AP Principles 2.2 2.3 2.4 MCQ only

1. A programmer is writing a program that is intended to be able to process large amounts of data. Which of the following considerations is LEAST likely to affect the ability of the program to process larger data sets?

    (A)   How long the program takes to run

    (B)   How many programming statements the program contains

    (C)   How much memory the program requires as it runs

    (D)   How much storage space the program requires as it runs

2. A cable television company stores information about movie purchases made by subscribers. Each day, the following information is summarized and stored in a publicly available database.

    • The day and date each movie was purchased

    • The title of each movie purchased

    • The cities where subscribers purchased each movie

    • The number of times each movie was purchased by subscribers in a given city

    A sample portion of the database is shown below. The database is sorted by date and movie title.

Day and Date	Movie Title	City	Number of Times Purchased
Sat 01 / 05 / 2014	Movie A	Houston, Texas	1
Sat 01 / 05 / 2014	Movie A	Detroit, Michigan	2
Sat 01 / 05 / 2014	Movie B	Houston, Texas	1
Sat 01 / 05 / 2014	Movie C	Anchorage, Alaska	1
Sun 01 / 06 / 2014	Movie A	Wichita, Kansas	3

    Which of the following CANNOT be determined using only the information in the database?

    (A)   The date when a certain movie was purchased the greatest number of times

    (B)   The number of movies purchased by an individual subscriber for a particular month

    (C)   The total number of cities in which a certain movie was purchased

    (D)   The total number of movies purchased in a certain city during a particular month

AP Principles 2.2 2.3 2.4 MCQ only

3. A certain social media Web site allows users to post messages and to comment on other messages that have been posted. When a user posts a message, the message itself is considered data. In addition to the data, the site stores the following metadata.

- The time the message was posted

- The name of the user who posted the message

- The names of any users who comment on the message and the times the comments were made

For which of the following goals would it be more useful to analyze the data instead of the metadata?

(A) To determine the users who post messages most frequently

(B) To determine the time of day that the site is most active

(C) To determine the topics that many users are posting about

(D) To determine which posts from a particular user have received the greatest number of comments

4. Which of the following is an advantage of a lossless compression algorithm over a lossy compression algorithm?

(A) A lossless compression algorithm can guarantee that compressed information is kept secure, while a lossy compression algorithm cannot.

(B) A lossless compression algorithm can guarantee reconstruction of original data, while a lossy compression algorithm cannot.

(C) A lossless compression algorithm typically allows for faster transmission speeds than does a lossy compression algorithm.

(D) A lossless compression algorithm typically provides a greater reduction in the number of bits stored or transmitted than does a lossy compression algorithm.

5. A large data set contains information about all students majoring in computer science in colleges across the United States. The data set contains the following information about each student.

- The student's gender

- The state in which the student attends college

- The student's grade point average on a 4.0 scale

Which of the following questions could be answered by analyzing only information in the data set?

(A) Do students majoring in computer science tend to have higher grade point averages than students majoring in other subjects?

(B) How many states have a higher percentage of female computer science majors than male computer science majors attending college in that state?

(C) What percent of students attending college in a certain state are majoring in computer science?

(D) Which college has the highest number of students majoring in computer science?

AP Principles 2.2 2.3 2.4 MCQ only

6. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

    A programmer is developing a word game. The programmer wants to create an algorithm that will take a list of words and return a list containing the first letter of all words that are palindromes (words that read the same backward or forward). The returned list should be in alphabetical order. For example, if the list contains the words `["banana", "kayak", "mom", "apple", "level"]`, the returned list would contain `["k", "l", "m"]` (because `"kayak"`, `"level"`, and `"mom"` are palindromes).

    The programmer knows that the following steps are necessary for the algorithm but is not sure in which order they should be executed.

Step	Explanation
Shorten	Takes a list of words and returns a new list that contains only the first letter of each word from the input list
Keep palindromes	Takes a list of words and returns a list that contains only the palindromes from the input list
Sort	Takes a list of words and returns a copy of the list in alphabetical order

    Executing which of the following sequences of steps will enable the algorithm to work as intended?

    > I. First shorten, then keep palindromes, then sort

    > II. First keep palindromes, then shorten, then sort

    > III. First sort, then keep palindromes, then shorten

    (A) I only

    (B) II only

    (C) I and III

    (D) II and III

7. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

    A library of e-books contains metadata for each book. The metadata are intended to help a search feature find books that users are interested in. Which of the following is LEAST likely to be contained in the metadata of each e-book?

    (A) An archive containing previous versions of the e-book

    (B) The author and title of the e-book

    (C) The date the e-book was first published

    (D) The genre of the e-book (e.g., comedy, fantasy, romance, etc.)

AP Principles 2.2 2.3 2.4 MCQ only

8. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

A media librarian at a movie studio is planning to save digital video files for archival purposes. The movie studio would like to be able to access full-quality videos if they are needed for future projects. Which of the following actions is LEAST likely to support the studio's goal?

(A)   Using video file formats that conform to published standards and are supported across many different devices

(B)   Using lossy compression software to reduce the size requirements of the data being stored

(C)   Using storage media that can be expanded for additional data capacity

(D)   Using a system that incorporates redundancy to handle disk failure

9. A digital photo file contains data representing the level of red, green, and blue for each pixel in the photo. The file also contains metadata that describes the date and geographic location where the photo was taken. For which of the following goals would analyzing the metadata be more appropriate than analyzing the data?

(A)   Determining the likelihood that the photo is a picture of the sky

(B)   Determining the likelihood that the photo was taken at a particular public event

(C)   Determining the number of people that appear in the photo

(D)   Determining the usability of the photo for projection onto a particular color background

10. A group of students take hundreds of digital photos for a science project about weather patterns. Each photo file contains data representing the level of red, green, and blue for each pixel in the photo. The file also contains metadata that describes the date, time, and geographic location where the photo was taken. For which of the following goals would analyzing the metadata be more appropriate than analyzing the data?

Select <u>two</u> answers.

A   Determining the chronological order of the photos

B   Determining the number of clouds in a particular photo

C   Determining whether a photo is suitable for printing in black-and-white

D   Determining whether two photos were taken at the same location on different days

11. An Internet service provider (ISP) is considering an update to its servers that would save copies of the Web pages most frequently visited by each user.

