

2009-2010 CSAP Results for 9th Grade Students

Prepared for



COLORADO
Department of Education

Prepared by

Nicholas Nguyen

May 6, 2024

Introduction (Ask)

This report seeks to understand how test scores across the state changed between the years 2009 and 2010, specifically on the CSAP math scores for 9th-grade students. During this same time, the CDE was working on implementing new initiatives in order to increase young Coloradans proficiency in math. This report seeks to answer the following questions that may help to give insight on the CDE's math initiatives:

- Was there improvement in math scores between 2009 and 2010?
 - If so, where there any changes in performance in reading and writing?
- Was there any association between a school's math passing rate and its passing rates on reading and writing?
 - How accurately can we predict a school's math passing rate from its passing rates on reading and writing?

Note that in this report, we attempt to find trends in the data itself. Even though some trends may seem like we are able to generalize them in such a way that we are able to attribute them to the CDE's initiatives, we cannot say with certainty the CDE's initiatives alone caused any trend to appear. This isn't to say that the CDE's initiatives surely have no effect on the trends we may uncover, but rather that there may be confounding variables out of the control of the CDE that may also affects the trends we uncover.

Data Acquisition and Analysis

Acquisition and Cleaning (Acquire)

Data for CSAP scores in 2009-2010 was collected from the Colorado Information Marketplace.¹ In order to ensure that the data we have acquired is suitable for use in our analyses, we must first "clean" the data. Data is cleaned in order to ensure that:

1. each row of data is a school with 31 or more students enrolled for the years 2009 and 2010,
 2. each column of data is a valid data type (categorical or numerical), and
 3. each row contains data only for one school, year (2009 or 2010), and subject (reading, writing, or math).
- This will make it so that there are six total observations for any given school.

Cleaning the data this way allows us to later easily sort the data by school, year, or subject. It also ensures that each observation we are using is valid, i.e. it belongs to a school and is not an observation that represents district or state totals.

Analysis (Analyze)

Change in scores from 2009 to 2010

The first analysis we will perform will be to determine if there were any statistically significant changes to the CSAP scores between 2009 and 2010. Throughout our analysis we will be working with a sample of 120 randomly selected schools from our cleaned data. Firstly, allow us to analyze the average change in passing rates².

Figure 1 shows us the changes in each CSAP subject area from 2009-2010. Initially we do see that, on average, math scores increased by 3.19%, writing scores decreased by -4.34%, and reading scores remained stagnant with a difference of 0.51%. However, we must perform further analysis in order to come to a more concrete conclusion. We will utilize a statistical method known as "bootstrapping" in order to help us come to said conclusion. It will allow us to construct confidence intervals that will help us understand the variability in our data. In particular, it will help us understand if the difference in testing scores are due to there truly being a difference in testing scores, or if it is simply variability in our sampling/data collection that causes us to see a difference in scores.

¹https://data.colorado.gov/K-12/CSAP-School-And-District-Summary-Results-2010/44v8-6fzj/about_data

²The passing rate is calculated by taking the sum of the number of proficient and advanced scoring students and dividing it by the total number of students who have scores at any given school.

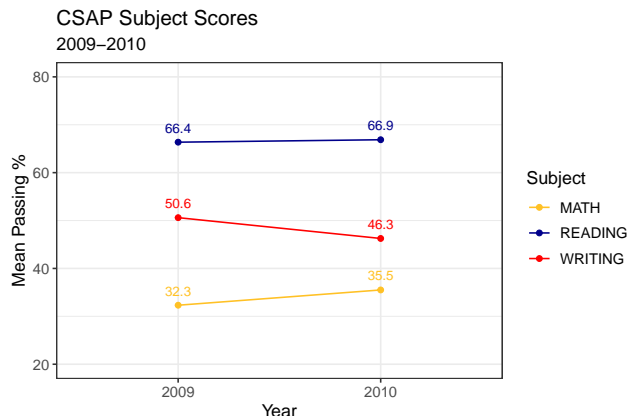


Figure 1: CSAP subject scores from 2009-2010

Figure 2 represents bootstrap sampling distributions with 95% confidence intervals. These confidence intervals give us an idea where the true mean of difference in passing rates between 2009 and 2010 lies³. If we find that the confidence interval contains 0, we can assume that there is no meaningful difference in the passing rates between 2009 and 2010. However, if we find that the confidence interval does not contain 0, we can assume that there is indeed a meaningful difference in the passing rates between 2009 and 2010⁴. From these ideas, we can conclude that math scores have increased, writing scores have decreased, and reading scores have no statistically significant change between 2009 and 2010.

