Project Writeup - Nelson Pollard

nwea/ai-academy-nelson-pollard

Description

Use machine learning to predict distribution of students across Achievement Level Descriptors (ALDs). Explore what features increase or decrease classification accuracy, with an interest in how political affiliation of a state may affect ALDs. Note that to simplify data collection and processing, only states that have adopted Common Core were included.

The ALDs are:

1. Below Basic (BB)
2. At Basic (BA)
3. At Proficient (PR)
4. At Advanced (AD)

Hypothesis

The original hypothesis was that political affiliation of a state, as well as the number of years since the last standard setting, would greatly influence the distribution of students across ALDs for a given year/grade/subject. The idea was that there was some sort of bimodal “ideal” distribution, with one normal curve for right-leaning states and another for left-leaning states. There might be some drift between standard settings, but immediately after a standard setting was performed we’d see the distribution across ALDs “snap back” into an approximation of the “ideal” distribution.

Features

- State
- Year
- Political affiliation (Republican, Democrat, Independent, or Unknown) of candidate voted for by state in most recent presidential election
- Political affiliation (Republican, Democrat, Independent, or Unknown) of governor of state
- Years since most recent standard setting

Data Sources

- ALDs: NAEP Data Service API
- Presidents: GitHub - zonination/election-history: US Presidential Elections since 1789
- Governors: Governor (state executive office) - Ballotpedia

Algorithms

- Linear regression: sklearn.linear_model.LinearRegression
- Multi-layer Perceptron regressor: sklearn.neural_network.MLPRegressor
  - hidden_layer_sizes=(30,30,30)
  - max_iter=1000

Results

Machine learning

Linear regression

Linear regression resulted in accuracy rates just below 80% across the board, regardless of the combination of features in/excluded. Truncated output:

```
------------------------------------------------------------
LinearRegression()
POLITICS | STATES | YEARS | YEARS SINCE STANDARD SETTING
0.7877199030960381
------------------------------------------------------------
LinearRegression()
POLITICS | STATES | YEARS | !YEARS SINCE STANDARD SETTING
0.7854942455283795
------------------------------------------------------------
LinearRegression()
POLITICS | STATES | !YEARS | YEARS SINCE STANDARD SETTING
0.7843786682897814
------------------------------------------------------------
LinearRegression()
POLITICS | STATES | !YEARS | !YEARS SINCE STANDARD SETTING
0.7922950734581131
------------------------------------------------------------
LinearRegression()
POLITICS | STATES | YEARS | YEARS SINCE STANDARD SETTING
0.7859662929352996

...
```

Multi-layer Perceptron

The highest accuracy was achieved when including the “State” feature and excluding the “Year” feature, just under 97%, with the inclusion of the number of years since the last standard setting increasing accuracy by less than half a percent:

```
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | STATES | !YEARS | !YEARS SINCE STANDARD SETTING
0.9676490507049749
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | STATES | !YEARS | !YEARS SINCE STANDARD SETTING
0.9677645069243545
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | STATES | !YEARS | YEARS SINCE STANDARD SETTING
0.9691524415578044
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | STATES | !YEARS | YEARS SINCE STANDARD SETTING
0.9693550880790959
```

Excluding the “State” feature yielded accuracy in the high 80s:

```
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | !STATES | !YEARS | !YEARS SINCE STANDARD SETTING
```

```
0.8711898454740707
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | !STATES | !YEARS | !YEARS SINCE STANDARD SETTING
0.8822530246594544
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | !STATES | !YEARS | YEARS SINCE STANDARD SETTING
0.8720625957121539
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | !STATES | !YEARS | YEARS SINCE STANDARD SETTING
0.8783474478269986
```

Including “Year” decreased accuracy to the 60s-70s range, likely due to overfitting:

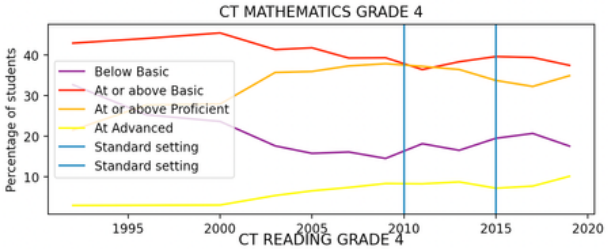
```
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | STATES | YEARS | !YEARS SINCE STANDARD SETTING
0.6893920744030189
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | STATES | YEARS | !YEARS SINCE STANDARD SETTING
0.6590560372043667
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | STATES | YEARS | YEARS SINCE STANDARD SETTING
0.6123630826575253
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | STATES | YEARS | YEARS SINCE STANDARD SETTING
0.7103966368277528
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | !STATES | YEARS | !YEARS SINCE STANDARD SETTING
0.7718847283209895
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | !STATES | YEARS | !YEARS SINCE STANDARD SETTING
0.6974333051365792
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
!POLITICS | !STATES | YEARS | YEARS SINCE STANDARD SETTING
0.6973731449238002
------------------------------------------------------------
MLPRegressor(hidden_layer_sizes=(30, 30, 30), max_iter=1000)
POLITICS | !STATES | YEARS | YEARS SINCE STANDARD SETTING
0.6813192856326488
```

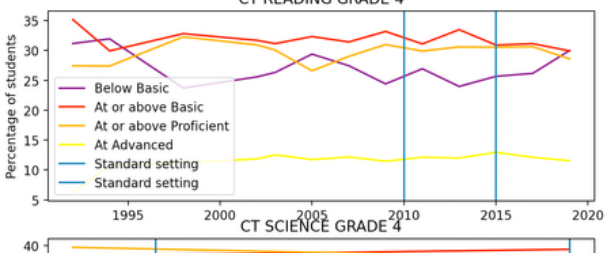
The inclusion or exclusion of political affiliation and the years since last standard setting appeared to have a negligible affect on accuracy.

