Chief Reader Report on Student Responses:

2025 AP® Statistics Free-Response Questions

- Number of Students Scored 267,690
- Number of Readers 1435
- Score Distribution

Exam Score	N	%At
5	45,457	17.0
4	57,349	21.4
3	58,483	21.8
2	42,684	15.9
1	63,717	23.8

- Global Mean 2.92

The following comments on the 2025 free-response questions for AP® Statistics were written by the Chief Reader, Dr. Barb Barnet, Department Chairperson and Professor, University of Wisconsin-Platteville. They give an overview of each free-response question and of how students performed on the question, including typical student errors. General comments regarding the skills and content that students frequently have the most problems with are included. Some suggestions for improving student preparation in these areas are also provided. Teachers are encouraged to attend a College Board workshop to learn strategies for improving student performance in specific areas.

Question 1

Task: Exploring Data
Topic: Mileage in Two Countries
Max Score: 4.00
Mean Score: 1.75

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) use side-by-side boxplots to compare the distributions for two different countries, (2) determine the relationship between the mean and the median based on the previous comparisons of the side-by-side boxplots, and (3) calculate the range and possible median of a newly constructed boxplot from the side-by-side boxplots for two different countries.

This question primarily assesses skills in Skill Category 2: Data Analysis and Skill Category 4: Statistical Argumentation. Skills required for responding to this question include 2.A: Describe data presented numerically or graphically; 2.C: Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response; 2.D: Compare distributions or relative positions of points within a distribution; and 4.B: Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers the content from Unit 1: Exploring One-Variable Data of the course framework in the *AP Statistics Course and Exam Description*. Refer to Topics 1.7, 1.8, and 1.9 and Learning Objectives UNC-1.I, UNC-1.J, UNC-1.K, UNC-1.M, and UNC-1.N.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

- In part A most responses indicated the value of the medians for both boxplots, but several failed to provide comparative language to indicate the center of the distribution of gas mileage for the sample of cars manufactured in Country A was lower than for the sample of cars manufactured in Country B. Some responses incorrectly identified the center of the boxplot as the mean. In some responses, an incorrect formula or understanding of the range was given, providing an incorrect comparison of spread. Most responses indicated the presence of an outlier, but some failed to communicate which distribution contained an outlier. Some responses failed to provide proper context to represent the variable.
- In part B most responses indicated the mean of the distribution of gas mileage for the sample of cars manufactured in Country A was expected to be greater than 18 mpg, but many failed to link the 18 mpg to the median of Country A's distribution. Part B required a justification based on the presence of an upper outlier or a right-skewed distribution; however, many responses failed to communicate the direction of the skewness or where the outlier was located.
- In part C most responses correctly determined a range of 26 mpg for the combined data set and justified their response with an argument identifying 40 as the maximum and 14 as the minimum. However, some responses incorrectly represented the range as an interval of two values instead of as the range of 26 mpg. Some responses were successful in identifying 24 mpg as the median of the combined data, but many responses provided an incorrect justification for the median of the combined data by averaging the median of Country A and the median of Country B. Other responses provided an insufficient justification for the median of the combined data by failing to adequately communicate that half of the combined data values would be below 24 and half above 24.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part A the response did not provide comparative language to compare the centers of the distributions of gas mileage of the sample of cars manufactured in Country A and Country B. • The response described the center of a boxplot as the mean or average.	• The median gas mileage for the sample of cars manufactured in Country A (18 mpg) is lower than the median gas mileage for the sample of cars manufactured in Country B (32 mpg).
• In part A the response indicated an incorrect understanding of the range of a distribution by calculating the range after eliminating the outlier, rather than as the maximum minus the minimum. • The response did not provide identification as to which value (range or IQR) was being used to compare the spreads.	• The range of the gas mileages for the sample of cars manufactured in Country A (24 mpg is the maximum of 38 minus the minimum of 16 mpg) is higher than the range of the gas mileages for the sample of cars manufactured in Country B (22 mpg is the maximum of 40 minus the minimum of 18 mpg). However, the IQR of Country A (8 mpg) is less than the IQR of Country B (12 mpg).
• In part A the response provided reference to a shape that cannot be determined from boxplots, such as "normal" or "unimodal," or reference to an incorrect shape, such as "the distribution of gas mileages for Country A is left skewed" or "the distribution of gas mileages for Country B is right skewed."	• The shape of the distribution of the gas mileage for cars manufactured in Country A can be described as right skewed. The shape of the distribution of the gas mileage for cars manufactured in Country B can be described as left skewed.
• In part B the response provided a justifiction that fails to provide a direction for the skewness or where the outlier was located. • The response provided a justifiction that fails to link the value of the median to 18 mpg.	• The mean of the distribution of gas mileage for the sample of cars manufactured in Country A is expected to be greater than 18 mpg, which represents the median. This is due to the skewed right distribution, which will pull the mean above the median. This effect can also be due to the presence of an upper outlier at 38 mpg.
• In part C the response referred to the range as an interval of two values. • The response provided a maximum and minimum value that is inconsistent with the values in the boxplots.	• The range of the combined data set is 26 mpg. This is because the maximum gas mileage from Country B minus the minimum gas milage from Country A is $40 \text{ mpg} - 14 \text{ mpg} = 26 \text{ mpg}$.
• In part C the response provided a calculation of the value of the median for the combined data as the average of the medians from the two individual data sets. • The response provided an incomplete justification with weak communication to relate the median of the combined data set, 24, with values of Q3 and Q1 from Country A and Country B, respectively.	• The median value of the combined data set is 24 mpg. Q3 for Country A is 24 mpg, which means that 75% of cars manufactured in Country A have gas mileages at or below 24 mpg. Q1 for Country B is 24 mpg, which means that 25% of cars manufactured in Country B have gas mileages at or below 24 mpg. Because both samples have a sample size of 100 cars, when combined, 50% of the 200 cars (100 cars) have gas mileages at or below 24 mpg, making 24 mpg the median of the combined data set.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Encourage students to make a direct comparison of center and spread when comparing the distributions of a quantitative variable.
- Encourage students to review the definition and application of Q1, median, Q3, IQR, and range.
- Encourage students to identify the acceptable and unacceptable descriptors for the shape of a boxplot. Show examples of histograms (e.g., right skewed, left skewed, and symmetric) and have them match the appropriate boxplot. Repeat this exercise with nonstandard distributions (e.g., bimodal, symmetric, uniform, and right skewed) to illustrate the idea that modality cannot be determined from a boxplot.
- Encourage students to clearly communicate the meaning of the quartiles of a boxplot.
- Encourage students to provide complete context in their responses, particularly when describing a distribution or comparing one or more distributions.
- Encourage students to clearly communicate the direction of skewness and the position of outliers in the distribution of a data set.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Pages 227–228 and 232 of the CED provide examples of key questions and instructional strategies designed to develop Skill 2.A (describe data presented numerically or graphically), Skill 2.C (calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response), Skill 2.D (compare distributions or relative positions of points within a distribution), and Skill 4.B (interpret statistical calculations and findings to assign meaning or assess a claim).
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Create Representations strategy, for example, may allow students to have a deeper understanding of the differences in quantitative graphs.
- AP Classroom provides three videos focused on the content and skills needed to answer this question.
    - The daily video 1 for Topic 1.9 explores a 2015 Exploring Data free-response question (FRQ) that was the same task as part A in this FRQ. The relevant takeaway from the video to this question is comparing the measures of centers and spread (part A).
    - The daily video 1 for Topic 1.8 discusses the relationship of the median and mean based on the distribution of the boxplot (see UNC-1.M.2). The relevant takeaway from the video to this question is the explanation of how the mean is nonresistant to outliers, while the median is resistant (part B).
    - The daily video 1 for Topic 1.7 explains how to determine the median (see UNC-1.I.3). A key takeaway from this video is that the middle value of an ordered data set is the median. Students needed to think about both boxplots combined as 200 gas mileages. The next step was to identify the median by the 100 gas mileages below the median and 100 gas mileages above the median to determine the answer.
- AP Classroom also provides topic questions for the formative assessment of Topics 1.7, 1.8, and 1.9, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate information about boxplots and their shape, search for the keywords "boxplot and shape" in the discussion boards. Discussions about what information boxplots provide and what they do not provide is some of the information available.

