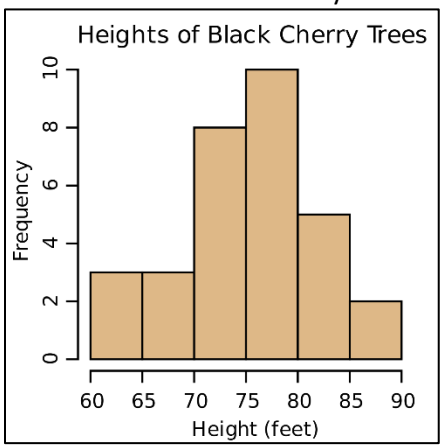
The following dataset lists the lifespan in years of some species of eagles:
14, 22, 21, 14, 9, 8, 21
Create a histogram to represent this dataset using a bin width of 5

Discuss the following questions with your group:

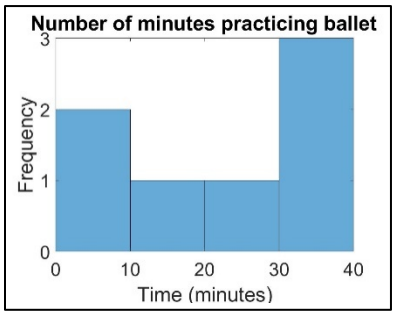
- What are some similarities and differences between histograms and dot plots?
- When do you think it would be better to use a dot plot versus a histogram?
- Which type of plot would you use for the following datasets? Why?
  Dataset 1: 8, 8, 8, 9, 10, 15, 15, 20
  Dataset 2: 14, 11, 53, 63, 24, 44, 47, 75, 83, 91, 52, 89, 98, 81, 3, 13, 11, 42, 56, 6, 12, 32, 47, 46, 35, 55, 92, 88, 94, 47, 51, 91, 77, 78, 31, 91, 13, 4, 38, 68, 55, 33, 59, 43, 23, 74
  Dataset 3: blue, blue, brown, black, green, green, yellow

Guidance, Resources, Representations, and Interactions

As the students are going through the worksheet, project an example histogram on the screen so that students can make sure they have the key elements:

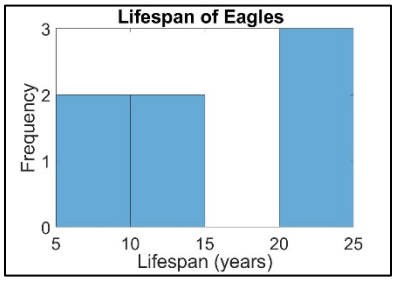


Additionally, the instructor can walk around the room, listen in for points of confusion, and take questions. He or she can steer conversations back on track if necessary. After most groups have finished up, bring the class back together and show solutions on slides. Ask students to self-check that they have all of the key elements and cleared up the muddy points.



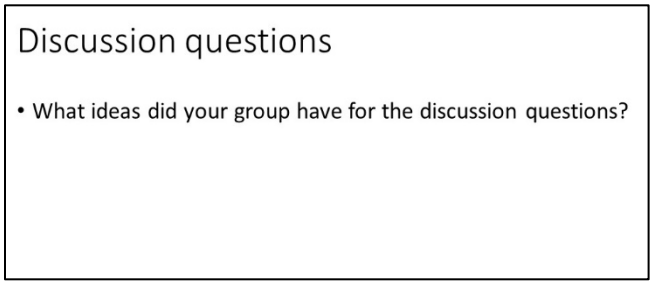
Instructor Notes

Ask students to self-check their graphs with this one. Did anyone have trouble counting 0 in the 0-10 range?



Instructor Notes

Ask students to self-check their graphs with this one. Did everyone handle the fact that there were no values in the 15-20 range correctly?



Instructor Notes

Ask students about their thoughts on the discussion questions. Add their ideas to the slide or write them on the board as you go.