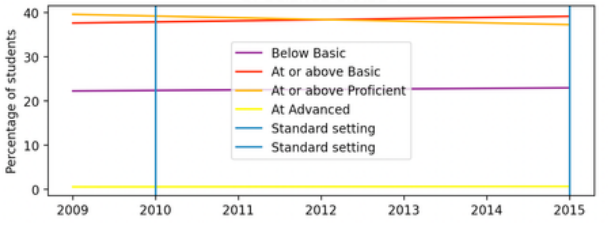
Pretty graphs

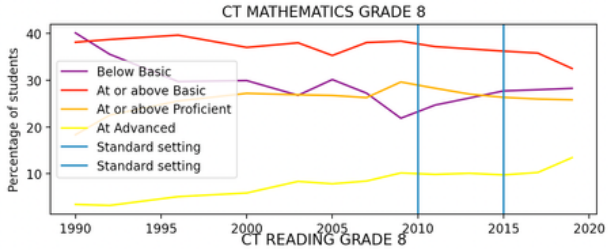
In addition running machine learning algorithms against the data, graphs were generated to visually check for trends, especially in relation to the years since last standard setting:

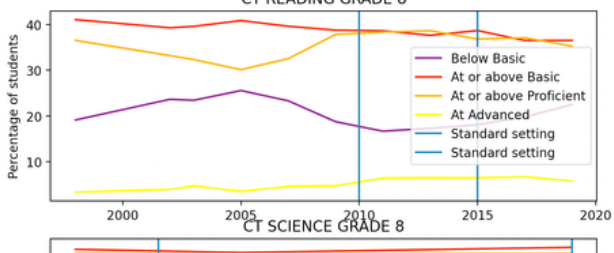
Left leaning states

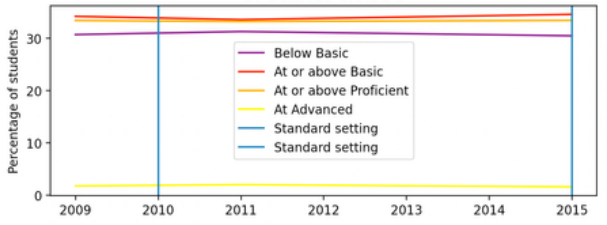


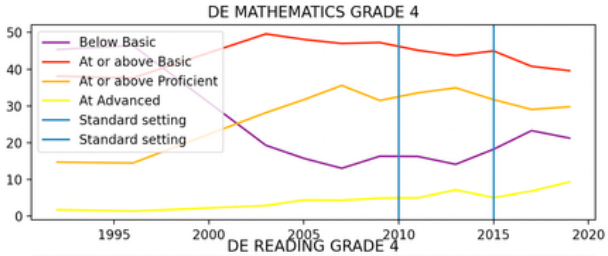


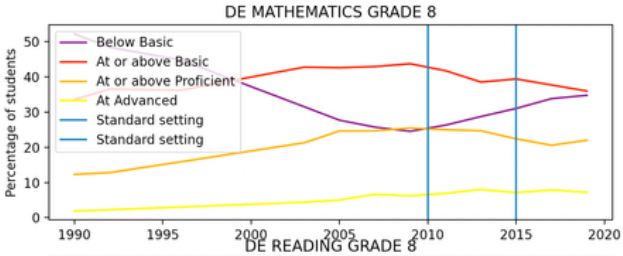


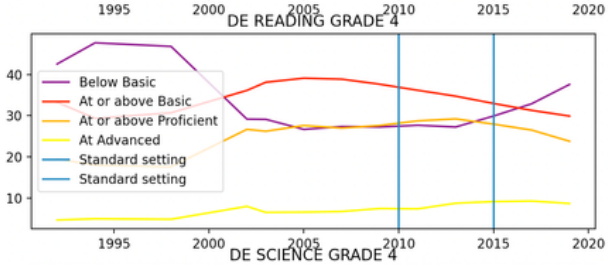


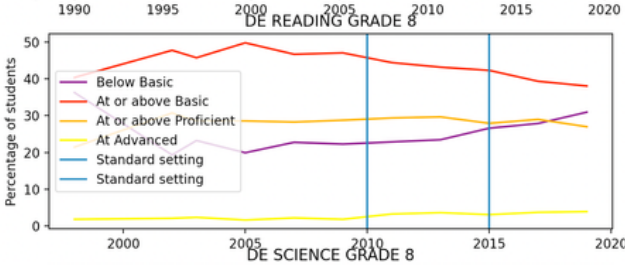


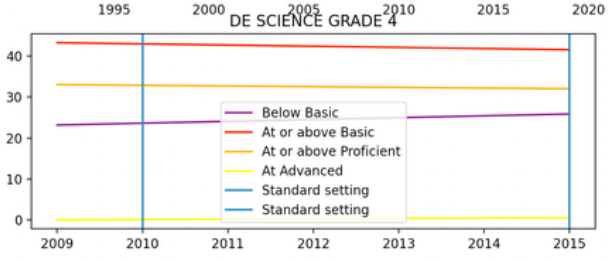


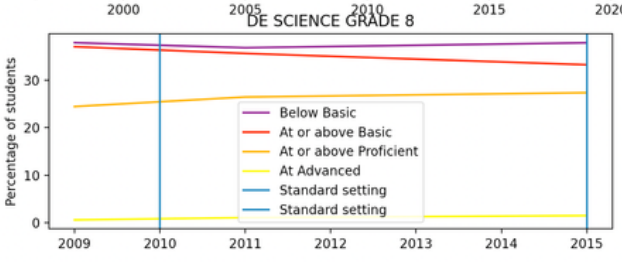

DE MATHEMATICS GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

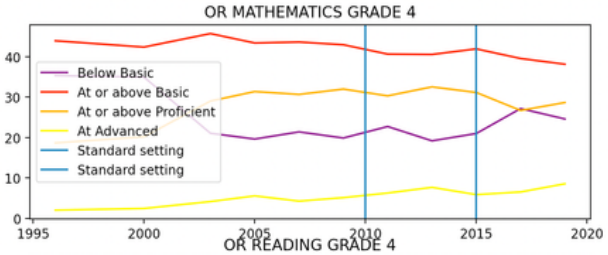

DE MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

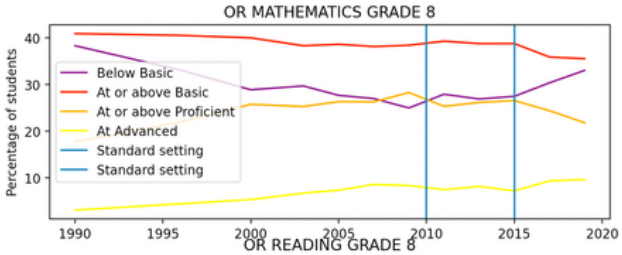

DE READING GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

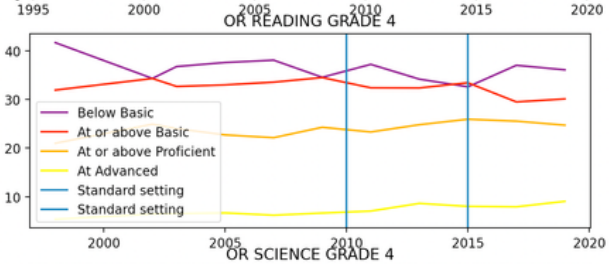

DE READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