Question 2

Task: Sampling
Topic: Aphids
Max Score: 4.00
Mean Score: 2.10

What were the responses to this question expected to demonstrate?

The primary goals of the question were to assess a student's ability to (1) explain whether a sampling method is an appropriate sampling method for a farmer to use to estimate the proportion of cabbage plants in the field that are damaged by aphids, (2) determine and justify whether a different sampling method is likely to provide an overestimate or an underestimate of the proportion of cabbage plants in the field that are damaged by aphids, and (3) describe how to implement another sampling method that requires random selection.

This question primarily assesses skills in Skill Category 1: Selecting Statistical Methods. The skill required for responding to this question is 1.C: Describe an appropriate method for gathering and representing data.

This question covers content from Unit 3: Collecting Data of the course framework in the *AP Statistics Course and Exam Description*. Refer to Topic 3.3 and Learning Objectives DAT-2.C and DAT-2.D.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

- In Part A almost all responses correctly indicated that sampling method I is not appropriate. However, many responses did not provide a correct justification for this. Some responses gave an incorrect justification, such as small sample size, confounding, or nonresponse bias. Many students attempted to explain that region 3 is not representative of the entire cabbage field by explaining how region 3 is different from regions near the river without explicitly stating that region 3 is not representative. Many responses provided answers in context, often including the phrase "the proportion of cabbage plants in the field," which indicated both parameter and population.
- In Part B almost all responses correctly indicated that sampling method II will likely produce an overestimate of the proportion of cabbage plants in the field damaged by aphids. Additionally, most responses provided correct justification by indicating that row E is the row closest to the river, and that the farmer believes that regions closer to the river have greater damage.
- In Part C most responses described how to implement a sampling method; however, some responses indicated the need to generate random numbers without describing a process to accomplish that. Many responses fully described a correct sampling method, while some of the responses failed to include details on which region to select after the random numbers are generated, how many random numbers are generated for each row, or that papers in a hat need to be mixed before a paper is selected.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part A some responses did not include sufficient context from the problem.	• Sampling method I is not an appropriate method to estimate the proportion of cabbage plants in the field that are damaged by aphids.
• In part A many responses included extraneous reasons why sampling method I is not appropriate, such as undercoverage, counfounding, or nonresponse bias.	• Sampling method I is a convenience sample where region 3 was not selected randomly.
• In part A some responses inferred that the farmer wanted to test their belief, rather than using the belief to justify why sampling method I is inappropriate.	• Because region 3 is far from the river, where the farmer thinks the most damage occurs, it will not be representative of the entire field.
• In part B several responses failed to fully justify the reason that sampling method II would likely produce an overestimate of the proportion of cabbage plants in the field that are damaged by aphids, by either failing to explicitly state that row E is closest to the river or by failing to state that regions near the river may experience greater damage.	• Sampling method II is likely to overestimate the proportion of cabbage plants damaged by aphids because the farmer believes that the damage will be greater near the river and row E is the row closest to the river.
• In part C several responses did not provide a process for randomly selecting the regions; for example, a response indicated that random numbers were needed, but it didn't provide a process to accomplish it.	• To randomly select a region in row A, use a random number generator to generate a number between 1 and 5 and select that region as part of the sample.
• In part C many responses did not use the region labels provided in the diagram and generated a random number from 1–5 for every row. Some of those responses did not clearly indicate which region should then be selected in row B, for example, if the number 3 were randomly generated.	• Write the region numbers from row A, 1–5, onto same-size slips of paper, and then put the numbers into a hat, mix well, and select one of the numbers. Repeat this process using the region numbers in each of the other rows (row B, 6–10; row C, 11–15; etc.).