Accommodations for different learners

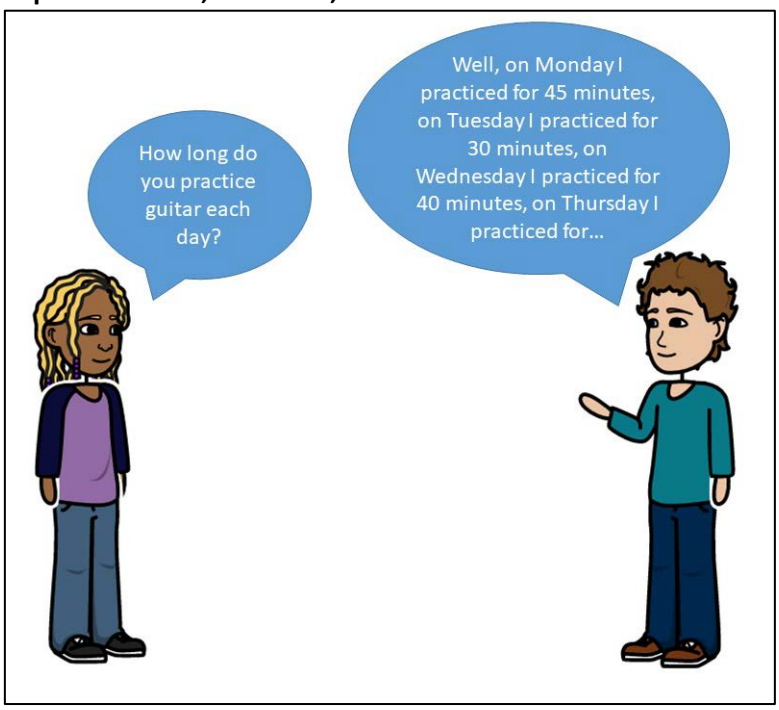
If one group finishes particularly quickly, the instructor can ask them more in-depth questions about the distinction between dot plots and histograms. Extroverts are probably more likely to offer ideas in the full-class discussion, but introverts will have a chance to speak in the small groups. Those who are really shy or can't keep up with the pace of the discussion will have more practice during the eLearning assessments (and have already had previous lectures as well).

Instruction Type 3: Lecture (New Topic)

Context

- Focus: Center and mean [C.C.3.1, C.C.4.1, C.P.4.1]
- Given on day 5
    - Students have already completed eLearning pre-assessment on topic, but never received explicit instruction
- Type of content: facts, basic explanations, procedures
- Key approach: didactic

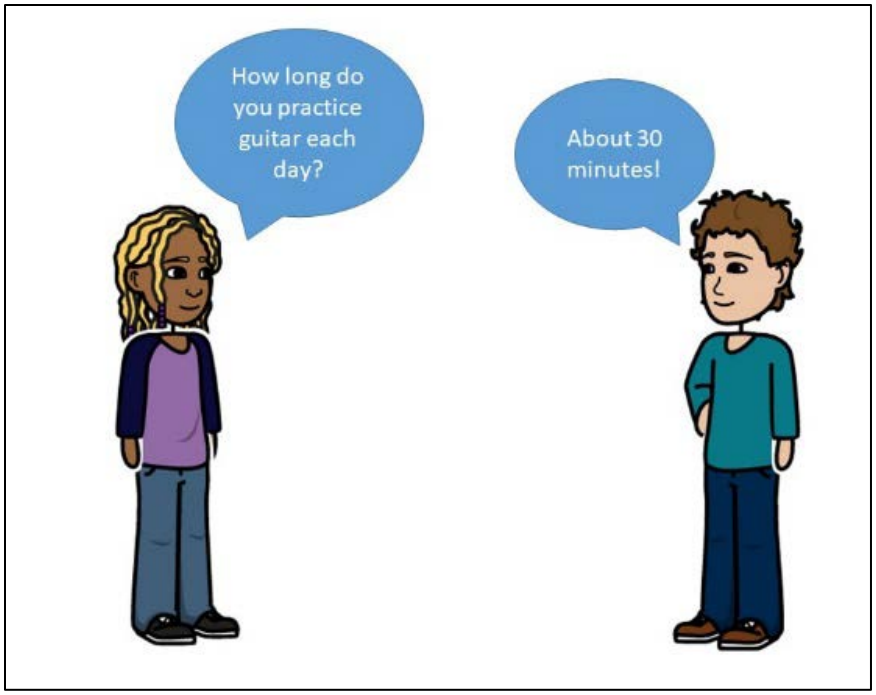
Representations, Guidance, and Interactions