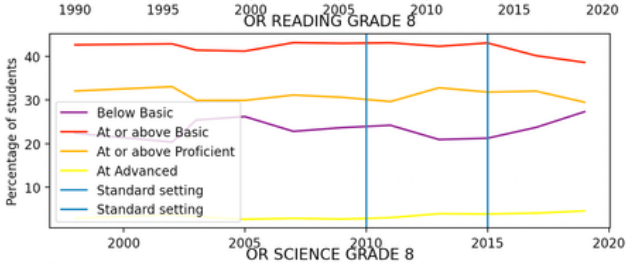

DE SCIENCE GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

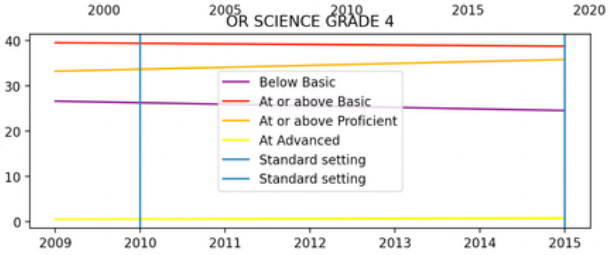

DE SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

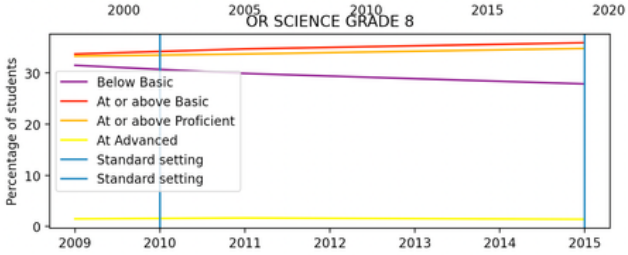

OR MATHEMATICS GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

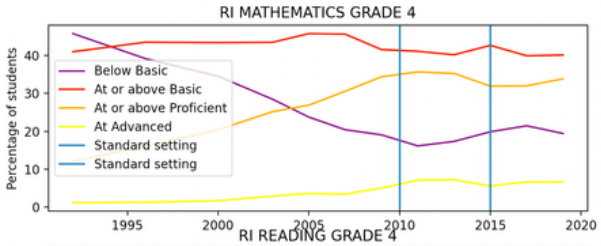

OR MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

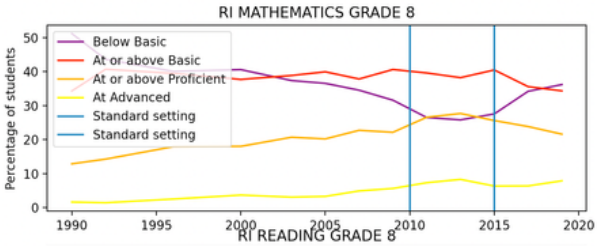

OR READING GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

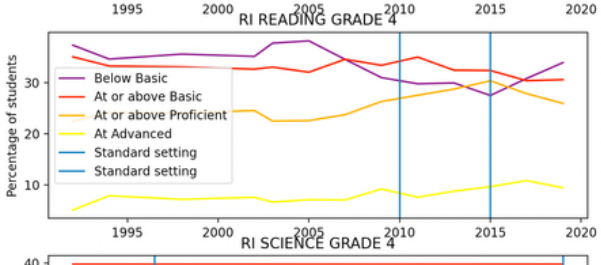

OR READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

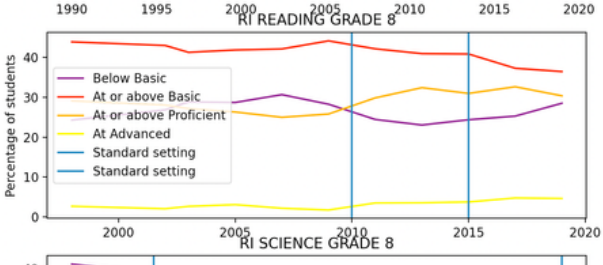

OR SCIENCE GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

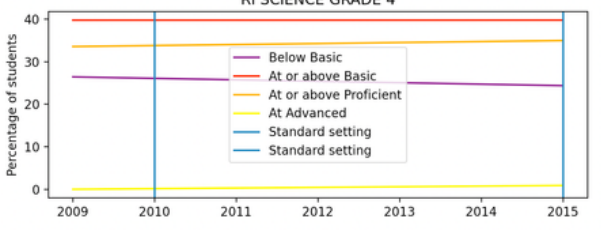

OR SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

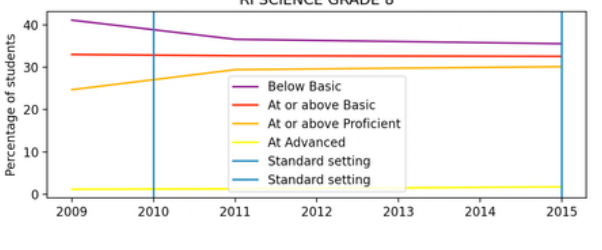

RI MATHEMATICS GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

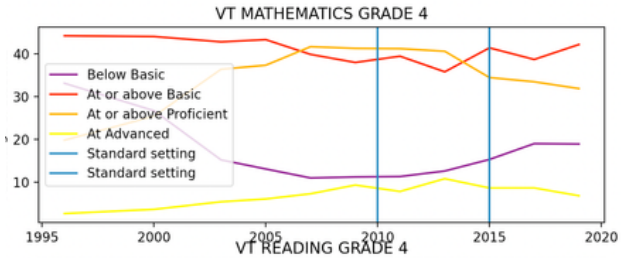

