Cayla J., et al., v. State of California, et al.
Los Angeles Superior Court Case No. RG20084386

EXPERT REPORT OF ANDREW HO, Ph.D.

Table of Contents

Qualifications

1. I am a psychometrician, appointed as the Charles William Eliot Professor of Education at Harvard University. I have a Ph.D. in Educational Psychology and an M.S. in Statistics from Stanford University. I serve as a member of the management committee for the Journal of Educational and Behavioral Statistics, the leading journal in educational statistics. In 2012 and 2016, I was appointed by two different U.S. Secretaries of Education to the National Assessment Governing Board in the role of Testing and Measurement Expert. On that board, I chaired the standing committee on Standards, Design, and Methodology from 2015 to 2020. My curriculum vitae is attached as Exhibit A.
2. I am the President-Elect of the National Council on Measurement in Education and a trustee of the Carnegie Foundation for the Advancement of Teaching. I have also served on numerous editorial boards, including the Journal of Research on Educational Effectiveness and the Harvard Data Science Review. I serve on Technical Advisory Committees for state assessment systems in Texas, New York, Rhode Island, Massachusetts, and California. I am a member of the American Educational Research Association, the National Council on Measurement in Education, and the Psychometric Society. I am an elected member of the National Academy of Education.
3. I am recognized within my profession as a leading expert on the use of test scores in educational monitoring and accountability systems.
4. I served as an expert witness once previously in 2021 supporting the Plaintiffs regarding Delaware Public Schools Litigation.
5. I was asked by Plaintiffs' counsel in November 2021 to consult with them, and I have been retained by Morrison Foerster LLP as an expert witness for this case. Plaintiffs' counsel asked me to expand my consultation in February 2023. A statement for the initial and expanded consultation is attached in Exhibit B.

Assignment and Summary of Conclusions

6. Counsel have asked me to offer my professional opinion regarding the meaning of test score results in California through the COVID-19 pandemic. Do percentages of students "meeting standards" in California indicate percentages of students with adequate

knowledge and skills for college and career success? Do declines in these percentages from pre-pandemic baselines indicate declines in percentages of students with adequate knowledge and skills for college and career success? Do increases in achievement gaps from pre-pandemic baselines indicate increasing inequality in adequate knowledge and skills for college success? Counsel have also asked me to comment on test score trends for the plaintiffs' schools and districts, as well as whether the plaintiffs' student report cards are valid indicators of student progress through the pandemic.

7. I summarize ten conclusions from my review of available test score evidence for California students before and through the pandemic. I review these conclusions here and then provide an argumentative and evidentiary basis for these conclusions in ten subsequent sections, respectively.
    1. Average college and career readiness as indicated by the California Assessment of Student Performance and Progress (CAASPP) has declined substantially among California students from a 2019 prepandemic baseline to the most recent publicly available state results in 2022 in grades 3-8. Average declines were largest in mathematics, where the percentage of students meeting standards fell 6 to 7 percentage points. Declines in English Language Arts (ELA) were 5-6 percentage points in early grades and 2-3 percentage points in later grades.
    2. Educational inequality also increased in California from 2019 to 2022. Achievement gaps widened between average White students and average Black students and between average White students and average Hispanic students. Average achievement gaps also widened between low-income and high-income White, Black, and Hispanic students in lower grades. The magnitude of learning loss was greater in California school districts that serve more low-income students.
    3. A state calculation appears to contradict my assertion that race-based and early-grade socioeconomic gaps have increased (Motion for Summary Judgment, 2023). I demonstrate why this calculation is fundamentally flawed. The state chose a biased calculation of achievement gaps that leads to an incorrect conclusion that achievement gaps have stayed the same through the pandemic. This bias is easily corrected by my preceding analysis that shows increasing gaps consistent with disparate impact.

4. An “interpretation guide” for CAASPP test scores from 2021 to 2022 appears to show accelerated recovery (California Department of Education, 2022a). I explain that this analysis only applies to students who elected to take the state test in 2021. This amounts to 1 out of every 5 students. I also show that this subpopulation is particularly high-scoring, has higher average socioeconomic status, and has more White students than the full state population. A similar analysis that the state conducted from 2019 to 2022 but was not included in the interpretation guide shows substantial and sustained academic learning loss.
5. California neglected to release straightforward “longitudinal” analyses of school- and district-level CAASPP test scores that could have documented the magnitude of academic learning loss through the pandemic. The limitations of available analyses that the state did make available, as well as testimony from state officials who admitted ignorance or confusion about the magnitude of academic learning loss, are consistent with a pattern of state disinterest in existing test scores and what they could measure through the pandemic.
6. CAASPP test score trends from available data in the plaintiffs’ districts, Los Angeles Unified School District and Oakland Unified School District, show average test score declines in elementary school grades from 2019 to 2022. There are inconsistent patterns in middle school grades and in plaintiffs’ schools within these districts. Absent longitudinal data that the state could have analyzed for these districts and schools, I cannot draw firmer conclusions about school- and district-level learning loss.
7. Absent longitudinal test score data, can student report cards enable valid conclusions about academic progress through the pandemic? I explain that report cards rely on teacher judgments that are not rigorously standardized from year to year. Report card data are thus not defensible evidence for the educational progress of students in this case.
8. California results from the National Assessment of Educational Progress (NAEP) from 2019 to 2022 show average California students declined in Mathematics in grades 4 and 8 and in Reading in grade 4. I contextualize this finding in terms of California’s standing on NAEP prior to the pandemic and the relatively limited

scope and relevance of NAEP to California state standards. I conclude that the CAASPP test scores and not NAEP results are the relevant results for this case.

9. I review national research findings that complement findings of academic learning loss in California. These studies show that academic learning loss is generally larger in mathematics, disproportionate for historically marginalized groups, and associated with remote learning.
10. I affirm the relevance of California state test score results by reviewing the content of the tests, the legitimacy of state content standards, the process the state used to adopt proficiency standards, and psychometric evidence from technical manuals. California's adopted performance standards are legitimate and reasonable, and there is convincing evidence that state test scores are unbiased overall and for subgroups.

8. Triangulating across state data sources, I conclude that average college and career readiness has declined substantially among California elementary and middle school students from 2019 to 2022, particularly in mathematics but also in ELA, particularly in early grades, and particularly for Black and Hispanic students. Throughout these analyses, I draw from my knowledge of academic publications acquired throughout my education and career, and the professional standards of my field, known as the Standards for Educational and Psychological Testing. These are the consensus standards of three professional organizations, the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (2014).

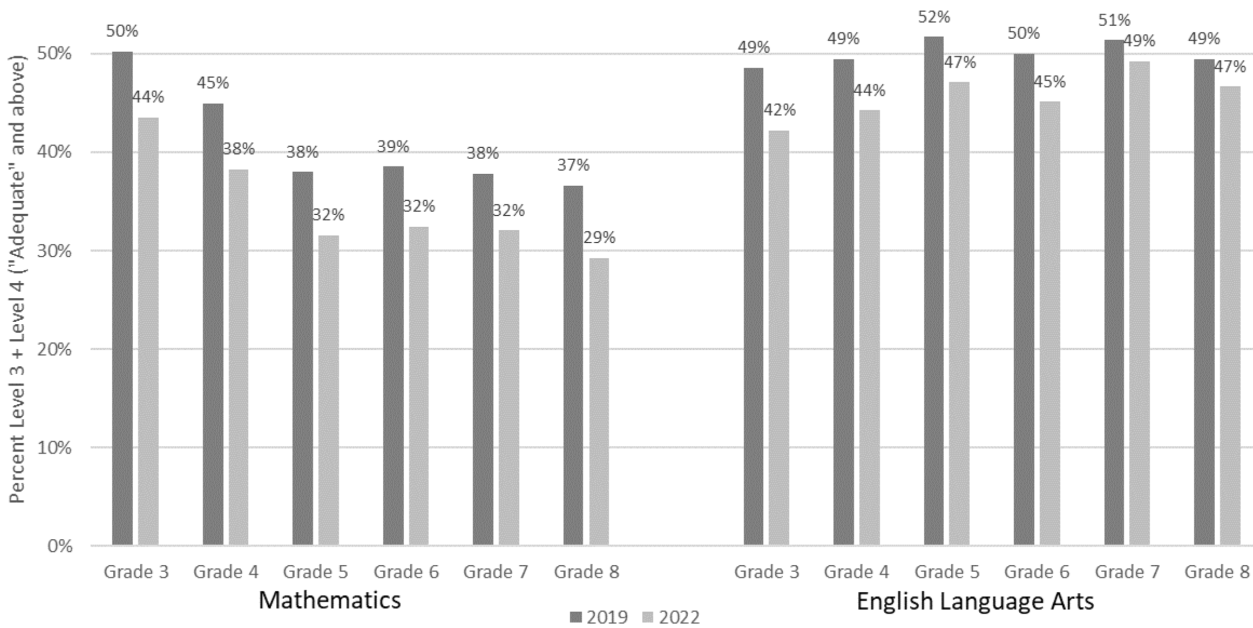
1) California Test Score Trends from 2019 to 2022

9. In the spring of 2022, California testing was sufficiently widespread to support robust inferences about statewide educational performance. The number of test scores in 2022 was 94% of the 2019 total across subjects (English Language Arts and Mathematics) and grades (3-8 and 11). Figure 1 shows that the 2022 percentages of Level 3 and Level 4 students (meeting or exceeding expectations) continues to trail 2019 percentages in both subjects and all grades, with particularly substantial losses in Mathematics and in early grade ELA. Mathematics declines were 6-7 percentage points across grades 3-8, and ELA

declines were 5 to 6 percentage points in grades 3-6 and 2 and 3 percentage points in grades 7 and 8 respectively.

10. As I explain later, the percent of students who are at Level 3 and Level 4 are important substantively, politically, and legally, as California has adopted a policy definition where a Level 3 student "demonstrates adequate understanding of English Language Arts (ELA) and mathematics and the ability to apply the knowledge and skills for his or her grade level that are associated with college and career readiness" (California Department of Education, 2021, p. 104).

Figure 1. Declining percentages of Level 3 and Level 4 students in California from prepandemic 2019 to 2022 on the California State Test: Smarter Balanced Assessment Consortium.



11. Table 1 shows that average test score trends are negative for major racial and ethnic groups. Because proficiency percentages like those shown in Figure 1 can bias group comparisons (Ho, 2008), the left-hand side of Table 1 uses scale score points to enable comparison of trends across groups. For example, in grade 3 ELA, Black students lost 18 scale score points in 2022 compared to their 2019 cohort, Hispanic students lost 22 points, and White students lost 11 points. In grade 3 math, these declines were 21, 22, and 9 points respectively. Appendix A shows trends for economically disadvantaged students.

Table 1. Trends in California test scores from prepandemic 2019 to spring 2022 in scale score points (left) and a rough "months of learning" conversion (right).

		2019-2022 Trend (SS)			Standard	2019-2022 Trend (months)		
Subject	Grade	Black	Hispanic	White	Deviation	Black	Hispanic	White
ELA	Grade 3	-18	-22	-11	92	-5	-7	-3
	Grade 4	-12	-16	-9	97	-3	-4	-2
	Grade 5	-10	-13	-10	101	-3	-4	-3
	Grade 6	-9	-11	-11	98	-2	-3	-3
	Grade 7	-5	-6	-6	103	-1	-2	-2
	Grade 8	-6	-8	-9	102	-2	-2	-2
Math	Grade 3	-21	-22	-9	83	-7	-7	-3
	Grade 4	-21	-23	-10	86	-7	-7	-3
	Grade 5	-21	-22	-14	95	-6	-6	-4
	Grade 6	-16	-18	-14	110	-4	-4	-3
	Grade 7	-17	-18	-18	115	-4	-4	-4
	Grade 8	-17	-22	-23	124	-4	-5	-5

Note: Negative numbers indicate decreases in average scores from 2019 to 2022 in scale scores (SS) or months of learning. ELA = English Language Arts. Hispanic students are Hispanic or Latino students. Standard deviations from 2017 are shown as references to estimate effect sizes. Informal "months of learning" interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year.

12. The magnitude of scale score points tends to be difficult for general audiences to understand. A common conversion known as an "effect size" expresses the change in terms of "standard deviation units." I use standard deviation units published by the California Department of Education (2018). These in turn can be roughly converted to "months of learning" interpretations by multiplying by 27 as a rule of thumb. This amounts to simple conversion, like expressing a difference in centimeters instead of inches. The right-hand side of Table 1 shows these trends in months of learning units, along with the standard deviations of the scale scores shown as reference. The calculation is simply: $27 * \frac{\text{scale score trend}}{\text{standard deviation}}$. This assumes learning occurs steadily on average at the rate of one third of a standard deviation in 9 months of schooling. Under this assumption, Table 1 shows how many months of learning students from each subgroup have lost relative to their cross-cohort peers from 2019 to 2022.

