Andrew Ho Rebuttal Report

In Re Delaware Public Schools Litigation, C.A. No. 2018-0029-VCL, State Track

Assignment

1. I have been asked to review the Expert Report of Gregory J. Cizek in support of the defense, assess the appropriateness of the analyses and arguments therein, and comment on how these affect my opinions, which I previously outlined in my first report.
2. In conducting my review, I have relied on a variety of sources of information summarized in the footnotes of the first report. In addition, to replicate and extend an analysis in the Cizek report, I use comprehensive data from the National Assessment of Educational Progress (NAEP).[1]

Summary of opinions

3. In my original report, I concluded:

> "Substantial evidence supports a straightforward interpretation of Delaware's reported test results for percentages of students scoring at the different levels defined by the Delaware Department of Education. A Level 3 student 'demonstrates adequate understanding of and ability to apply the English language and literacy (mathematics) knowledge and skills needed for success in college and career, as specified in the Common Core State Standards.' Level 1 and 2 students demonstrate 'minimal' and 'partial' understanding and ability, respectively."
> (p. 2).

Prof. Cizek's report does not affect this conclusion. In addition, two claims that Prof. Cizek makes are unsubstantiated and incorrect.

Delaware performance levels describe performance appropriately. There is no evidence of exhortatory purpose.

4. It is not disputed that, as both Prof. Cizek and I have said, performance levels for Delaware have been established using contemporary best practices for standard setting. Prof. Cizek recognizes that Delaware assessments "have been evaluated and judged according to professional psychometric standards

---

to be valid for obtaining high quality, dependable information about student achievement" (p. 6). He also agrees that, "the performance levels for Delaware state assessments have been established using contemporary best practices for standard setting common to state and national student achievement testing programs" (p. 30). Nevertheless, he seeks to discount the performance levels by arguing, without supporting evidence, that they are exhortatory "stretch goals" (p. 32).

5. As I explained in Section 3 of my report, it is a non sequitur to assert that any exhortatory purpose of performance levels invalidates the interpretation that percentages of Level 3 students are percentages with adequate understanding and ability. Level 3 students are those who have "adequate understanding of and ability to apply… knowledge and skills needed."[2] As I noted in my original report, "inadequate understanding is inadequate" (p. 14). SBAC Level 3 standards are lower than others like NAEP Proficient. Adequacy is not a "stretch goal."

6. Further, nothing shows that either the Delaware Department of Education or the standard setting panelists intended performance standards to be exhortatory. Indeed, the design of the achievement level setting procedure protected against undue influence from any subjective or political inclination toward exhortation.

7. Prof. Cizek correctly notes that "panelists were presented with and explicitly considered the anticipated percentages of students that would meet the Smarter Balanced performance standards they were recommending" (p. 30). He concludes that, "panelists were explicitly recommending exhortatory performance standards" (p. 31). However, neither empirical data nor logic support Prof. Cizek's conclusion.

8. Prof. Cizek neglects to mention when panelists were presented with these anticipated percentages. Following common practice, the percentages were presented between the second and third round of panelist judgments. There were no changes following presentation of those percentages that show an exhortatory purpose. The SBAC achievement level setting data[3] show no consistent patterns in the level or convergence of recommended standards between the second and final rounds of cut score selection.

---

[2] Delaware System of Student Assessments (DeSSA) Executive State Summary, https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/111/DeSSA%20Executive%20State%20Summary%202018-FINAL.pdf, p. 10.

[3] Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report. https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf

9. In Grades 3-8, English Language Arts (ELA) panelists lowered their recommended standard by 1.4 points on average, and Mathematics panelists raised their recommended standard by 2 points on average. These are common average shifts from round to round that are consistent with shifts from earlier rounds. There was no change in the spread of judgments from the second to third round in ELA, and mathematics judgments narrowed by a few points as measured by the interquartile range. In sum, there is no evidence in the empirical data that panelists prioritized exhortatory purpose over descriptive purpose.

10. The data show that panelists continued to build consensus about the cut score that best distinguished between adequate and inadequate understanding and ability. The selected cut score is rightly grounded in the academic content measured by the items, not any notion of what the "correct percentage" of Level 3 students should be. This is the reason moderators present impact data to standard setting panelists between the second and third rounds, rather than before the first round. The standard setting protocol is designed to prevent an exhortatory purpose from biasing the judgment of panelists away from their primary goal: to operationalize educational performance levels.

Delaware's educational progress this century is nearly the worst in the nation.