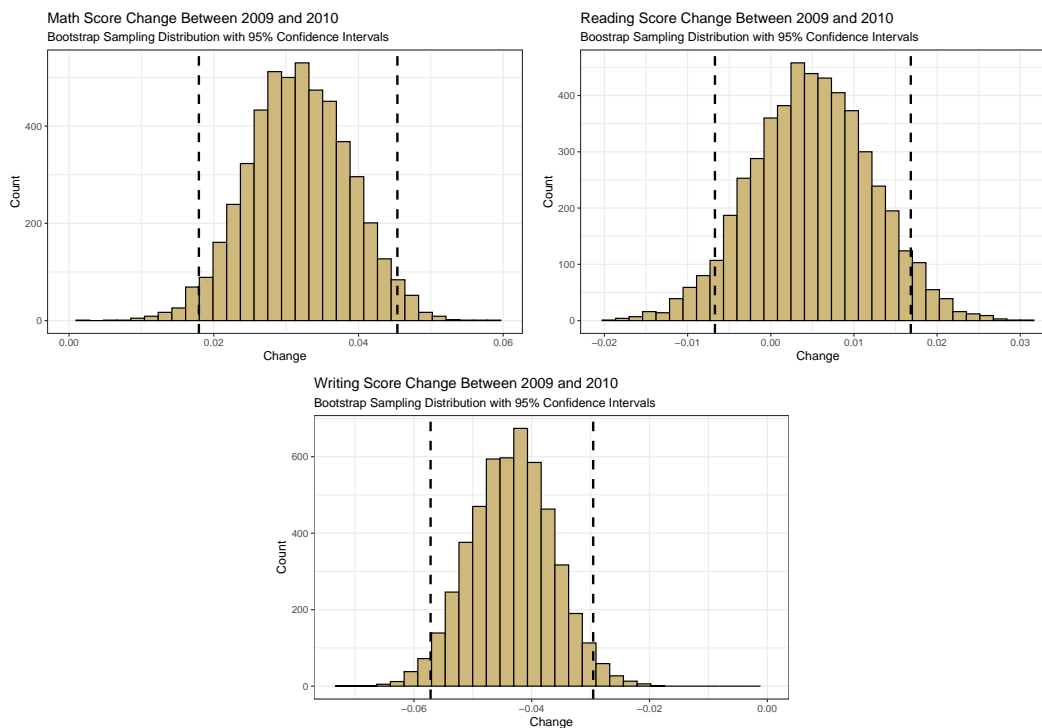


Figure 2: Bootstrap sampling distributions with 2000 resamples each and 95% confidence intervals

³Note that there is much more nuance in interpreting confidence intervals than is mentioned here. Particularly, a 95% confidence interval formally only tells us that 95% of samples we collect are going to be within a certain range.

⁴Formally, if the confidence interval does not contain 0, we would say that “we reject the idea that there is no difference in scores from 2009 to 2010”. Consequently, we can’t say that the CDE’s initiatives have affected scores, but rather that scores have indeed changed and that it should be investigated further.

It naturally follows that we now wonder if there is any difference in the change in performance in writing for schools that did better vs. worse on the math section of the CSAP. Figure 3 compares the changes in writing scores between schools that performed better and schools that performed worse on the CSAP math section. The writing score difference plots and confidence intervals are generated by first sorting schools depending on whether they did better or worse on the math section between 2009 and 2010. After math-improved ($n = 86$) and -worsened ($n = 34$) schools are sorted, 2000 bootstrap resamples for calculating mean difference between 2009 and 2010 are generated for each category and plotted. The difference of score changes between math-improved and -worsened schools is calculated by randomly pairing bootstrap-calculated means from each category with each other, and finding the difference. The resulting values are then plotted and the confidence interval is then calculated from these values. As seen from the plot representing the difference in mean scores between math-improved and -worsened schools, the confidence interval does not contain 0. From this we can conclude that there is statistically significant difference in change between 2009 and 2010 for writing scores between schools that did better and schools that did worse on the math section of the CSAP⁵.