Which of the following is LEAST likely to occur as a result of the update?

(A)   Average response time for user requests might decrease.

(B)   Privacy of users might be negatively affected.

(C)   Storage requirements for the servers might increase.

(D)   Web sites that are not visited frequently might no longer be accessible to users.

AP Principles 2.2 2.3 2.4 MCQ only

12. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

Grades in a computer science course are based on total points earned on a midterm exam and a final exam. The teacher provides a way for students to improve their course grades if they receive high scores on the final exam: if a student's final exam score is greater than the student's midterm exam score, the final exam score replaces the midterm exam score in the calculation of total points.

The table below shows two students' scores on the midterm and final exams and the calculated total points each student earns.

Student Name	Midterm Exam Score	Final Exam Score	Total Points Calculation
Khalil	90	80	90 + 80 = 170
Josefina	70	90	90 + 90 = 180

- Khalil does better on the midterm exam than on the final exam, so his original midterm and final exam scores are added to compute his total points.
- Josefina does better on the final exam than on the midterm exam, so her final exam score replaces her midterm exam score in the total points calculation.

A programmer is writing a procedure to calculate a student's final grade in the course using the score replacement policy described. The student's exam scores are stored in the variables `midtermExam` and `finalExam`. The procedure `Max (a, b)` returns the larger of `a` and `b`.

Which of the following could be used in the procedure to calculate a student's total points earned in the course and store the result in the variable `adjustedTotal`?

(A) `adjustedTotal ← Max (midtermExam, finalExam)`

(B) `adjustedTotal ← Max (midtermExam, finalExam) + finalExam`

(C) `adjustedTotal ← Max (midtermExam, finalExam) + midtermExam`

(D) `adjustedTotal ← 2 * Max (midtermExam, finalExam)`

AP Principles 2.2 2.3 2.4 MCQ only

13. A retailer that sells footwear maintains a single database containing records with the following information about each item for sale in the retailer's store.

- Item identification number

- Footwear type (sneakers, boots, sandals, etc.)

- Selling price (in dollars)

- Size

- Color

- Quantity available

Using only the database, which of the following can be determined?

(A) Which items listed in the database are not currently in the store

(B) Which colors are more popular among men than women

(C) Which type of footwear is most popular among adults

(D) The total number of shoes sold in a particular month

14. ASCII is a character-encoding scheme that uses 7 bits to represent each character. The decimal (base 10) values 65 through 90 represent the capital letters A through Z, as shown in the table below.

Decimal	ASCII Character	Decimal	ASCII Character
65	A	78	N
66	B	79	O
67	C	80	P
68	D	81	Q
69	E	82	R
70	F	83	S
71	G	84	T
72	H	85	U
73	I	86	V
74	J	87	W
75	K	88	X
76	L	89	Y
77	M	90	Z

What ASCII character is represented by the binary (base 2) number 1001010 ?

(A) H

(B) I

(C) J

(D) K

AP Principles 2.2 2.3 2.4 MCQ only

15. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

A search engine has a trend-tracking feature that provides information on how popular a search term is. The data can be filtered by geographic region, date, and category. Categories include arts and entertainment, computers and electronics, games, news, people and society, shopping, sports, and travel. Which of the following questions is LEAST likely to be answerable using the trends feature?

(A) In what month does a particular sport receive the most searches?

(B) In which political candidates are people interested?

(C) What is the cost of a certain electronics product?

(D) Which region of the country has the greatest number of people searching for opera performances?

16. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

A smartphone stores the following data for each photo that is taken using the phone.

- The file name of the photo
- The date and time the photo was taken
- The geographic location where the photo was taken

Assume that all of the photos that have been taken on the phone are accessible. Which of the following can be determined using the photo data described above?

    I. The number of photos that were taken at a particular geographic location
   II. The number of photos that were taken in the last year
  III. The name of the person who took the most recent photo

(A) III only

(B) I and II only

(C) I and III only

(D) I, II, and III

AP Principles 2.2 2.3 2.4 MCQ only

A file storage application allows users to save their files on cloud servers. A group of researchers gathered user data for the first eight years of the application's existence. Some of the data are summarized in the following graphs. The line graph on the left shows the number of registered users each year. The line graph on the right shows the total amount of data stored by all users each year. The circle graph shows the distribution of file sizes currently stored by all users.

REGISTERED USERS BY YEAR

TOTAL AMOUNT OF DATA STORED

FILE SIZE DISTRIBUTION
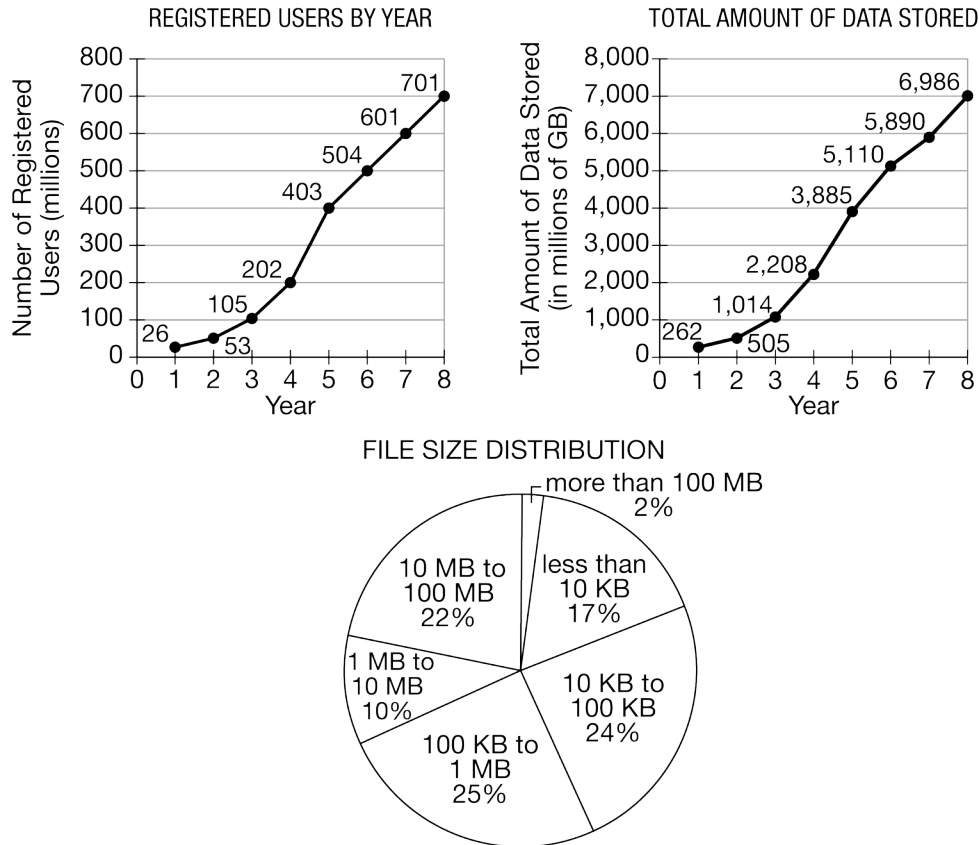
(note: 1 MB = 1,000 KB)