Instructor Notes

[Hook students with need for measure of central tendency and how students use it naturally in their everyday lives.]

When someone asks you a question like this do you respond with this kind of detail? No!



Instructor Notes
You usually give a single number that represents the typical value. What does this have to do with statistics? If we have a dataset, we usually want to know the typical value of it.

Center

- Definition: A single number that indicates the typical value of a dataset
- Example: The typical age of 6th graders at Kutz is 11.5 years.

Instructor Notes
We call this center—a single number that indicates the value of a dataset. For example, the typical age of 6th graders at this school is 11.5 years old.

How do you think we can calculate center?

- "Use the most common number"
- "I would tell them any number, but not the smallest or largest."
- [insert some other student responses to the pre-assessment here]

Instructor Notes
Last night, for the eLearning homework, you gave some ideas about how we could calculate center.

Mean

- Step 1: Add all of the values in the dataset
- Step 2: Divide the total by the number of data points in the dataset

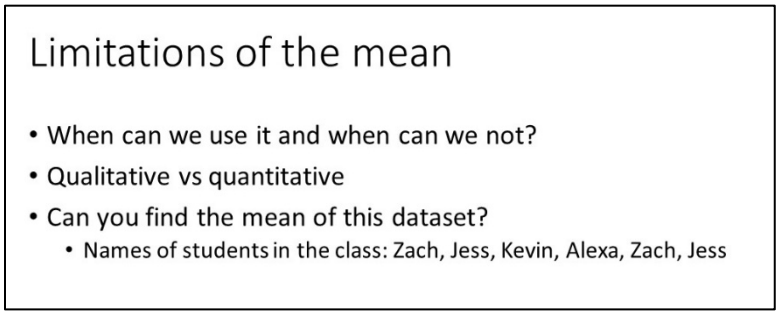
Example

Find the mean of: 2, 5, 3, 2 ← 4 data points

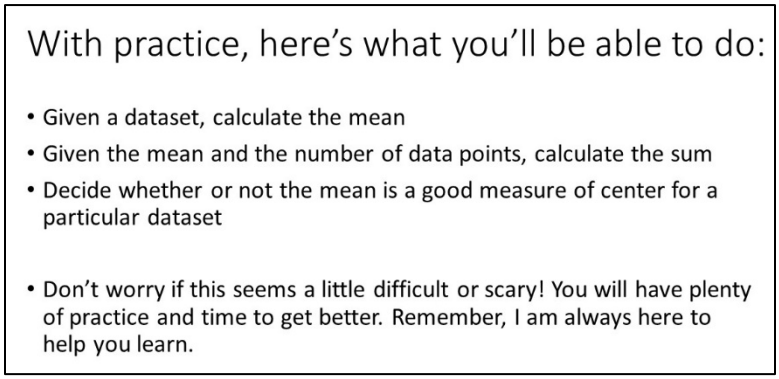
- Step 1
  - 2 + 5 + 3 + 2 = 12
- Step 2
  - 12 ÷ 4 = 3

Instructor Notes
These are all reasonable choices. In statistics, we have a few ways of calculating center. One is called the "mean" or "average."

[Bullets and boxes appear one at a time. Walk through definition and example]



Instructor Notes
[Bullets appear one at a time. Ask students for ideas before providing answer.]