11. Prof. Cizek provides an analysis of Delaware scores on the National Assessment of Educational Progress (NAEP) and concludes that, "Delaware has made steady educational progress in reading and mathematics since the 1990s." This is not correct. In fact, Delaware's educational progress in this century is nearly the worst among the 50 states, second only to Alaska. Prof. Cizek commits a "baseline fallacy" in his analysis by choosing an arbitrary baseline and assuming that progress from that point is steady. The data show that Delaware has in fact made steady declines in educational progress this century.

12. Prof. Cizek sets baseline years in the 1990s to measure progress in Math and Reading. However, through much of the 1990s, NAEP did not allow accommodations for students with disabilities and English Learners. Because accommodations were allowed on later NAEP administrations, and because accommodations have different and unpredictable effects on scores and trends in different states, the National Center for Education Statistics denotes all assessment years where accommodations were not permitted

with a superscript as shown in Exhibit A.[4] To avoid this confound, I restrict the analysis to administrations this century, where accommodations are offered consistently. I also display full trend data for each state. This avoids the baseline fallacy by allowing readers to judge whether a linear trend appropriately reflects state progress. I look at how average educational progress in Delaware compares to other states this century on NAEP.

13. On subsequent pages, Figures 1-4 show educational progress this century for all 50 states and the nation. Delaware is in the upper right. I fit a simple trend line to describe progress for each jurisdiction. One line includes a baseline year in the 1990s, as Prof. Cizek does. The other line uses only data this century. The figure shows that Prof. Cizek's choice of baseline dramatically misstates recent progress in Delaware. The simpler story is that Delaware's 4th and 8th graders have declined in educational proficiency this century.

14. The figures show that each state's progress this century can be roughly described by a straight trend line. By estimating the steepness of this line, we obtain a single number that describes the magnitude of annual state progress by subject and grade. Table 1, below, shows each state's ranking on this NAEP progress this century, from 2000-2019. Delaware has the 2nd worst performance in the nation in Grade 4 Math and Reading. In Grade 8, Delaware's performance is the 2nd worst in Math and the 3rd worst in Reading. Across subjects and grades, this is the second worst portfolio in the nation next to Alaska.[5] Inspection of Figures 1-4 on the following pages show that Delaware's declines are concentrated in recent years.

15. Prof. Cizek also cites Delaware's apparent achievement gap reduction for at-risk students, using NAEP data for students eligible for the National School Lunch Program (NSLP). However, recent changes to the NSLP program allow eligible schools to provide NSLP services to students regardless of their economic status. The National Center for Education Statistics concludes that, "readers should interpret NSLP trend results with caution."[6]

---

[4] Exhibit A to this report is a screenshot from the NAEP Data Explorer, https://nces.ed.gov/nationsreportcard/data/, showing Delaware data.

[5] We can rank the "portfolios" of states across subjects and grades either by taking the average of ranks across grades and subjects, or by averaging the slopes of the trend lines across grades and subjects. Delaware is worst in the nation by average rank and second worst next to Alaska by average slopes.

[6] National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/reading/interpret_results.aspx

Prof. Cizek does not account for these changes in his analysis, and his conclusions are not warranted.

Conclusion

16. I stress, however, that Prof. Cizek's flawed conclusions about exhortatory purposes and Delaware's relative and absolute performance on NAEP are not relevant to the conclusion I draw in my original report. That conclusion remains: Delaware has defined a standard of adequate education that large percentages of disadvantaged students do not meet.

Table 1. NAEP state rankings by educational progress from 2000-2019 (50 = best progress, 1 = worst declines)

State	Grade 4		Grade 8	
	Math	Reading	Math	Reading
Alabama	35	44	27	39
Alaska	1	1	1	9
Arizona	45	47	43	45
Arkansas	30	27	44	21
California	42	48	47	50
Colorado	14	23	10	32
Connecticut	10	12	13	44
Delaware	2	2	2	3
Florida	29	49	25	46
Georgia	41	41	46	47
Hawaii	49	46	49	49
Idaho	27	26	34	36
Illinois	28	30	35	25
Indiana	33	36	26	41
Iowa	21	11	6	10
Kansas	6	18	5	6
Kentucky	48	35	30	19
Louisiana	25	42	31	31
Maine	18	4	9	7
Maryland	39	37	32	42
Massachusetts	31	28	45	35
Michigan	8	10	11	23
Minnesota	36	15	15	13
Mississippi	50	50	50	27
Missouri	23	16	33	12
Montana	24	5	4	5
Nebraska	46	29	22	15
Nevada	32	45	38	48
New Hampshire	3	19	29	29
New Jersey	12	34	42	37

