Section I: Getting Started

Investigation
Questioning Through the Investigative Process

Overview

As described in the foreword, statistics is recognized as a necessary component in the K–12 mathematics curriculum that is reflected in current state standards, many based on the Common Core State Standards (CCSS) and high-stakes national assessments such as SAT, ACT, and NAEP.

To support and further elaborate on the Statistics and Probability standards, the American Statistical Association (ASA) produced *Guidelines for Assessment and Instruction in Statistics Education (GAISE): A Pre-K–12 Curriculum Framework,* which was approved by the ASA in 2005 (*www.amstat. org/education/gaise*). The ASA/NCTM Joint Committee on Curriculum in Statistics and Probability in 2007 worked with the authors of the GAISE Framework to incorporate final editing and provide funding for printing the report in book format. GAISE 2 will be released in 2020, keeping the spirit of the original GAISE but updating with respect to advances in technology, the wealth of big data, and the importance of the statistical problem-solving process particularly related to the role of questioning in statistics.

Goals of the GAISE and GAISE 2 Framework are the following:

» Present the statistics curriculum for grades Pre-K–12 as a cohesive and coherent curriculum strand (e.g., the progression of the mean from elementary to middle to secondary)

» Promote and develop statistical literacy for all students before graduating from secondary school

» Provide links with the NCTM 2000 *Principles and Standards for School Mathematics* and 2018 *Catalyzing Change: Initiating Critical Conversations*

» Discuss differences between mathematical and statistical thinking, particularly the importance of context and variability within statistical thinking

» Clarify the role of probability in statistics

» Illustrate concepts associated with the statistical problem-solving process

The framework stresses hands-on active learning and emphasizes that statistical analysis is an investigative process that turns loosely formed ideas into scientific studies by doing the following:

» Formulating a question that can be answered with data

» Designing a plan to collect appropriate data

» Analyzing the collected data utilizing graphical and numerical methods

» Interpreting the results to reflect insight on the original question

The investigative process requires the investigator to formulate questions throughout the statistical problem-solving steps that will

be the focus of the study. There are **research** questions that motivate the study, **investigative** questions that can be answered with data, **survey** questions to collect the data, **analysis** questions to prompt which graphs and calculations to perform, and **interpretation** questions to help focus the drawing of conclusions.

The purpose of this lesson is to help your students learn how to question throughout the statistical problem-solving process and learn how to formulate an investigative question—a question that can be answered with data.

Note: The Common Core State Standards for Mathematics (CCSSM) uses the vocabulary

---

Learning Goal

Understand how to use questioning throughout the statistical problem-solving process and how to construct a good investigative question.

Mathematical Practices Through a Statistical Lens

MP1. Make sense of problems and persevere in solving them.

Statistically proficient students understand how to carry out the four steps of the statistical problem-solving process.

MP2. Reason abstractly and quantitatively.

Statistically proficient students reason abstractly about the investigative problem at hand, understanding this requires the clarity of the variables that need to be measured in the data-collection process.

MP6. Attend to precision.

Statistically proficient students understand needing to be precise about the words used to ensure the intent of the investigative question is clear, needing to name variables and populations correctly.

Materials

Student worksheets are available at *www.statisticsteacher.org/statistics-teacher-publications/focus.*

» Student Worksheet 1: Questioning Throughout the Investigative Process

» Student Worksheet 2: Investigative Questions

» Student Worksheet 3 (Optional): Investigative Process

» Exit Ticket

Estimated Time

One 50-minute class period

"statistical question." Standard 6.SP.1 states, "Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers." The GAISE 2 Report, to be published in 2020, uses the term investigative question rather than statistical question. Whichever term is used, questions to analyze need to be answered with data. Throughout the investigations in *Focus on Statistics*, the term statistical question has been used.

Instructional Plan

Explain to your students that statistics is an investigative process guided by the following four statistical problem-solving steps:

1. Formulating a question that can be answered with data

2. Designing a plan to collect appropriate data

3. Analyzing the collected data utilizing graphical and numerical methods

4. Interpreting the results to reflect insight on the original question

Note: You may want to post these four steps.