2) California Average Achievement Gap Trends from 2019 to 2022

13. As Table 1 implies, California test scores show that racial inequality increased in both subjects and early grades. Economic inequality also increased within racial and ethnic categories in early grades. Table 2a shows that average score differences between White and Black and between White and Hispanic or Latino students generally increased from 2019 to 2022 in early grades. In grade 3, for example, the White-Black gap increased by 6 points in ELA and 12 points in math, and the White-Hispanic gap increased by 11 points in ELA and 13 points in math. I also compare trends in socioeconomic gaps within racial categories. In grade 3, for example, economically disadvantaged Black students lost 4 more points than their economically non-disadvantaged Black peers in ELA and 7 more points in math. In the same grade, economically disadvantaged White students lost 4 more points than their economically non-disadvantaged White peers in ELA and 5 more points in math. Appendix A shows gap trends for economically disadvantaged students.

Table 2a. Gap trends in California test scores from prepandemic spring 2019 to spring 2022 in scale score points.

		Race/Ethnicity Gap Trend (SS)		Economic Gap Trend (SS)		
Subject	Grade	White-Black	White-Hispanic	Black	Hispanic	White
ELA	Grade 3	6	11	4	4	4
	Grade 4	3	7	3	-1	0
	Grade 5	1	3	1	-2	1
	Grade 6	-2	0	1	-5	-2
	Grade 7	-1	-1	2	-4	-1
	Grade 8	-3	0	1	-5	-1
Math	Grade 3	12	13	7	5	5
	Grade 4	11	13	8	3	4
	Grade 5	7	9	0	1	3
	Grade 6	2	4	-2	-4	-1
	Grade 7	-1	0	2	-5	-3
	Grade 8	-6	-1	-1	-7	-4

Note: Positive numbers indicate increases in average scale score (SS) differences from 2019 to 2022 in scale score points. ELA = English Language Arts. Hispanic students are Hispanic or Latino students. Economic gaps are those between Economically Non-Disadvantaged students and Economically Disadvantaged students within the racial and ethnic categories noted.

14. Table 2b shows how many months of learning students from certain subgroups have lost relative to their cross-cohort peers from 2019 to 2022. Black and Hispanic students third graders fell months behind their White peers in ELA and math, and economic gaps within racial/ethnic categories also widened in early grades. Although some gaps closed slightly in higher grades for some groups, like for Hispanic/Latino students, the gap closure was very small compared to the amount that the entire group lost, as shown in Table 1. Appendix A shows gap trends for economically disadvantaged students.

Table 2b. Gap trends in California test scores from prepandemic spring 2019 to spring 2022 in "months of learning" units.

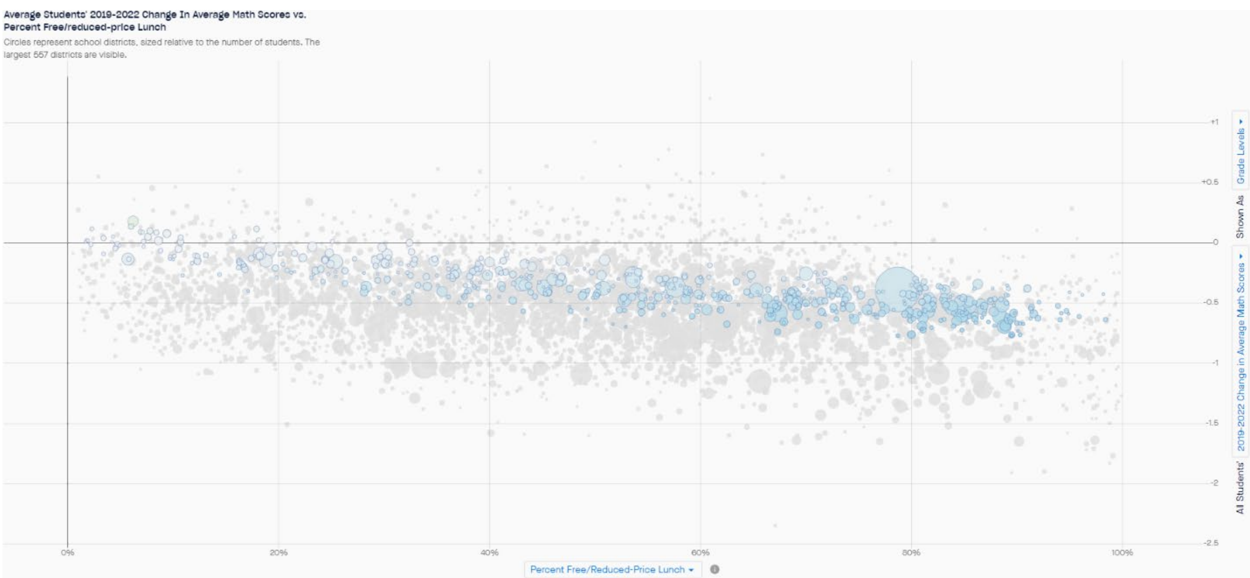
		Race Gap Trend (months)		Economic Gap Trend (months)			Standard
Subject	Grade	White-Black	White-Hispanic	Black	Hispanic	White	Deviation
ELA	Grade 3	2	3	1	1	1	92
	Grade 4	1	2	1	0	0	97
	Grade 5	0	1	0	0	0	101
	Grade 6	-1	0	0	-1	-1	98
	Grade 7	0	0	1	-1	0	103
	Grade 8	-1	0	0	-1	0	102
Math	Grade 3	4	4	2	2	2	83
	Grade 4	3	4	3	1	1	86
	Grade 5	2	2	0	0	1	95
	Grade 6	1	1	-1	-1	0	110
	Grade 7	0	0	0	-1	-1	115
	Grade 8	-1	0	0	-2	-1	124

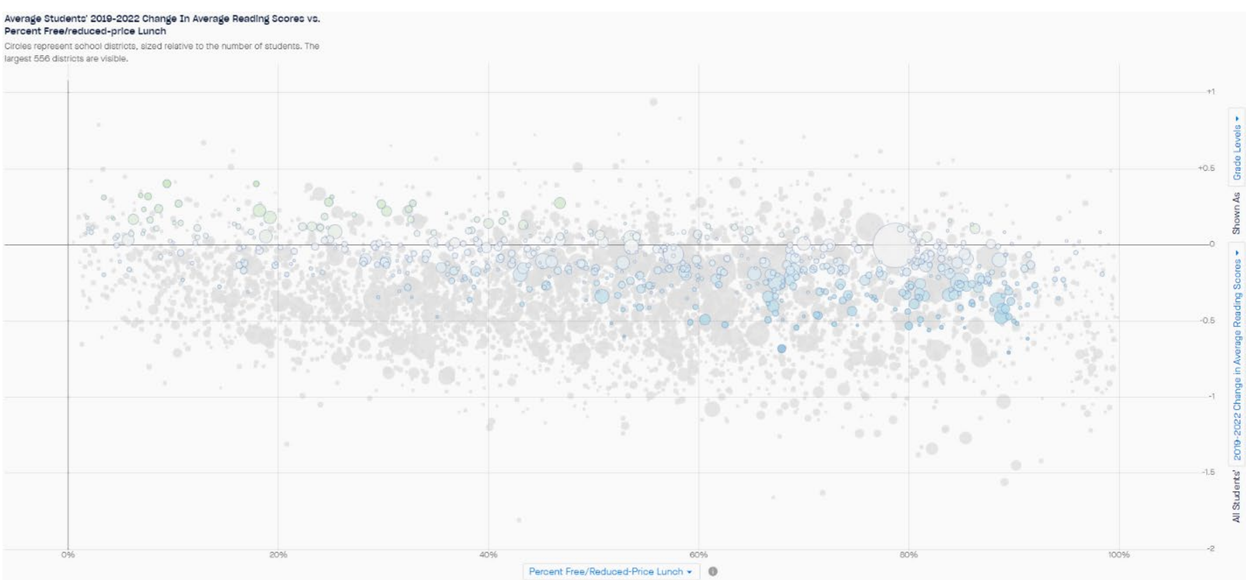
Note: Positive numbers indicate increases in average months-of-learning disparities from 2019 to 2022. ELA = English Language Arts. Hispanic students are Hispanic or Latino students. Economic gaps are those between Economically Non-Disadvantaged students and Economically Disadvantaged students within the racial and ethnic categories noted.

15. Additional insight to socioeconomic achievement gap trends comes from district-level data from my project with collaborators known as the Stanford Education Data Archive (Reardon et al., 2023). The figures on the following page, which are also available at our website http://edopportunity.org, show California districts in the foreground in light blue. Academic learning losses in Mathematics or ELA are on the vertical axis, and the percentage of students in a district who are eligible for free and reduced-price lunch, a proxy for lower socioeconomic status, is on the horizontal axis. The downward sloping

pattern from left to right means that districts with lower socioeconomic status also had more academic learning losses, more than half a grade level for the lowest socioeconomic status districts in California in Mathematics and roughly a quarter of a grade level for the lowest socioeconomic status districts in California in Reading Language Arts. In contrast, the highest socioeconomic status districts in California at the left had no average academic learning losses at all.

Figure 2. Data from the Stanford Education Data Archive (Reardon et al., 2023) show the socioeconomic gradient of district-level academic learning losses in California (in light blue), Mathematics above and Reading Language Arts below.





16. The patterns shown in Tables 1 and 2 show that substantial losses and worsening inequality are particularly severe in early grades. This pattern is similar to those from studies (e.g., Goldhaber Kane, McEachin, & Morton, 2022) as well as reviews across many state and national studies (e.g., Cohodes et al., 2022). These patterns are also consistent with the theory that learning remotely outside of school is particularly difficult for younger and economically disadvantaged children who are not accustomed to digital devices or cannot afford them.

3) A State Calculation for Achievement Gap Trends is Fundamentally Flawed

17. In support of the defendant's motion for summary judgment (2023), authors cited an analysis and drew a conclusion that seems to contradict my preceding analysis: "pre- and post-distance-learning scores on the statewide Smarter Balanced Assessment in ELA and mathematics show that the pre-pandemic achievement gaps between Black and Latinx students as compared to white students, and between students of differing economic status, remained about the same, instead of widening, as would be expected if plaintiffs' allegations of disparate impact were correct" (p. 23). In support of this conclusion, the motion used the following formulation, "while statewide the overall percentage of students meeting or exceeding standards declined between the 2018-19 and 2021-22 school years, that decline was consistent between the relevant student groups" (Motion for Summary Judgment, 2023, p. 23).
18. This formulation for achievement gaps has known technical flaws that I have outlined in my research (Ho, 2008). My colleague Morgan Polikoff summarized this in an open letter to the U.S. Department of Education as follows, "percent proficient… is a very poor measure of performance gaps between subgroups, because percent proficient will be affected by how a proficiency cut score on the state assessments is chosen (Ho, 2008; Holland, 2002). Indeed, prior research suggests that using percent proficient can even reverse the sign of changes in achievement gaps over time relative to if a more accurate method is used (Linn, 2007)" (Polikoff, 2016). Professor Polikoff's open letter was signed by over 50 educational researchers and experts, including me and California State Board of Education President Linda Darling-Hammond.

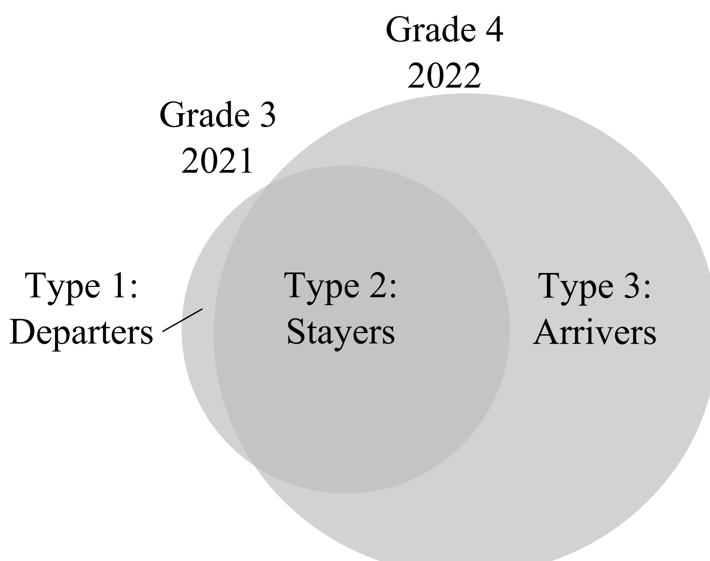
19. The “more accurate method” that Professor Polikoff refers to is the method that I present in the preceding section and Appendix A that shows that racial achievement gaps increased in early grades and for economically disadvantaged students. The problem with the state’s approach to using “percent meeting standards” as a basis for achievement gap trends begins with the fact that student test-score distributions are “bell shaped,” with greater percentages of students near the middle of the test-score distribution than at the extremes. When a proficiency standard happens to be near the middle of a distribution, any movement of that distribution will show large trends due to the density of students near the proficiency standard. A proficiency-based trend is thus a biased mixture of the movement of a distribution and the density of students near the proficiency cut score.
20. These biases are particularly severe and predictable when proficiency percentages for one subgroup are near 50% and proficiency percentages for another subgroup are much higher or lower. In these cases, the group near 50% will show greater trends even when the actual movement of their test-score distribution is minimal. This is exactly what happened with the state’s calculation for achievement gaps. The percent of White students who are proficient is closer to 50%. Thus, when the White student score distribution shifted slightly, it showed larger (relatively biased) negative trends. The Black and Hispanic student test score distributions declined more, but the state’s chosen metric masked this decline. The correct conclusion is that racial achievement gaps widened in early grades and for socioeconomically disadvantaged students in California from 2019 to 2022, as I showed in the preceding section and Appendix A.

