

Question 1: Focus on Exploring Data

4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

Model Solution	Scoring
A The distribution of gas mileage for the sample of cars manufactured in Country A has a lower center than the distribution of gas mileage for the sample of cars manufactured in Country B. The median gas mileage for the sample of cars manufactured in Country A (18 mpg) is less than the median gas mileage for the sample of cars manufactured in Country B (32 mpg). The range of the gas mileages for the sample of cars manufactured in Country A (24 mpg) is slightly greater than the range of the gas mileages for the sample of cars manufactured in Country B (22 mpg). However, the IQR of the gas mileages for the sample of cars manufactured in Country A (8 mpg) is less than the IQR of the gas mileages for the sample of the cars manufactured in Country B (12 mpg). The car manufactured in Country A with 38 mpg (the maximum of the sample of cars manufactured in Country A) is an outlier, while the distribution of gas mileage for the sample of cars manufactured in Country B has no outliers.	Essentially correct (E) if the response satisfies at least three of the following four components: - Directly compares the center for the two distributions - Directly compares the spread (either IQR or range) for the two distributions - Indicates that the gas mileage of one of the cars manufactured in Country A is an outlier - Provides sufficient context, which includes the manufacturing countries (“Country A” and “Country B”) AND the dependent variable (“gas mileage” or “mpg”) Partially correct (P) if the response satisfies only two of the four components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- The response need not include specific numerical values to satisfy any given component or to score E on part A.
- A response that only uses the “means” in the comparison of center would not satisfy component 1.
- Any acceptable mention of shape in the response should be ignored because complete shape information cannot be determined from a boxplot. Acceptable mentions of shape include:
 - The shape of the distribution of the gas mileage for cars manufactured in Country A can be described as skewed, positively skewed, or right skewed.
 - The shape of the distribution of the gas mileage for cars manufactured in Country B can be

- If the response describes the shape of either distribution as just “symmetric,” “normal,” “unimodal,” or an incorrect shape (e.g., “the distribution of gas mileages for Country A is left skewed” or “the distribution of gas mileages for Country B is right skewed”), then part A cannot be scored E.
- A response that only lists values for center and/or spread and does not directly compare them does not satisfy components 1 and/or 2.
- A response that just refers to “A” or “B” AND the dependent variable (“gas mileage” or “mpg”) may satisfy component 4.

Model Solution	Scoring
B 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, the median of the distribution. Because the distribution of gas mileage for the sample of cars manufactured in Country A has an outlier to the right (or is skewed to the right), the mean of the distribution (which is not resistant) is expected to be pulled above the median (which is resistant) toward the higher values of gas mileage.	Essentially correct (E) if the response satisfies the following three components: - Indicates that 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 - Indicates 18 as the value of the median - Provides a reasonable justification for the relationship between the mean and the median based on a right-skewed distribution or the expectation that an outlier in the right tail will pull the mean above the median Partially correct (P) if the response satisfies component 1 AND either component 2 or component 3. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- A response that states the median is 18 mpg in either part A or part B satisfies component 2.
- A response that provides a weak justification (e.g., “because the distribution is skewed”) does not satisfy component 3. However, a response that provides a justification based on the distribution being right skewed may satisfy component 3.
- Component 3 may be satisfied with an appropriate numerical argument that shows that the mean must be greater than 18 mpg.
 - An example of an inappropriate numerical argument would be averaging portions of the five-number summary.