17. Which of the following best describes the average amount of data stored per user for the first eight years of the application's existence?

   (A) Across all eight years, the average amount of data stored per user was about 10 GB.

   (B) Across all eight years, the average amount of data stored per user was about 100 GB.

   (C) The average amount of data stored per user appears to increase by about 10 GB each year.

   (D) The average amount of data stored per user appears to increase by about 100 GB each year.

18. Which of the following observations is most consistent with the information in the circle graph?

AP Principles 2.2 2.3 2.4 MCQ only

    (A)   Over 75% of the files stored are 1 MB in size or less.

    (B)   Over 75% of the files stored are 10 MB in size or less.

    (C)   Over 75% of the files stored are at least 100 KB in size.

    (D)   Over 75% of the files stored are at least 1 MB in size.

19. Which of the following best describes the growth in the number of registered users for the first eight years of the application's existence?

    (A)   The number of registered users increased at about a constant rate each year for all eight years.

    (B)   The number of registered users increased at about a constant rate for years 1 to 5 and then about doubled each year after that.

    (C)   The number of registered users about doubled each year for years 1 to 5 and then increased at about a constant rate after that.

    (D)   The number of registered users about doubled each year for all eight years.

20. A video-streaming Web site uses 32-bit integers to count the number of times each video has been played. In anticipation of some videos being played more times than can be represented with 32 bits, the Web site is planning to change to 64-bit integers for the counter. Which of the following best describes the result of using 64-bit integers instead of 32-bit integers?

    (A)   2 times as many values can be represented.

    (B)   32 times as many values can be represented.

    (C) 2^{32} times as many values can be represented.

    (D) 32^2 times as many values can be represented.

21. Biologists often attach tracking collars to wild animals. For each animal, the following geolocation data is collected at frequent intervals.

    • The time

    • The date

    • The location of the animal

Which of the following questions about a particular animal could NOT be answered using only the data collected from the tracking collars?

    (A)   Approximately how many miles did the animal travel in one week?

    (B)   Does the animal travel in groups with other tracked animals?

    (C)   Do the movement patterns of the animal vary according to the weather?

    (D)   In what geographic locations does the animal typically travel?

AP Principles 2.2 2.3 2.4 MCQ only

22. A bookstore has a database containing information about each book for sale in the store. A sample portion of the database is shown below.

Author	Title	Selling Price	Genre	Quantity Available
J. M. Barrie	*Peter and Wendy*	$6.99	Fantasy	3
L. Frank Baum	*The Wonderful Wizard of Oz*	$7.99	Fantasy	2
Arthur Conan Doyle	*The Hound of the Baskervilles*	$7.49	Mystery	4
Mary Shelley	*Frankenstein*	$7.99	Horror	4
Jules Verne	*Twenty Thousand Leagues Under the Sea*	$6.99	Science Fiction	3
H. G. Wells	*The War of the Worlds*	$4.99	Science Fiction	3

A store employee wants to calculate the total amount of money the store will receive if they sell all of the available science fiction books. Which columns in the database can be <u>ignored</u> and still allow the employee to perform this calculation?

Select <u>two</u> answers.

A   Author

B   Title

C   Genre

D   Quantity Available

AP Principles 2.2 2.3 2.4 MCQ only

23. A large spreadsheet contains the following information about the books at a bookstore. A sample portion of the spreadsheet is shown below.

	A Book Title	B Author	C Genre	D Number of Copies in Stock	E Cost (in dollars)
1	Little Women	Louisa May Alcott	Fiction	3	13.95
2	The Secret Adversary	Agatha Christie	Mystery	2	12.95
3	A Study in Scarlet	Arthur Conan Doyle	Mystery	0	8.99
4	The Hound of the Baskervilles	Arthur Conan Doyle	Mystery	1	8.95
5	Les Misérables	Victor Hugo	Fiction	1	12.99
6	Frankenstein	Mary Shelley	Horror	2	11.95

An employee wants to count the number of books that meet all of the following criteria.

- Is a mystery book
- Costs less than $10.00
- Has at least one copy in stock

For a given row in the spreadsheet, suppose `genre` contains the genre as a string, `num` contains the number of copies in stock as a number, and `cost` contains the cost as a number. Which of the following expressions will evaluate to `true` if the book should be counted and evaluates to `false` otherwise?

(A) `(genre = "mystery") AND ((1 ≤ num) AND (cost < 10.00))`

(B) `(genre = "mystery") AND ((1 < num) AND (cost < 10.00))`

(C) `(genre = "mystery") OR ((1 ≤ num) OR (cost < 10.00))`

(D) `(genre = "mystery") OR ((1 < num) OR (cost < 10.00))`

AP Principles 2.2 2.3 2.4 MCQ only

24. A researcher is analyzing data about students in a school district to determine whether there is a relationship between grade point average and number of absences. The researcher plans on compiling data from several sources to create a record for each student.

The researcher has access to a database with the following information about each student.

- Last name
- First name
- Grade level (9, 10, 11, or 12)
- Grade point average (on a 0.0 to 4.0 scale)

The researcher also has access to another database with the following information about each student.

- First name
- Last name
- Number of absences from school
- Number of late arrivals to school

Upon compiling the data, the researcher identifies a problem due to the fact that neither data source uses a unique ID number for each student. Which of the following best describes the problem caused by the lack of unique ID numbers?