4) California’s Analysis of Test Score Trends from 2021 to 2022 is Short-Sighted

21. In October of 2022, California released an “Interpretation Guide to the 2021–22 Statewide Assessment Results” (California Department of Education, 2022a). A contributed article by California State Board President Linda Darling-Hammond in December of 2022, excerpted one of the figures from this guide and provided one interpretation of the results, “if students continue to learn at this accelerated pace, they will not only close the gaps with prior cohorts, but will move ahead of them in the years to come.” I describe here how this analysis is short-sighted and supports rosier interpretations of the educational recovery than is warranted.

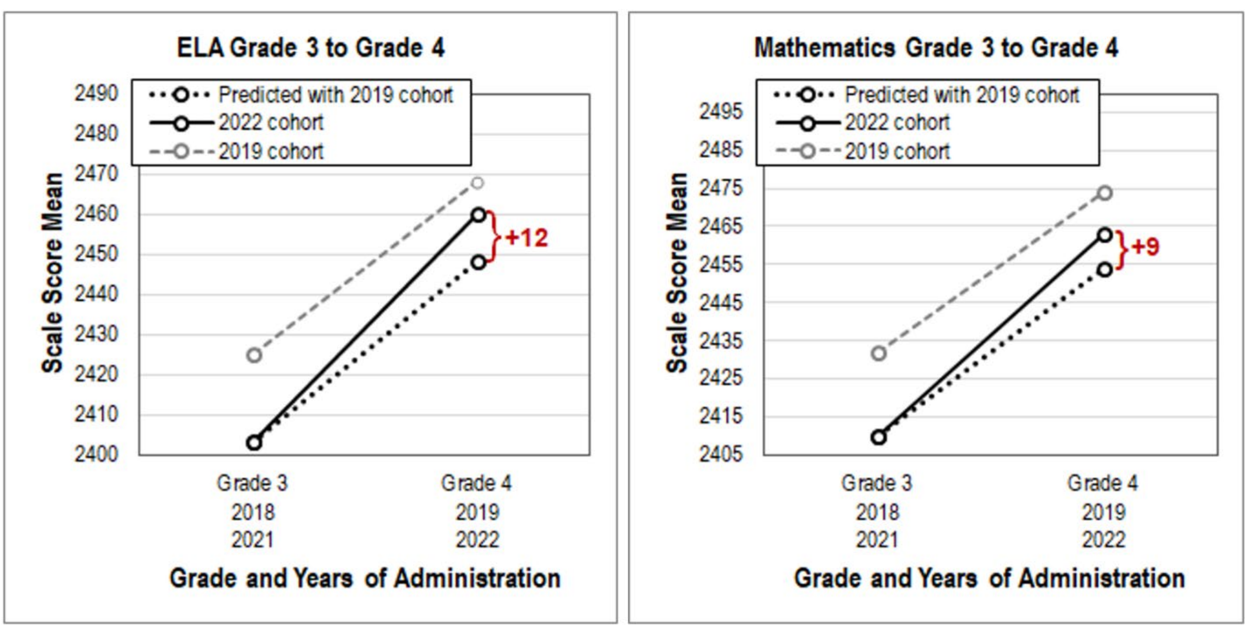
22. Test score trends require careful estimation when populations change. In 2021, when the tested population was only 23% of the size of the 2019 tested population, a simple comparison of scores from 2019 to 2021 would have confounded true changes in state proficiency with a change in the proficiency of the students that happened to take the test. Comparing a near-complete population from 2019 to 2022, as I did in the previous sections, is in contrast more defensible. Still, as I describe in my research (Ho, 2021), it is possible to conduct responsible "longitudinal" analyses of test scores by following individual student test scores over time. The state used a functionally similar method to one of the methods that I recommend in my research that I call the "fair trend." However, they did not implement the necessary complementary method that I also recommended, called the "equity check" (Ho, 2021). Why does this result in a short-sighted and overly optimistic perspective on educational progress?
23. Figure 3 below is adapted from my research and helps to illustrate why the state's report is incomplete. With roughly a quarter of California students testing in the spring of 2021 but roughly 95% testing in 2022, a large and nonrepresentative population of students rejoined the cohort that year. The state analysis (California Department of Education, 2022a) asks a relevant question: Did students who tested in 2021 and 2022 make greater than expected progress? Figure 3 describes these students as "Type 2 students."

Figure 3. Illustrating challenges to trend interpretations when populations change as they did in California from 2021 to 2022. From a smaller and nonrepresentative cohort of third graders in 2021, a few students may depart (Type 1), and most likely stay (Type 2). Many previously untested students rejoin the cohort in fourth grade in 2022 (Type 3).



24. The state analysis uses historical longitudinal data to show that these “Type 2 students” are making greater than expected progress in many subjects and grades. Figure 4 on the following page, excerpted from Figures 8 and 12 of the California Department of Education (2022a) report, show a case where these Type 2 students are making substantially greater progress than expected and less progress than expected, respectively.
25. This analysis is remarkably incomplete. As Figure 3 makes clear, this inference is only accurate for the considerable minority of students who took the test in 2021. Although both the report and Board President Darling-Hammond acknowledge this, the implication is that we can draw inferences from the progress of these students to the substantial majority of other students for whom 2021 data does not exist. The report already shows that these Type 3 students are likely to be more educationally disadvantaged and are proportionally less White and more Hispanic. In addition, the state could have included a simple calculation that shows that Type 3 students are also lower scoring. I provide this calculation in Table 3 on the following page.

Figure 4. Excerpts from a 2022 California Department of Education Interpretation Guide showing above-expected and below-expected progress by a minority of California students from 2021 to 2022. Source: https://www.cde.ca.gov/ta/tg/ca/documents/assessmentresultsguide22.docx



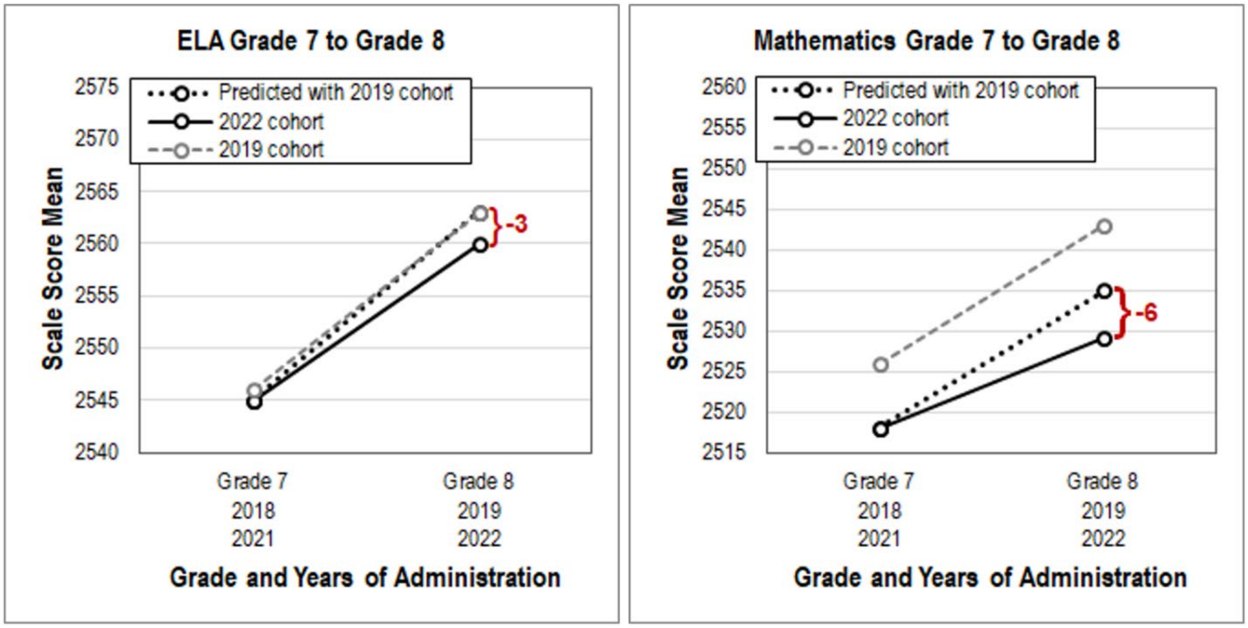


Table 3. Students who take tests in 2021 and 2022 (Type 2) are not representative of and lower scoring than those who take tests only in 2022 (Type 3).

Subject-Grade	# Type 2 Students	# Students in 2022	% Type 2 Students	Type 2 Avg SS	Type 3 Avg SS	Score Diff (SS)	Standard Deviations	"Months of Learning"
ELA Grade 4	75519	419781	18%	2460	2453	-7	97	-2
ELA Grade 5	78160	427741	18%	2498	2490	-8	101	-2
ELA Grade 6	78722	425490	19%	2520	2513	-7	98	-2
ELA Grade 7	81904	416984	20%	2550	2541	-9	103	-2
ELA Grade 8	91408	435110	21%	2560	2551	-9	102	-3
Math Grade 4	80407	421578	19%	2463	2452	-11	86	-3
Math Grade 5	83103	429046	19%	2485	2474	-11	95	-3
Math Grade 6	83550	426374	20%	2507	2495	-12	110	-3
Math Grade 7	81803	417683	20%	2521	2506	-15	115	-3
Math Grade 8	91276	435449	21%	2529	2515	-14	124	-3

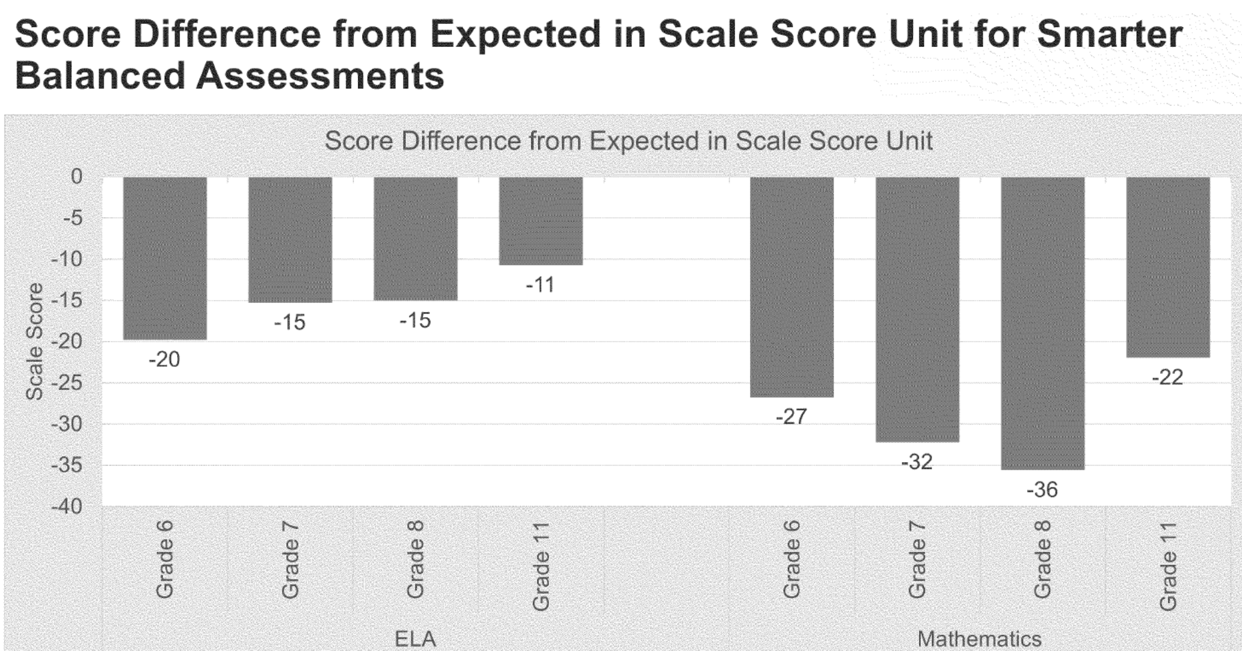
Note: Negative scale score (SS) differences indicate that students who take tests only in 2022 score lower than students who take tests in both 2021 and 2022. ELA = English Language Arts. Standard deviations from 2017 are shown as references to estimate effect sizes. Informal “months of learning” interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year.

26. Combining data from the California report and published state test score results, Table 3 shows that the students analyzed in the California Department of Education interpretation

guide represent only 18-21% of tested students in 2022. These students are substantially higher scoring than the students who test only in 2022, with average score differences ranging from 7-9 points in ELA and 11-15 points in math. In terms of the “months of learning” calculation I describe above, this means that the students that the state analysis is ignoring are 2-3 months behind the students they include.