Instructor Notes
[Go over expectations of what students should eventually be able to do. Be reassuring, not intimidating.]

Go over a few more examples on the board.

Accommodations for different learners
Students who do not learn well from lectures will still be able to view lessons and do formative assessments on the eLearning module. There will also be group discussions in the next class.

INSTRUCTION TYPE 4: ELEARNING LESSONS

Context

- Focus: IQR [C.C.5.1, C.P.5.1]
- Given on day 8
    - Students already had previous lecture introducing this topic
- Type of content: refining and understanding skills
- Key approach: coaching

Directions, Guidance, Resources, and Accommodation
Students go through formative assessment questions about the IQR. These provide scaffolding and instant feedback automatically tailored to the students' needs. See Assessment Type 1: Formative eLearning in Part C for an example. While in class, students can ask instructor and peers for help. As they go through the exercises, the students can refer to videos and worksheets similar to those on Khan Academy

- https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/cc-6th/v/calculating-interquartile-range-iqr
- https://www.khanacademy.org/math/probability/data-distributions-a1/summarizing-spread-distributions/a/range-and-interquartile-range-worksheet

Part E: Evaluation Research Design

Research to Evaluate Educational Implementation

Fidelity check of teacher following the proposed instruction and assessment design

Teacher Fidelity Check 1: Inter-Rater Reliability of Rubric

Description

The efficacy of this instructional intervention depends a great deal on the teacher's assessment of the statistics project, both for the intermediate steps and the final product. Thus, it's a good idea to ensure the instructor uses the rubric consistently and reliably. If two courses are running at the same time (with different instructors, of course), we could test the inter-rater reliability by having each instructor additionally assess student projects from the other class (resulting in 2 teacher assessments per project).

Data to be collected

- Scores for statistics project (both for the milestones and the final submission) from 2 instructors, each teaching a different class. Each teacher rates their own students' work, as well as the students from another class. If this is too time-intensive, a random sample of milestones could be chosen to have 2 raters.

Forms to be used

- Milestone rubrics
- Final submission rubric

Scoring plan

- Correlate the instructors' rubric ratings for all assessments with 2 instructor ratings.

Teacher Fidelity Check 2: Adapting Instruction for Muddy Points

Description

The instruction involves the teacher using the results of the summative eLearning to identify muddy points and alter the instruction based on those findings. We should check to make sure instructors are adapting the lesson plans accordingly.

Data to be collected

- Student scores on eLearning summative assessments
- Teacher notes/diary/logs that indicate
    - Identification of muddy points based on each eLearning summative assessment
    - Notes on how the review session was adapted to target the muddy points
- Adapted slides or other instructional material

Forms to be used

- For each day with a review lecture, instructors fill out a form with the following fields
    - Topic to review:
    - Muddiest point(s):
    - Evidence for student difficulty:
    - Adaptations made to review lecture based on muddiest point:

Scoring plan

- For each review lesson, teachers will receive a 1 or a 0 in the following categories
    - Identified muddy point
    - Provide sufficient evidence for (lack of) difficulty in this area
    - Made appropriate adaptations to relevant instructional material

Teacher Fidelity Check 3: Running Group Activities

Description

Classroom observations by administrators or researchers can help ensure the teachers is handling the group activities appropriately.

Data to be collected

- For 3-5 class days, multiple researchers (exact number depends on class size; ~1 per 10 students) will observe classroom dynamics and take notes based on what they see.

Forms to be used

In addition to any comments or suggestions they see for improving administration of group activities, observers will use a rubric to measure:

- Instructor engagement
    - 3: Instructor actively intervenes with distracting groups, answers questions to the best of his or her ability, and gives further prompts to groups who finish early
    - 2: Instructor handles distracting groups mostly well and answers some student questions
    - 1: Instructor has little involvement in group activity after giving directions to students.
- Student focus
    - 3: Most groups on topic for the majority of the session
    - 2: Some distractions and difficulty staying on task
    - 1: Discussions mostly off-topic

Scoring plan

Sum of points in each scoring category can provide overall measure of classroom engagement. Researcher notes can help provide specific recommendations for how to improve students' focus during in class activities.

