Colorado Student Assessment Program (CSAP) Analysis

Ryan McElvain

2024-12-16

Contents

Introduction

The Colorado Department of Education (CDE) seeks to analyze the performance of 9th graders in math, reading, and writing for the years 2009 and 2010. This report addresses two key questions:

1. **Was there an improvement in passing rates for the math exam between 2009 and 2010? Did reading and writing performance change as well?**
2. **Was there an association between passing rates on the math exam and the reading and writing exams? How accurately could math passing rates be predicted from the reading and writing passing rates?**

By employing the **5A Method** (Ask, Acquire, Analyze, Advise, Answer), this report provides answers to these questions.

---

1. Ask: Understanding the Problem

CDE administrators aim to improve math performance among high school freshmen while monitoring potential trade-offs in reading and writing scores. The questions of interest are:

- Did math performance improve between 2009 and 2010?
- Were declines observed in reading and writing during the same period?
- How strong is the association between math performance and other subjects?
- How accurately can reading and writing performance predict math passing rates?

---

2. Acquire: Data Import and Preparation

Data Acquisition

The data was obtained via the Colorado Information Marketplace API for the 2009 and 2010 school years, covering math, reading, and writing scores for 9th graders.

Data Wrangling

Key steps in data wrangling included:

1. Removing invalid rows (e.g., schools with `school_no == 0`).
2. Filtering schools with fewer than 31 students in either year.
3. Reshaping the dataset to align test scores with their respective school IDs
4. Randomly sampling 120 school IDS without replacement

The final dataset includes:

- Rows: 720 (120 schools × 2 years × 3 subjects)
- Columns: 8 (e.g., `year`, `subject`, and performance categories).

---

3. Analyze: Exploratory and Inferential Analysis

Changes in Math Passing Rates

A bootstrap analysis was conducted to compare mean passing rates for math between 2009 and 2010. The bootstrap distribution allowed for the construction of 95% confidence intervals for the mean difference in passing rates.

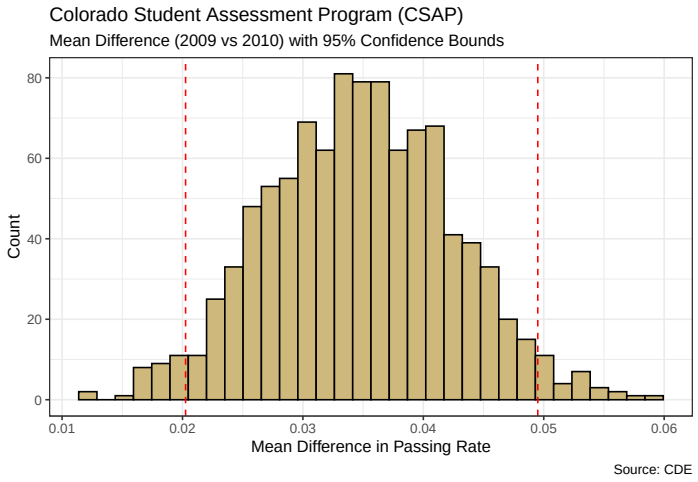

Figure 1: Bootstrap Distribution of Math Passing Rates

Results:

- The mean difference in passing rates was approximately 0.0344642.
- The 95% confidence interval for the mean difference was approximately [0.0202557, 0.0495046].

Changes in Reading and Writing Passing Rates

Similarly, the bootstrap procedure was applied to assess differences in passing rates for reading and writing between 2009 and 2010.

Mean Difference for Reading Passing Rates

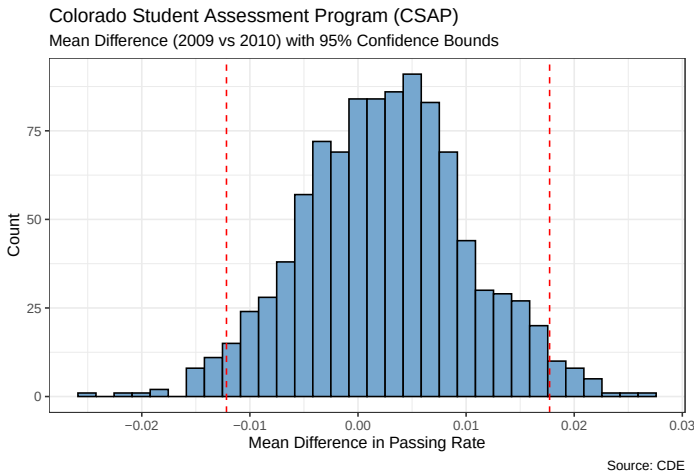

Figure 2: Bootstrap Distribution of Reading Passing Rates

Results:

- The mean difference in passing rates was approximately 0.002428.
- The 95% confidence interval for the mean difference was approximately [-0.012168, 0.017723].