27. Notably, in a presentation at a Regional Assessment Network Meeting in 2022 (Bacher & Liang, 2022, DEFS-111769), the state also presented results for middle-school students from 2019 to 2022, rather than limiting their analysis to progress from 2021 to 2022. Figure 5 shows the results of this analysis below, which reveal substantial declines, greater than the cross-sectional trends suggest. Showing slight recovery in some grades from 2021-2022 while ignoring the substantial declines from 2019-2022 robs readers of the full context of academic learning loss through the pandemic. As a researcher, I am also disappointed that the state has the ability to conduct these state-level longitudinal analyses but does not report them widely nor perform these analyses for schools and districts to enable them and their constituents to understand and address academic learning loss.

Figure 5. From DEFS-111804, results from a “matched cohort analysis” that the state presented to a regional network shows substantial academic learning loss for Grades 6-8 and 11 from 2019 to 2022.



5) Patterns of state neglect of academic learning data and appropriate methods of analysis

28. My assessment of academic learning loss in sections 1 and 2 was hindered by my inability to find longitudinal analyses that addressed changing features of the tested population through the pandemic. The state made only some of the appropriate longitudinal analyses available in "interpretation guides" (see the previous section) that only focused on state-level results, with no disaggregation by subgroups, districts, or schools. In my review of transcripts from depositions of state officials, I find numerous responses that indicate to me a lack of awareness of or interest in data that could enable accurate estimates of academic learning loss. The responses also indicate to me ignorance of standard statistical and psychometric methods that could enable accurate estimates of learning loss, including those that I have proposed (Ho, 2021) and those that Pier et al. (2021) implemented in CORE districts. I select and present these quotes in Exhibit 1.
29. For example, Director of the Student Achievement and Support Division, Lindsay Tornatore, does not answer a question about whether estimates of learning loss exist, deferring to others. Chief Deputy Superintendent Mary Nicely is not aware of research that estimates learning loss in California and thinks that "it would be hard even for researchers to come to any conclusions." As I have shown and other researchers have demonstrated, estimating academic learning loss is, in fact, possible from California's available data.

Exhibit 1: Selected quotes from depositions relating to measuring academic learning loss.

Question posed to Lindsay Tornatore on October 13, 2021: "Does CDE have an estimate of the amount of learning loss that California public school students suffered as consequence of the pandemic?"
Lindsay Tornatore: "And I defer to other programmatic content expert colleagues that would be able to speak to that, to answer your question."
Question: "Well, my first question is: Do you know in fact whether or not such estimates exist?
Answer: "Again, I defer to other CDE colleagues."
Question: "Do you know whether or not such estimates exist?"
Answer: "I'm not comfortable answering this question because it is outside of my programmatic expertise."

Question posed to Mary Nicely on October 27, 2021: "Is CDE aware, to your knowledge, as to estimates of learning loss as consequence of the pandemic in California or across the country that other entities or institutions or researchers have prepared?"

Mary Nicely: CDE does not. And without those assessments, I think it would be hard even for researchers to come to any conclusions."
Question: Does CDE know if there is, in fact, any research out there that estimates learning loss as consequence of the pandemic either in California or across the nation?
Answer: "Yeah, I do not know."

Question posed to Cheryl Cotton on January 24, 2022: "...to your knowledge, has anyone in the CDE made any investigation or inquiry into the extent of achievement gaps during the COVID pandemic?"
Cheryl Cotton: "Not that I'm aware."
Question: "Do you know if anyone else within the CDE that would have that information or knowledge?
Answer: "No at this time."
…Question posed to Cheryl Cotton on January 24, 2022: "What was your conclusion, based on your review of the test scores?"
Cheryl Cotton: "I think that the challenges with test administration at that time does not allow for broad conclusions to be drawn about student -- student results or student data."

Question posed to Mindi Parsons on January 27, 2022: "To your knowledge, has anybody else at CDE conducted an inquiry into the extent of learning loss experienced by students during the 2020-21 school year?
Mindy Parsons: "Not to my knowledge."

Question posed to Joshua Strong on February 4, 2022: "... does it seem right that you've not seen a way that this learning loss can be measured?"
Joshua Strong: "That's not actually part of my responsibility or the responsibility of the office."
Question: "Do you know whose responsibility it is?"
Answer: "It's the responsibility of the LEA itself."

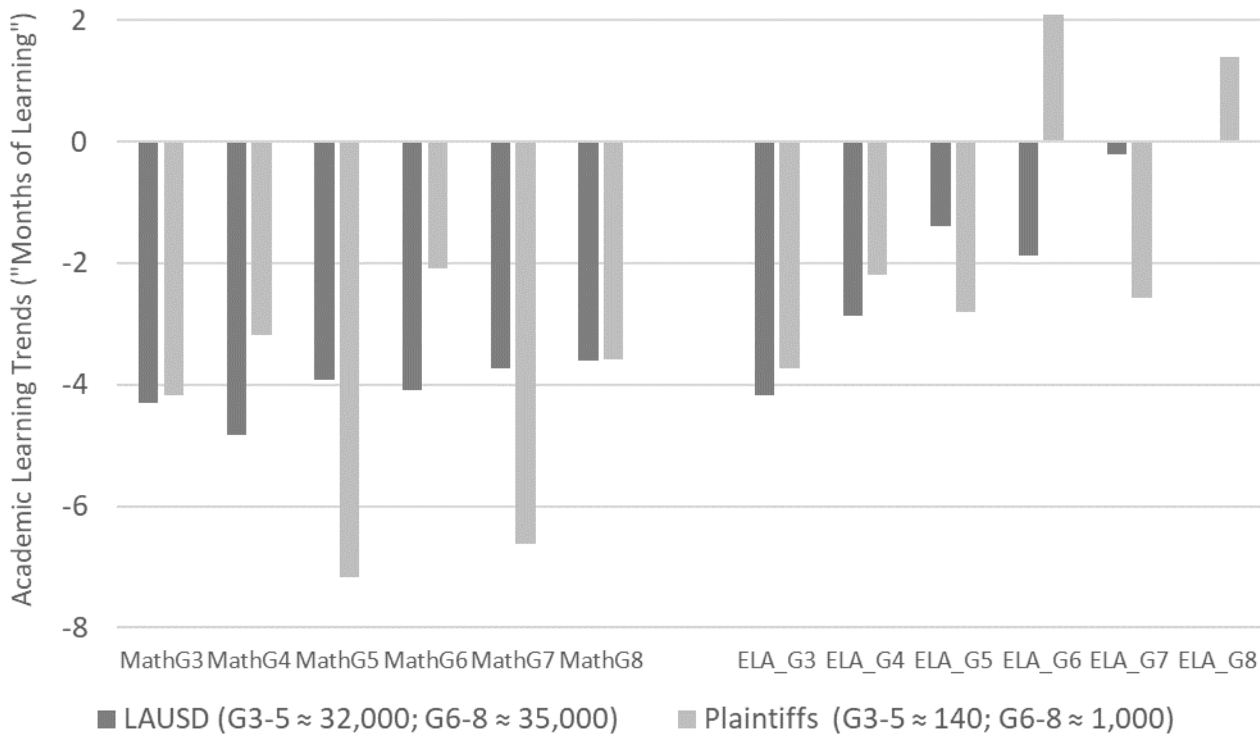
30. Exhibit 1 shows that Deputy Superintendent of Public Instruction and Measurement Cheryl Cotton was not aware of any investigation or inquiry into the extent of achievement gaps during the COVID-19 pandemic. Like Chief Deputy Superintendent Nicely, Deputy Superintendent Cotton suggests that measurement challenges are insurmountable. I disagree. Administrator of the Integrated Student Support and Programs Office Mindi Parsons is not aware of academic learning loss research, either. And Administrator of the System of Support Office Joshua Strong abdicates responsibility of measuring learning loss to local agencies.
31. These statements, along with the lack of appropriate analyses of existing test score data, suggest to me that, by neglect or by intent, the state made their unsubstantiated belief that academic test score data could not be useful into a self-fulling prophecy. Data currently

exist in state repositories to answer questions about the magnitude of academic learning loss for jurisdictions and subgroups in the state of California, untapped.

6) Plaintiff School and District Results on the California State Test

32. Counsel provided me with a list of the plaintiffs' schools and asked for my conclusions about academic learning trends in these schools from 2019 to 2022 from available test score data. Unfortunately, as I mention in the previous section, the state does not provide longitudinal analyses to enable me to draw precise conclusions. The ideal analyses require student-level longitudinal data that the state possesses, not aggregate-level statistics that they make available to the public. Without student-level longitudinal data, it is difficult to distinguish student academic progress from changes in the composition of schools. For example, in most standalone elementary and middle schools, there are essentially no students who were in the tested grades 3-5 or 6-8 in 2019 who are still in the school in 2022. Changes in test scores can thus reflect changes in student composition.
33. State- and district-level trend analyses tend to be more robust because students move less frequently between states and districts than they do between schools (Reardon et al., 2019). For these reasons, and because between-school movement has been particularly substantial during the pandemic, I expect simple aggregate statistics that I present in this section to be less accurate at the school level than they may be at the district level.
34. Figure 6 shows academic learning trends from the spring of 2019 to the spring of 2022 for the Los Angeles Unified School District (LAUSD) and for plaintiffs' schools within the same district. Darker bars show LAUSD trends and lighter bars show the weighted average of plaintiffs' school results. For example, LAUSD shows declines of roughly 4 months of learning in mathematics, and the plaintiffs' schools with available data show, on average, the same declines, with slightly larger declines in grades 5 and 7 and smaller declines in other grades. In ELA, LAUSD shows larger declines in early grades and no changes in higher grades. The plaintiffs' schools show more scattered results; for example, there are gains in grades 6 and 8 and a larger decline in grade 7. Smaller sample sizes and a greater likelihood of school-level compositional change contribute to this imprecision.

Figure 6. Comparing 2019-2022 California test score trends for Los Angeles Unified School District (LAUSD) and plaintiffs' schools with available data within the same district.

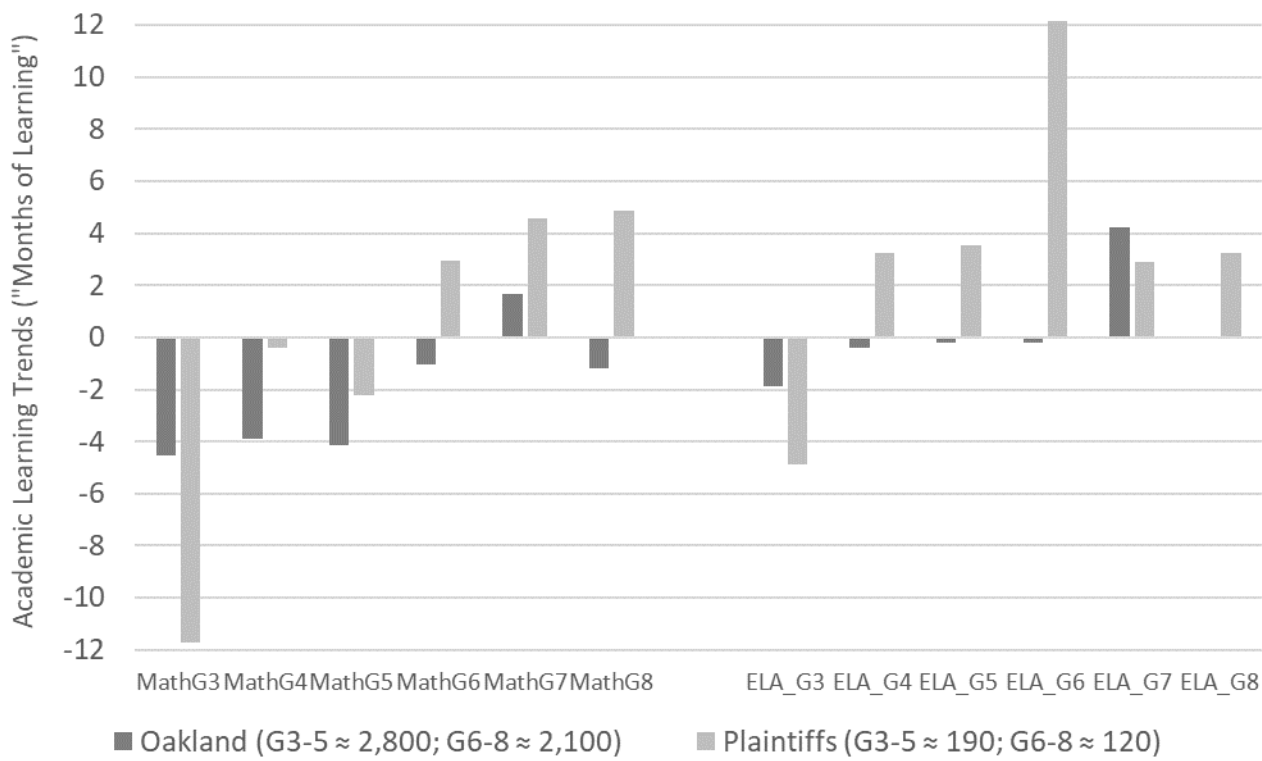


Note: Informal "months of learning" interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year. Plaintiffs' schools with available data include five elementary schools with a total of roughly 140 students per subject and grade (59th Street, Paseo del Rey, San Miguel, 61st Street, and Middleton Street Primary) and three middle schools with a total of roughly 1000 students per subject and grade each year (John Muir, Orville Wright, and Paul Revere Charter). Overall, Los Angeles has roughly 37,000 students per subject and grade in elementary schools and roughly 32,000 students per subject and grade in middle schools each year.