(A) Students who have the same name may be confused with each other.

(B) Students who have the same grade point average may be confused with each other.

(C) Students who have the same grade level may be confused with each other.

(D) Students who have the same number of absences may be confused with each other.

25. A team of researchers wants to create a program to analyze the amount of pollution reported in roughly 3,000 counties across the United States. The program is intended to combine county data sets and then process the data. Which of the following is most likely to be a challenge in creating the program?

(A) A computer program cannot combine data from different files.

(B) Different counties may organize data in different ways.

(C) The number of counties is too large for the program to process.

(D) The total number of rows of data is too large for the program to process.

26. A student is creating a Web site that is intended to display information about a city based on a city name that a user enters in a text field. Which of the following are likely to be challenges associated with processing city names that users might provide as input?

Select two answers.

A Users might attempt to use the Web site to search for multiple cities.

B Users might enter abbreviations for the names of cities.

C Users might misspell the name of the city.

D Users might be slow at typing a city name in the text field.

AP Principles 2.2 2.3 2.4 MCQ only

27. The owner of a clothing store records the following information for each transaction made at the store during a 7-day period.

   • The date of the transaction
   • The method of payment used in the transaction
   • The number of items purchased in the transaction
   • The total amount of the transaction, in dollars

   Customers can pay for purchases using cash, check, a debit card, or a credit card.

   Using only the data collected during the 7-day period, which of the following statements is true?

   (A) The average amount spent per day during the 7-day period can be determined by sorting the data by the total transaction amount, then adding the 7 greatest amounts, and then dividing the sum by 7.

   (B) The method of payment that was used in the greatest number of transactions during the 7-day period can be determined by sorting the data by the method of payment, then adding the number of items purchased for each type of payment method, and then finding the maximum sum.

   (C) The most expensive item purchased on a given date can be determined by searching the data for all items purchased on the given date and then sorting the matching items by item price.

   (D) The total number of items purchased on a given date can be determined by searching the data for all transactions that occurred on the given date and then adding the number of items purchased for each matching transaction.

AP Principles 2.2 2.3 2.4 MCQ only

28. Two lists, *list1* and *list2*, contain the names of books found in two different collections. A librarian wants to create *newList*, which will contain the names of all books found in either list, in alphabetical order, with duplicate entries removed.

For example, if *list1* contains

["Macbeth", "Frankenstein", "Jane Eyre"]

and *list2* contains

["Frankenstein", "Dracula", "Macbeth", "Hamlet"],

then *newList* will contain

["Dracula", "Frankenstein", "Hamlet", "Jane Eyre", "Macbeth"].

The following procedures are available to create *newList*.

Procedure	Explanation
Sort (list)	Sorts `list` in alphabetical order and returns the resulting list.
Combine (list1, list2)	Creates a new list consisting of the entries from `list1` followed by the entries from `list2`. The resulting list is returned.
RemoveDuplicates (list)	Iterates through `list`. If any two or more entries have the same value, the duplicate entries are removed so that any entry appears at most once. The resulting list is returned.

Which of the following code segments will correctly create *newList* ?

AP Principles 2.2 2.3 2.4 MCQ only

(A)
newList ← Combine (list1, list2)
newList ← Sort (newList)
newList ← RemoveDuplicates (newList)

(B)
list1 ← Sort (list1)
list2 ← Sort (list2)
newList ← Combine (list1, list2)
newList ← RemoveDuplicates (newList)

(C)
list1 ← RemoveDuplicates (list1)
list2 ← RemoveDuplicates (list2)
newList ← Combine (list1, list2)
newList ← Sort (newList)

(D)
list1 ← RemoveDuplicates (list1)
list1 ← Sort (list1)
list2 ← RemoveDuplicates (list2)
list2 ← Sort (list2)
newList ← Combine (list1, list2)

29. Each student at a school has a unique student ID number. A teacher has the following spreadsheets available.

- Spreadsheet I contains information on all students at the school. For each entry in this spreadsheet, the student name, the student ID, and the student's grade point average are included.
- Spreadsheet II contains information on only students who play at least one sport. For each entry in this spreadsheet, the student ID and the names of the sports the student plays are included.
- Spreadsheet III contains information on only students whose grade point average is greater than 3.5. For each entry in this spreadsheet, the student name and the student ID are included.
- Spreadsheet IV contains information on only students who play more than one sport. For each entry in this spreadsheet, the student name and the student ID are included.

The teacher wants to determine whether students who play a sport are more or less likely to have higher grade point averages than students who do not play any sports. Which of the following pairs of spreadsheets can be combined and analyzed to determine the desired information?

(A) Spreadsheets I and II

(B) Spreadsheets I and IV

(C) Spreadsheets II and III

(D) Spreadsheets III and IV

AP Principles 2.2 2.3 2.4 MCQ only

30. Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.

Participants in a survey were asked how many hours per day they spend reading, how many hours per day they spend using a smartphone, and whether or not they would be interested in a smartphone application that lets users share book reviews.

The data from the survey are represented in the graph below. Each ✕ represents a survey participant who said he or she was interested in the application, and each ○ represents a participant who said he or she was not interested.
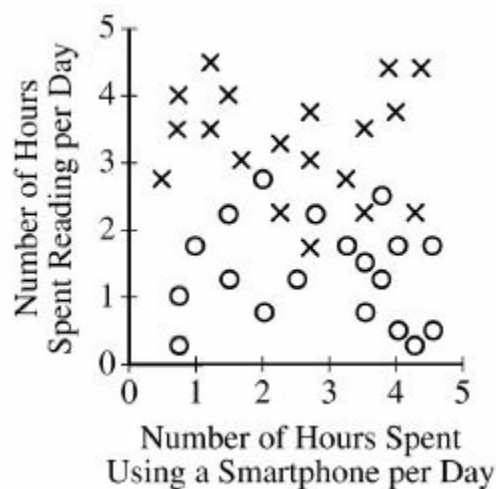
Which of the following hypotheses is most consistent with the data in the graph?

(A) Participants who read more were generally more likely to say they are interested in the application.

(B) Participants who read more were generally less likely to say they are interested in the application.

(C) Participants who use a smartphone more were generally more likely to say they read more.

(D) Participants who use a smartphone more were generally less likely to say they read more.