Mean Difference for Writing Passing Rates

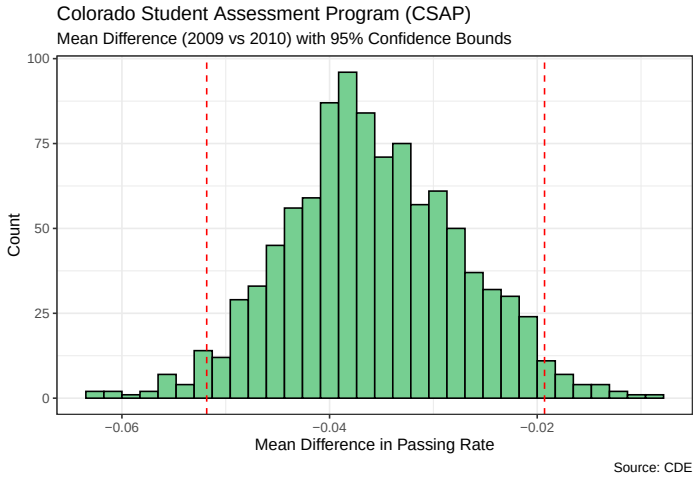

Figure 3: Bootstrap Distribution of Writing Passing Rates

Results:

- The mean difference in passing rates was approximately -0.036.
- The 95% confidence interval for the mean difference was approximately [-0.051835, -0.019295].

Interpretation of Findings: Changes in Passing Rates

Based on the bootstrap analysis:

- **Math Passing Rates**: The mean passing rate for math increased from 2009 to 2010, with the confidence interval for the mean difference not crossing zero. This indicates a significant improvement in math performance.
- **Reading and Writing Passing Rates**: Both reading and writing passing rates exhibited contrasting trends from 2009 to 2010. Reading scores showed a slight increase, while writing scores experienced a large decrease, with confidence intervals for the mean differences in writing excluding zero, proving its statistical significance. These patterns suggest a potential trade-off as math passing rates improved during the same period.

In summary, while math passing rates improved significantly, there may have been a cost to performance in reading and writing.

Predicting Math Passing Rates Using Reading and Writing

A linear regression model was built using 2009 data to predict math passing rates based on reading and writing passing rates. The models were then tested on 2010 data to evaluate their predictive accuracy.