35. Similarly, Figure 7 shows academic learning trends for Oakland Unified School District and for plaintiffs' schools within the same district. Darker bars show Oakland trends, and lighter bars show the weighted average of plaintiffs' school results. Oakland is a substantially smaller school district than LAUSD, and plaintiffs' middle school populations are also much smaller than LAUSD middle schools. Oakland's elementary school mathematics trends show similar declines to those in LAUSD. (Note that the scale of Figure 6 is different from Figure 7 to accommodate Oakland's broader range of trends.) The academic performance of Oakland middle school students was similar in 2019 and 2022 in both ELA and mathematics. Plaintiffs' schools generally had higher scores in

2022 than they did in 2019. Again, smaller sample sizes and a greater likelihood of school-level compositional change can swing school-level results and motivate student-level longitudinal analyses that the state has not to my knowledge provided.

Figure 7. Comparing 2019-2022 California test score trends for Oakland Unified School District and plaintiffs' schools with available data within the same district. (Note that the vertical axis range differs from Figure 6.)

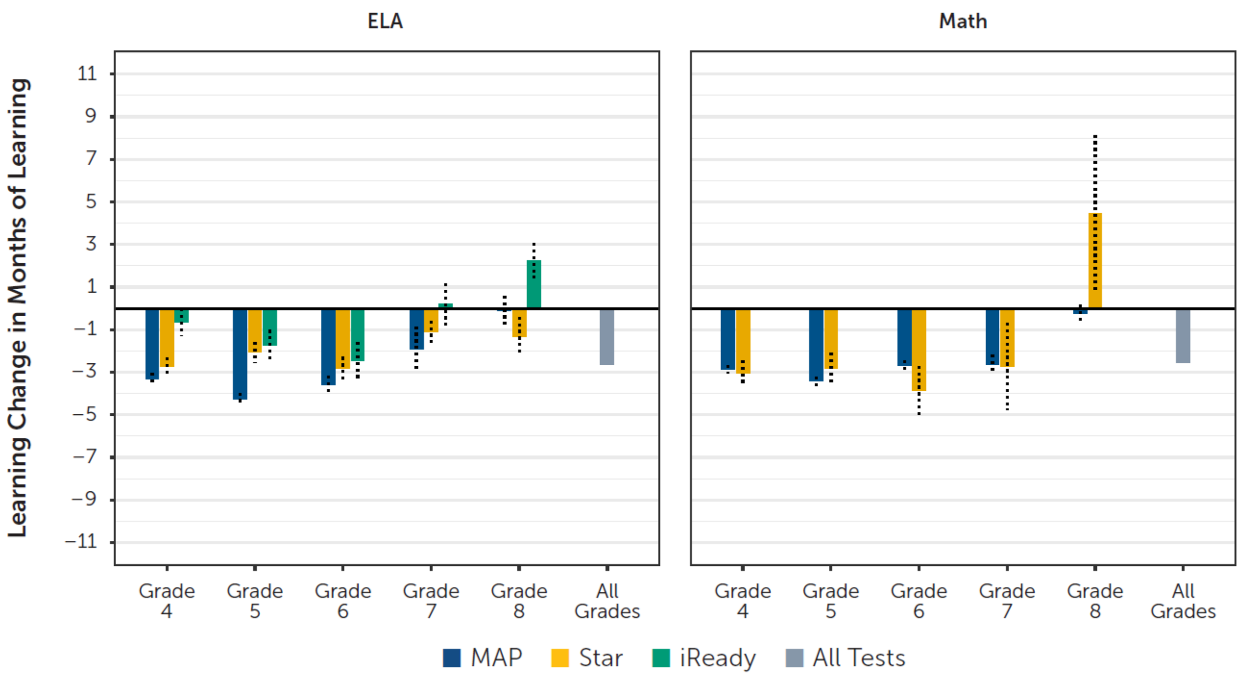


Note: Informal "months of learning" interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year. Plaintiffs' schools with available data four elementary schools with a total of roughly 190 students per subject and grade each year (Bridges, Emerson, Glass Valley, and La Escuelita) and two middle schools with a total of roughly 120 students per subject and grade each year (Westlake and the aforementioned La Escuelita). Overall, Oakland has roughly 2,800 students per subject and grade in elementary schools and roughly 2,100 students per subject and grade in middle schools each year.

36. An additional study that includes LAUSD and Oakland is Pier, Christian, Tymeson, and Meyer (2021). The authors used available locally determined assessment data from roughly 10,000 students per grade in 19 of the California "CORE" districts and agencies, a collaborative that shares educational data and a common reporting system. They use

longitudinal adjustment methods like those I recommend above to control for population shifts. Figure 8 shows their topline findings for this small population, that students in these districts and on these tests showed and learning loss of over two months on average across subjects and grades.

Figure 8. Aggregate learning change on locally determined assessments in 19 CORE districts from Fall 2019 to Winter 2021, from Pier, Christian, Tymeson, and Meyer (2021).



7) Student Report Cards are Inaccurate Indicators of Student Progress Over Time

37. Counsel have also provided me with the student report cards of nine plaintiffs and asked me whether these report cards support inferences about the adequacy of the educational progress of these nine students through the pandemic. They do not. The purpose of student report cards is not to monitor educational progress accurately but for teachers to provide students, parents, and other teachers with qualitative information about student academic strengths. These teachers' judgments are useful informally, for their intended classroom-level purposes. However, they are not standardized, which means that there is no evidentiary basis for their comparability across different teachers and over time.
38. When a measure is unstandardized, it means that a student's score may rise or fall not because of their actual educational progress but because one teacher has different

standards than another. A student's score may also rise or fall because teachers are referencing their performance to their relative position in a cohort. In a pandemic, a teacher may not notice if a single student is falling behind, because the entire cohort is also falling behind. Like swimmers treading water in a receding current, they may not appear to be moving unless there is an absolute standard to reference. The many factors that may cloud and bias teacher judgments of student performance and progress are the reasons why California has a state standardized testing program, to measure student and school progress fairly over time. Teacher-scored report cards have no evidentiary basis for measuring educational progress accurately over time.

8) California Results from the National Assessment of Educational Progress

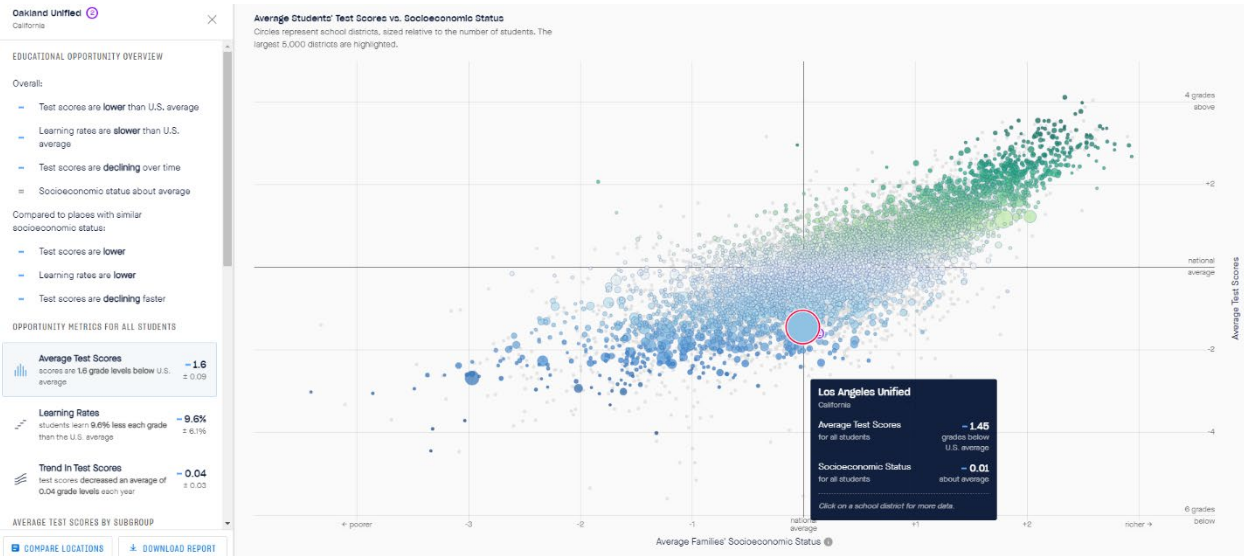
39. The fall of 2022 also brought test score results from the National Assessment of Educational Progress (NAEP) for the nation, all 50 states, and many large urban districts, including Los Angeles and San Diego. NAEP, also known as the Nation's Report Card, differs from state tests in many ways, including content, sampling, and stakes (Ho, 2007). For example, NAEP measures reading, whereas California's SBAC measures Reading, Writing, Listening, and Research/Inquiry. NAEP measures only a sample of students from selected schools in grades 4 and 8, whereas SBAC is a census test administered to almost all public school students. And NAEP is intended as a low-stakes test for high-level educational monitoring. No state, including California, adopts NAEP content standards nor NAEP performance standards.
40. California's recent NAEP scores prior to the pandemic have been very low, particularly when accounting for the relatively high socioeconomic status of California families. As our Stanford Education Data Archive shows (Reardon et al., 2021), California's NAEP scores were .7 grade levels below the national average over this period, whereas family socioeconomic status was .34 standard deviations above the national average. Figure 9a allows visualization of this in a scatterplot that shows all 50 states arranged by relative NAEP scores and family socioeconomic status. The arc of the plot from the bottom left to the upper right shows that a state's average family income is positively correlated with state NAEP scores. However, the location of California toward the bottom of this plot shows that it has lower NAEP scores than family income would suggest.

41. Similarly, Figure 9b shows that many California school districts have underperformed in state and NAEP scores from 2009 to 2018, including two of the districts related to this case, Los Angeles and Oakland Unified School Districts. In a plot with over 12,000 school districts in this country, these two districts are again toward the bottom of this plot, with scores 1.5 grade levels below the national average. This is much lower educational opportunity than expected, given that family socioeconomic status in these districts is near the national average.

Figure 9a. From 2009 to 2018, California had much lower educational opportunity than expected given family socioeconomic status.

Source: https://edopportunity.org/explorer/#/chart/none/states/avg/ses/all/5.6/37.42/-119.27/0622710,34.205,-118.5+0628050,37.768,-122.178+06,40.084,-121.915

Figure 9b. From 2009 to 2018, Los Angeles and Oakland Unified School Districts had much lower educational opportunity than expected given family socioeconomic status.
Source: https://edopportunity.org/explorer/#/chart/none/districts/avg/ses/all/12/37.77/-122.18/0622710,34.205,-118.5+0628050,37.768,-122.178



42. Against this backdrop, through the pandemic, the NAEP results for California showed overall declines in educational achievement from 2019 to 2022 in all subjects and grades except for Grade 8 Reading. Table 4 below summarizes the results from SBAC and NAEP for the same subjects, grades, and years. Side-by-side results are available for California as a whole, as well as for Los Angeles and San Diego. I focus on these two large districts because they are the only two for which NAEP reports official scores. Both NAEP and SBAC scores show declines for California overall. For example, in fourth grade mathematics, NAEP shows a decline of 4 "months of learning," and SBAC shows a decline of 5 months of learning. NAEP scores show less of a decline than SBAC, underestimating SBAC declines by roughly 1 "month of learning" in mathematics and roughly 2 "months of learning" in Reading/ELA.
43. Table 4 and Figure 5 show that these cross-test trend discrepancies are more extreme in Los Angeles. There, SBAC trends show declines in grade 8 mathematics and grade 4 reading, and NAEP shows slight gains. San Diego's results are more commensurate, with NAEP declines agreeing with SBAC declines in Reading and slightly exceeding SBAC declines in math. What should we make of NAEP results in Los Angeles, where one test is

showing declines and another is showing gains? Testing experts must often resolve these discrepancies, which we call a “two watches” problem: “A person with one watch knows what time it is; a person with two watches is never quite sure” (Brennan, 2001).