31. A user wants to save a data file on an online storage site. The user wants to reduce the size of the file, if possible, and wants to be able to completely restore the file to its original version. Which of the following actions best supports the user's needs?

(A) Compressing the file using a lossless compression algorithm before uploading it

(B) Compressing the file using a lossy compression algorithm before uploading it

(C) Compressing the file using both lossy and lossless compression algorithms before uploading it

(D) Uploading the original file without using any compression algorithm

AP Principles 2.2 2.3 2.4 MCQ only

32. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

Byte pair encoding is a data encoding technique. The encoding algorithm looks for pairs of characters that appear in the string more than once and replaces each instance of that pair with a corresponding character that does not appear in the string. The algorithm saves a list containing the mapping of character pairs to their corresponding replacement characters.

For example, the string `"THIS_IS_THE_BEST_WISH"` can be encoded as `"%#_#_%E_BEST_W#H"` by replacing all instances of `"TH"` with `"%"` and replacing all instances of `"IS"` with `"#"`.

Which of the following statements about byte pair encoding is true?

(A) Byte pair encoding is an example of a lossy transformation because it discards some of the data in the original string.

(B) Byte pair encoding is an example of a lossy transformation because some pairs of characters are replaced by a single character.

(C) Byte pair encoding is an example of a lossless transformation because an encoded string can be restored to its original version.

(D) Byte pair encoding is an example of a lossless transformation because it can be used to transmit messages securely.

33. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

Computers are often used to search through large sets of data to find useful patterns in the data. Which of the following tasks is NOT an example where searching for patterns is needed to produce useful information?

(A) A credit card company analyzing credit card purchases to identify potential fraudulent charges

(B) A grocery store analyzing customers' past purchases to suggest new products the customer may be interested in

(C) A high school analyzing student grades to identify the students with the top ten highest grade point averages

(D) An online retailer analyzing customers' viewing habits to suggest other products based on the purchasing history of other customers

AP Principles 2.2 2.3 2.4 MCQ only

34. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

Grades in a computer science course are based on total points earned on a midterm exam and a final exam. The teacher provides a way for students to improve their course grades if they receive high scores on the final exam: if a student's final exam score is greater than the student's midterm exam score, the final exam score replaces the midterm exam score in the calculation of total points.

The table below shows two students' scores on the midterm and final exams and the calculated total points each student earns.

Student Name	Midterm Exam Score	Final Exam Score	Total Points Calculation
Khalil	90	80	90 + 80 = 170
Josefina	70	90	90 + 90 = 180

- Khalil does better on the midterm exam than on the final exam, so his original midterm and final exam scores are added to compute his total points.
- Josefina does better on the final exam than on the midterm exam, so her final exam score replaces her midterm exam score in the total points calculation.

The teacher has data representing the scores of thousands of students. For each student, the data contain the student name, the midterm exam score, the final exam score, and the result of the total points calculation. Which of the following could be determined from the data?

I. The average total points earned per student
II. The average increase in total points per student as a result of the score replacement policy
III. The proportion of students who improved their total points as a result of the score replacement policy

- (A) III only
- (B) I and II only
- (C) I and III only
- (D) I, II, and III

AP Principles 2.2 2.3 2.4 MCQ only

35. **Directions: For the question or incomplete statement below, two of the suggested answers are correct. For this question, you must select both correct choices to earn credit. No partial credit will be earned if only one correct choice is selected. Select the two that are best in each case.**

Two different schools maintain data sets about their currently enrolled students. No individual student is enrolled at both schools. Each line of data contains information, separated by commas, about one student.

East High School stores the data in the following format.

Data format	School Name, Last Name, First Name, Age, ZIP Code, Days Absent
Sample line of data	East H.S., Ableson, Carlton, 16, 98987, 8

West High School stores the data in the following format.

Data format	Student ID, First Name, Last Name, Age, Days Absent
Sample line of data	87667721, Kendra, Angelos, 14, 12

The two schools would like to combine their data to make a single data set. Which of the following can be done with the combined data?

Select two answers.

A   The schools can create a single list of student names, sorted by last name.

B   The schools can determine the average number of days students are absent.

C   The schools can determine which ZIP code is represented by the most students.

D   The schools can determine the student ID of the student with the greatest number of absences.

36. Which of the following is a true statement about data compression?

(A) Data compression is only useful for files being transmitted over the Internet.

(B) Regardless of the compression technique used, once a data file is compressed, it cannot be restored to its original state.

(C) Sending a compressed version of a file ensures that the contents of the file cannot be intercepted by an unauthorized user.

(D) There are trade-offs involved in choosing a compression technique for storing and transmitting data.

AP Principles 2.2 2.3 2.4 MCQ only

37. A company uses the following data files.

File Name	Description	Contents
Customers	A list of customers	Customer ID Customer address Customer e-mail address Customer phone number
Products	A list of products available for purchase from the company	Product ID Product name Type of battery used by the product, if any
Purchases	A list of customer purchases	Product ID Product serial number Customer ID

A new rechargeable battery pack is available for products that use AA batteries. Which of the following best explains how the data files in the table can be used to send a targeted e-mail to only those customers who have purchased products that use AA batteries to let them know about the new accessory?

AP Principles 2.2 2.3 2.4 MCQ only

(A) Use the customer IDs in the purchases file to search the customers file to generate a list of e-mail addresses

(B) Use the product IDs in the purchases file to search the products file to generate a list of product names that use AA batteries

(C) Use the customers file to generate a list of customer IDs, then use the list of customer IDs to search the products file to generate a list of product names that use AA batteries

(D) Use the products file to generate a list of product IDs that use AA batteries, then use the list of product IDs to search the purchases file to generate a list of customer IDs, then use the list of customer IDs to search the customers file to generate a list of e-mail addresses

38. A database of information about shows at a concert venue contains the following information.

- Name of artist performing at the show
- Date of show
- Total dollar amount of all tickets sold

Which of the following additional pieces of information would be most useful in determining the artist with the greatest attendance during a particular month?

(A) Average ticket price

(B) Length of the show in minutes

(C) Start time of the show

(D) Total dollar amount of food and drinks sold during the show

39. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

Digital images are often represented by the red, green, and blue values (an RGB triplet) of each individual pixel in the image. A photographer is manipulating a digital image and overwriting the original image. Which of the following describes a lossless transformation of the digital image?