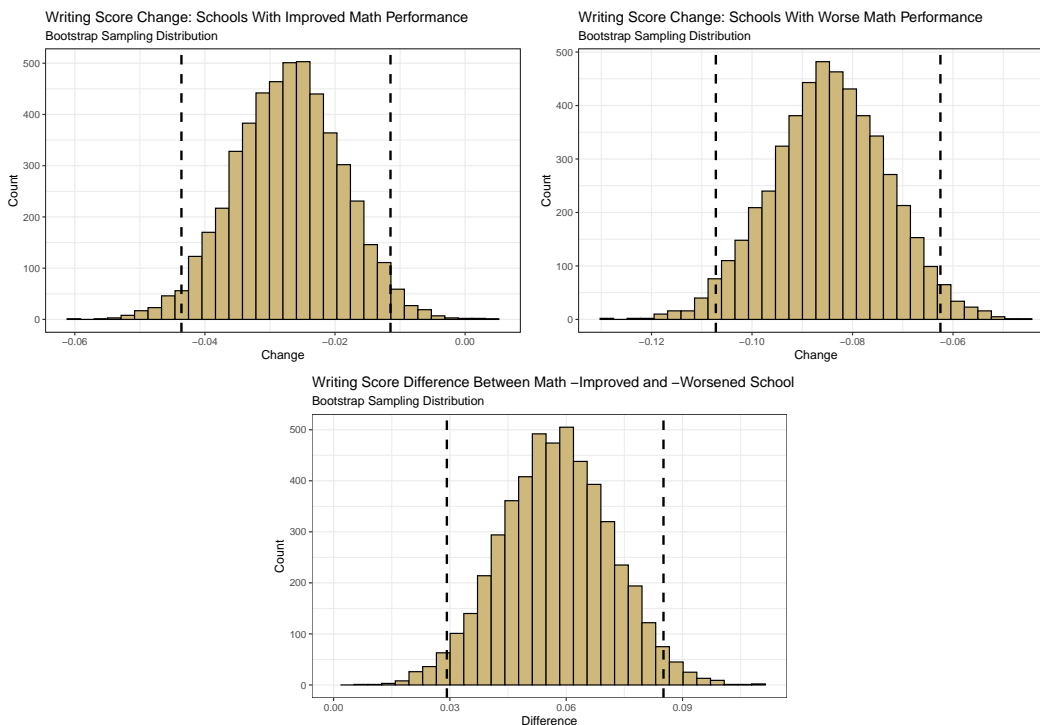


Figure 3: Comparing written scores between schools with improved and worsened math scores

Association between subject pairs and their passing rates

We will now generate models in order to see if there is any association between the passing scores of each subject pair. We will use the data from 2009 in order to generate our models. Figure 4 shows the linear regression for each pair of subjects math/writing, math/reading, and reading/writing for the year 2009. Each pair of subjects has a high coefficient of correlation ($r = 0.92$, $r = 0.88$, and $r = 0.96$ respectively) which shows that passing rates between each subject pair have strong, positive correlations. Using these linear regression models “trained” on the data from 2009, we can use the model and apply it to data from 2010 and test how accurate it is.

⁵From this observation however, we cannot conclude that there is a correlation between difference in writing and difference in math scores between 2009 and 2010. We can only conclude that there is a difference between change in writing scores between math-improved and -worsened schools.