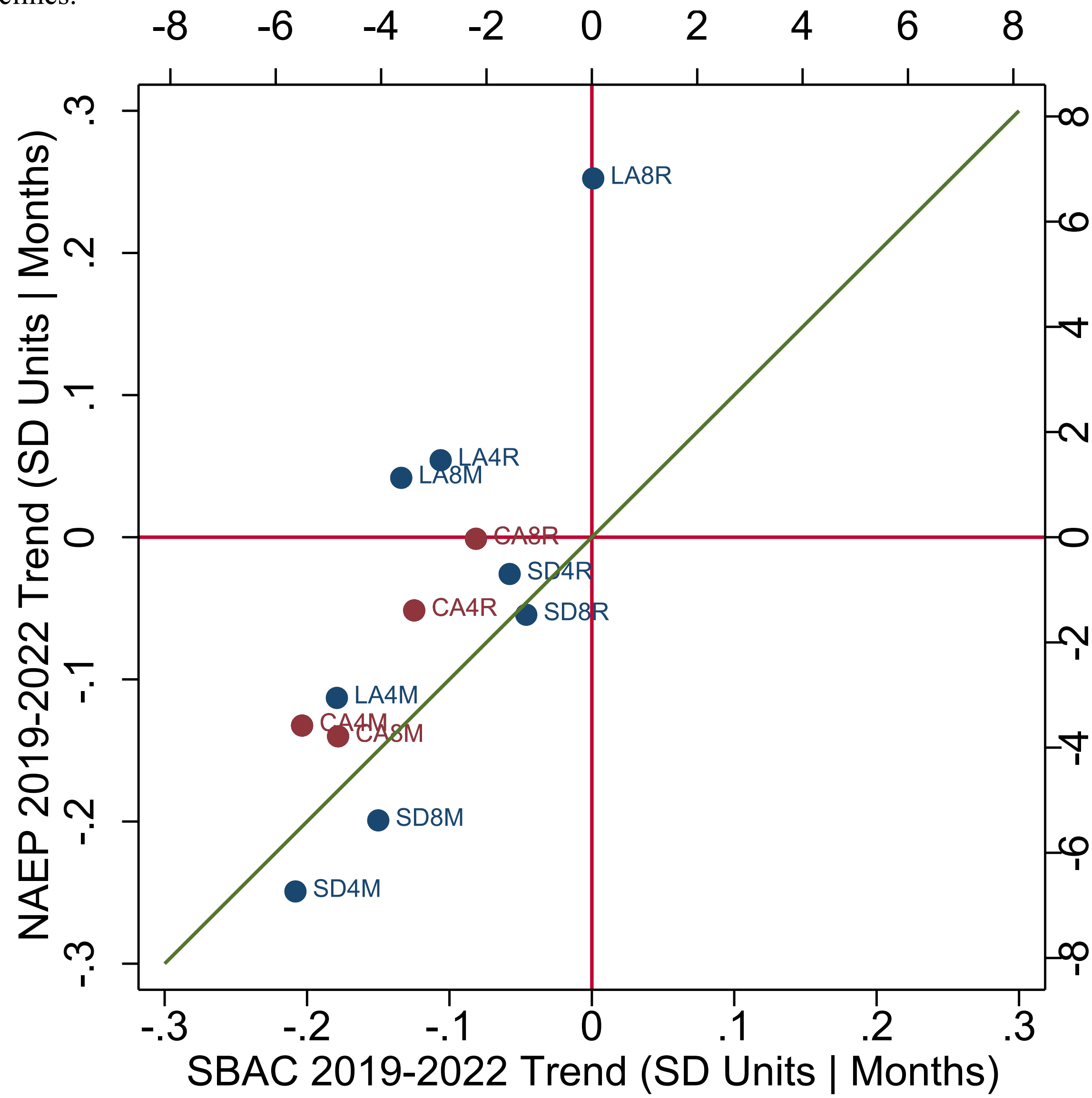
Table 4. Comparing California’s 2019-2022 test score trends for the National Assessment of Educational Progress and the state test, the Smarter Balanced Assessment Consortium. The table shows results in 2017 standard deviation units and a linear “months of learning” conversion.

			Standard Deviations		"Months of Learning"	
Jurisdiction	Subject	Grade	NAEP	SBAC	NAEP	SBAC
California	Math	4	-0.13	-0.20	-4	-5
California	Math	8	-0.14	-0.18	-4	-5
California	ELA/R	4	-0.05	-0.12	-1	-3
California	ELA/R	8	0.00	-0.08	0	-2
LAUSD	Math	4	-0.11	-0.18	-3	-5
LAUSD	Math	8	0.04	-0.13	1	-4
LAUSD	ELA/R	4	0.05	-0.11	1	-3
LAUSD	ELA/R	8	0.25	0.00	7	0
San Diego	Math	4	-0.25	-0.21	-7	-6
San Diego	Math	8	-0.20	-0.15	-5	-4
San Diego	ELA/R	4	-0.03	-0.06	-1	-2
San Diego	ELA/R	8	-0.05	-0.05	-1	-1

Note: Informal “months of learning” interpretations assume linear learning rates of 1/3 standard deviations over 9 months of schooling per year.

44. Table 4 and Figure 10 show that these cross-test trend discrepancies are more extreme in Los Angeles. There, SBAC trends show declines in grade 8 mathematics and grade 4 reading, and NAEP shows slight gains. San Diego’s results are more commensurate, with NAEP declines agreeing with SBAC declines in Reading and slightly exceeding SBAC declines in math. What should we make of NAEP results in Los Angeles, where one test is showing declines and another is showing gains? Testing experts must often resolve these discrepancies, which we call a “two watches” problem: “A person with one watch knows what time it is; a person with two watches is never quite sure” (Brennan, 2001).

Figure 10. Comparing California's 2019-2022 test score trends for the National Assessment of Educational Progress and the state test, the Smarter Balanced Assessment Consortium. Results are expressed in 2017 standard deviation units and a linear "months of learning" conversion. The diagonal line indicates agreement. Scores in the bottom left quadrant indicate agreement about declines.



Note: Informal "months of learning" interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year.

45. In my opinion, the most relevant trends for this case are those for California's adopted SBAC test, not NAEP. The SBAC samples more students, covers a broader range of content within English Language Arts, and holds the authority and relevance of a test that the California state board adopted with deliberation and intention. California's career and

college readiness standards that I defended in my initial report are those of the SBAC, not those of NAEP.

46. There are also technical reasons to be skeptical of taking the Los Angeles trends at face value. Daniel McGrath, Branch Chief of the Assessments Division at the National Center for Education Statistics, explained that a change in school sampling may explain part of the anomalous increase. In an article for the LA School Report, he explained that, in 2019, the Los Angeles sampling plan did not include any of its relatively high-performing public charter schools. Including five of these schools in the 2022 sample may have accounted for the disproportionately positive Los Angeles test score trends (Jacobsen, 2022).

9) Leveraging national data to draw stronger inferences about California results

47. California's available learning data through the early stages of the pandemic in 2020 and 2021 is particularly limited due to low test participation rates in 2021. Because 2021 scores could be confounded by population changes, I believe it is helpful to draw upon data from other sources, as many are finding the same broad patterns for which it is difficult for me to imagine California being an exception. For example, over 200,000 students in each grade took one of three alternative district-level assessments in 2021. Although California-specific results are not directly available, these three test vendors have reported on national results for their tests, and all of them show three general patterns similar to the test score results in earlier exhibits.
48. First, all show relative declines and slower growth in mathematics when compared to reading. For example, for NWEA, Lewis, Kuhfeld, Ruzek, and McEachin (2021) show percentile rank differences from 2019 to 2021 of -5 and -3 in grades 4 and 8 Reading and -11 and -8 in grades 4 and 8 Mathematics, respectively. For iReady (Curriculum Associates, 2021), compared to historical averages, the percentage of students on grade level declined 4 percentage points and 2 percentage points in grades 4 and 8 Reading compared to 16 percentage points and 5 percentage points in grades 4 and 8 Mathematics. For STAR (Renaissance Learning, 2021), primary-aged students scored .03 standard deviation units lower in Reading and .20 standard deviation units lower in Mathematics. Second, the results in the previous paragraph also indicate that, like earlier exhibits in California, declines are greater in magnitude in earlier grades than later grades. Third, all

three reports note increases in inequality in terms of areas and groups with less average economic advantage scoring not just lower on average but lower than expected given previous results.

49. I include two figures from a recent report by Goldhaber, Kane, McEachin, and Morton (2022) that illustrate the magnitude of academic learning loss using national NWEA data using an approach analogous to the "fair trend" I discussed earlier. When establishing appropriate expectations for populations, Figure 11 shows that learning losses are substantial and unequal, with declines in math scores in particular in early grades and for Black and Hispanic students. Figure 12 shows that losses are also largest in high-poverty schools. However, the smallest average losses measured are still substantial. Recall that a standard deviation unit is roughly 3 grade levels, so a standard deviation unit change of 0.2 is roughly 5 months of learning. Reading losses are roughly half of mathematics losses.

Figure 11. From Figure 17 in Goldhaber, Kane, McEachin, and Morton (2022), difference between median predicted mathematics growth for the pandemic cohort and pre-pandemic growth rates, by race and grade.

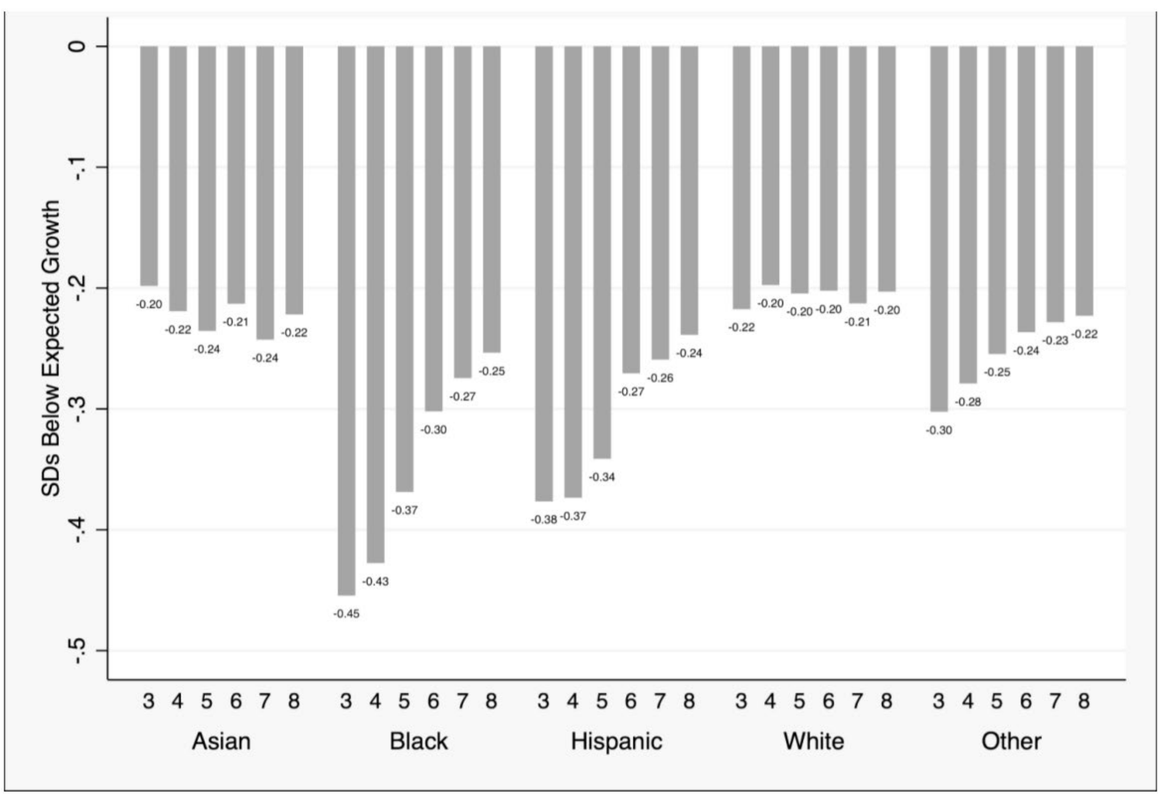
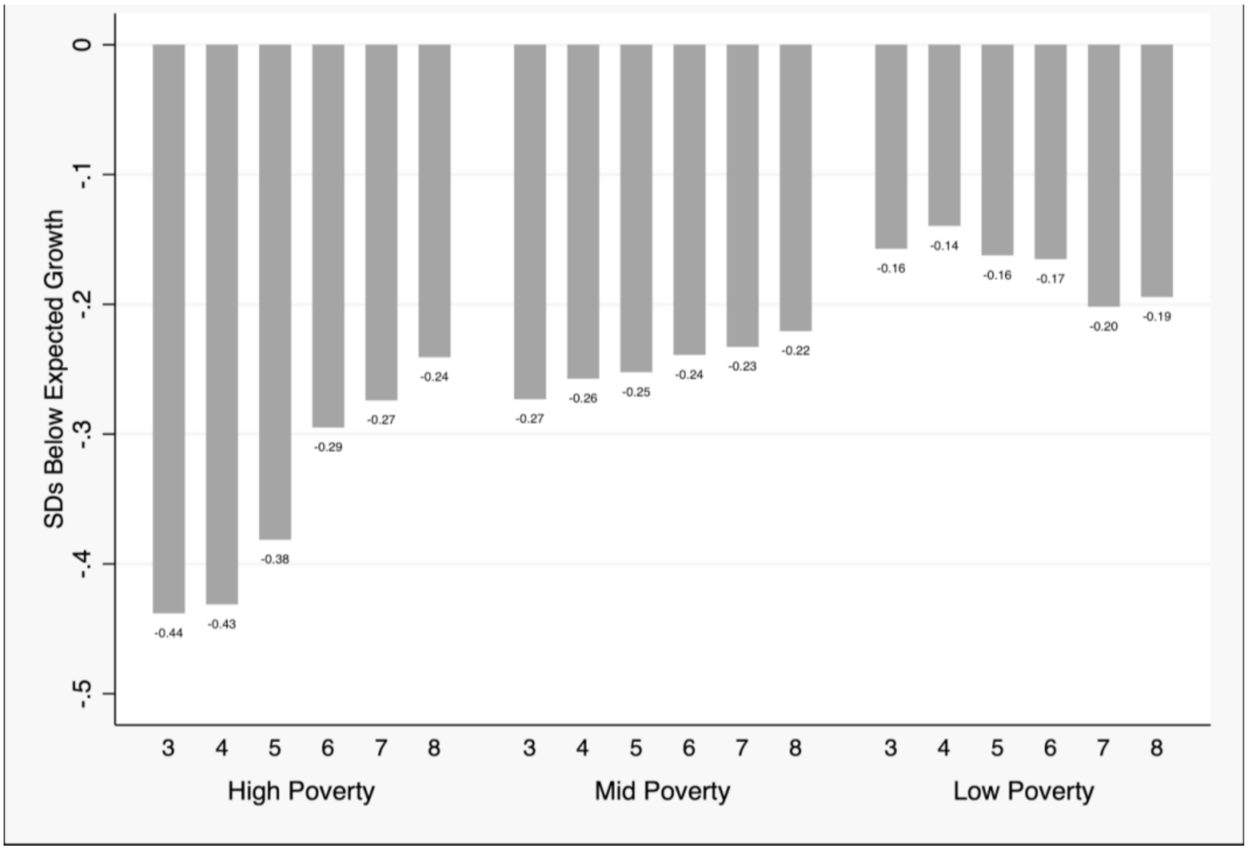