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Discuss statistical terms related to biased sampling methods, such as nonrepresentative, nonrandom, and convenience samples. Students often had a correct understanding of the concepts but lacked the statistical language to provide correct justifications.
- Provide more experiences for students to describe how to implement different sampling methods, such as stratified, cluster, and systematic, as well as simple random samples, in different contexts. Ask students to describe a sampling procedure and then ask other students to implement the process exactly as written.
- Encourage students to write responses so they are unique to the context of the problem. For instance, the response should not be so general that it could be applied to another problem without change.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description*, effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Page 225 of the CED provides examples of key questions and instructional strategies designed to develop Skill 1.C (describe an appropriate method for gathering and representing data).
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Paraphrasing strategy, for example, may be helpful in developing students' abilities to describe different sampling methods and determining the differences between them. The scoring guidelines have different examples that are broken down that could help guide students' learning.
- AP Classroom provides one video focused on the content and skills needed to answer this question.
    - The daily video 1 for Topic 3.3 demonstrates how to determine the type of study being conducted (see DAT-2.C.2 and DAT-2.D.1). The key takeaway from this video is that random samples, when well executed, tend to provide representative samples.
- AP Classroom also provides topic questions for formative assessment of Topic 3.3, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give students practice verifying conditions, enter the keywords "sampling methods" in the search bar. Worksheets, practice questions, and activities are among the resources available.

Question 3

Task: Probability
Topic: Music Selection
Max Score: 4.00
Mean Score: 1.55

What were the responses to this question expected to demonstrate?

The primary goals of the question were to assess a student's ability to (1) calculate a probability, (2) calculate a probability using a previous calculated probability, (3) define the random variable of interest and how the random variable is distributed, (4) describe the random variable as having a binomial distribution including the number of trials and probability of success, (5) calculate the expected value of the random variable, (6) calculate a probability for a binomial distribution, and (7) interpret the probability to justify a claim.

This question primarily assesses skills in Skill Category 3: Using Probability and Simulation, and Skill Category 4: Statistical Argumentation. Skills required for responding to this question include 3.A: Determine relative frequencies, proportions, or probabilities using simulation or calculations; 3.B: Determine parameters for probability distributions; and 4.B: Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from Unit 4: Probability, Random Variables, and Probability Distributions in the *AP Statistics Course and Exam Description*. Refer to Topics 4.3, 4.6, 4.10, and 4.11 and Learning Objectives UNC-3.A, UNC-3.B, UNC-3.C, VAR-4.A, VAR-4.B, and VAR-4.E.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

- In part A most responses correctly calculated the probabilities and included adequate work. However, some responses mistakenly calculated the probability in part A (ii) by assuming dependent events.
- In part B (i) many responses correctly defined the random variable as the number of rock songs played in one hour. However, some responses defined the random variable as simply "rock songs," which is not correct. In part B (i), many responses did not attempt to state how the random variable is distributed. In part B (ii) most responses correctly calculated the expected value. However, some students did not label the parameters or adequately show their work.
- In part C (i) many responses used the binomial formula to calculate the probability. Some responses incorrectly calculated the probability of exactly 4 or more than 4. In Part C (ii) the explanation often did not show clear linkage to the probability.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part A (ii) multiplying $\frac{100}{1,000}$ by $\frac{99}{999}$ does not take into consideration that the song selections are independent.	• $P(\text{Both Rock Songs}) = (0.10)(0.10) = 0.01$
• In part B (i) many responses defined the random variable as simply "rock songs," the mean of rock songs, or the proportion of rock songs.	• X is the number of rock songs played in one hour.
• In part B (i) many responses did not notice the second half of the task: stating how the random variable is distributed. Others responded with a statement such as "the variable is distributed randomly."	• The distribution is binomial with $n = 20$ and $p = 0.10$.
• In part B (i) some responses stated that the random variable followed a normal distribution.	• The distribution is binomial with $n = 20$ and $p = 0.10$.
• In part C (i) some responses incorrectly calculated either the $P(X = 4)$ or $1 - P(X \le 4)$.	$P(X \ge 4) = 1 - P(X \le 3) = 1 - \left[\binom{20}{0}(0.10)^0(0.90)^{20} + \binom{20}{1}(0.10)^1(0.90)^{19} + \binom{20}{2}(0.10)^2(0.90)^{18} + \binom{20}{3}(0.10)^3(0.90)^{17}\right]$
• In part C (i) some responses wrote $P(X \ge 4) = 1 - \text{binomcdf}(20,\ 0.1,\ 3)$ without labelling the parameters.	• $P(X \ge 4) = 1 - \text{binomcdf}(n = 20,\ p = 0.1,\ x = 3)$

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Have students practice identifying and explaining why events are independent or dependent. Then determine how this would affect the probability calculations and results.
- When working with binomial, geometric, and normal probabilities, have students define the random variable. A visualization is often helpful. Have students provide examples of what could happen during a specific hour. For example, from 9–10 a.m. there might be 5 rock songs played, or from 10–11 a.m. there might be 3 rock songs played. The random variable is the number of rock songs played in one hour.
- When stating how a random variable is distributed (describing a distribution), students should always include the parameters of the distribution. Once several distributions have been studied, have students in groups practice distinguishing between different distributions given several prompts. Discussion should include identifying the parameters and justifying their decisions.