Model Solution	Scoring
C i. The maximum value in the combined data is 40 mpg because 40 mpg is the maximum gas mileage for the sample of cars manufactured in Country B, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country A are less than 40 mpg. The minimum value in the combined data is 14 mpg, because 14 mpg is the minimum mpg for the sample of cars manufactured in Country A, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country B are greater than 14. Thus, the range of the combined data set is $40 - 14 = 26$ mpg. ii. In the combined data, there are 200 gas mileages. The median is a value where at least half, or 100, of the gas mileages in the combined data are less than or equal to the median value and at least half, or 100, of the gas mileages in the combined data are greater than or equal to the median value. From the boxplot for the sample of cars manufactured in Country A, the third quartile, Q3, is 24 mpg indicating there are at least 75 gas mileages less than or equal to 24 mpg and at least 25 gas mileages greater than or equal to 24 mpg. From the boxplot for the sample of cars manufactured in Country B, the first quartile, Q1, is 24 mpg indicating there are at least 25 gas mileages less than or equal to 24 mpg and at least 75 gas mileages greater than or equal to 24 mpg. Thus, in the combined data set, there are at least 100 gas mileages less than or equal to 24 mpg and at least 100 gas mileages greater than or equal to 24 mpg, which implies 24 is the value of the median of the combined data set.	Essentially correct (E) if the response satisfies the following four components: 1. Correctly calculates 26 mpg as the range of the combined data set 2. Provides a justification for the range with an argument based on identifying 40 as the maximum of the combined data set and identifying 14 as the minimum of the combined data set 3. Calculates 24 mpg as a possible value of the median of the combined data set 4. Provides a justification for the median by indicating that at least half of the combined data values are at most 24 and at least half of the combined data values are at least 24 Partially correct (P) if the response satisfies only two or three of the four components. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- A response that solely indicates a range of values (e.g., “the gas mileages range from 14 mpg to 40 mpg”) does not satisfy component 1 but may satisfy component 2.
- A response that states the range as a value of 25 or 27 from misinterpreting the minimum value in the distribution of gas mileages for Country A as 13 or 15 may be considered a minor error and satisfies components 1 and 2.

- A response that gives the correct value for the median (24) and justifies the value by looking at the quartiles on one side of 24 (e.g., identifying that one-fourth of Country B cars and three-fourths of Country A cars have mpg less than or equal to 24) satisfies component 4.
- A response that provides justification of the median based on a value near the 100th data value satisfies component 4.
- A response that provides justification of the median based on counting the segments in the boxplots may satisfy component 4. The response is not required to specify equal sample sizes.
 - For example, one-half of the boxplot segments are no greater than 24 mpg (four of eight segments—three segments from Country A and one segment from Country B) and one-half of the boxplot segments are at least 24 mpg (four of eight segments—one segment from Country A and three segments from Country B), which indicates that 24 mpg is the median of the combined data set.

Scoring for Question 1	Score
Complete Response Three parts essentially correct	4
Substantial Response Two parts essentially correct and one part partially correct	3
Developing Response Two parts essentially correct and no part partially correct <i>OR</i> One part essentially correct and one or two parts partially correct <i>OR</i> Three parts partially correct	2
Minimal Response One part essentially correct and no parts partially correct <i>OR</i> No part essentially correct and two parts partially correct	1

Question 6: Investigative Task

4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

Model Solution	Scoring
(a) (i) The inference procedure that should be used to estimate the mean price, in dollars (\$), of this type of whistle at all stores that sell the whistle is a one-sample t-interval for a population mean. (ii) The parameter of interest is the mean whistle price, in dollars (\$), of this type of whistle at all stores that sell the whistle.	Essentially correct (E) if the response satisfies the following three components: - In part (a-i) the response identifies a one-sample t-interval for a population mean by name (e.g., “one sample t-interval”) - In part (a-ii) the response identifies the correct parameter by including reference to the mean - In part (a-ii) the response provides sufficient context for the parameter by including reference to the population (all stores that sell the whistle) and the variable of interest (the price of this type of whistle) Partially correct (P) if the response only satisfies one or two of the three components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- In part (a-i) a response that refers to a “test” does not satisfy component 1.
- In part (a-i) a response that refers to “a mean,” singular, satisfies the one-sample aspect of component 1. If the response states “one-sample,” it is not required to also include “for a mean” to satisfy component 1.
- A response that addresses “mean” in part (a-i) does not satisfy component 2 unless it is also addressed in part (a-ii).
- In part (a-ii) a response may satisfy the variable of interest aspect of component 3 with a minimum of “price” and “whistle.”
- In part (a-ii) a response may satisfy the population aspect of component 3 by using words such as: “population,” “all,” or “true,” or by using the symbol μ .
- In part (a-ii) if the response clearly refers to the sample mean instead of the population mean using words or symbols (e.g., $\bar{x}$), then component 3 is not satisfied unless the symbol used is defined as the population mean.
- For responses that are scored P, the number of components satisfied should be considered if holistic scoring is required.