Figure 12. From Figure 19 in Goldhaber, Kane, McEachin, and Morton (2022), difference between median predicted mathematics growth for the pandemic cohort and pre-pandemic growth rates, by grade and school poverty status (above 75% free or reduced-price lunch, between 25% and 75%, and below 25%).



10) The relevance and authority of tested content and performance standards

50. As the pandemic has elevated concerns about physical and social-emotional health, it is important to remember why test scores are meaningful and what role they play. The Standards for Educational and Psychological Testing (hereafter, "the Standards") define validity as, "the degree to which evidence and theory support the interpretations of test scores" (AERA, APA, & NCME, 2014, p. 11). I have been asked to evaluate the assertion that percentages of Level 3 and 4 students on the California assessment tests indicate percentages of students on track for college readiness, and that percentages of Level 1 and Level 2 students indicate percentages not on track. The Standards list five common

sources of validity evidence. The first of these is "evidence based on test content." Do test questions assess skills relevant to college and careers?

51. To evaluate test content, I reviewed documents referenced in the SBAC Technical Report, the Common Core State Standards in English Language Arts (ELA) and Mathematics, and the SBAC Content Specifications in ELA and Math . I determined from these documents that the content is relevant to college and career readiness. Evidence includes a survey of 1815 postsecondary instructors who evaluated the applicability of the Common Core State Standards to their courses and alignment studies that assess whether items measure these standards.
52. As an illustration of the relevance of the tested content to postsecondary outcomes, I include in Exhibit 2 the 4 domains that the SBAC test specifications define for each subject, Mathematics and ELA (Smarter Balanced Assessment Consortium, 2019). These claims show that performance on the SBAC does not only require memorization and understanding, but application, analysis, and evaluation.

Exhibit 2: Mathematics and ELA Claims from SBAC Test Specifications (SBAC, 2019).

• Mathematics Claim #1: Concepts and Procedures "Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency."
• Mathematics Claim #2: Problem Solving "Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies."
• Mathematics Claim #3: Communicating Reasoning "Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others."
• Mathematics Claim #4: Modeling and Data Analysis "Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems."

• ELA/Literacy Claim #1: Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.

• ELA/Literacy Claim #2: Students can produce effective and well-grounded writing for a range of purposes and audiences.
• ELA/Literacy Claim #3: Students can employ effective listening skills for a range of purposes and audiences.
• ELA/Literacy Claim #4: Students can engage in research/inquiry to investigate topics, and to analyze, integrate, and present information.

53. Chapter 1 in the SBAC Technical Report reviews dozens of studies and reports that represent accumulated evidence, both conceptual and empirical, for the relevance of the Common Core State Standards to college and career readiness, and for SBAC as an assessment that produces scores informative about student readiness. These studies include, for example, the judgment of postsecondary instructors that these standards are relevant to postsecondary outcomes, the findings of trained interviewers that examinees answering items correctly were doing so with cognitively appropriate strategies, and the judgment of subject matter experts that test items are aligned to the content standards (Doorey & Polikoff, 2016).
54. I also reviewed test items, including California practice tests for the SBAC in grades 3-8, in ELA and Mathematics, available at the CDE website. As noted above, others who have analyzed the alignment of SBAC to the Common Core State Standards have concluded that these items have good alignment to standards. The depth of knowledge that I observed test items to require in my review of items leads me to the same conclusion. The SBAC Technical Report also reviews the test design process, demonstrating how test questions were developed in a manner aligned with content and complexity appropriate for each grade and across grades. I conclude from this evidence that tested content is relevant to postsecondary readiness.

a) The achievement levels were set according to widely accepted professional assessment standards

55. In educational assessment, we distinguish between content standards and performance standards. Content standards describe what is measured. Performance standards describe

levels of achievement. The process of choosing "cut scores" that demarcate achievement levels is called achievement level setting or, briefly, standard setting.

56. I explained in the previous section why the California assessment system measures content relevant for determining whether students demonstrate adequate understanding of and ability to apply the knowledge and skills needed for postsecondary success. Standard setting begins with achievement level descriptors and uses evidence-based judgments from teachers and subject-matter experts to select cut scores demarcating achievement levels. The SBAC developed achievement level descriptors and set cut scores with teachers and subject-matter experts, including 66 in-person participants from California, 29 for math and 37 for ELA (Smarter Balanced Assessment Consortium, 2015).

57. I conclude on the basis of SBAC achievement level setting documentation that the SBAC set achievement levels in an authoritative, transparent, and appropriate manner consistent with the Standards for Educational and Psychological Testing. I agree with the collected statements of support from my colleagues in the field of educational measurement, including those from an independent external auditor and two separate 10-member panels. The Achievement Level Setting Advisory Panel stated in their letter of support, that "the design and procedures for the Achievement Level Setting and the Vertical Articulation were implemented as planned, represent a valid process that is consistent with best practices in standard setting, and support the defensibility of the content-based performance standards" (Smarter Balanced Assessment Consortium, 2014, p. 2). I agree with this assessment. The clarity of the achievement level descriptors, the representativeness of the judges, and the documentation of consensus all indicate a legitimate process that meets professional standards.

58. The technical documentation describes a well-established standard setting process. There is a clear description of a borderline Level 3 student that teachers and subject-matter experts crafted. Panelists used evidence from examinee testing responses to anchor this borderline description to the test score scale in each subject and grade. Logically, this anchored scale score location is the cut score that represents the lower border of the Level 3 achievement level. I describe this process in more detail here to emphasize its logic and legitimacy.

59. Rather than defining adequacy relative to some grade level average, the standard setting process begins with an explicit "policy definition" that is broad in scope and applicable across subjects and grades. For example, a student who meets the SBAC Level 3 definition, "demonstrates adequate understanding of ELA and mathematics and the ability to apply the knowledge and skills for his or her grade level that are associated with college and career readiness." Educators and subject matter experts then ground this high-level policy definition in the content learned in each subject and grade. The resulting achievement level descriptors are much more specific and connected to the claims in Exhibit 1 above. Take the Exhibit 1 claim under "Concepts and Procedures" as an example. In Grade 6 mathematics, this claim has one of many nested targets, "Statistics and Probability." For this, there is a specific achievement level description that teachers and subject-matter experts crafted for Level 3 students in the 6th grade. They "should be able to pose statistical questions and understand that the responses to a statistical question have a distribution described by its center, spread, and overall shape…" (Smarter Balanced Assessment Consortium, 2015, p. 388). There are numerous statements like this for each target and claim.

60. Next, because these statements describe a range, teachers and experts craft an additional description of a student who would be exactly at the lower borderline, or threshold, of Level 3. Teachers and experts consider this borderline description as they review test questions in a booklet. The items in the booklet are ordered by their known difficulty levels, based upon student responses in field trials. The question posed to these panelists was, "Would a student at the threshold have at least a 50% chance of earning this point" (Smarter Balanced Assessment Consortium, 2015, p. 60)? Each panelist proceeds through the booklet, from easier items to harder items, until they reach an item that they believe the threshold student would no longer answer correctly. Logically, this question defines the Level 3 cut score. Because the threshold student description is linked to the boundary of the achievement level, and the question is linked empirically to the score scale, this process establishes empirically a specific Level 3 cut score for each panelist.

61. Naturally, panelist judgments differ. The SBAC technical documentation reports distributions of cut scores from panelists as they converge through three rounds of the consensus-building process. Panelists can be expected to disagree slightly about their final

recommended cut scores. When the median was proposed as the consensus, the vast majority (405 out of 438, 92%) of panelists agreed with the statement, "I am confident about the defensibility and appropriateness of the final recommended cut scores" (Smarter Balanced Assessment Consortium, 2015). These results are consistent with an authoritative, evidence-based process that resulted in legitimate cut scores.

62. After my review of well-specified achievement level descriptors, anchored via evidence from student responses to a score scale, in a judgmental process with strong expert consensus and confidence, I conclude that SBAC achievement levels meet professional standards. California's active participation in the SBAC achievement level setting process and the Board's adoption of SBAC achievement levels reinforces the legitimacy of these standards. California's reported percentages of Level 3 and higher students who demonstrate, "adequate understanding of ELA and mathematics and the ability to apply the knowledge and skills for his or her grade level that are associated with college and career readiness," are authoritative.

b) The achievement levels were set at the appropriate level.

63. Standards are sometimes criticized as being inaccurate on the basis that they are set "aspirationally," that SBAC cut scores represent aspirational performance levels designed to push for greater student achievement and, as a result, the percentages of students scoring at Level 1 and Level 2 do not show that those percentages of students are inadequately educated. That is a non sequitur. California teachers and subject-matter experts participated in the SBAC standard setting process and came to consensus about what constituted adequacy, from their experiences with the California students they taught. This consensus definition of adequacy has legitimacy and meaning regardless of any design for greater achievement.

64. One way in which the achievement levels might be "too high" would be if the policy definition for Level 3 were too stringent. One could recognize, as I explained in the previous section, that a Level 3 student demonstrates "adequate understanding," but argue that adequacy is not the standard we hope for students, and that instead some lower level of understanding, perhaps Level 2 (partial) or Level 1 (minimal), is sufficient. As an illustrative contrast, consider an achievement level descriptor that is considered by some

to be “too high,” that of the National Assessment of Educational Progress (NAEP), “Students performing at or above the NAEP Proficient level on NAEP assessments demonstrate solid academic performance and competency over challenging subject matter.” A reasonable person could argue that “competency over challenging subject matter” is a standard that is too high. In other words, even if the description of a student scoring NAEP Proficient were accurate, someone could argue that NAEP Proficient is an unnecessarily high standard.

65. I reject that argument in the case of California, because California’s definitions for students in grades 3-8 state that Level 3 students are those who have “adequate understanding of and the ability to apply … knowledge and skills….” The achievement levels for Level 3 were intended to describe adequate understanding and, as described above, they were set at appropriate levels for that purpose. Inadequate understanding is inadequate.

66. NAEP mapping studies that show the relative stringency of state performance standards reveal that SBAC Level 3 standards are lower than the NAEP Proficient standards. The less rigorous SBAC Level 3 cut score is consistent with the difference between “NAEP Proficient” as defined by “competency over challenging subject matter” and “SBAC Level 3” as defined by “adequate understanding and ability.” If this SBAC Level 3 description were intended to be an aspirational long-term goal, I cannot explain why it was described as “adequate,” nor why a higher aspiration was not set, closer to that of NAEP Proficient.

67. The standards I have reviewed are California state standards. The state participated directly in the achievement level setting process. The state legitimized these standards and descriptions as those of California, not some external entity.

c) The test scores show no evidence of bias against selected groups.

68. A common concern with tests is that they inflate or misrepresent differences among groups rather than measure differences accurately. This could happen if tests measured irrelevant content with which higher-scoring groups had disproportionate skill or experience. If this were the case, the relatively low percentages Level 3 and above students who are, for example, Black, Hispanic, English Learners, Low Income, or students with disabilities would be even lower than they would be had a more relevant test

been given. Colloquially, this is often called “test bias.” It is best detected when a more relevant and representative test could be given as an audit and reveals different patterns of group or item performance.