RI MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

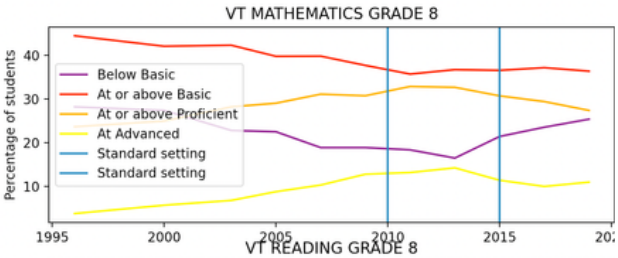

RI READING GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

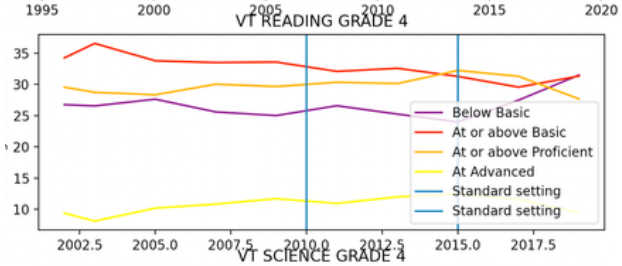

RI READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

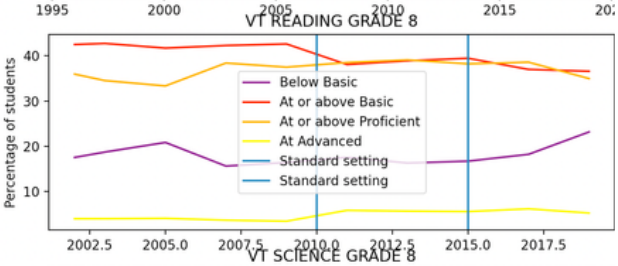

RI SCIENCE GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

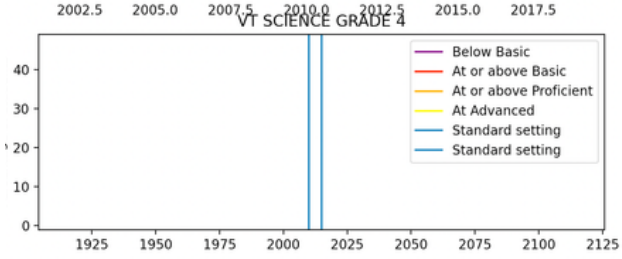

RI SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

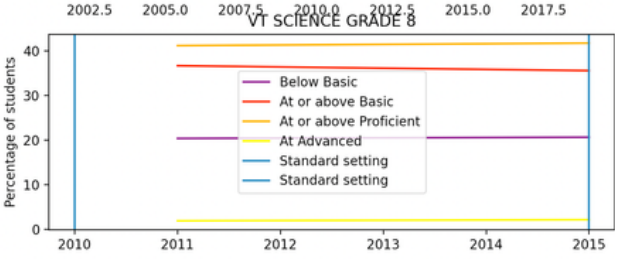

VT MATHEMATICS GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

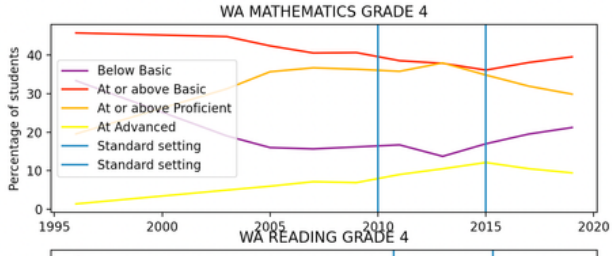

VT MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

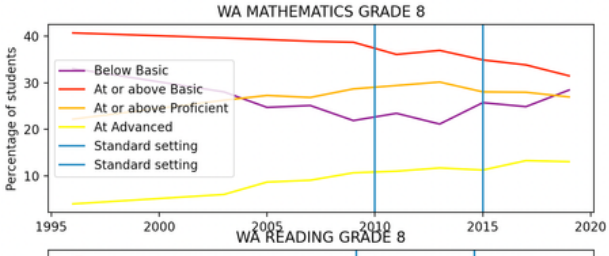

VT READING GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

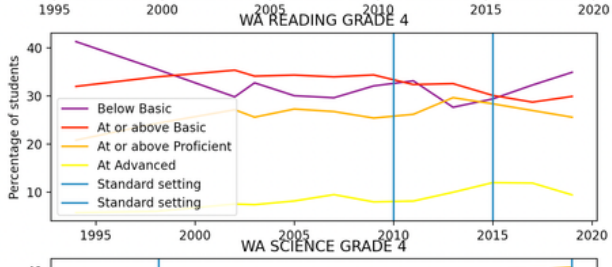

VT READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

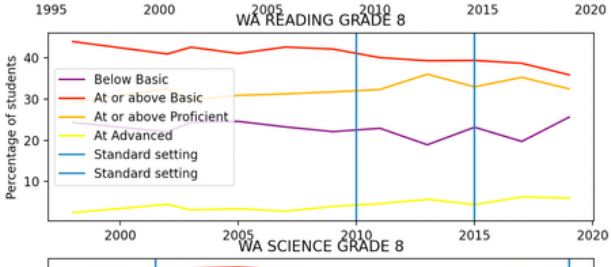

VT SCIENCE GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

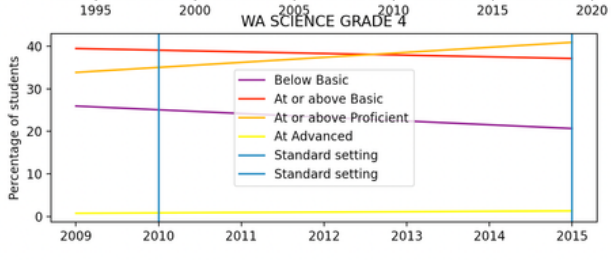