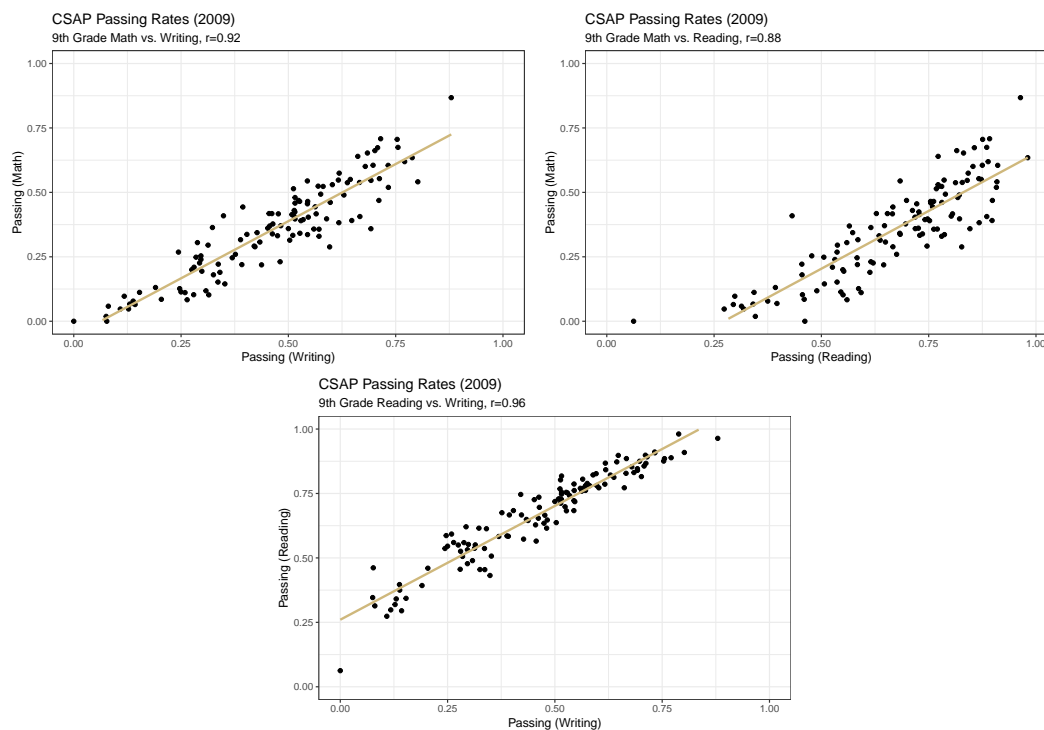


Figure 4: Linear regression for each pair of subjects

In order to quantify how accurate our models are when tested on data from 2010, we will employ a performance indicator known as Root Mean Square Error (RMSE). This is a performance indicator will give us an estimate for how much error we can expect when using our model. Table 1 displays the RMSE for each pair of subjects when their respective models were tested on data from 2010. For example, when Math/Writing model is used to predict 2010 math scores from writing scores, we can expect an error of about 10 percent points.

	Math/Writing	Math/Reading	Reading/Writing
RMSE	0.1003566	0.0968392	0.0687498

Table 1: RSME for different subject pairs' linear regression model

Conclusion (Answer)

Thus far, we have been able to come to understand the following:

- Math scores have increased, writing scores have decreased, and reading scores have had no meaningful change,
- There is a difference in the change in writing scores between schools who have improved in math vs those who have gotten worse in math,
- We are able to predict passing rates for any subject with relatively high accuracy. In math specifically, we are able calculate math scores from writing and reading scores with an error of about 10.0 and 9.7 percent points. This indicates that schools scoring highly in some subjects tend to score highly in other subjects as well.

However, it is important that these results are not misinterpreted. For example, although some scores have had a statistically significant change, we cannot come to any conclusion about why those scores changed. Similarly, while there is a difference in the change in writing scores between the two different sets of schools described above, we cannot conclude if initiatives to improve math education directly led to the decline in writing scores.

Moving forward (Advise)

While our current data doesn't allow us to come to any conclusions about the CDE's math initiatives and its association to score changes, that doesn't mean the conclusions we arrived to have no value. These conclusions tell us that it would be worthwhile to perform further studies/analysis on the effects of the CDE initiatives and CSAP scores. In the future the CDE can collect data differently in order to come to more rigorous conclusions about these associations. Specifically, if the CDE wishes to quantify the association between the CDE's math initiative and score changes, the CDE can perform a study containing a control group and a group that engages in the CDE initiatives. By doing this, we will have two groups that will allow us to perform hypothesis testing in order to quantify any association between the CDE's initiatives and changes in CSAP scores.