Model Solution	Scoring
(b) (i) The distribution of the sample of whistle prices appears slightly skewed to the right, because the mean is slightly higher than the median. (ii) Based on the $1.5 \times \text{IQR}$ rule, there are not whistle prices in this sample that would be considered outliers. A whistle price is an outlier using this method if it is more than $1.5 \times \text{IQR}$ below the first quartile (Q_1) or more than $1.5 \times \text{IQR}$ above the third quartile (Q_3). Because $Q_1 - 1.5 \times \text{IQR} = 4.51 - 1.5(5.475 - 4.51) = 3.0625,$ and the minimum value (4.25) is greater than 3.0625, there are no outliers to the left. Because $Q_3 + 1.5 \times \text{IQR} = 5.475 + 1.5(5.475 - 4.51) = 6.9225,$ and the maximum value (6.58) is less than 6.9225, there are no outliers to the right.	Essentially correct (E) if the response satisfies the following four components: - In part (b-i) the response indicates the distribution is skewed to the right or approximately symmetric - In part (b-i) the response justifies the response to component 1 with appropriate reasoning based on the summary statistics - In part (b-ii) the response provides a justification for stating there are no outliers by correctly calculating the lower and upper outlier criteria with work shown - In part (b-ii) the response states whether there are outliers based on the calculated lower and upper outlier criteria Partially correct (P) if the response satisfies three of the four components required for E. <i>OR</i> satisfies components 1 and 2 <i>OR</i> components 3 and 4 of components 1–4 required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

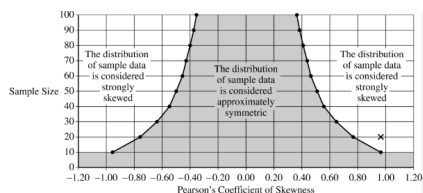
- If the response states skewed right, to satisfy component 2 the response should refer to the median being less than the mean. If the response states approximately symmetric, the response should refer to the median and mean being close together.
- A response that indicates skewed left does not satisfy components 1 or 2.
- If the response says the distribution is skewed to the right, examples of responses that satisfy component 2 include:
 - the mean is higher than the median,
 - the ratio between the mean and median is greater than 1,
 - the difference between the maximum and median is greater than the difference between the median and minimum, and
 - the difference between the maximum and third quartile is greater than the difference between the first quartile and minimum.
- For responses that score P, the number of components satisfied should be considered if holistic scoring is required.

Model Solution

Scoring

- (c) (i) Pearson's coefficient of skewness is $\frac{3(5.12 - 4.885)}{0.743} \approx 0.949$.

(ii)



Essentially correct (E) if the response satisfies the following two components:

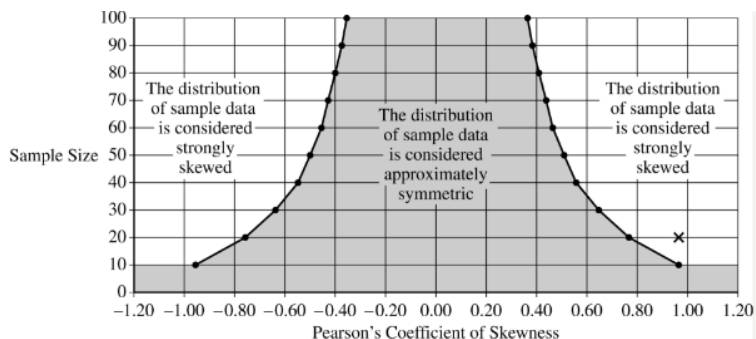
1. In part (c-i) the response calculates Pearson's coefficient of skewness with work shown
2. In part (c-ii) the response includes an x at approximately (0.949, 20) or an x consistent with the calculation in part (c-i)

Partially correct (P) if the response satisfies only one of the two components required for E.

Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- Responses must include an x that is located on the line for sample size of 20 and located between 0.90 and 1.00 on the horizontal axis or be consistent with the value given in part (c-i) to satisfy component 2.



Model Solution

Scoring

- (d) (i) Looking at the graph in part (c), for a sample size of 20, and a skewness coefficient of 0.949, this point falls in "the distribution of sample data is considered strongly skewed" region. Therefore, we would consider the shape of the distribution of the sample of whistle prices to be strongly skewed.
- (ii) No, based on the response to part (d-i), Julio's data would not satisfy the normality condition because neither of the criteria listed are met. Julio only has a sample size of 20, which is less than 30, and Pearson's coefficient of skewness indicates the distribution of sample data is strongly skewed.

Essentially correct (E) if the response satisfies the following five components:

1. In part (d-i) the response indicates the sample is strongly skewed or makes a statement consistent with the graph in part (c)
2. In part (d-i) the response uses or refers to the graph or the Pearson coefficient of skewness to make the correct determination of skewness
3. In part (d-ii) the response states that the condition of normality is not met or makes a statement consistent with the explanation in part (d-i)
4. In part (d-ii) the response explains the sample size of 20 is less than 30
5. In part (d-ii) the response explains that the distribution of sample data is strongly skewed or makes a statement consistent with the explanation in part (d-i)

Partially correct (P) if the response does not meet the criteria for E but satisfies three or four of components 1–5 required for E.

Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that determines the sample is strongly skewed, but says the normality condition is satisfied, should be scored I.
- In part (d-i) a response that indicates the distribution is strongly skewed to the left does not satisfy component 1.
- If a response in part (c-ii) has an x on the curve, the response in part (d-i) must indicate the sample is skewed (not strongly skewed or approximately symmetric), or indicate it is not possible to choose whether it is strongly skewed or approximately symmetric to satisfy component 1.
- In part (d-i) a response that indicates the distribution is skewed, but not strongly skewed, because the Pearson coefficient of skewness is not that far from the curve on the graph satisfies component 1 and component 2.
- In part (d-i) a response that indicates the distribution is skewed, but not strongly skewed, may satisfy components 3–5 if in part (d-ii) the response indicates the conditions are met because the distribution is not very skewed and there are no outliers. The response does not need to indicate the sample size is less than 30.
- In part (d-ii) a response that indicates the sample size is small without explicitly referring to the values 20 and 30 may satisfy component 4.
- The quality of communication, or the number of components satisfied for responses with score P, should be considered if holistic scoring is required.

Scoring for Question 6

Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as $\frac{1}{2}$ point.

	Score
Complete Response	4
Substantial Response	3
Developing Response	2
Minimal Response	1

If a response is between two scores (for example, 2 $\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.