VT SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

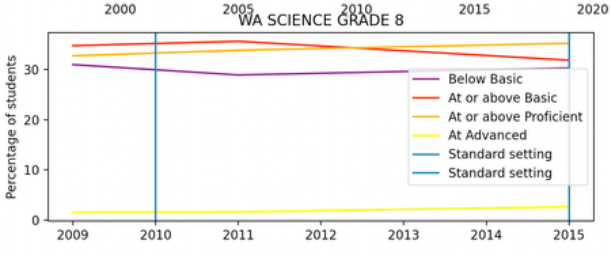

WA MATHEMATICS GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting


WA MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

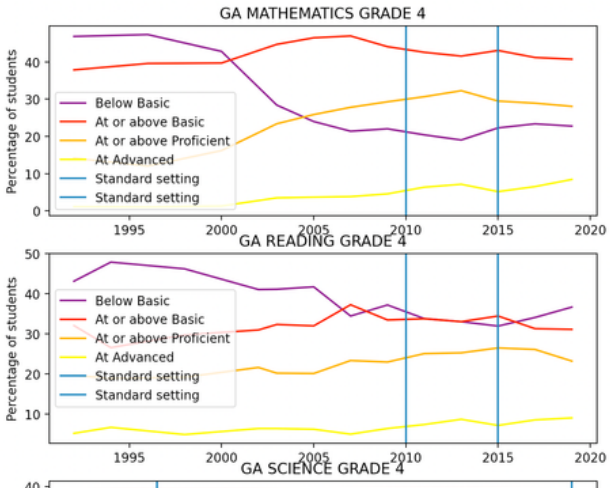

WA READING GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

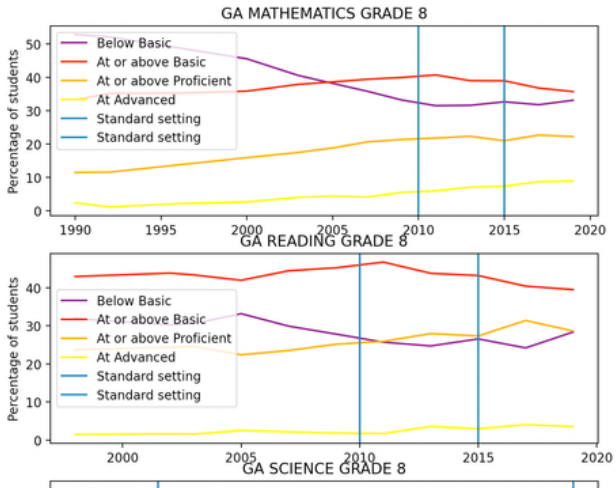

WA READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

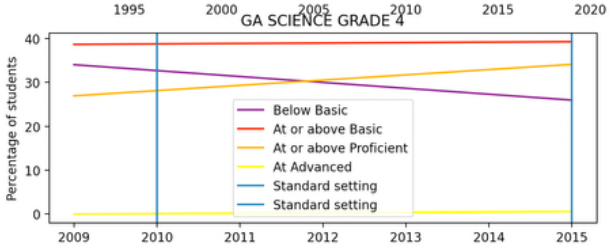

WA SCIENCE GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

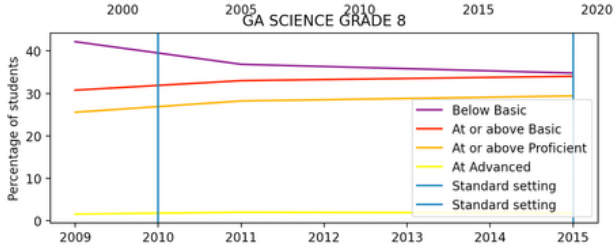

WA SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

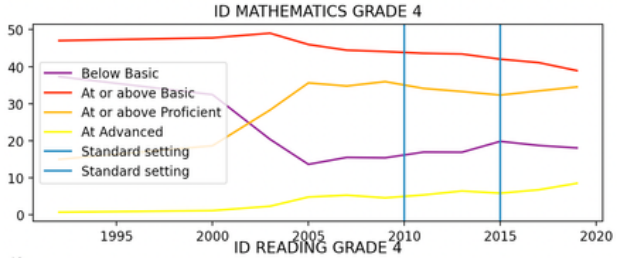
Right leaning states

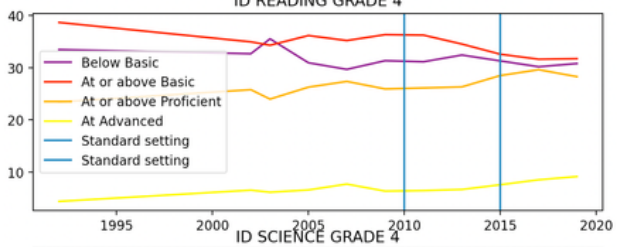


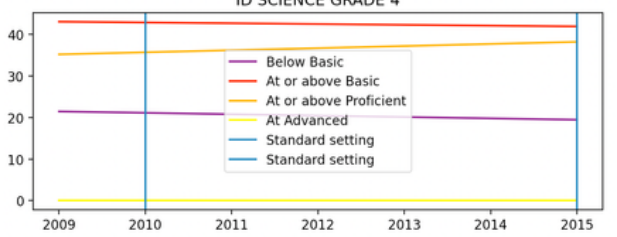


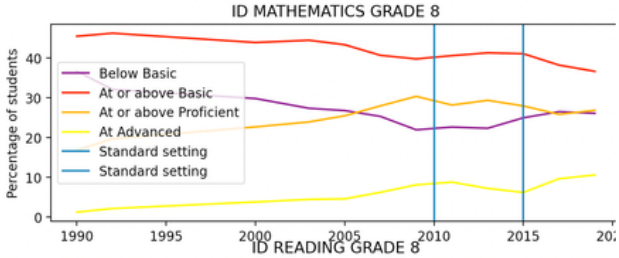


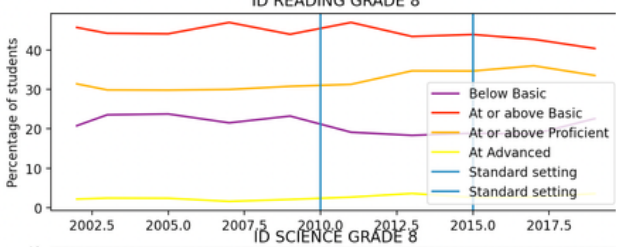


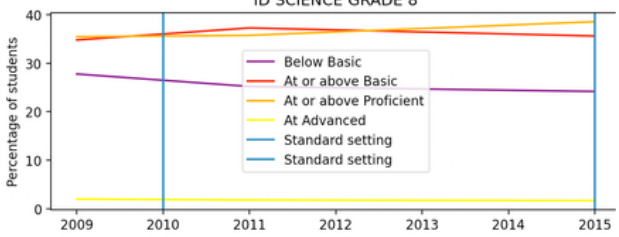

ID MATHEMATICS GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