- In part B (ii) many students did not label *n* or *p* when calculating the expected value $20(0.1) = 2$. Although this satisfied component 4, the response could not recoup component 3 if they failed to address the parameters of the binomial distribution. Have students clearly label all mathematical work.
- After completing a probability calculation, have students classify the probability as large, small, likely, or unlikely and describe what this may mean in the context of the prompt.
- Practice calculating different inequalities using the binomial formula.
- Encourage students to communicate clearly, use correct statistical notation, and show formulas with correct values included.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description*, effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Pages 229, 230, and 232 of the CED provide examples of key questions and instructional strategies designed to develop Skill 3.A (determine relative frequencies, proportions, or probabilities using simulation or calculations), Skill 3.B (determine parameters for probability distributions), and Skill 4.B (interpret statistical calculations and findings to assign meaning or assess a claim).
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Notation Read Aloud strategy, for example, may be helpful in developing students' abilities to verbalize probability interpretations.
- AP Classroom provides four videos focused on the content and skills needed to answer this question.
    - The daily video 1 for Topic 4.3 shows how to calculate probabilities and interpret them (see VAR-4.A.2 and VAR-4.B.1). The key takeaways from this video show how to solve a probability by hand and with the calculator (part A (i) and part C (ii)).
    - The daily video 1 for Topic 4.6 explains how to determine if two events are independent (see VAR-4.E.2). The key takeaways from this video show how to determine independence of events several different ways so that students can choose a method they are comfortable with (part A (ii)).
    - The daily video 1 for Topic 4.10 shows how to define a random variable and solve a probability for a binomial distribution (see UNC-3.A.2 and UNC-3.B.1). The key takeaways from this video show how to identify if a distribution is binomial (part B (i) and part C (i)).
    - The daily video 1 for Topic 4.11 shows how to calculate the expected value for a binomial distribution (see UNC-3.A.2). The key takeaway from this video shows how an expected value for a binomial distribution is calculated (part B (ii)).
- AP Classroom also provides topic questions for formative assessment of Topics 4.3, 4.6, 4.10, and 4.11, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give students practice verifying conditions, enter the keywords "binomial distributions" in the search bar, and then select the drop-down menu for "Resource Library." Filter for "Classroom-Ready Materials" to find worksheets, data sets, practice questions, and guided notes, among other resources.

Question 4

Task: Inference
Topic: Homework App
Max Score: 4.00
Mean Score: 1.65

What were the responses to this question expected to demonstrate?

The primary goals of the question were to assess a student's ability to (1) identify an appropriate procedure for conducting a hypothesis test for a one-sample z-test for a population proportion, (2) identify the correct hypotheses for conducting a one-sample z-test for a population proportion, (3) check the conditions for the hypothesis test for a one-sample z-test for a population proportion, (4) calculate the test statistic and p-value for a one-sample z-test for a population proportion, (5) compare the p-value to a significance level to make a decision regarding the hypotheses, and (6) determine an appropriate conclusion for a hypothesis test for a one-sample z-test for a population proportion.

This question primarily assesses skills in Skill Category 1: Selecting Statistical Methods, Skill Category 3: Using Probability and Simulation, and Skill Category 4: Statistical Argumentation. Skills required for responding to this question include 1.E: Identify an appropriate inference method for significance tests; 1.F: Identify null and alternative hypotheses; 3.E: Calculate a test statistic and find a p-value, provided conditions for inference are met; 4.C: Verify that inference procedures apply in a given situation; and 4.E: Justify a claim using a decision based on significance tests.

This question covers the content from Unit 6: Inference for Categorical Data: Proportions of the course framework in the *AP Statistics Course and Exam Description*. Refer to Topics 6.4, 6.5, and 6.6 and Learning Objectives VAR-6.D, VAR-6.E, VAR-6.F, VAR-6.G, and DAT-3.B.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

This question was scored in three sections. The first section includes statements of the null and alternative hypotheses and identification of the appropriate hypothesis test. The second section includes verifying conditions for the test identified in the first section and calculating the value of the test statistic and corresponding p-value. The third section includes the conclusion for the test identified in the first section.

- In section 1 most responses indicated that a one-sample sample z-test for a population proportion should be used. Most responses indicated that the null hypothesis involved equality with 0.22 and that the alternative hypothesis was greater than the value provided in the null hypothesis. Some responses did not set up the hypotheses in terms of proportions or used nonstandard notation. A somewhat common error was an incomplete description of the parameter through a lack of sufficient context for the population.
- In section 2 most responses made attempts to verify conditions. Many responses did not clearly indicate that a random sample was selected, as opposed to a random assignment. Many responses attempted to check the 10% condition but used either no context or their population context contradicted what they had already identified as the population. Most responses addressed the large counts condition by correctly comparing observed numbers of successes and failures to a standard criterion (e.g., 5 or 10). Some responses incorrectly compared the counts to 30 or solely compared the sample size to $30 (130 > 30)$. Most responses correctly reported the value of the z-statistic and p-value. Some responses used a confidence interval to conduct the test, but this test had a one-sided alternative hypothesis, and almost no responses either constructed a one-sided interval or compared the hypothesized value to only one side of the confidence interval.

- In section 3 most responses made an explicit comparison between the *p*-value and the stated $\alpha = 0.05$ and included a correct decision to reject the null hypothesis, although some struggled with the inequality symbols. Some responses used incorrect, definitive language (e.g., "Karen's school is greater than 22%" or "We have proof that the proportion of students at Karen's school who use the app to help with their homework is greater than 0.22 ") in stating the conclusion. Most responses had a conclusion stated (correctly) in terms of the alternative hypothesis. Many students had insufficient context in the conclusion, lacking clear communication about what the population, the proportion, the group, and/or the variable of interest were.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
In Section 1: • Wrote the null and alternative hypotheses in nonstandard and nondefined notation (e.g., $H_0 : x = 0.22$ and $H_a : x > 0.22$). • Wrote an incomplete name for the inference procedure (one sample *z*-test). • Defined the population parameter with incomplete context for the question (*p* is the proportion of students that use the application at least once a week).	In Section 1: • $H_0 : p = 0.22$ and $H_a : p > 0.22$. • One-sample *z*-test for a population proportion. • *p* is the true proportion of students at Karen's high school that use the application at least once a week.