Reading Prediction

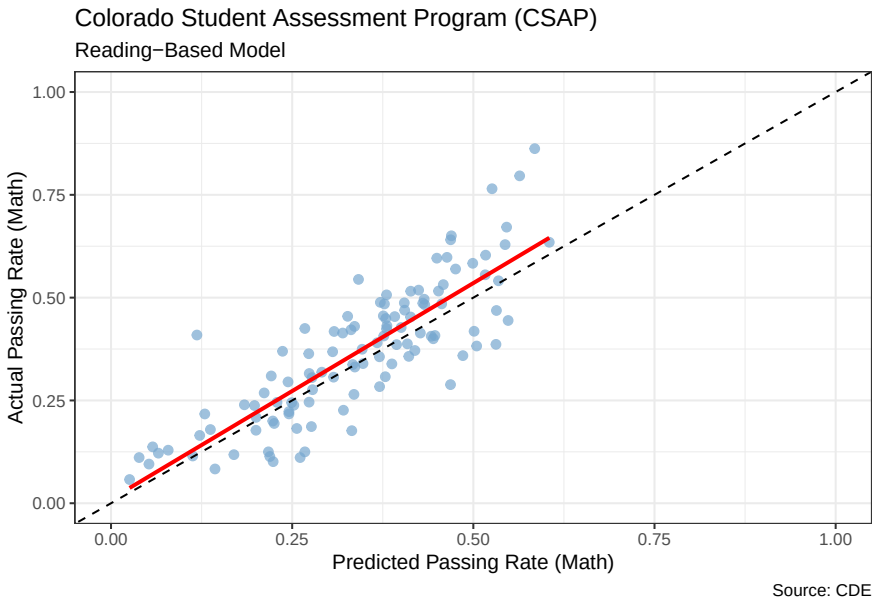

Figure 4: Actual vs. Predicted Math Passing Rates (2010) Using Reading-Based Model. The dashed black line represents perfect prediction (actual = predicted), while the red line shows the best-fit linear regression line for the observed data.

Results:

- **RMSE (Root Mean Squared Error)**: The RMSE for the reading-based model is approximately 0.096085, representing the average prediction error when using reading scores to predict math passing rates.
- **Regression Equation:**

$$\text{MATH} = -0.263961 + 0.885803 \cdot \text{READING}$$

Writing Prediction

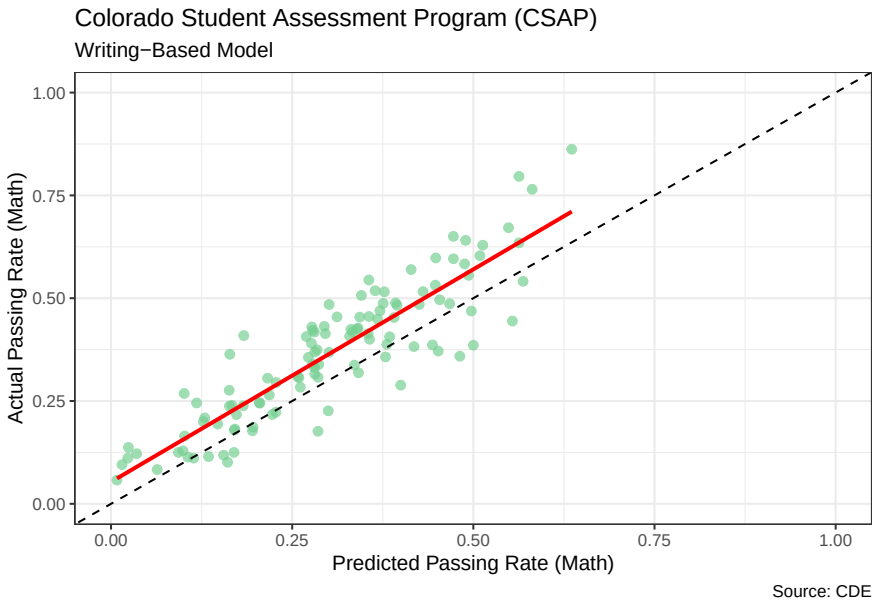

Figure 5: Actual vs. Predicted Math Passing Rates (2010) Using Writing-Based Model. The dashed black line represents perfect prediction (actual = predicted), while the red line shows the best-fit linear regression line for the observed data.

Results:

- **RMSE (Root Mean Squared Error)**: The RMSE for the writing-based model is approximately 0.098506, representing the average prediction error when using writing scores to predict math passing rates.
- **Regression Equation:**

$$\text{MATH} = \text{-}0.105994 + 0.848505 \cdot \text{WRITING}$$

Interpretation of Findings: Predicting Math Passing Rates

The predictive analysis aimed to assess how well reading and writing scores could predict math passing rates for the 2010 data, based on models trained using 2009 data.

Reading-Based Model

The reading-based model's regression equation was:

$$\text{MATH} = \text{-}0.263961 + 0.885803 \cdot \text{READING}$$

This equation implies:

- For every 1-unit increase in the reading passing rate, the math passing rate is predicted to increase by **0.8858** units.

The **RMSE** (Root Mean Squared Error) for this model was approximately **0.096085**, indicating that the average prediction error when using reading scores to predict math passing rates was about **9.6 percentage points.**

Writing-Based Model

The writing-based model's regression equation was:

$$\text{MATH} = \text{-0.105994} + 0.848505 \cdot \text{WRITING}$$

This equation suggests:

- For every 1-unit increase in the writing passing rate, the math passing rate is predicted to increase by **0.8485** units.