69. I find no evidence for bias in this case, and I conclude that percentages of students in each achievement level are comparable across reporting categories. The strongest argument against this bias is the relevance of the content itself, as I presented in the first section. I find it difficult to argue that a more relevant and representative test exists for the purpose of measuring these subjects. Beyond content, the SBAC technical report shows that developers employ many standard procedures for ensuring fairness and accessibility, including measuring only the intended construct, establishing bias and sensitivity review committees, and ensuring that items and administration conditions do not offend or distract students.
70. Secondary evidence exists in the form of the National Assessment of Educational Progress, which reports results from California in Reading and Mathematics, in grades 4 and 8, in odd years through 2019. The NAEP results show that California’s achievement gaps between traditionally advantaged and disadvantaged groups are commensurate with those reported by the SBAC.
71. There is also a standard method known as Differential Item Functioning (DIF) that was applied in SBAC development to assess whether individual items appeared to favor certain groups over others. Items that exhibited DIF were flagged for a content review panel to evaluate for possible exclusion. The SBAC technical manual presents item DIF statistics, and there is nothing from this presentation that raises any concern about interpreting Level 3 percentages across groups as intended.
72. I also find solid evidence about the fairness of score reporting for students with disabilities and English Learners in the SBAC Technical Manual. The goal in these cases is to allow students to provide evidence of their proficiency while removing “construct-irrelevant barriers” (SBAC, 2019, p. 66). A list of appropriate accommodations was compiled from the research literature and are deployed depending on the construct or subject being measured. For example, for an English Learner, translations are deployed for mathematics, where inexperience with the English language is a construct-irrelevant barrier to measuring mathematics proficiency. However, translations are not deployed when

assessing English Language Arts, because English fluency is the target of measurement. This supports the validity of interpretations of mathematics and English Language Arts scores for English Learners. The technical manual also compares indicators of measurement quality, like standard errors of measurement, across groups and finds them comparable.

Conclusion

73. I conclude that, in California, the weight of evidence, including 2019, 2021, and 2022 test score results, strong validity evidence for the California state test, and supporting national evidence from the locally determined tests administered in California, confirms that large and increasing numbers of California students, including disproportionately high numbers of Black and Hispanic students, had reduced opportunities to learn academic skills during the pandemic. I also conclude from deposed testimony and the paucity of released statistical and psychometric analyses that the state neglected to commission or conduct longitudinal research that could document academic learning loss and elevate attention to addressable inequalities in academic opportunities. The state should recognize this as the educational emergency that it is and rise to meet this challenge armed with all appropriate data and support: Substantial and growing proportions of California students are not on track.

ANDREW DEAN HO

Appendix A. Trends and gap trends (Tables 1 and 2) for Economically Disadvantaged Students.

Table A1. Trends in California test scores from prepandemic 2019 to spring 2022 in scale score points (left) and a rough "months of learning" conversion (right) for economically disadvantaged students.

		2019-2022 Trend (SS)			Standard	2019-2022 Trend (months)		
Subject	Grade	Black	Hispanic	White	Deviation	Black	Hispanic	White
ELA	Grade 3	-19	-25	-14	92	-6	-7	-4
	Grade 4	-14	-17	-8	97	-4	-5	-2
	Grade 5	-11	-14	-9	101	-3	-4	-3
	Grade 6	-10	-11	-9	98	-3	-3	-2
	Grade 7	-6	-6	-5	103	-2	-2	-1
	Grade 8	-8	-8	-7	102	-2	-2	-2
Math	Grade 3	-24	-25	-13	83	-8	-8	-4
	Grade 4	-24	-25	-13	86	-8	-8	-4
	Grade 5	-22	-24	-15	95	-6	-7	-4
	Grade 6	-16	-19	-13	110	-4	-5	-3
	Grade 7	-18	-18	-15	115	-4	-4	-3
	Grade 8	-18	-21	-19	124	-4	-5	-4

Note: Negative numbers indicate decreases in average scores from 2019 to 2022 in scale scores (SS) or months of learning. ELA = English Language Arts. Hispanic students are Hispanic or Latino students. Standard deviations from 2017 are shown as references to estimate effect sizes. Informal "months of learning" interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year.

Table A2. Gap trends in California test scores from prepandemic spring 2019 to spring 2022 in scale score points (left) and months of learning (MoL, right) for economically disadvantaged students.

		Race/Ethnicity Gap Trend (SS)		Race/Ethnicity Gap Trend (MoL)	
Subject	Grade	White-Black	White-Hispanic	White-Black	White-Hispanic
ELA	Grade 3	6	11	2	3
	Grade 4	5	9	1	3
	Grade 5	2	5	1	1
	Grade 6	1	2	0	1
	Grade 7	1	1	0	0
	Grade 8	1	1	0	0
Math	Grade 3	11	12	4	4
	Grade 4	12	12	4	4
	Grade 5	7	9	2	3
	Grade 6	4	6	1	1
	Grade 7	3	4	1	1
	Grade 8	-1	2	0	1

Note: Positive numbers indicate increases in average differences from 2019 to 2022 in scale scores or months of learning (MoL). ELA = English Language Arts. Hispanic students are Hispanic or Latino students. Informal “months of learning” interpretations assume linear learning rates of 1/3 of a standard deviation over 9 months of schooling per year.

Errata:

1) In a previous version of this report, Tables 1 and 2 for English Language Arts (ELA) trends and gap trends for Black, White, and Hispanic students did not specify that results were for economically disadvantaged students. See above for corrected and supplemental tables.

2) In a previous version of this report, a reference to Vella (2023) should have been a reference to the Motion for Summary Judgment (2023). The full reference list is included below.

3) In a previous version of this report, tables did not reference the source for standard deviation units. The source was the publicly available Technical Report for the state test (California Department of Education, 2018): https://www.cde.ca.gov/ta/tg/ca/documents/sbac17techrpt.pdf The full reference list is included below.

References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Retrieved from https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Betebenner, D. W., & Van Iwaarden, A. (2022). *COVID-19 Academic Impact in Rhode Island*. National Center for the Improvement of Educational Assessment. Retrieved from https://www.ride.ri.gov/Portals/0/Uploads/Documents/News/Rhode_Island_Academic_Impact_Report_FINAL.pdf

Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement, 38,* 295-317. Retrieved from https://www.jstor.org/stable/1435452

California Department of Education (2018). California Assessment of Student Performance and Progress Smarter Balanced Technical Report, 2016–17 Administration. Retrieved from https://www.cde.ca.gov/ta/tg/ca/documents/sbac17techrpt.pdf

California Department of Education. (2021, June 2). *California Assessment of Student Performance and Progress. Smarter Balanced 2019-2020 Technical Report.* ETS Contract No. CN150012. Retrieved from https://www.cde.ca.gov/ta/tg/ca/documents/sbactechrpt1920.pdf

California Department of Education. (2022a). *Test Results for California's Assessments.* Retrieved from https://caaspp-elpac.cde.ca.gov/caaspp/

California Department of Education. (2022b). *California Assessment of Student Performance and Progress. Smarter Balanced 2020-2021 Technical Report.* ETS Contract No. CN150012.

California Department of Education. (2022c). Interpretation Guide to the 2020-2021 Statewide Assessment Results. Retrieved from https://www.cde.ca.gov/ta/tg/ca/documents/assessmentresultsguide21.docx

Center on Reinventing Public Education. (2022). CRPE 2021-22 State Response Database. Retrieved from https://docs.google.com/spreadsheets/d/1AFyMob7ATHlrDDvWXsGVNUSrj_vpJkFdqU49zhb4TJU/

Cohodes, S., Goldhaber, D., Hill, P., Ho, A. D., Kogan, V., Polikoff, M., Sampson, C., & West, M. (2022). Student achievement gaps and the pandemic: A new review of evidence from 2021-2022. Center on Reinventing Public Education. Retrieved from https://files.eric.ed.gov/fulltext/ED622905.pdf

Comprehensive Center Network. (2022). *Spring 2021 Performance on Three Locally Determined Assessments in California.* Region 15 Comprehensive Center at WestEd.

Retrieved from https://csaa.wested.org/wp-content/uploads/2022/01/CA_Spr2021_LocalAssessment_FINAL-ADA.pdf

Curriculum Associates. (2021). Academic Achievement at the End of the 2020-2021 School Year. Retrieved from https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-understanding-student-needs-paper-spring-results-2021.pdf

Doorey, N., & Polikoff, M. (2016). Evaluating the content and quality of next-generation assessments. Thomas B. Fordham Institute. Retrieved from https://fordhaminstitute.org/sites/default/files/%2802.09%20-%20Final%20Published%29%20Evaluating%20the%20Content%20and%20Quality%20of%20Next%20Generation%20Assessments.pdf

Goldhaber, D., Kane, T. J., McEachin, A., & Morton, E. (2022). *A Comprehensive Picture of Achievement Across the COVID-19 Pandemic Years: Examining Variation in Test Levels and Growth Across Districts, Schools, Grades, and Students*. National Center for Analysis of Longitudinal Data in Education Research. Retrieved from https://caldercenter.org/sites/default/files/CALDER%20Working%20Paper%20266-0522_0.pdf

Ho, A. D. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice, 26*(4), 11-20. Retrieved from https://doi.org/10.1111/j.1745-3992.2007.00104.x

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, *37*, 351-360. Retrieved from https://www-leland.stanford.edu/~hakuta/Courses/Ed205X%20Website/Resources/Ho%20The%20Problem%20with%20Proficiency%20ER%20v37%20n6.pdf

Ho, A. D. (2021). *Three Test-Score Metrics that All States Should Report.* Technical Issues in Large-Scale Assessment, State Collaborative on Assessment and Student Standards. Council of Chief State School Officers. Retrieved from https://scholar.harvard.edu/files/andrewho/files/threemetrics.pdf

Holland, P. W. (2002). Two Measures of Change in the Gaps Between the CDFs of Test-Score Distributions. *Journal of Educational and Behavioral Statistics, 27*, 3–17.

Jacobson, L. (2022, November 7). L.A. vs. the Wonks: District's 8th-Grade Reading Miracle on NAEP Draws Scrutiny. *The74*. Retrieved from https://www.the74million.org/article/l-a-vs-the-wonks-districts-8th-grade-reading-miracle-on-naep-draws-scrutiny/

Lewis, K., Kuhfeld, M., Ruzek, E., & McEachin, A. (2021). Learning during COVID-19: Reading and math achievement in the 2020-2021 school year. Retrieved from https://www.nwea.org/content/uploads/2021/07/Learning-during-COVID-19-Reading-and-math-achievement-in-the-2020-2021-school-year.research-brief-1.pdf

Linn, R. L. (2008). Educational accountability systems. In K. E. Ryan & L. A. Shepard. (Eds.) *The Future of Test-Based Accountability, 3-24.* New York, NY: Routledge.

Modan, N. (2021, October 29). 2021 testing participation varies widely – what will data mean for districts? *K-12 Dive*. Retrieved from https://www.k12dive.com/news/2021-testing-participation-varies-widely-what-will-data-mean-for-district/608939/

Motion for Summary Judgment. (2023, August 4). Defendants' Memorandum of Points and Authorities in Support of Motion for Summary Judgment or, in the Alternative, Summary Adjudication. Reservation No. 284458662516. Case No. RG20084386.

Pier, L., Christian, M., Tymeson, H., & Meyer, R. H. (2021). COVID-19 Impacts on Student Learning: Evidence from Interim Assessments in California. Policy Analysis for California Education. Retrieved from https://edpolicyinca.org/sites/default/files/2021-06/r_pier_jun2021.pdf

Reardon, S. F., Kalogrides, D., & Ho A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics, 46*(2), 138–167. Retrieved from https://doi.org/10.3102%2F1076998619874089

Reardon, S.F., Papay, J.P., Kilbride, T., Strunk, K.O., Cowen, J., An, L., & Donohue, K. (2019). Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive. (CEPA Working Paper No.19-08). Retrieved from http://cepa.stanford.edu/wp19-08

Reardon, S. F., Fahle, E. M., Ho, A. D., Shear, B. R., Kalogrides, D., Saliba, J. & Kane, T. J. (2023). Stanford Education Data Archive (Version SEDA 2022 2.0). Retrieved from http://purl.stanford.edu/db586ns4974.

Renaissance Learning. (2021). Understanding Progress in the 2020/21 Academic Year. Retrieved fromhttps://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1062293/Understanding_Progress_in_the_2020_21_Academic_Year_Initial_findings_from_the_spring_term.pdf

Smarter Balanced Assessment Consortium. (2014). Statements of Support. Retrieved from https://web.archive.org/web/20170818012308/https://portal.smarterbalanced.org/library/en/statements-of-support-achievement-level-setting-for-the-smarter-balanced-assessments.pdf

Smarter Balanced Assessment Consortium. (2015). Achievement Level Setting Final Report. Retrieved from https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf

Smarter Balanced Assessment Consortium. (2019). 2018-19 Summative Technical Report. Retrieved from https://technicalreports.smarterbalanced.org/2018-19_summative-report/_book/report-interp.html

Utah State Board of Education. (2021). Exploring the effects of the COVID-19 pandemic on student achievement in Utah. Retrieved from https://www.schools.utah.gov/file/b3c018ec-ccbb-4565-8199-248646c79524