In Section 2:
• Failed to check both parts of the independence condition fully:
  ○ Not specifying a random sample (random – *check*)
  ○ Not fully indicating the population in context for the 10% condition (sample is < 10% of population)
• Failed to check the large counts conditions correctly:
  ○ Comparing np and $n(1-p)$ to numbers other than 5 or 10
  ○ Incorrectly comparing $130 > 30$
• Reported an incorrect *z*-statistic $(z = 1.81)$ from placing $\hat{p}\ (0.2923)$ in the standard error instead of $p\ (0.22)$.
• Reported an incorrect *p*-value by using the wrong tail (direction).
	In Section 2:
	• Checked both parts of the independence condition fully:
	  ○ Data was obtained from a simple random sample of 130 high school students from Karen's school.
	  ○ There are more than 2000 students at Karen's school $(130 < 0.10(2000))$.
	• Correctly checked the large counts condition
	  ○ $np = 130(0.22) = 28.6 > 10$
	  ○ $np = 130(1-0.22) = 101.4 > 10$
	• Reported the correct *z*-statistic $z = \frac{0.2923 - 0.22}{\sqrt{\frac{0.22(1-0.22)}{130}}} \approx 1.99$
	• Reported the correct *p*-value $P(z > 1.99) \approx 0.023$

In Section 3: • Used the wrong inequality when comparing the *p*-value to alpha $(0.023 > 0.05)$. • Wrote a conclusion in definitive language (e.g., "We have proof that the proportion of students at Karen's school who use the app to help with their homework is greater than 0.22 ").	In Section 3: • Because $0.023 < 0.05$, we reject H_0. • There is convincing statistical evidence to suggest that the true proportion of students at Karen's school who use the app to help with their homework is greater than 0.22.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Encourage students to clearly define all parameters in context of the problem and not provide contradictory definitions ("proportion of all students in Country W," "proportion of all students," or "proportion of all students at Karen's school").
- When labeling parameters, encourage students to use correct notation (e.g., "p" or "μ").
- Encourage students to clearly communicate the correct procedure. (They are more likely to do so when communicating in words and not via formulas.)
- When checking conditions, encourage students to write "random sample" instead of using abbreviations such as "random."
- Encourage students to verify all conditions and state in words that they have been satisfied.
- Encourage students to conduct a significance test when asked to determine if the data provides convincing evidence for a claim, rather than using a confidence interval.
- Encourage students to provide a test statistic AND a *p*-value when conducting a hypothesis test.
- Encourage students to avoid calculator syntax when calculating a *p*-value.
- Encourage students to make a direct comparison of the *p*-value to the provided alpha level and use the result of that comparison to support their conclusion.
- Encourage students to clearly communicate that the data provide (or do not provide) convincing evidence to support the alternative hypothesis.
- Encourage students to provide complete context in their conclusion, including population, parameter, group, and variable of interest.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description*, effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Pages 226, 230, and 232 of the CED provide examples of key questions and instructional strategies designed to develop Skill 1.E (identify an appropriate inference method for significance tests), Skill 1.F (identify null and alternative hypotheses), Skill 3.E (calculate a test statistic and find a *p*-value, provided conditions for inference are met), Skill 4.C (verify that inference procedures apply in a given situation), and Skill 4.E (justify a claim using a decision based on significance tests).

- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Build the Model Solution strategy, for example, may be helpful in developing students' abilities to use precise language when justifying a claim based on a significance test.
- AP Classroom provides five videos focused on the content and skills needed to answer this question.
    - The daily video 1 for Topic 6.4 discusses how to state the null and alternative hypotheses for a one-sample *z*-test for a population proportion (see VAR-6.D.3 and VAR-6.D.4). The relevant takeaways from the video to this question involve identifying the hypotheses.
    - The daily video 2 for Topic 6.4 explores how to identify the procedure and check the conditions for performing a one-sample *z*-test for a population proportion (see VAR-6.E.1, and VAR-6.F.1). The key takeaway from this video is the explanation for how to take a random sample correctly.
    - The daily video 1 for Topic 6.5 describes how to calculate a test statistic and *p*-value for a one-sample *z*-test for a population proportion (see VAR-6.G.3). The key takeaway from the video shows students the relationship between the test statistic and the alternative hypothesis sign direction (i.e., less than, greater than, and not equal to).
    - The daily video 1 for Topic 6.6 demonstrates how to correctly state a conclusion for a one-sample *z*-test for a population proportion (see DAT-3.B.2, DAT-3.B.3, and DAT-3.B.4). A key takeaway from this video is the explanation for how to correctly conclude the hypotheses test.
    - The daily video 2 for Topic 6.6 demonstrates how to complete a hypotheses test for a one-sample *z*-test for a population proportion (see VAR-6.D.3, VAR-6.D.4, VAR-6.E.1, VAR-6.F.1, VAR-6.G.3, DAT-3.B.2, DAT-3.B.3, and DAT-3.B.4). The key takeaway for the video helps students understand how to identify the steps for the entire process of completing the test and what is expected.
- AP Classroom also provides topic questions for formative assessment of Topics 6.4, 6.5, and 6.6, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give students practice verifying conditions, enter the keywords "one-sample *z*-test for a population proportion" in the search bar, and then select the drop-down menu for "Resource Library." Filter for "Classroom-Ready Materials" to find worksheets, data sets, practice questions, and guided notes, among other resources.

Question 5

Task: Multifocus
Topic: Number of Bedrooms
Max Score: 4.00
Mean Score: 1.43

What were the responses to this question expected to demonstrate?

The primary goals of the question were to assess a student's ability to (1) calculate a probability, (2) calculate an expected value, (3) identify the correct hypotheses for conducting a one-sample *t*-test for a population mean, (4) explain what a Type I error would be for the one-sample *t*-test for a population mean, and (5) make a conclusion for the one-sample *t*-test for a population mean based the results of a one-sample *t*-interval for a population mean.