The **RMSE** for this model was approximately **0.098506**, representing an average prediction error of about **9.9 percentage points.**

Comparison and Insights

- Both models performed similarly, with slightly better predictive accuracy observed for the reading-based model (lower RMSE).
- The coefficients for both models suggest a strong positive association between math passing rates and reading/writing passing rates, as expected.
- The relatively small RMSE values suggest that these models can be useful for predicting math passing rates.

In summary, both reading and writing scores are strong predictors of math performance, but reading scores exhibited slightly stronger predictive power.

---

4. Advise: Interpretation of Results

The analysis of the 2009–2010 CSAP results provides key insights into student performance trends and the relationships between core academic subjects. This section's aim is to interpret these findings.

Improvements in Math Performance

- **Finding**: The mean difference in math passing rates between 2009 and 2010 was approximately 0.0344642. The associated 95% confidence interval (0.0202557, 0.0495046) did not include zero, indicating a statistically significant improvement in math performance across the sampled schools.
- **Interpretation**: These results show the success of the improved math curriculum initiatives during this time. The steady improvement across schools suggests the new curriculum is effective and could work on a larger scale.

Trade-offs in Reading and Writing Performance

- **Finding**: The mean differences in passing rates for reading and writing between 2009 and 2010 were approximately `0.0024284` and `-0.0359996`, respectively. While reading showed a slight improvement, writing experienced a decline over this period. Confidence intervals for the differences reflect this trend, with reading including mostly positive values and writing including mostly negative values.
- **Interpretation**: The results suggest that efforts to improve student performance may have led to a slight increase in reading passing rates, but the decline in writing performance is concerning. Writing is a key part of literacy and academic success, making it important to address this imbalance.

Association Between Subjects

- **Finding**: Strong positive relationships were observed between math and the other two subjects, as indicated by the regression models. For the reading-based model, the slope coefficient was 0.8858033, and the RMSE was 0.096085. The writing-based model showed similar results, with a slope of 0.8485048 and an RMSE of 0.098506.
- **Interpretation**: These results imply that schools with higher literacy rates also tend to have better math outcomes. Notably, reading appears to be a slightly stronger predictor of math performance than writing, suggesting a foundational role for reading skills.

Predictive Models: Use and Limitations

- **Strengths**:
  - The predictive models demonstrated reasonable accuracy, particularly the reading-based model with an RMSE of 0.096085. These models can be used to identify schools at risk of underperforming in math by analyzing their literacy scores.
- **Limitations**:
  1. **Data Scope**: The analysis uses a sample of 120 schools, which is representative but might not show all the differences in education across the state.
  2. **Causality**: The relationships seen here are correlative, not causal, and shouldn't be taken as proof that one subject directly affects performance in another.
  3. **Extrapolation**: The predictive accuracy only applies to the observed data range (2009–2010). Using these models for future years might need extra validation or adjustments.

---

5. Answer: Addressing the Research Questions

The analysis conducted provides clear answers to the Colorado Department of Education's (CDE) research questions. Below, I summarize my findings.

Research Question 1: Was there an improvement in passing rates for the math exam between 2009 and 2010? Did reading and writing performance change as well?

- **Answer**: The mean difference in passing rates for math was approximately 0.034464, with a 95% confidence interval of 0.0202557, 0.0495046, showing a statistically significant improvement. However, while reading passing rates showed a slight increase (0.002428), writing passing rates dropped by about -0.036. These results suggest that the focus on improving math performance may have unintentionally affected literacy outcomes.

Research Question 2: Was there an association between passing rates on the math exam and the reading and writing exams? How accurately could math passing rates be predicted from the reading and writing passing rates?

- **Answer**: Strong positive relationships were found between math and the other two subjects. The regression models showed that reading passing rates were slightly better at predicting math performance, with an RMSE of 0.096085, compared to writing, which had an RMSE of 0.098506. These findings emphasize the connection between literacy and math skills.

Conclusion

The analysis shows that while math performance improved significantly from 2009 to 2010, the drop in writing passing rates and small changes in reading raise concerns about unintended effects. Strong links between

subjects highlight the need for strategies that support all core areas of learning. These findings provide a basis for CDE to develop balanced and effective educational programs in the future.