(A) Compressing the image in a way that may lose information but will suffer only a small loss of image quality.

(B) Creating the gray scale of an image by averaging the amounts of red, green, and blue in each pixel and assigning this new value to the corresponding pixel in the new image. The new value of each pixel represents a shade of gray, ranging from white to black.

(C) Creating the negative of an image by creating a new RGB triplet for each pixel in which each value is calculated by subtracting the original value from 255. The negative of an image is reversed from the original; light areas appear dark, and colors are reversed.

(D) Modifying part of the image by taking the pixels in one part of the picture and copying them to the pixels in another part of the picture.

40. Historically, it has been observed that computer processing speeds tend to double every two years. Which of the following best describes how technology companies can use this observation for planning purposes?

AP Principles 2.2 2.3 2.4 MCQ only

(A)   Technology companies can accurately predict the dates when new computing innovations will be available to use.

(B)   Technology companies can plan to double the costs of new products each time advances in processing speed occur.

(C)   Technology companies can set research and development goals based on anticipated processing speeds.

(D)   Technology companies can spend less effort developing new processors because processing speed will always improve at the observed rate.

A large spreadsheet contains the following information about local restaurants. A sample portion of the spreadsheet is shown below.

	A Restaurant Name	B Price Range	C Number of Customer Ratings	D Average Customer Rating	E Accepts Credit Cards
1	Joey Calzone's Pizzeria	lo	182	3.5	false
2	78th Street Bistro	med	41	4.5	false
3	Seaside Taqueria	med	214	4.5	true
4	Delicious Sub Shop II	lo	202	4.0	false
5	Rustic Farm Tavern	hi	116	4.5	true
6	ABC Downtown Diner	med	0	-1.0	true

In column B, the price range represents the typical cost of a meal, where `"lo"` indicates under $10, `"med"` indicates $11 to $30, and `"hi"` indicates over $30.
In column D, the average customer rating is set to `-1.0` for restaurants that have no customer ratings.

41. A student wants to count the number of restaurants in the spreadsheet whose price range is $30 or less and whose average customer rating is at least `4.0`. For a given row in the spreadsheet, suppose `prcRange` contains the price range as a string and `avgRating` contains the average customer rating as a decimal number.

    Which of the following expressions will evaluate to `true` if the restaurant should be counted and evaluates to `false` otherwise?

    (A)   `(avgRating ≥ 4.0) AND ((prcRange = "lo") AND (prcRange = "med"))`

    (B)   `(avgRating ≥ 4.0) AND ((prcRange = "lo") OR (prcRange = "med"))`

    (C)   `(avgRating ≥ 4.0) OR ((prcRange = "lo") AND (prcRange = "med"))`

    (D)   `(avgRating ≥ 4.0) OR ((prcRange = "lo") OR (prcRange = "med"))`

AP Principles 2.2 2.3 2.4 MCQ only

42. A student is developing an algorithm to determine which of the restaurants that accept credit cards has the greatest average customer rating. Restaurants that have not yet received any customer ratings and restaurants that do not accept credit card are to be ignored.

Once the algorithm is complete, the desired restaurant will appear in the first row of the spreadsheet. If there are multiple entries that fit the desired criteria, it does not matter which of them appears in the first row.

The student has the following actions available but is not sure of the order in which they should be executed.

Action	Explanation
Filter by number of ratings	Remove entries for restaurants with no customer ratings
Filter by payment type	Remove entries for restaurants that do not accept credit cards
Sort by rating	Sort the rows in the spreadsheet on column D from greatest to least

Assume that applying either of the filters will not change the relative order of the rows remaining in the spreadsheet.

Which of the following sequences of steps can be used to identify the desired restaurant?

    I. Filter by number of ratings, then filter by payment type, then sort by rating
    II. Filter by number of ratings, then sort by rating, then filter by payment type
    III. Sort by rating, then filter by number of ratings, then filter by payment type

(A) I and II only

(B) I and III only

(C) II and III only

(D) I, II, and III

---

43. Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.

Byte pair encoding is a data encoding technique. The encoding algorithm looks for pairs of characters that appear in the string more than once and replaces each instance of that pair with a corresponding character that does not appear in the string. The algorithm saves a list containing the mapping of character pairs to their corresponding replacement characters.

For example, the string `"THIS_IS_THE_BEST_WISH"` can be encoded as `"%#_#_%E_BEST_W#H"` by replacing all instances of `"TH"` with `"%"` and replacing all instances of `"IS"` with `"#"`.

For which of the following strings is it NOT possible to use byte pair encoding to shorten the string's length?

AP Principles 2.2 2.3 2.4 MCQ only

(A) `"BANANA"`

(B) `"LEVEL_UP"`

(C) `"MEET_ME_LATER"`

(D) `"NEITHER_HERE_NOR_THERE"`

44. Delivery trucks enter and leave a depot through a controlled gate. At the depot, each truck is loaded with packages, which will then be delivered to one or more customers. As each truck enters and leaves the depot, the following information is recorded and uploaded to a database.

   - The truck's identification number
   - The truck's weight
   - The date and time the truck passes through the gate
   - Whether the truck is entering or leaving the depot

   Using only the information in the database, which of the following questions CANNOT be answered?

   (A) On which day in a particular range of dates did the greatest number of trucks enter and leave the depot?

   (B) What is the average number of customer deliveries made by each truck on a particular day?

   (C) What is the change in weight of a particular truck between when it entered and left the depot?

   (D) Which truck has the shortest average time spent at the depot on a particular day?

45. A library system contains information for each book that was borrowed. Each time a person borrows or returns a book from the library, the following information is recorded in a database.

   - Name and the unique ID number of the person who was borrowing the book
   - Author, title, and the unique ID number of the book that was borrowed
   - Date that the book was borrowed
   - Date that the book was due to be returned
   - Date that the book was returned (or 0 if the book has not been returned yet)

   Which of the following CANNOT be determined from the information collected by the system?

   (A) The total number of books borrowed in a given year

   (B) The total number of books that were never borrowed in a given year

   (C) The total number of books that were returned past their due date in a given year

   (D) The total number of people who borrowed at least one book in a given year

46. A camera mounted on the dashboard of a car captures an image of the view from the driver's seat every second. Each image is stored as data. Along with each image, the camera also captures and stores the car's speed, the date and time, and the car's GPS location as metadata. Which of the following can best be determined using only the data and none of the metadata?