This question primarily assesses skills in Skill Category 1: Selecting Statistical Methods, Skill Category 3: Using Probability and Simulation, and Skill Category 4: Statistical Argumentation. Skills required for responding to this question include 1.B: Identify key and relevant information to answer a question or solve a problem; 1.F: Identify null and alternative hypotheses; 3.A: Determine relative frequencies, proportions, or probabilities using simulation or calculations; 3.B: Determine parameters for probability distributions; and 4.D: Justify a claim based on a confidence interval.

This question covers the content from Unit 4: Probability, Random Variables, and Probability Distributions; Unit 6: Inference for Categorical Data: Proportions; and Unit 7: Inference for Quantitative Data: Means in the *AP Statistics Course and Exam Description*. Refer to Topics 4.6, 4.8, 6.7, 7.3, and 7.4 and Learning Objectives UNC-4.T, UNC-5.A, VAR-4.E, VAR-5.C, and VAR-7.C.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

- In part A (i) most responses addressed and understood the probability of a compound event. Most responses correctly calculated the probability and showed work by adding the two individual probabilities to attain a correct value for the event ("fewer than 3"). A small number of responses misinterpreted the event as "3 or fewer" and calculated the probability of the event as 0.62. In part A (ii) most responses correctly calculated the mean of a discrete random variable. Several responses, however, failed to consider the weighted probabilities assigned to each value and incorrectly calculated the mean of the integer values instead.
- In part B (i) many responses provided the correct hypotheses for a one-sample *t*-test for a population mean. Several responses, however, either assumed the test was a two-sample *t*-test or assumed the scalar value used in the hypotheses could be represented by a variable. These responses incorrectly reported hypotheses that mirrored a two-sample *t*-test for means. Most responses correctly identified, through words or symbols, the two-sided alternative hypothesis. A few responses noted the sample mean of 3.1 bedrooms from part A (ii) and used it as justification to report a right-tailed alternative hypothesis. Many responses included the correct context of the hypotheses by defining the parameter used in their notation. Where a response failed to provide enough context, it was due primarily to the lack of a complete description of the population.
- In part B (ii) most students failed to properly define a Type I error in context. These responses typically erred by either misunderstanding the conditional aspect of the Type I error (dependent upon H_o being true), interpreting the error as a Type II error, or failing to cite the response variable "number of bedrooms" as context.

- In part C most responses provided a conclusion to the hypothesis test. Several responses reported this conclusion using deterministic language like "conclude," "there *is* a difference," or "Rodney is correct." A few responses provided a conclusion that was not in the context of Rodney's alternative hypothesis and, instead, referred to the null hypothesis. These responses referred to rejecting the null hypothesis in context. Most responses correctly justified their conclusion by referencing the value of the null hypothesis (2.9) falling outside of the boundaries of the given confidence interval $(3.01,\ 3.19)$.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part A (i) the response demonstrated a misunderstanding the compound event of "fewer than 3" as "3 or fewer."	• $P(\text{fewer than 3}) = P(1 \text{ or } 2) = 0.12 + 0.22 = 0.34$
• In part A (ii) the calculation of the mean of a discrete random variable used the weighted probabilities associated with each value.	• $\mu = 1(0.12) + 2(0.22) + \ldots + 6(0.02) = 3.1$ bedrooms
• In part B (i) the description of the hypotheses of a one sample *t*-test was in terms of a two-sample *t*-test ($H_0 : \mu_{2017} = \mu_{2024}$ vs $H_a : \mu_{2017} \neq \mu_{2024}$).	• $H_0 : \mu = 2.9$ vs. $H_a : \mu \neq 2.9$ where μ is the population mean number of bedrooms in newly built houses in 2024 • $H_0 : \mu_{2024} = \mu_{2017}$ vs. $H_a : \mu_{2024} \neq \mu_{2017}$ where $\mu_{2017} = 2.9$ and μ_{2024} is the population mean number of bedrooms in newly built houses in 2024
• In part B (i) the response did not identify the parameter in a one-sample *t*-test for the mean as a population mean. For example, "H_0: the mean number of bedrooms in newly built houses in 2024 is equal to 2.9. vs. H_a: the mean number of bedrooms in newly built houses in 2024 is not equal to 2.9."	• H_0: the true mean number of bedrooms in newly built houses in 2024 is equal to 2.9 vs. H_a: the true mean number of bedrooms in newly built houses in 2024 is not equal to 2.9.
• In part B (ii) the response interpreted a Type I error without the conditional component of the error. For example, "A type I error would be concluding that the mean number of bedrooms in newly built houses in country B is not 2.9."	• A Type I error would be finding that the population mean number of bedrooms in newly built houses in 2024 in Country B is not equal to 2.9, when the mean really is 2.9.
• In part B (ii) the response interpreted a Type I error as a Type II error. For example, "A type I error would be concluding that the mean number of bedrooms in newly built houses is 2.9, when it truly is not 2.9."	• A Type I error would be determining the population mean number of bedrooms in newly built houses in 2024 in Country B is not equal to 2.9, when the mean really is 2.9.

• In part C the conclusion of the hypothesis test was stated in terms of the null hypothesis, instead of the alternative hypothesis. For example, "We don't have sufficient evidence to suggest the true mean number of bedrooms in newly built houses in 2024 is 2.9."	• We have sufficient evidence to suggest that the true mean number of bedrooms in newly built houses in 2024 is not equal to 2.9.
• In part C the response used the interpretation of the confidence interval as the conclusion of a hypothesis test. For example, "I am 97% confident that the true mean number of bedrooms in newly built houses in 2024 is captured in the interval from 3.01 to 3.19."	• We have sufficient evidence to suggest that the true mean number of bedrooms in newly built houses in 2024 is not equal to 2.9.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Provide opportunities for students to not only solve but also write out in words a probability event.
- Provide opportunities for students to carry out practical expected value problems based on real-word contexts like games of chance, raffles, and insurance premiums to encourage retention of the concept.
- Provide opportunities for students to practice identifying inference problems given a written description of the context.
- Encourage students to provide complete context for the parameter when defining hypotheses. This includes the parameter type (e.g., mean or proportion), response variable, sampling units, and a reference to the population.
- Provide opportunities for students to practice writing Type I and Type II errors in various types of hypotheses testing situations.
- Encourage students to use a 2 x 2 table to visualize, organize, and retain the concepts of statistical errors and power.
- Provide opportunities for students to make connections between confidence intervals and their corresponding hypothesis tests.
- Encourage students to clearly communicate that the data provide (or do not provide) convincing evidence to support the alternative hypothesis (and not just state that there is evidence to reject the null hypothesis).