Question 6: Investigative Task**4 points****General Scoring Notes**

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

Model Solution	Scoring
(a) The boxplots reveal that the team tended to have a higher average per-game attendance during years in the new stadium than during years in the old stadium because the median value of about 25,000 attendees per game during years in the new stadium is greater than the median value of about 16,000 attendees per game during years in the old stadium. The interquartile ranges (IQRs) are similar, which indicates that variability in average per-game attendance is similar during the years in the two stadiums, but the range of average per-game attendance is slightly larger during the years in the new stadium. There are no years with average per-game attendance identified as an outlier for either stadium.	Essentially correct (E) if the response satisfies the following three components: - Indicates that the median average per-game attendance is greater in the new stadium <i>OR</i> Indicates that the average per-game attendance is usually (typically) greater in the new stadium - Indicates that the variability in average per-game attendance is greater in the new stadium <i>OR</i> Indicates that the variability in the two stadiums is roughly the same according to the IQRs - Provides sufficient context by including the response variable (average attendance or attendance) or the units of the response variable (thousands of people or people) Partially correct (P) if the response satisfies only two of the three components <i>OR</i> if the response satisfies only component 1 <i>OR</i> if the response satisfies only component 2. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 1, the response may refer only to “center” and does not need to specify which measure of center (e.g., median, mean) is being used.

- To satisfy component 2, a response may refer only to “variability” or “spread” and does not need to specify which measure of variability (e.g., range or IQR) is being used. However, if the response states that the variability is about the same, the response must explicitly refer to the IQRs.
- To satisfy components 1 and 2, the stadiums must be identified (e.g., “old,” “new”), and an explicit comparison phrase (e.g., “greater than,” “about the same as”) must be used. Separate lists of characteristics alone or summary statistics alone do not count as a comparison.
- To satisfy components 1 and 2, numerical values are not required. However, if they are included, they should be reasonably correct. Numerical values can be reported in units of people (e.g., median = 16,000) or in thousands of people (e.g., median = 16).
- Any mention of shape is ignored in the scoring of part (a) because complete shape information cannot be obtained from a boxplot. Statements about shape not clearly supported by the boxplots (e.g., “the old stadium distribution is approximately normal”) should be considered a negative in terms of holistic scoring. However, statements about shape that are supported by the boxplots (e.g., “the old stadium distribution is roughly symmetric,” “the new stadium distribution is skewed to the left”) should be considered a positive.

Model Solution	Scoring
(b) During the years in the new stadium, the average per-game attendance increases linearly, from about 16,000 people in 2000 to about 27,000 people in 2016. However, during the years in the old stadium, there is no obvious increasing or decreasing trend over time for the average per-game attendance. The average attendance appears to vary about an average of approximately 16,000 attendees per game from 1970 to 1999.	Essentially correct (E) if the response satisfies the following three components: 1. Describes the direction of the trend in average per-game attendance in the new stadium as increasing (positive) 2. Describes the direction of the trend in average per-game attendance in the old stadium as relatively constant (e.g., “no association,” “flat”) <i>OR</i> Describes the direction of the trend in the old stadium as positive but less steep (e.g., “less positive,” “flatter”) than the trend in the new stadium 3. Provides sufficient context by including the two groups (old stadium, new stadium) AND the explanatory variable (time or year) AND the response variable (average attendance or attendance) or the units of the response variable (thousands of people or people) Partially correct (P) if the response satisfies only two of the three components Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- Only describing an association or correlation as “strong” or “weak” addresses strength and not the direction of the trend and does not satisfy components 1 or 2.
- Only describing an association as “linear” or “non-linear” addresses form and not the direction of the trend and does not satisfy components 1 or 2.
- Numerical values, including years (e.g., “from 2000 to 2016”), are not required for any component. However, a response that includes years in numerical form (e.g., “1970”) satisfies the context requirement for the explanatory variable in component 3.
- Component 1 can also be satisfied if the response provides an estimated value for the correlation for the new stadium that is positive.
- Component 2 can also be satisfied if the response provides an estimated value for the correlation for the old stadium of 0 (or approximately 0). Providing positive correlations for both stadiums does not satisfy component 2 because it is impossible to compare the steepness of the trends using their correlations.