   (A) The average number of hours per day that the car is in use

   (B) The car's average speed on a particular day

   (C) The distance the car traveled on a particular day

   (D) The number of bicycles the car passed on a particular day

AP Principles 2.2 2.3 2.4 MCQ only

47. A teacher sends students an anonymous survey in order to learn more about the students' work habits. The survey contains the following questions.

- On average, how long does homework take you each night (in minutes) ?
- On average, how long do you study for each test (in minutes) ?
- Do you enjoy the subject material of this class (yes or no) ?

Which of the following questions about the students who responded to the survey can the teacher answer by analyzing the survey results?

   I. Do students who enjoy the subject material tend to spend more time on homework each night than the other students do?
  II. Do students who spend more time on homework each night tend to spend less time studying for tests than the other students do?
 III. Do students who spend more time studying for tests tend to earn higher grades in the class than the other students do?

(A) I only

(B) III only

(C) I and II

(D) I and III

48. A person wants to transmit an audio file from a device to a second device. Which of the following scenarios best demonstrates the use of lossless compression of the original file?

(A) A device compresses the audio file before transmitting it to a second device. The second device restores the compressed file to its original version before playing it.

(B) A device compresses the audio file by removing details that are not easily perceived by the human ear. The compressed file is transmitted to a second device, which plays it.

(C) A device transmits the original audio file to a second device. The second device removes metadata from the file before playing it.

(D) A device transmits the original audio file to a second device. The second device plays the transmitted file as is.

49. In which of the following situations would it be most appropriate to choose lossy compression over lossless compression?

(A) Storing digital photographs to be printed and displayed in a large format in an art gallery

(B) Storing a formatted text document to be restored to its original version for a print publication

(C) Storing music files on a smartphone in order to maximize the number of songs that can be stored

(D) Storing a video file on an external device in order to preserve the highest possible video quality

AP Principles 2.2 2.3 2.4 MCQ only

50. When a cellular telephone user places a call, the carrier transmits the caller's voice as well as the voice of the person who is called. The encoded voices are the data of the call. In addition to transmitting the data, the carrier also stores metadata. The metadata of the call include information such as the time the call is placed and the phone numbers of both participants. For which of the following goals would it be more useful to computationally analyze the metadata instead of the data?

    I. To determine if a caller frequently uses a specific word

    II. To estimate the number of phone calls that will be placed next Monday between 10:30 A.M. and noon.

    III. To generate a list of criminal suspects when given the telephone number of a known criminal

(A) I only

(B) II only

(C) II and III only

(D) I, II, and III

51. A list of binary values (0 or 1) is used to represent a black-and-white image. Which of the following is LEAST likely to be stored as metadata associated with the image?

(A) Copyright information for the image

(B) The date and time the image was created

(C) The dimensions (number of rows and columns of pixels) of the image

(D) A duplicate copy of the data

AP Principles 2.2 2.3 2.4 MCQ only

52. A large spreadsheet contains information about the photographs in a museum's collection. A sample portion of the spreadsheet is shown below.

	A	**B**	**C**	**D**
	Photographer	**Subject**	**Year**	**Publicly Available**
1	Steven Greene	Geyser Eruption	2004	true
2	Linda James	Giant Sloth Fossil	-1	true
3	Yajaira Lopez	Diplodocus Skull	1997	false
4	Masahiro Higashi	Sea Turtle	1989	true
5	(unknown)	Solar Eclipse	2011	false
6	(unknown)	Giant Sequoia	-1	true

- In column A, each unknown photographer is set to `"(unknown)"`.
- In column C, each unknown year is set to `-1`.

A student is developing an algorithm to determine the name of the photographer who took the oldest photograph in the collection. Photographs whose photographer or year are unknown are to be ignored.

Once the algorithm is complete, the desired entry will appear in the first row of the spreadsheet. If there are multiple entries that meet the desired criteria, then any of them can appear in the first row.

The student has the following actions available.

Action	Explanation
Filter by photographer	Removes entries whose photographer is `"(unknown)"`
Filter by year	Removes entries whose year is `-1`
Sort by subject	Sorts the rows in the spreadsheet on column B alphabetically from A to Z
Sort by year	Sorts the rows in the spreadsheet on column C from least to greatest

Assume that applying either of the filters will not change the relative order of the rows remaining in the spreadsheet.

Which of the following sequences of steps can be used to identify the desired entry?

Select <u>two</u> answers.

AP Principles 2.2 2.3 2.4 MCQ only

A | Filter by photographer, then filter by year, then sort by subject

B | Filter by photographer, then filter by year, then sort by year

C | Sort by subject, then sort by year, then filter by photographer

D | Sort by year, then filter by year, then filter by photographer

53. A large spreadsheet contains information about the schedule for a college radio station. A sample portion of the spreadsheet is shown below.

	A Show Name	B Genre	C Day	D Start Time	E End Time
1	Dot Dot Dash	rock	Sunday	11:00 A.M.	1:00 P.M.
2	New Afternoon Show	talk	Sunday	1:00 P.M.	3:00 P.M.
3	Thursday Beats	hip-hop	Thursday	7:00 P.M.	9:00 P.M.
4	Gossip Time	talk	Friday	4:00 P.M.	6:00 P.M.
5	Campus Chat	talk	Saturday	6:00 P.M.	8:00 P.M.
6	Jazz Brunch	jazz	Saturday	12:00 P.M.	3:00 P.M.

A student wants to count the number of shows that meet both of the following criteria.

Is a talk show
Is on Saturday or Sunday

For a given row in the spreadsheet, suppose `genre` contains the genre as a string and `day` contains the day as a string. Which of the following expressions will evaluate to `true` if the show should be counted and evaluates to `false` otherwise?

(A) `(genre = "talk") AND ((day = "Saturday") AND (day = "Sunday"))`

(B) `(genre = "talk") AND ((day = "Saturday") OR (day = "Sunday"))`

(C) `(genre = "talk") OR ((day = "Saturday") AND (day = "Sunday"))`

(D) `(genre = "talk") OR ((day = "Saturday") OR (day = "Sunday"))`

AP Principles 2.2 2.3 2.4 MCQ only

54. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

A social media site allows users to send messages to each other. A group of researchers gathered user data for the first 10 years of the site's existence. Some of the data are summarized in the table below, along with some of the company milestones.