Discuss that the statistical investigative process requires the investigator to use questions throughout the four steps. There are **research** questions that motivate the study, **investigative** questions that can be answered with data, **survey** questions to collect the data, **analysis** questions to prompt which representations such as graphs to construct and numerical methods to perform, and **interpretation** questions to help focus the drawing of conclusions.

Hand out Student Worksheet 1: Questioning Throughout the Investigative Process. Ask your students to read the summary, and then discuss the different types and examples of questions—investigative questions, survey questions, and analysis questions.

Student Worksheet 1: Questioning Throughout the Investigative Process

Administrators and teachers within a large school district are concerned about the perceived lack of sleep among middle-school students. Students more than ever are seemingly tired in class and struggling to stay focused. There have been complaints from parents that too much homework is being assigned. The school district believes there are several potential factors contributing to sleep deprivation among the students that are not academic. These include the number of extracurricular activities middle-school students are actively involved in and time spent on the internet and electronic devices. The school district decides to conduct a survey of selected middle-school students in the large district to investigate potential factors contributing to sleep deprivation among their middle-school students.

Formulate a question that can be answered with data. A possible *investigative question* that will assist the school district with this scenario is: What are the number of extracurricular activities middle-school students in the school district are actively involved in throughout the school year?

Collect data. A possible *survey question* for gathering data to help answer the investigative question is: How many extracurricular activities do you actively participate in during the school year?

Note: At this stage, ask your students whether the data will assist in answering the investigative question and whether there are other survey questions that need to be asked.

Analyze the data. Possible *analysis questions* are: What is an appropriate graphical display to show the distribution for the number of extracurricular activities for middle-school students? What is the shape of the distribution of the number of extracurricular activities? How much does the number of extracurricular activities vary in the distribution? What is a typical interval for the number of extracurricular activities, and where does the number of extracurricular activities cluster? What is the mean number of extracurricular activities? Are there unusual values?

Interpret the results. Ask if the analysis makes sense within the context of the situation and investigative question. Connect the results of the analysis questions to the context of the posed investigative question and make a conclusion.

Explain to your students that we now want to elaborate on how to recognize and write good investigative questions—a question that can be answered with data.

Ask your students what might be important to have in an investigative question—a question that can be answered with data.

Collate their ideas on the board, testing them as you go against the criteria for what makes a good investigative question.

Prompts you might use include the following:

» What data are the question about? – Leading to the variable(s) of interest needing to be clear in the investigative question

» Who is the question about? – Leading to the population or group of interest needing to be clear in the investigative question

» What sort of analysis does the question suggest we do with the data? – Leading to the intent of the investigative question

being clear (Is it wanting to investigate summarizing data for one variable, comparing two or more groups with respect to one variable, or looking at the association between two variables?)

» What data will we collect to answer the data? – Leading to the data is/will be available to answer the investigative question (or if we are using secondary data, the investigative question can be answered with the data we have).

» Is the investigative question useful? Does it have a purpose? Is it interesting?

» Does the investigative question allow for statistical analysis to the whole group you are interested in studying?

Share with your students that criteria to consider for a well-written investigative question are the following:

» The investigative question clearly states the variable(s) of interest.

» The investigative question clearly states the population of interest.

» The investigative question clearly states the intent (summary, comparison, or association investigation).

» It should be clear from the investigative question whether the data (measurements) can be collected (called primary data) to help answer the question or if the data are already available (called secondary data).

» The investigative question is worth investigating, it is interesting and/or it has a purpose, and it assists in answering the research question.

» The investigative question allows for statistical analysis to be made of the whole group.

Refer to the investigative question listed in Step 1 on Student Worksheet 1 Questioning Throughout the Investigative Process.

1. What is the number of extracurricular activities middle-school students in the coastal school district are actively involved in throughout the school year?

Ask the students to consider the criteria for what makes a good investigative question. They will also need to decide what data need to be collected.