Model Solution	Scoring
(c) (i) Graph I indicates a strong, positive, linear relationship between average per-game attendance and the number of games won during the 47 years of the team's existence. Average per-game attendance increases linearly, with an average increase of about 500 attendees per game for each additional game won. Variation about the linear trend in attendance is relatively small and about the same for any number of games won. (ii) No. Graph II suggests that the rates at which average per-game attendance increases as the number of games won increases are about the same for the two stadiums. A line drawn through the points for the old stadium has about the same slope as (or may have a slightly larger slope than) a line drawn through the points for the new stadium.	Essentially correct (E) if the response satisfies four or five of the following five components: 1. In part (c-i) describes the direction of the relationship as positive 2. In part (c-i) describes the form of the relationship as linear or nearly linear 3. In part (c-i) describes the strength of the relationship as very strong, strong, or moderately strong 4. In part (c-ii) indicates that the rates are about the same for the two stadiums (or slightly larger for the old stadium) 5. In part (c-ii) provides an explanation that indicates that if a line were drawn through the points for the old stadium, the slope would be roughly the same (or slightly greater than) the slope of a line through the points for the new stadium Partially correct (P) if the response satisfies only three of the five components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that provides an estimated value of the correlation satisfies component 1 if the estimated correlation is positive, but an estimated correlation cannot satisfy components 2 or 3.
- A response need not include the word “positive” to satisfy component 1. For example, “the average attendance is higher when the team has more wins” satisfies component 1. Likewise, a response need not include the word “strong” to satisfy component 3. For example, “variation about the linear trend in attendance is relatively small” satisfies component 3.
- Correct comments on homogeneous variation in part (c-i) (i.e., the variation about the linear trend in attendance is about the same for any number of games won) should be considered a positive in terms of holistic scoring.
- Responses that satisfy all 5 components should be considered a positive in terms of holistic scoring.
- Context is not required in part (c) because it has already been assessed in parts (a) and (b).

Model Solution	Scoring
(d) The number of games won could be a confounding variable for assessing the potential effect of opening the new stadium on average per-game attendance. The boxplots in part (a) show that average per-game attendance tended to be higher for games in the new stadium than for games in the old stadium, but the cause of the increase may actually be that attendees were more excited about attending games for teams that were better at winning. The scatterplots in part (c) show that average per-game attendance has a strong positive correlation with games won, and the team tended to win more games playing in the new stadium than in the old stadium.	Essentially correct (E) if the response provides an explanation that satisfies the following four components: 1. States that there is an association between attendance and one of the explanatory variables (stadium, year, wins) 2. States that there is an association between attendance and a different one of the explanatory variables (stadium, year, wins) 3. States that there is an association between the two explanatory variables (stadium, year, wins) identified in components 1 and 2 4. Explains the idea of confounding by describing that the variable identified as a potential confounding variable could be the cause of the association between attendance and the other explanatory variable identified in components 1 and 2 OR Explains the idea of confounding by stating that it is impossible to know which of the two explanatory variables identified in components 1 and 2 may be the cause of the increase in attendance Partially correct (P) if the response satisfies only three of the four components OR if the response satisfies only two of the four components and justifies at least one of the statements in components 1 through 3 by referring to the appropriate graph from parts (a) through (c) (e.g., “based on the boxplots,” “in part (b)”). Incorrect (I) if the response does not otherwise meet the criteria for E or P.

Additional Notes:

- The response can use any combination of 2 of the 3 explanatory variables (stadium and wins, stadium and year, year and wins). The response cannot introduce a new variable (e.g., weather, having popular players) to satisfy any component.
- An incorrectly described association cannot be used to satisfy components 1 through 3.
- To satisfy component 4 the response must discuss all three variables: the response variable (attendance)
- The strength of the response to part (d), especially communication in component 4, can be considered in holistic scoring. A well-communicated response to part (d) should be considered a positive. A poorly-communicated response to part (d), should be considered a negative.

Scoring for Question 6

Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as $\frac{1}{2}$ point.

Score**Complete Response****4****Substantial Response****3****Developing Response****2****Minimal Response****1**

If a response is between two scores (for example, 2 $\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.