Year	Milestone	Total Number of Registered Users (in millions)	Average Number of Active Daily Users (in millions)	Average Number of Daily Messages Sent Per User	Average Number of Characters Per Message
1	Web site launched	25.4	0.8	3.6	360
2		26.0	0.7	3.5	362
3		26.5	0.6	3.5	358
4		26.9	0.6	3.4	360
5	Mobile app released	27.4	0.9	3.3	269
6		28.0	0.9	3.4	242
7		28.6	1.1	3.5	195
8		29.1	1.2	3.5	176
9		29.6	1.1	3.6	104
10		30.2	1.1	3.6	96

The researchers noticed that the total number of registered users appears to be increasing at about a constant rate. If this pattern continues, which of the following best approximates the total number of registered users, in millions, in year 12 (two years after the last entry in the table) ?

(A) 30.6

(B) 31.2

(C) 31.8

(D) 32.4

AP Principles 2.2 2.3 2.4 MCQ only

55. **Directions: The question or incomplete statement below is followed by four suggested answers or completions. Select the one that is best in each case.**

A social media site allows users to send messages to each other. A group of researchers gathered user data for the first 10 years of the site's existence. Some of the data are summarized in the table below, along with some of the company milestones.

Year	Milestone	Total Number of Registered Users (in millions)	Average Number of Active Daily Users (in millions)	Average Number of Daily Messages Sent Per User	Average Number of Characters Per Message
1	Web site launched	25.4	0.8	3.6	360
2		26.0	0.7	3.5	362
3		26.5	0.6	3.5	358
4		26.9	0.6	3.4	360
5	Mobile app released	27.4	0.9	3.3	269
6		28.0	0.9	3.4	242
7		28.6	1.1	3.5	195
8		29.1	1.2	3.5	176
9		29.6	1.1	3.6	104
10		30.2	1.1	3.6	96

Which of the following hypotheses is most consistent with the data in the table?

(A) The mobile app release did not have any effect on the average number of daily messages sent per user.

(B) The mobile app release discouraged new user registration on the site.

(C) The mobile app release led to users being less frequently active on the site.

(D) The mobile app release led to users tending to write shorter messages.

56. The table below shows the time a computer system takes to complete a specified task on the customer data of different-sized companies.

Task	Small Company (approximately 100 customers)	Medium Company (approximately 1,000 customers)	Large Company (approximately 10,000 customers)
Backing up data	2 hours	20 hours	200 hours
Deleting entries from data	100 hours	200 hours	300 hours
Searching through data	250 hours	300 hours	350 hours
Sorting data	0.01 hour	1 hour	100 hours

Based on the information in the table, which of the following tasks is likely to take the longest amount of time when scaled up for a very large company of approximately 100,000 customers?

AP Principles 2.2 2.3 2.4 MCQ only

    (A)   Backing up data

    (B)   Deleting entries from data

    (C)   Searching through data

    (D)   Sorting data

57. A city maintains a database of all traffic tickets that were issued over the past ten years. The tickets are divided into the following two categories.

- Moving violations
- Nonmoving violations

The data recorded for each ticket include only the following information.

- The month and year in which the ticket was issued
- The category of the ticket

Which of the following questions CANNOT be answered using only the information in the database?

    (A)   Have the total number of traffic tickets per year increased each year over the past ten years?

    (B)   In the past ten years, were nonmoving violations more likely to occur on a weekend than on a weekday?

    (C)   In the past ten years, were there any months when moving violations occurred more often than nonmoving violations?

    (D)   In how many of the past ten years were there more than one million moving violations?

58. A programmer is developing software for a social media platform. The programmer is planning to use compression when users send attachments to other users. Which of the following is a true statement about the use of compression?

    (A)   Lossless compression of video files will generally save more space than lossy compression of video files.

    (B)   Lossless compression of an image file will generally result in a file that is equal in size to the original file.

    (C)   Lossy compression of an image file generally provides a greater reduction in transmission time than lossless compression does.

    (D)   Sound clips compressed with lossy compression for storage on the platform can be restored to their original quality when they are played.

AP Principles 2.2 2.3 2.4 MCQ only

59. A wildlife preserve is developing an interactive exhibit for its guests. The exhibit is intended to allow guests to select the name of an animal on a touch screen and display various facts about the selected animal.

For example, if a guest selects the animal name "wolf," the exhibit is intended to display the following information.

- Classification: mammal
- Skin type: fur
- Thermoregulation: warm-blooded
- Lifestyle: pack
- Average life span: 10–12 years
- Top speed: 75 kilometers/hour

The preserve has two databases of information available to use for the exhibit. The first database contains information for each animal's name, classification, skin type, and thermoregulation. The second database contains information for each animal's name, lifestyle, average life span, and top speed.

Which of the following explains how the two databases can be used to develop the interactive exhibit?

(A) Only the first database is needed. It can be searched by animal name to find all the information to be displayed.

(B) Only the second database is needed. It can be searched by animal name to find all the information to be displayed.

(C) Both databases are needed. Each database can be searched by animal name to find all information to be displayed.

(D) The two databases are not sufficient to display all the necessary information because the intended display information does not include the animal name.

A student's overall course grade in a certain class is based on the student's scores on individual assignments. The course grade is calculated by dropping the student's lowest individual assignment score and averaging the remaining scores.

For example, if a particular student has individual assignment scores of 85, 75, 90, and 95, the lowest score (75) is dropped. The calculated course grade is $(85 + 90 + 95)/3 = 90$.

60. An administrator at the school has data about hundreds of students in a particular course. While the administrator does not know the values of each student's individual assignment scores, the administrator does have the following information for each student.

- The student name
- A unique student ID number
- The number of assignments for the course
- The average assignment score before the lowest score was dropped
- The course grade after the lowest score was dropped

Which of the following CANNOT be determined from this data alone?

(A) For a given student, the value of the highest assignment score

(B) For a given student, the value of the lowest assignment score

(C) For a given student, the change in course grade as a result of dropping the lowest score

(D) The proportion of students who improved their course grade as a result of dropping the lowest score