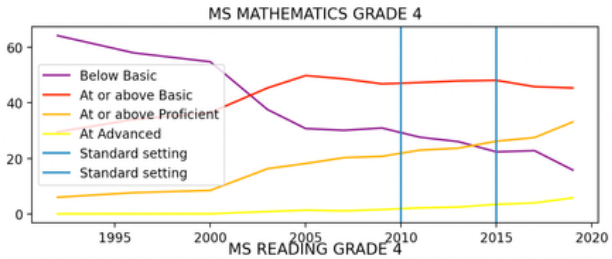

ID READING GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

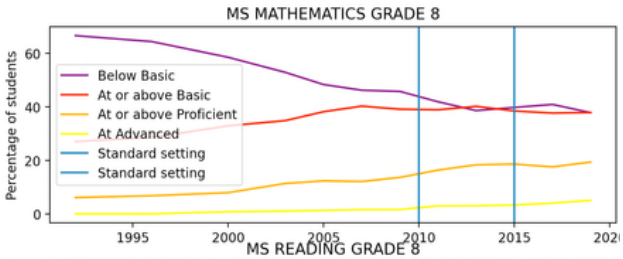

ID SCIENCE GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

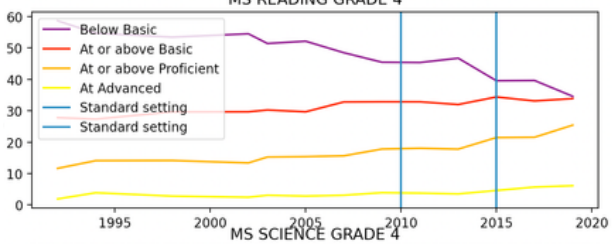

ID MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

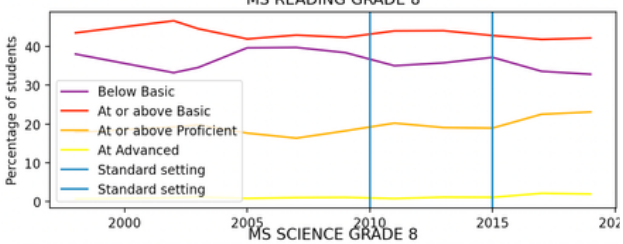

ID READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

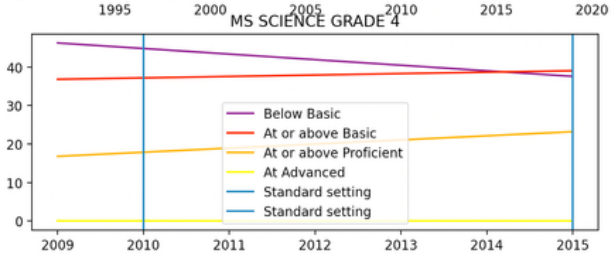

ID SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

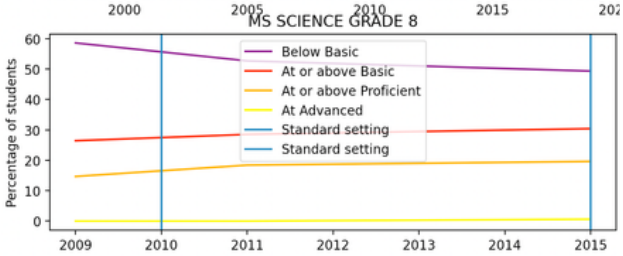

MS MATHEMATICS GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

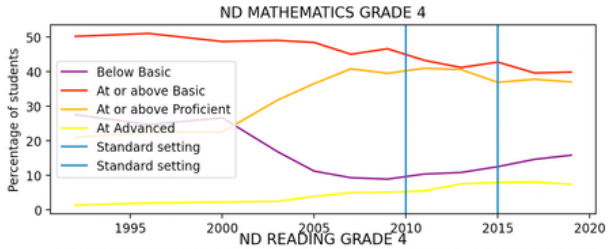

MS MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

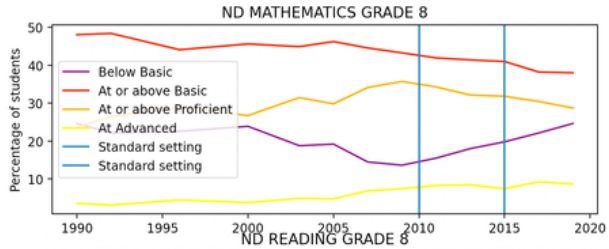

MS READING GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

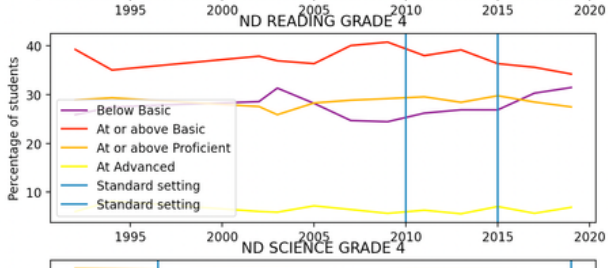

MS READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

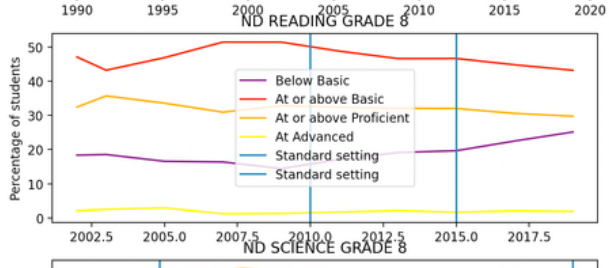

MS SCIENCE GRADE 4