Answers:

Considering the criteria

1. *Variable of interest: The number of extracurricular activities students are actively involved in throughout the school year*

2. *Population: Middle-school students in the school district*

3. *Intent is clear: This will be a summary or descriptive analysis for one variable.*

4. *Data: The data are the number of extracurricular activities the sampled students actively participate in during the school year.*

5. *Interesting/purposeful: This area of the investigation is of interest to the school district, especially if they want to provide evidence that it is not (or not only) homework contributing to students being tired and unable to focus.*

6. *About the whole group: This question considers the whole group of school district middle-school students.*

Pose a second example investigative question for your students to consider.

Example: For high-school district students, do algebra students who complete their homework tend to score better on an algebra test than algebra students who do not complete their homework?

Ask the students to check this question against the criteria for what makes a good investigative question. They will also need to decide what data need to be collected.

Answers:

Considering the criteria

1. *Variable of interest: The typical test score on an algebra test*

2. *Population: School district algebra students who complete their homework and school district algebra students who do not complete their homework*

3. *Intent is clear: This will be a comparative study of two groups' (one group completes homework and one group does not) test scores.*

4. *Data: For the sampled students, test scores on an algebra test and whether the students do their homework*

5. *Interesting/purposeful: The area of investigation is of interest to the school district, especially if they want to provide evidence that doing homework improves algebra test scores.*

6. *About the whole group: This question considers the whole of both groups of school district algebra students.*

The data to be collected are test scores on an algebra test and whether the students do their homework.

Pose a third example investigative question for your students to consider. This example

is not well written. Through considering the criteria, identify what is missing and how the investigative question could be improved.

Question: Are members of the boys' soccer team fitter than the members of the boys' football team?

Ask the students to consider the criteria for what makes a good investigative question. They will also need to decide what data need to be collected.

Answers:

Considering the criteria

1. *Variable of interest: Fitter. It is not clear what fitter means; the variable needs to be redefined or improved.*

2. *Population: Boys' soccer team and boys' football team, but we don't know if this is from one school or if it is primary, middle, or secondary school. The population needs to reflect the actual population from which data will be collected (e.g., the 8th-grade boys' soccer team and the 8th-grade boys' football team).*

3. *Intent is clear: This will be a comparative analysis, though the investigative question suggests the boys' soccer team would all be fitter than the boys' football team. To improve, use the idea of tendency (e.g., Do the members of the boys' soccer team tend to be fitter than the members of the boys' football team?).*

4. *Data: How fitter will be measured needs to be determined first. It is not clear what data need to be collected from the investigative question.*

5. *Interesting/purposeful: The area of investigation is of interest to the coaches and athletes, especially if they want to use the information to improve coaching and playing techniques.*

6. *About the whole group: Once the population is more clearly specified, this question can consider the whole of both groups.*

The data to be collected would be based on the criteria that are decided to judge fitness.

Ask students if this third example investigative question helps explain why middle-school students may lack a reasonable amount of sleep—the problem being investigated by the coastal school district.

Possible answer: *It is not clear how understanding whether the fitness of boys' soccer players is better than the fitness of boys' football players may or may not contribute to a lack of sleep.*

Hand out Student Worksheet 2: Investigative Questions. Place students into groups of three and ask them to complete Part I of the worksheet.

After students have completed Part I, review the answers. For the questions needing improvement, discuss possible investigative questions, and then ask your students to complete Part II.

Possible answers for Student Worksheet 2 are shown on the following pages.

Student Worksheet 2: Investigative Questions

Part I: For each investigative question, consider the criteria for a well-written investigative question. Give your reason for deciding whether the investigative question is well written or needs improvement.