Fidelity check of students participating in instruction and assessment as anticipated

Student Fidelity Check 1: Use of Hints Features in eLearning

Description

During the formative eLearning assessments, students have the option of requesting hints when they are stuck on certain steps. These hints eventually bottom-out to providing the answer. Because of this, students may potentially keep requesting hints without spending any time trying to understand how the solutions are derived.

Data to be collected

- Student click logs can track how many hints the student requested and the timestamps for each request

Forms to be used

- A form will be auto-populated by the eLearning platform that has the following columns
    - Student ID: unique string of anonymous student ID
    - Problem ID: unique number ID ; can be linked to document containing full problem
    - Problem Name: description of problem (e.g., Calculate IQR)
    - Problem Start Timestamp: time student first saw problem
    - Targeted Goals: comma-separated list of relevant goals from Part B
    - Step Name: description of intermediate step (e.g., Calculate Q1)
    - Total Possible Hints: number of hints available for this step
    - Hint Number: integer of hint number in hint sequence (1 row for each)
    - Hint Requested: 1 if hint number requested, 0 otherwise
    - Hint Timestamp: time hint number was requested if applicable

Scoring plan

- Develop a gaming-behavior model similar to that described in “Off-task behavior in the cognitive tutor classroom: when students "game the system"” (Baker, Corbett, Koedinger, & Wagner, 2004).
- Correlate likelihood of gaming the system with improvement on eLearning assessments; can also explore relationships between gaming the system and performance on statistics project

STUDENT FIDELITY CHECK 2: ENGAGEMENT IN REVISING AND REFLECTING

Description

As the unit goes on, students are expected to make revisions to their project milestones based on instructor feedback. While the teacher rates how well the students incorporate feedback on a rubric, the students might not have had experience with this form of instruction in the past. We should ensure they are engaging in this process at the anticipated level before beginning another iteration of the unit.

Data to be collected

- Teacher rubric ratings that are specifically related to incorporating feedback
- Samplings of student milestones

Forms to be used

- Rubrics similar to those in the project assessment sections of Part C

Scoring plan

- For students who received below the highest rubric score on the first iteration of the milestone, calculate the percentage of them who made a revision that resulted in an improved score.

Research to Evaluate Educational Impact

Research Question

Do motivational prompts in statistics eLearning modules lead to higher improvements of self-efficacy as measured by rubrics for project milestones? Including messages such as “If you try hard, you can learn to solve even the most difficult problems” in instruction and assessments seems like a good idea based on intuition. We can all use encouragement during more challenging eLearning problems. On the other hand, these messages take time to read and could be seen as a nuisance. Is it worth it to include them?

Experimental Design

Experimental Group

During the eLearning modules, a popup will display a motivational prompt focused on self-efficacy before the formative exercises and before the summative quizzes. Students will need to click "OK" to close the window and continue with the lesson.

- Example motivational prompt: "Believe in yourself! Know that with practice you can learn to do this."

Control Group

During the eLearning modules, a popup will display a short message about their progress before the formative exercises and before the summative quizzes. Students will need to click "OK" to close the window and continue with the lesson.