New Mexico	44	21	39	26
New York	11	9	19	11
North Carolina	16	31	14	20
North Dakota	26	8	8	2
Ohio	17	22	24	17
Oklahoma	34	32	20	16
Oregon	15	20	7	18
Pennsylvania	13	33	23	28
Rhode Island	37	40	37	38
South Carolina	20	24	12	30
South Dakota	4	6	3	1
Tennessee	47	43	48	40
Texas	22	14	28	14
Utah	40	38	40	43
Vermont	9	7	18	22
Virginia	43	25	41	4
Washington	5	17	16	34
West Virginia	19	3	17	8
Wisconsin	7	13	21	33
Wyoming	38	39	36	24

Date: May 28, 2020

_____

ANDREW DEAN HO

Figure 1. State and national trends on the National Assessment of Educational Progress, Grade 4 Reading. Linear trend lines shown including (dashed) and excluding (bold) a baseline year in the 1990s.
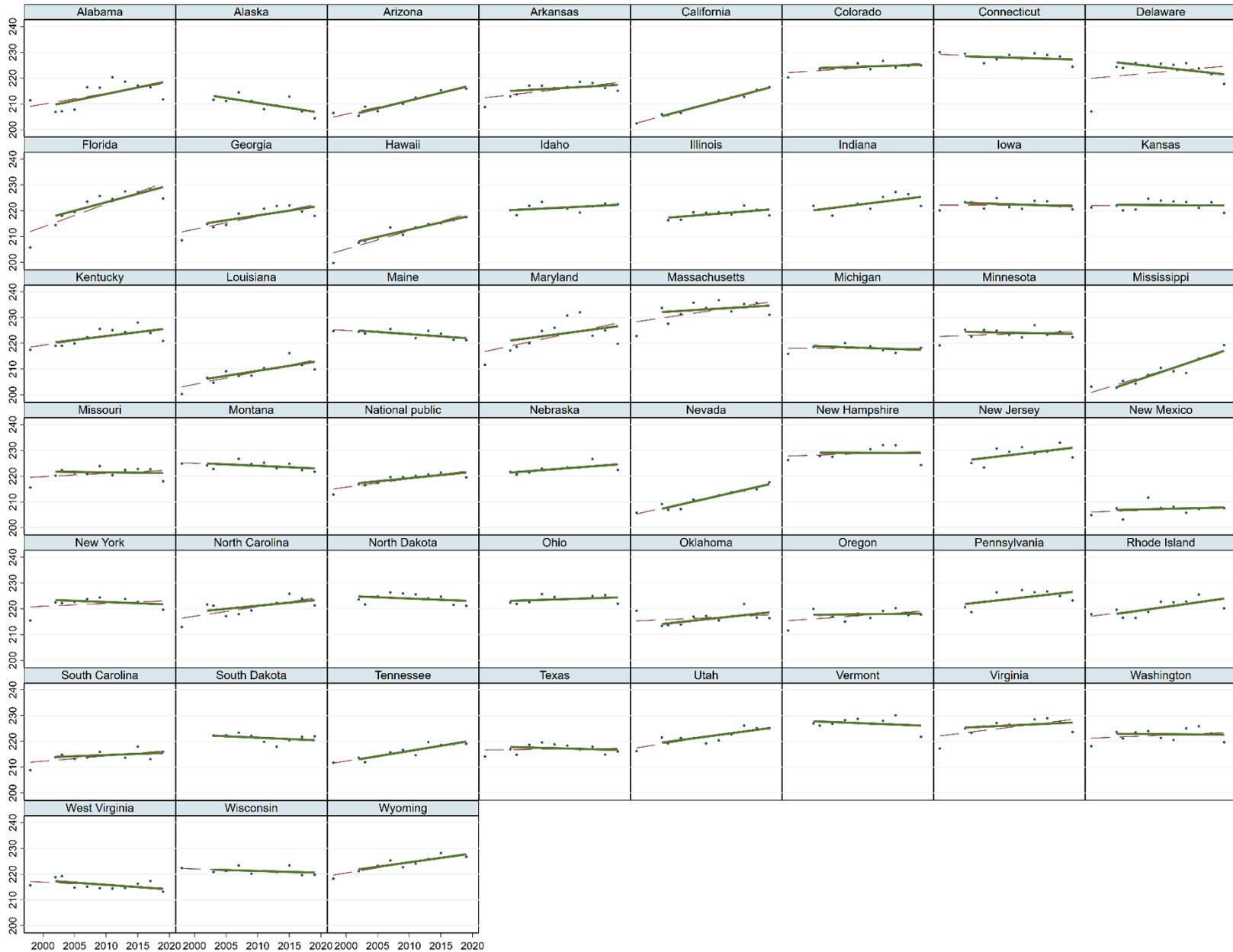
Figure 2. State and national trends on the National Assessment of Educational Progress, Grade 4 Mathematics. Linear trend lines shown including (dashed) and excluding (bold) a baseline year in the 1990s.
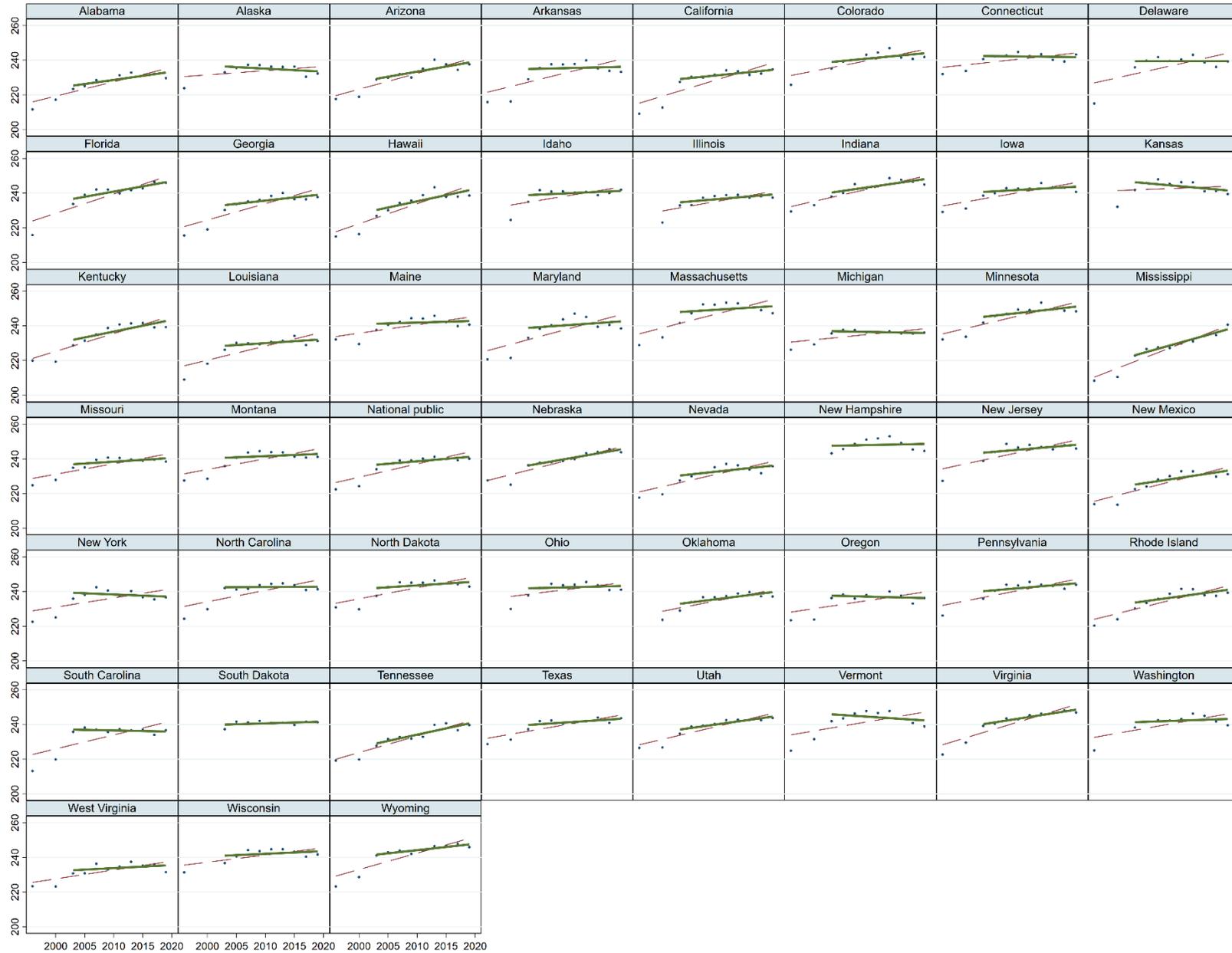
Figure 3. State and national trends on the National Assessment of Educational Progress, Grade 8 Reading. Linear trend lines shown including (dashed) and excluding (bold) a baseline year in the 1990s.
