

6. Stefan, a psychologist, conducted a study to investigate the effect of time of day on reading comprehension in children. One hundred children volunteered, with their parents' consent, to participate in the study. Fifty of the children were randomly assigned to read a story at 9 a.m. and then answer 25 questions about it. The remaining 50 children were assigned to read the same story at 3 p.m. and answer the same 25 questions. The reading comprehension for each child was measured by a reading score, which was determined by the number of questions that were answered correctly about the story. Stefan is interested in comparing the mean reading scores for the two times of day. Table 1 shows the results of Stefan's study.

Table 1: Summary Statistics of Reading Scores

	n	Mean	Standard Deviation
9 a.m.	50	15.2	4.12
3 p.m.	50	17.9	4.43

Stefan found the conditions for inference were met and conducted a two-sample t -test for the difference in two population means. Let μ_{AM} represent the mean reading score for all children, similar to those in the study, who would read the story at 9 a.m. Let μ_{PM} represent the mean reading score for all children, similar to those in the study, who would read the story at 3 p.m. Stefan's hypotheses are as shown.

$$H_0: \mu_{AM} = \mu_{PM}$$

$$H_a: \mu_{AM} \neq \mu_{PM}$$

- A. The p -value for Stefan's hypothesis test was 0.002. State an appropriate conclusion, at the 5 percent significance level, for Stefan's test in the context of the investigation. Justify your answer.
- B. Explain why it was appropriate for Stefan to conduct a two-sample t -test for the difference in two population means instead of a paired t -test for the population mean difference.

- C. Researchers are usually interested in the practical importance of their results as well as the statistical significance of the hypothesis test. The practical importance of the results indicates whether the observed results are meaningful in real life. For example, in an investigation of the heights of two groups of students, a difference in the two group means of 3.8 inches is much more meaningful, or has more practical importance, than a difference in the two group means of only 0.2 inches.

One indicator of practical importance is effect size. A common method for measuring effect size for the difference in two group means is Cohen's d coefficient. Cohen's d coefficient compares the absolute value of the difference in the means of the two groups to the pooled variability of the observed data values from the two groups.

Cohen's d coefficient can be calculated using $d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p}$, where s_p represents the pooled standard deviation, \bar{x}_1 represents the sample mean for the first group, and \bar{x}_2 represents the sample mean for the second group. When the sizes of the groups are equal, s_p is calculated as $s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$, where s_1 represents the sample standard deviation for the first group and s_2 represents the sample standard deviation for the second group.

Consider the summary statistics from Stefan's study in Table 1.

- Calculate Cohen's d coefficient for Stefan's study. Show your work.
- Higher values of Cohen's d indicate greater practical importance and lower values of Cohen's d indicate less practical importance. Typically, we use the intervals listed in Table 2 to help interpret practical importance.

Table 2: Guidelines for Interpreting Cohen's d Coefficient

Cohen's d Coefficient	Practical Importance
$0 \leq d \leq 0.20$	Not very meaningful in real life
$0.20 < d < 0.80$	Somewhat meaningful in real life
$d \geq 0.80$	Very meaningful in real life

Based on your answer to part C (i) and the information in Tables 1 and 2, describe the practical importance of Stefan's results, in context.