- Example message: "We will now begin the formative assessments"

Independent Variable

- Presence of motivational prompts before start of formative and summative assessments in eLearning module

Dependent Variable

- Improvement in rubric score for self-efficacy measures based on project milestones

Covariates

- Scores on eLearning summative assessments
- Rubric scores for formative project assessments (other than self-efficacy measures)
- Gender
- Grades in previous math classes
- Whether or not students spend time reading prompts (indicated by amount of time pop up remains on screen)

Method

Subjects

- 6^{th} grade students in statistics unit

Procedure

- Stratified random assignment to experimental group or control group
- Split-class design (students do not normally see each other's screens)

Materials

- Students complete eLearning modules on laptops or tablets, possibly loaned from school
- Verification of interaction with prompts automatically stored in student data logs

Timeline

- Experiment continues through entire unit

Data Collection and Scoring

- The eLearning module logs the amount of time the pop ups with messages remain on screen

- Self-efficacy is measured on a rubric based on responses to certain prompts for the statistics project. See Assessment Type 4: Summative Project in Part C and the corresponding rubric for an example.
- The exact prompt used for the initial prompt and final prompt will be counterbalanced across students to ensure that changes in efficacy score are not just due to inherent propertied of the question (e.g., one may sound more intimidating than the other)

- Improvement in self-efficacy = self-efficacy score on final milestone – self-efficacy score on initial milestone

Hypotheses and Related Predictions

Main hypotheses

- Both groups will see some gain in self-efficacy measures due to other instructional elements that promote self-efficacy
- Experimental group will see slightly more gains in self-efficacy (and will see them in relatively earlier milestones)
- Control group will see slightly less gains in self-efficacy (and will see them in relatively later milestones)

Related predictions

- Girls may begin with slightly lower self-efficacy scores than boys
- Students who have had higher math grades in previous courses/units will begin with slightly higher self-efficacy scores than those with lower grades
- Students who score higher on other assessments within this unit will end with slightly higher self-efficacy scores than those who score lower

Assessment of Design Quality

Sampling

As a whole, this population of students is not representative of all students in the US. The community is likely more affluent and less-diverse. In terms of assigning students to groups, we should make sure that (i) the proportion of girls to boys is about equal and (ii) that the proportion of high-achievers in previous math classes to low achievers is roughly the same.

Validity and reliability

See "Key Features" table for Assessment Type 4: Summative Project for details. The inter-rater reliability of the rubric for self-efficacy is checked in the fidelity check. When rating students in the other class, instructors are blind to which group the students belong to, as well as which milestone they are rating (initial or final).

Possible confounds

Variations in parental involvement may lead to more or less improvements in self-efficacy measures despite group assignment. This is difficult to measure and may bias the results if one group's parents are generally more invested in their children's education.

Project Reflection

Self-Assessment of the Project Product

How well aligned are your goals, assessment and instruction?
I would say they are generally well-aligned. The number system definitely helps keep track of the relevant goals when describing the assessments and instruction. I did not create *specific* assessments and instructional activities for each goal, but the general descriptions in parts C and D cover everything. I tried to use assessments of appropriate complexity given the intellectual depth of the learning goals while balancing the time constraints of the course.

How did your age level focus impact the design, compared to similar units that have been or could be designed for younger and / or older age levels?
My age level focus influenced the amount of detail and nuance regarding concepts in statistics that I incorporated into the design. 6th graders are starting to develop stronger abstract reasoning abilities, so I figured they could handle weighing multiple options and justifying the use of one method over the other. However, because this is likely their first unit in statistics, the design focuses on ideas that more or less make sense at face value. For example, mean absolute deviation is a lot more intuitive as a measure of spread than root mean square error or standard deviation. Given a dataset, older students might be able to compare more possible evaluation methods, but I was not concerned with whether or not 6th graders could identify the best *overall* measure of center or spread. Rather, I focused on equipping them with strategies that would allow them to recognize that RMSE is probably better than MAD once they learn about that topic later.

In what ways does your design exemplify course principles (i.e., utilize your big ideas)?
Major design decisions, such as they type of instruction/assessments to use, style of feedback to give to students, and timing of the unit, are justified with my Big Ideas. I included footnotes that link design elements to specific syntheses, each of which is backed up by sources from EGIA. For parts C and D, I explain how I exemplify the course principles in the general descriptions, and then apply them to specific assessment items and instructional activities.