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

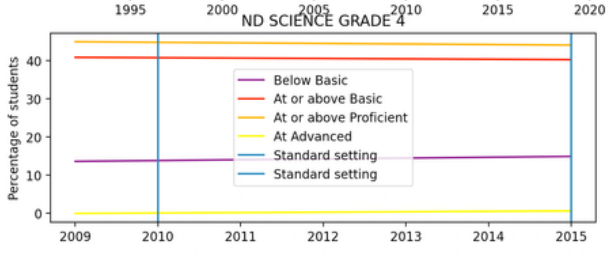

MS SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

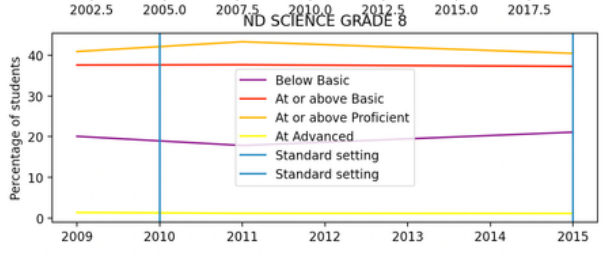

ND MATHEMATICS GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

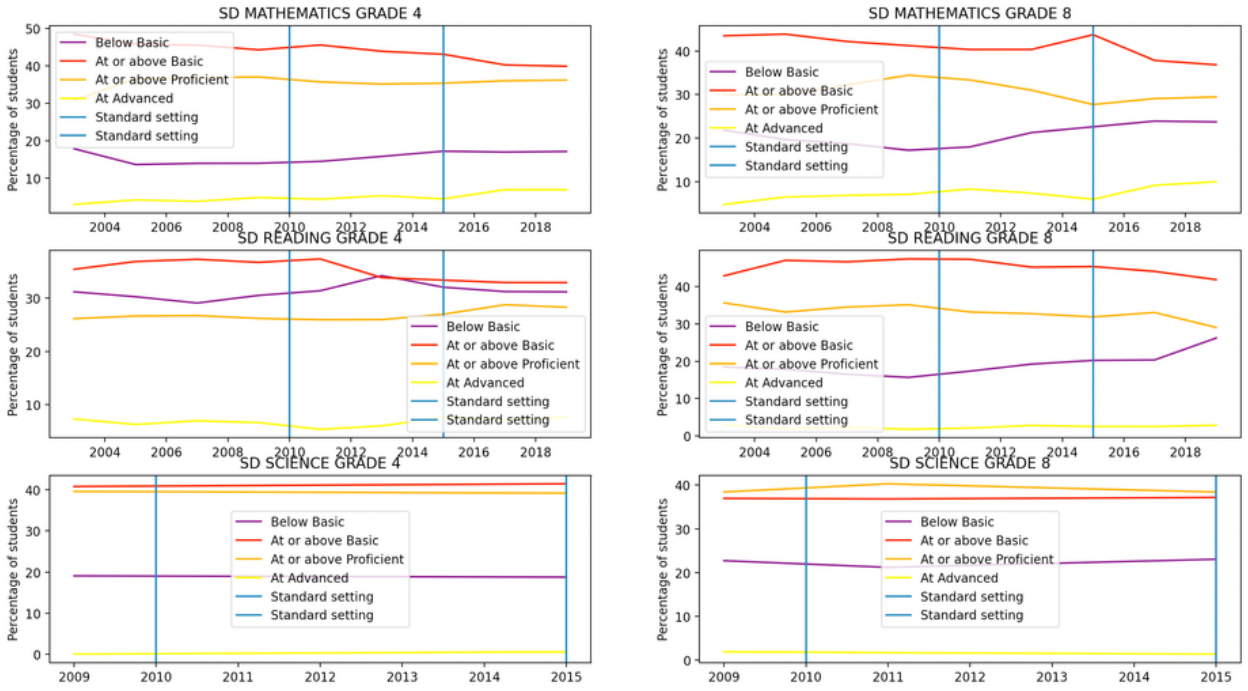

ND MATHEMATICS GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting


ND READING GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

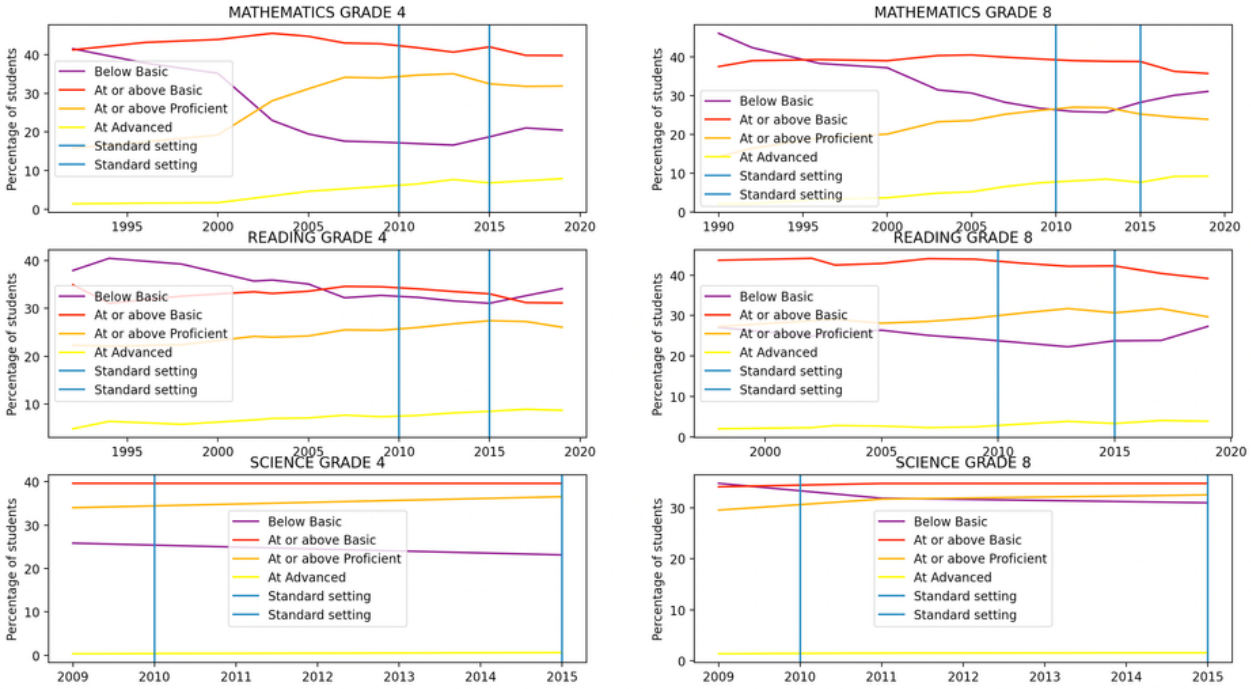

ND READING GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting


ND SCIENCE GRADE 4
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

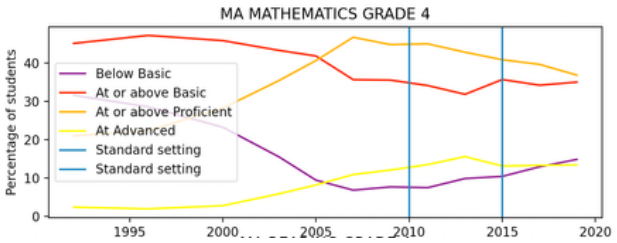

ND SCIENCE GRADE 8
Percentage of students
Below Basic
At or above Basic
At or above Proficient
At Advanced
Standard setting
Standard setting

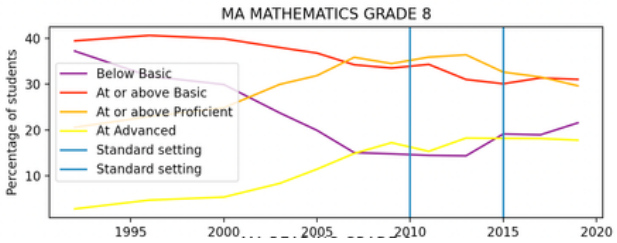


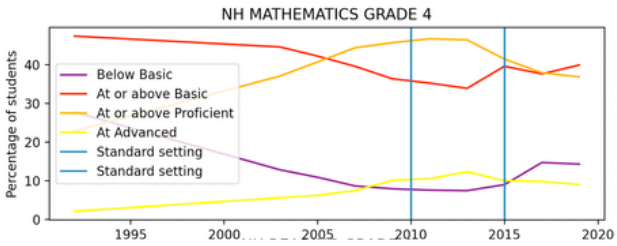
Since Common Core states have all adopted the same standards, an average across all states was also plotted:

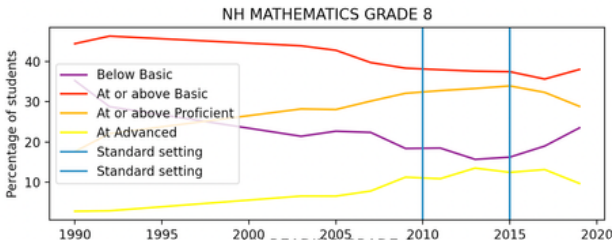


Two observable patterns are the relative sparsity of “Science” data and the almost universally low percentage of students in the “At Advanced” ALD. The only exceptions to this latter pattern are:









Conclusion

The graphs did not show any obvious “snapping back” to an “ideal” distribution following standard settings. There is also not an obvious difference in the distributions of the most left-leaning states versus the most right-leaning states. The fact that these are Common Core states, and thus are sharing standards, might explain why we do not see any obvious skewing in relation to political affiliation. Perhaps states that set their own standards display such skewing. Of course, it’s also possible that processing the data differently, using different machine learning algorithms, and/or visually plotting in a different manner would bring forth different results. But for this given approach to this set of data, it must be concluded that we have not proven the hypothesis.