Question	Criteria to Consider	Explain What Criteria the Question Does Not Meet, If Any
Are cars speeding in the posted school zone?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*'Speeding' needs to be defined; what school zone needs to be specified and time of day; is the intent to summarize the proportion of cars speeding or a typical speed over the speed limit?*
What proportion of seniors at the high school participate in school-sponsored activities that take place after the school day?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*The population needs to be more specific—what high school? Note: Since proportion is specified in question as the summative value desired, the variable of interest is categorical: yes or no as to whether the student participates. As written, it is not clear this question is about the whole group, as it is only asking about those senior students who participate in school-sponsored activities (one category of the categorical variable). A better question might be to ask about what activities all senior-high school students participate in. Include in the question school- and non-school-sponsored, and then the current given question would be an analysis question.*
How much money is spent on the daily lunch program?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*Population is not clear—elementary, middle, or high school, and where is the school located? What time period—per day, per month, per year?*

Table continued on next page

Table continued

Question	Criteria to Consider	Explain What Criteria the Question Does Not Meet, If Any
Does taking 500 mg of Vitamin C daily protect high-school students from catching a cold during the winter months?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*Population is not clear—what high-school students? Intent is not clear—is this to be a comparison in which an experiment is conducted with some students assigned to take Vitamin C and others not? Will we then compare the proportion who catch cold in each group, or look for an association based on observational data from surveyed students?*
What is the typical length of girls' hair?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*Population of girls needs to be specified, For example, age interval of girls and where the girls are from. This question would be better if it asked "what are typical lengths of girls' hair." The phrasing "typical" suggests an average for an answer and is therefore not about the whole group.*
Is there an association between gender and ability to roll the tongue?	1. Variable of interest is clear: 2. Population is clear: 3. Intent is clear: 4. Data: 5. Interesting/purposeful: 6. About the whole group:	*Population needs to be more specific.*

Part II: Write an investigative question for each of the following general topics:

1. Price of a new car

Possible answer: What are the prices of new 2020 mid-size SUVs?

2. Effect of listening to music and math test scores

Possible answer: Do students at our high school who listen to music while taking an algebra test tend to score better than those students at our high school who do not listen to music while taking an algebra test?

 Exit Ticket

1. A group of biology students asked the question, "What's the fastest animal in the world?"

» Explain why this is not a well-written investigative question.

Possible answer: There is only one answer—no variability in the data.

» Rewrite the question so it would be a well-written investigative question.

Possible answer: What are typical speeds of animals found in the continental United States?

2. A coffee house owner asked the question, "How much money is spent in my coffee house?"

» Explain why this is not a well-written investigative question.

Possible answer: Population is not clear. Define "my" coffee house. Also, is the coffee owner interested in a certain time frame the coffee shop is open?

» Rewrite the question so it would be a well-written investigative question

Possible answer: How much money do customers spend between the hours of 7 a.m. and 10 a.m. at the Java House?

3. A politician asked the question, "If the election were held today, whom would you vote for?"

» Explain why this is not a well-written investigative question.

Possible answer: This is an example of a survey question to obtain data for answering an investigative question. Population is not specified.

» Rewrite the question so it would be a well-written investigative question.

Possible answer: What proportion of likely voters in Wisconsin will vote for each presidential candidate in the upcoming election?

Extension

Student Worksheet 3: Investigative Process

Ms. Brown, an administrator at a local high school of 935 students, is interested in studying the effects of playing video games on the academic achievement of students.

Formulate an investigative question that can be answered with data. Remember to clearly identify the variables of interest, the population, and the data or measurements needed. Check your investigative question against the criteria for a well-written investigative question.

Possible answer: Is there an association between the number of hours per school day ninth grade students at the local high school play video games and their GPA at the end of the ninth grade school year?

Create a data-collection plan including any survey questions.

Possible answer: Randomly select 50 ninth grade students and ask each selected student how many hours per school day he/she usually plays video games. At the end of the school year, ask the principal for the GPAs of the 50 selected students.

Describe possible data analysis questions and types of graphs and calculations needed to be performed.

Possible answer: Construct dot plots of the amount of time playing video games and GPAs. Explore the distributions to describe the amount of time playing video games and GPAs. Construct a scatterplot of GPA vs. Video Game Hours for the 50 students. Explore the possible association of GPA and Video Game Hours by describing trend and strength (weak, moderate, strong).

Describe how Ms. Brown might interpret the results.

Possible answer: Does the scatterplot show a pattern/association between video game hours and GPA, or is there no pattern that emerges?