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description*, effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Pages 225–226, 229, and 232 of the CED provide examples of key questions and instructional strategies designed to develop Skill 1.B (identify key and relevant information to answer a question or solve a problem), Skill 1.F (identify null and alternative hypotheses), Skill 3.A (determine relative frequencies, proportions, or probabilities using simulation or calculations), Skill 3.B (determine parameters for probability distributions), and Skill 4.D (justify a claim based on a confidence interval).
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Graphic Organizer strategy, for example, may be helpful for students to think of how the content is aligned. This FRQ has a one-sample *t*-test for a population mean and a one-sample *t*-interval for a population mean linked together, and students could create a visual representation of the connection.

- AP Classroom provides five videos focused on the content and skills needed to answer this question.
    - The daily video 2 for Topic 4.6 demonstrates how to properly calculate the probability of the union of two events (see VAR-4.E.4). The key takeaway from this video is how to calculate the union of probabilities when the two events are mutually exclusive (part A (i)).
    - The daily video for Topic 4.8 explains how to solve the mean for a random variable (see VAR-5.C.2), which is relevant to part A (ii).
    - The daily video 1 for Topic 8.5 explores how to identify the hypotheses (see VAR-7.C.1) for a one-sample *t*-test for a population mean. The key takeaway from this video is how to identify the null value. This video will help students with part B (i).
    - The daily video 1 for Topic 6.7 explores Type I errors (see UNC-5.A.1). The key takeaway from this video is that students can easily learn the difference between the two types of errors and how to interpret them.
    - The daily video 1 for Topic 7.3 illustrates how to conclude the confidence interval (see UNC-4.T.1) for a one-sample *t*-test for a population mean. The key takeaway from this video is the explanation of what the values in the confidence interval mean, which is relevant to part C.
- AP Classroom also provides topic questions for formative assessment of Topics 4.6, 4.8, 6.7, 7.3, and 7.4, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give students practice verifying conditions, enter the keywords “confidence intervals” in the search bar, and then select the drop-down menu for “Resource Library.” Filter for “Classroom-Ready Materials” to find worksheets, data sets, practice questions, and guided notes, among other resources.

Question 6

Task: Investigative Task
Topic: Cohen's D
Max Score: 4.00
Mean Score: 1.78

What were the responses to this question expected to demonstrate?

The primary goals of the question were to assess a student's ability to (1) compare the *p*-value to a significance level to make a decision regarding the hypotheses, (2) determine an appropriate conclusion for a hypothesis test for a two-sample *t*-test for the difference in two population means, (3) explain why it was appropriate to conduct a two-sample *t*-test for the difference in two population means instead of a matched pairs *t*-test for the population mean difference, (4) calculate Cohen's *d* coefficient for the two-sample *t*-test for the difference in two population means using summary statistics for reading scores, (5) describe the practical importance of the two-sample *t*-test for the difference in two population means using the calculated Cohen's *d* coefficient and guidelines for interpreting Cohen's *d* coefficient, (6) describe the effect on the Cohen's *d* coefficient if both samples' standard deviations were over a certain value, and (7) determine the practical importance of the Cohen's *d* coefficient when both samples' standard deviations were over a certain value.

This question primarily assesses skills in Skill Category 1: Selecting Statistical Methods, Skill Category 2: Data Analysis, and Skill Category 4: Statistical Argumentation. Skills required for responding to this question include 1.E: Identify an appropriate inference method for significance tests; 2.A: Describe data presented numerically or graphically; 2.C: Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response; 4.A: Make an appropriate claim or draw an appropriate conclusion; 4.B: Interpret statistical calculations and findings to assign meaning or assess a claim; and 4.E: Justify a claim using a decision based on significance tests.

This question covers the content from Unit 7: Inference for Quantitative Data: Means of the course framework in the *AP Statistics Course and Exam Description*. Refer to Topics 7.8 and 7.9 and Learning Objectives DAT-3.H and VAR-7.F.

How well did the responses address the course content related to this question? How well did the responses integrate the skill(s) required on this question?

- In part A most responses provided a correct comparison of the *p*-value to alpha and provided a correct decision about the null and/or alternative hypothesis. Many responses did not provide a complete conclusion in context; some did not include a reference to means and others did not provide sufficient context (e.g., not referencing the groups, sampling units, and/or variable of interest). Of the responses that provided a conclusion, a majority correctly used nondefinitive language. Of the responses that chose to add an interpretation of the *p*-value, a majority of the interpretations were incorrect.
- In part B a few responses indicated the groups of children were independent; therefore, a two-sample *t*-test was appropriate. However, there were responses that indicated the groups of children were not paired in a meaningful way. Many responses provided an example of how the data could have been paired (e.g., the same children participated at 9 a.m. and 3 p.m.). However, of those responses, many seemed to imply their example was the only way the children could have been paired. Very few responses indicated the use of random assignment as a reason why the groups were independent.
- In part C (i) many responses correctly computed the value of Cohen's *d*. Some responses had arithmetic errors, some did not square the standard deviations in the pooled standard deviation formula, and some omitted the absolute value in the Cohen's *d* formula. In part C (ii) a majority of responses indicated the correct level of practical importance.