How clearly have you described all five sections of your design?
I would say the descriptions are mostly clear in each section, especially for parts A and B. There may be some ambiguity in my specific assessments and instructional activities. They make sense to me, but perhaps that's only because I am very familiar with the context of the entire unit. Classmates who have read them tell me that the descriptions make sense, especially when they have the relevant lecture slides handy. I did not have anyone from outside of EGIA review my work though, which might have brought unclear points to my attention.

What are the innovative aspects of your design?
I feel like my design combines eLearning and classroom instruction well. It takes advantage of the strengths of each—eLearning provides instant feedback and personalized practice, while the classroom instruction and statistics project enable the application of material to personal interest and target dispositional goals. Furthermore, the idea to have students continuously incorporate new material into a product is innovative (well, at least for a statistics course—I kind of stole it from the current project!).

How did you incorporate peer feedback to enhance your project product?
I incorporated some ideas from the poster session. For example, the idea to check the fidelity of how well the instructor monitors the group activities arose from a discussion on Part E. However, I received the bulk of my peer feedback from informal discussions with classmates. They suggested that I use a table to organize the types of reliability and validity in my assessments and to provide sample student responses to written prompts.

Self-Assessment of the Project Process

What were the strengths and weaknesses of your individual project design process?
I generally benefited from a good series of iterations, especially for parts A and B. I did my best to understand the instructor's feedback and incorporate it into my design. I also had some very long discussions with peers (shout out to Pankaj, Moe, Rog, Elizabeth, and Julia), which helped me clarify ideas and organize points in a more straightforward way.

The major weakness of my process has to do with the specific assessment and instruction activities. I did not submit a complete version of these parts in time to receive feedback from the professor. I also could have read the Understanding by Design handouts and Carver 2006 more thoroughly when they were assigned in class. I didn't put in too much effort to understand them until I needed them for the project, which definitely made completing parts C and D take significantly longer.

What challenges did you face?
I could take a lot of inspiration from open-source course materials regarding skills and procedures for 6th grade statistics, but specific instruction and assessments that target dispositions and metacognition were harder to come by. Most of the references I found discussed these topics at high levels, but did not go into detail about the validity, reliability, equitably of their materials. Furthermore, I also had trouble taking into account whether or not students could get a good score without understanding, especially in the eLearning assessments.

How did you overcome them and/or why do some remain?
I overcame the lack of dispositional instruction/assessment materials specifically for 6th grade statistics by looking over materials in other domains more thoroughly. I tried to pick out the key elements that target dispositions and metacognition and adapt them for statistics. Reviewing the readings from EGIA helped with this as well. To deal with the possibility of students doing well on assessments despite not understanding, I acknowledged it in my fidelity check of student interactions with the eLearning program.

How did the experience of giving and receiving peer feedback impact your project process?
This was very integral to my understanding of parts C and D. I didn't fully understand the different types of reliability and validity before (i) getting an explanation from a peer and (ii) helping a friend incorporate it into his project. Reading over others' projects, as well as having peers read over my mine, definitely helped me with consistency and alignment throughout my design.

What are your next steps, either with respect to this project if you plan to continue it, or with respect to other projects that could benefit from this approach?
I would like to have a more thorough understanding of the Understanding by Design process. I plan on going through that book more in depth and will try to apply those Ideas to other ventures. If I endeavor

on a similar project in the future (which is likely given my career goals), I will definitely take some of the approaches applied here. For example, it will help to lay out a clear set of objectives and milestones and do my best to stick to that timeline. I will also be sure to seek feedback from peers so I can iterate on my work using insight other than my own.

The next time you have an opportunity to begin a new project, how do you plan to proceed differently than you have on this project?

I think I suffered from trying to make everything too polished on my first iteration. As a result of that, I wasn't able to complete all parts in time to receive good feedback and do more revisions. Next time, I will try to at least get something down as a first draft so that peers can review my work in a timely manner.