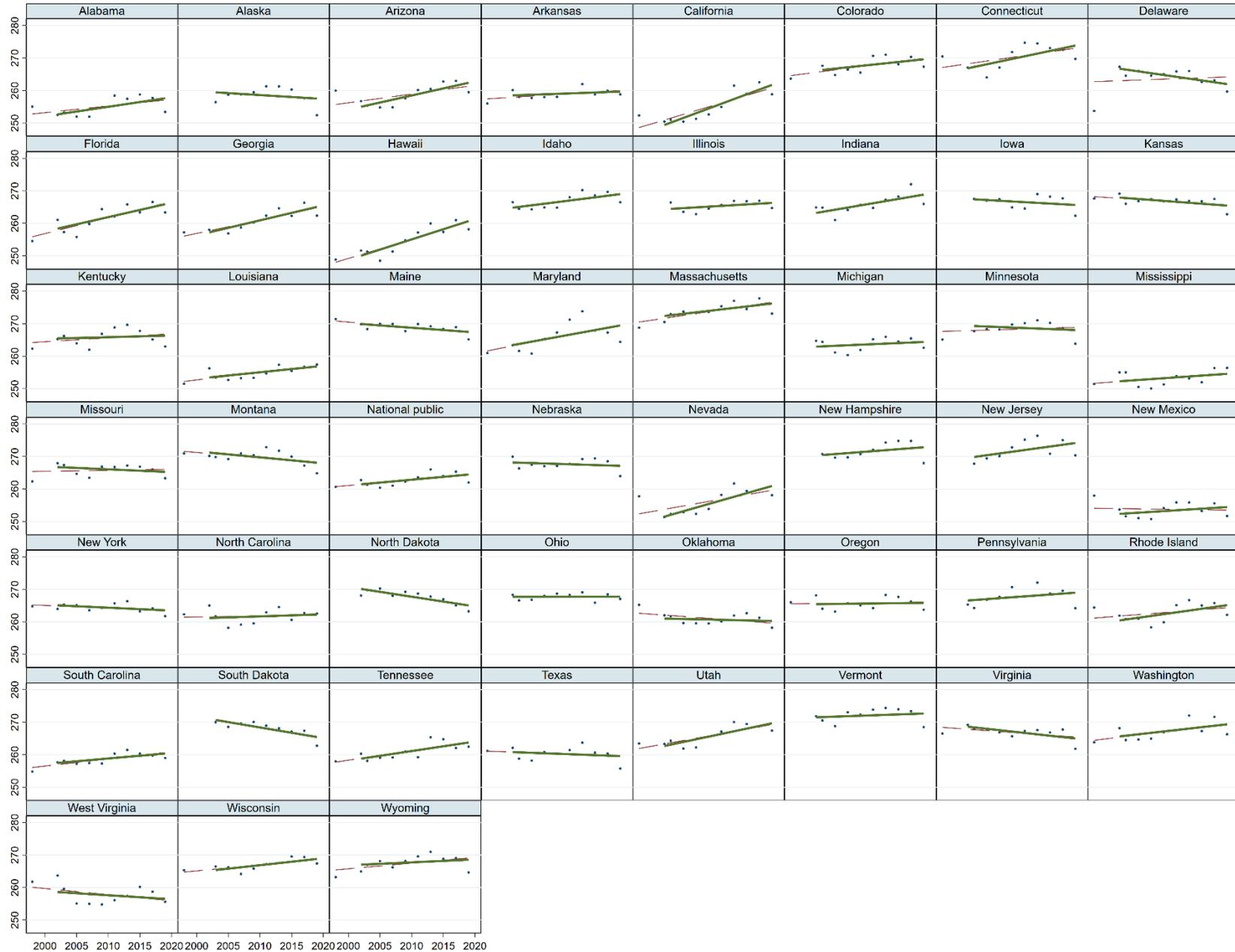
Figure 4. State and national trends on the National Assessment of Educational Progress, Grade 8 Mathematics. Linear trend lines shown including (dashed) and excluding (bold) a baseline year in the 1990s.
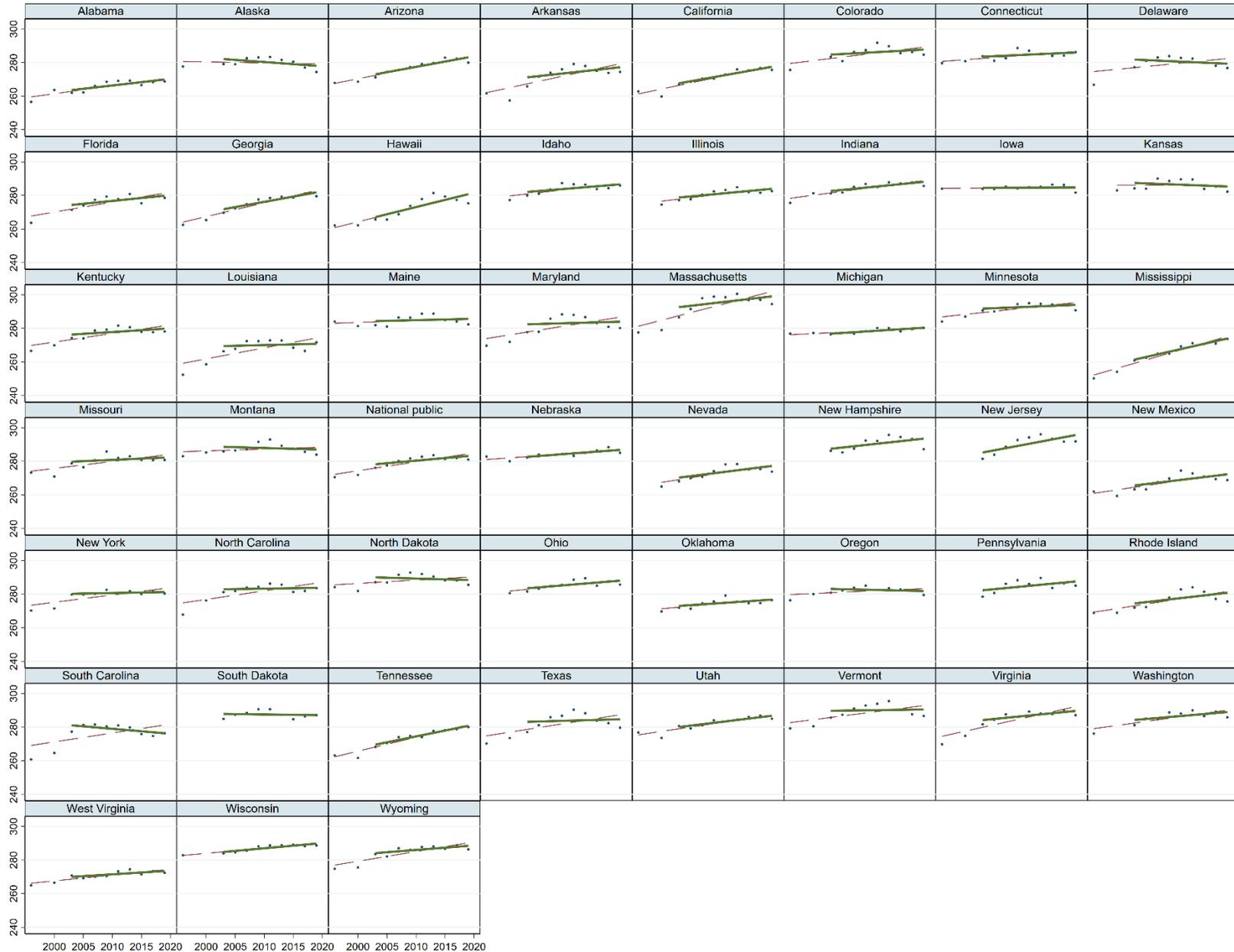
Exhibit A. An example of standard output from the NAEP Data Explorer shows superscripts for testing years without accommodations.

Average scale scores for grade 8 mathematics, by all students [TOTAL] and jurisdiction: 2019, 2017, 2015, 2013, 2011, 2009, 2007, 2005, 2003, 1996, 1992, and 1990

Year	Jurisdiction	All students	Average scale score
2019	Delaware	All students	277
2017	Delaware	All students	278
2015	Delaware	All students	280
2013	Delaware	All students	282
2011	Delaware	All students	283
2009	Delaware	All students	284
2007	Delaware	All students	283
2005	Delaware	All students	281
2003	Delaware	All students	277
1996[1]	Delaware	All students	267
1992[1]	Delaware	All students	263
1990[1]	Delaware	All students	261

[1] Accommodations were not permitted for this assessment.
NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017, and 2019 Mathematics Assessments.