- In part D (i) many responses stated the value of Cohen's *d* would be smaller and provided a correct reason based on the increase in pooled standard deviation. Of those that provided a correct reason, a majority focused on the denominator of the Cohen's *d* formula and few provided a conceptual reason (e.g., the pooled standard deviation would increase and therefore show a lower effect size). Some responses provided an example calculation of a new Cohen's *d* with increased standard deviations. In part D (ii) many responses stated that the decrease in Cohen's *d* would indicate there was less practical importance than the original study results. Some responses used the categories in Table 2 and incorrectly indicated the new Cohen's *d* would stay in the same category of practical importance.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part A some responses failed to explicitly compare the *p*-value to alpha.	• Because the *p*-value of 0.002 is less than the level of significance of 0.05, the null hypothesis should be rejected.
• In part A many responses failed to reference means in their conclusion.	• There is convincing statistical evidence of a difference between the mean reading score for all children, similar to those who participated in the study, who would read the story at 9 a.m. and the mean reading score for all children, similar to those who participated in the study, who would read the story at 3 p.m.
• In part A many responses did not provide sufficient context.	• There is convincing statistical evidence of a difference between the mean reading score for all children, similar to those who participated in the study, who would read the story at 9 a.m. and the mean reading score for all children, similar to those who participated in the study, who would read the story at 3 p.m.
• In part B many responses did not reference independence or random assignment.	• It was appropriate for Stefan to conduct a two-sample *t*-test instead of a paired *t*-test because the two groups are independent. Stefan used random assignment to place the 100 volunteer children into the two groups.
• In part B many responses provided an example of how the data could have been paired and implied the example was the only way the data could have been paired.	• There is no indication that the two groups of 50 children are paired in any meaningful way (e.g., age or reading comprehension level).
• In part C (i) some responses did not square the standard deviations in the pooled standard deviation formula.	• $s_p = \sqrt{\frac{(4.12)^2 + (4.43)^2}{2}} \approx 4.28$
• In part C (i) some responses omitted the absolute value in the Cohen's *d* formula.	• $d = \frac{\lvert 15.2 - 17.9 \rvert}{4.28} \approx 0.63.$

• In part C (i) some responses had arithmetic or transcription errors.	• $s_p = \sqrt{\frac{(4.12)^2 + (4.43)^2}{2}} \approx 4.28$ • $d = \frac{\lvert 15.2 - 17.9 \rvert}{4.28} \approx 0.63.$
• In part D (i) some responses provided a reason for the change in the value of Cohen's *d* based on the increase in the two group standard deviations rather than the pooled standard deviation.	• The value of Cohen's *d* would decrease. If the standard deviations for the a.m. group and for the p.m. group were both greater than 4.43, the pooled standard deviation would be greater than 4.28. With a larger value in the denominator and the same value in the numerator, the value of Cohen's *d* would be smaller than 0.63.
• In part D (ii) some responses used the categories in Table 2 and incorrectly indicated the new Cohen's *d* would stay in the same category of practical importance.	• The lower Cohen's *d* would indicate less practical importance than that of the original results of Stefan's study.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve student performance on the exam?

- Encourage students to make a direct comparison of the *p*-value to a (provided) significance level and use the result of that comparison to support their conclusion.
- Encourage students to clearly communicate that the data provide (or do not provide) convincing evidence to support the alternative hypothesis (and not just that there is not convincing evidence to reject the null hypothesis).
- Encourage students to provide complete context in their conclusion, including the sampling units, parameters, groups, and variable of interest.
- Encourage students to use nondefinitive language when stating a conclusion for a hypothesis test.
- Discourage students from providing an interpretation of the *p*-value as part of a hypothesis test procedure.
- Compare and contrast different study designs (e.g., completely randomized versus paired), and discuss how to link the appropriate inference procedures (e.g., two-sample *t*-test for the difference in two population means versus a paired *t*-test for the population mean difference) to each design.
- Encourage students to show their work when asked to do a computation.
- Encourage students to reflect on an answer to determine if the value is valid, such as noticing that a measure of effect size like Cohen's *d* should be positive when they are provided with a table that includes only positive ranges of Cohen's *d* values.
- Provide students with practice sets that require them to work on clear communication.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description*, effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Pages 226, 227, and 231–232 of the CED provide examples of key questions and instructional strategies designed to develop Skill 1.E (identify an appropriate inference method for significance tests), Skill 2.A (describe data presented numerically or graphically), Skill 2.C (calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response), Skill 4.A (make an appropriate claim or draw an appropriate conclusion), Skill 4.B (interpret statistical calculations and findings to assign meaning or assess a claim), and Skill 4.E (justify a claim using a decision based on significance tests).
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213–223 of the CED. The Marking the Text strategy, for example, may be helpful in getting students to think of the investigative task in diverse ways. This FRQ could be used as an example to help students identify valuable information in the text and make notes about the interpretation of tasks required and concepts to apply to reach a solution. It will also help students understand the format of the investigative task better.
- AP Classroom provides four videos focused on the content and skills needed to answer this question.
    - The daily video 2 for Topic 7.9 shows how to conclude a two-sample *t*-test for a difference of two population means (see DAT-3.H.1 and DAT-3.H.2). The key takeaway from this video is the explanation of how to compare the *p*-value to the level of significance leading to the decision about the null hypothesis (part A).
    - The daily video 2 for Topic 7.4 demonstrates how to identify a two-sample *t*-test for a difference of two population means (see VAR-7.F.1). The key takeaway from this video describes how to correctly ensure that both groups are independent (part B).
- AP Classroom also provides topic questions for formative assessment of Topics 7.8 and 7.9, as well as access to the question bank, which is a searchable database of past AP questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give students practice verifying conditions, enter the keywords "matched pairs *t*-test" in the search bar, and then select the drop-down menu for "Resource Library." Filter for "Classroom-Ready Materials" to find worksheets, data sets, practice questions, and guided notes, among other resources.