

A ship in harbor is safe, but that is not what ships are built for.

- John A. Shedd, 1928



2018 Accountability Update



2018 CAASPP Test Results,
Report to the Governing Board
November 13, 2018



Glen Webb, Director of Curriculum, Instruction and Assessment

The Shift to a Growth Model

Meeting all students where they are and supporting their growth...It's a statewide, multi-year journey.



California's progress towards a new accountability system

2016-17

 First year Smarter Balanced assessment scores used to calculate status and pre-populated into the evaluation rubrics by January 2017

 Second year Smarter Balanced assessment scores available (Fall 2016)

 First year of California Alternate Assessment scores available (Fall 2016)

 SBE approves a set of criteria to facilitate the selection of a growth model methodology

 CDE staff with assistance from technical experts will explore growth model methodologies that meet the SBE determined criteria

2017-18

 Third year Smarter Balanced assessment scores available (Fall 2017)

 Second year of California Alternate Assessment scores available (Fall 2017)

 SBE approves a growth model methodology

 CDE with assistance from technical experts will perform growth model calculations for the purpose of reporting to the SBE and finalizing business rules and requirements

It's a
Journey



2018-19



CDE with assistance from technical experts will perform growth model calculations for the purpose of public reporting

2019-2020



Growth model available for accountability and continuous improvement



OUR OBJECTIVES

- Gain perspective of the current status of the SBAC test and how to use the data in the most credible and reliable manner.
- Understand our areas of strengths and challenges as indicated by our CAASPP scores and validated with multiple measures.
- Understand how our current actions correlate to challenges as revealed by data.

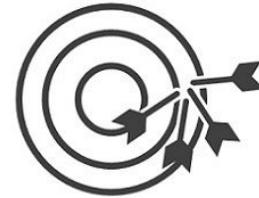
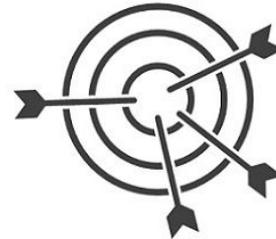
How is the Test Functioning?

STATUS



BIAS

VOLATILITY



Accuracy Vs Precision

Debated issues with the testing system and current methodologies for state's data presentation: The case for looking at **comparative status and cohort growth** data instead of **scaled status and annual change** (full slides can be found at end of presentation)

A ship in harbor is safe, but that is not what ships are built for.
- John A. Sheard, 1928

2018 Accountability Update

2018 CAASPP Test Results,
Report to the Governing Board
November 13, 2018

Glen Wilcox, Director of Curriculum, Instruction and Assessment

The Shift to a Growth Model

Meeting all students where they are and supporting their growth...It's a statewide, multi-year journey.

California's progress towards a new accountability system

2016-17: First year Statewide Balanced Assessment scores used to calculate annual and year-over-year growth model calculations for the 2017-18 school year.

2017-18: Third year Statewide Balanced Assessment scores available (Fall 2017). Second year of California Alternate Assessment scores available (Fall 2017). SBC approves a growth model methodology. CDE with assistance from technical experts will perform growth model calculations for the 2017-18 school year.

2018-19: CDE with assistance from technical experts will perform growth model calculations for the 2018-19 school year. SBC approves a set of criteria to facilitate the selection of a growth model methodology. CDE with assistance from technical experts will evaluate growth model methodologies that meet the SBC approved criteria.

2019-2020: Growth model available for accountability and continuous improvement.

It's a Journey

OUR OBJECTIVES

- Gain perspective of the current status of the SBAC test and how to use the data in the most credible and reliable manner.
- Understand our areas of strengths and challenges as indicated by our CAASPP scores and validated with multiple measures.
- Understand how our current actions correlate to challenges as revealed by data.

How is the Test Functioning?

STATUS **VOLATILITY** **BIAS** **Change** **GROWTH** **CERTAINTY** **Accuracy Vs Precision**

Volatility

How is the test looking for year to year Volatility?

Gain scores between the SBAC and PARCC systems appear more consistent than they were last year. However, gain scores as reported on the dashboard will show last year's very volatile results and will take time to stabilize. Our analysis will use comparative growth as a control for volatility.

Change

Reporting on the Dashboard... Status and Change

Five by Five reference charts are used to assign an indicator color that will be displayed on the state dashboard. This example for English Language Arts categorizes the change in scores from the previous year along the top axis and status measured as distance from level 3 along the vertical axis.

The scores being rated around the methodology is that 'change' is not 'growth' as it uses different graduating years of students for year to year comparisons rather than the same cohort to determine growth. Our analysis today will be based on cohorts to help control measurement volatility.

Scales

Reporting Status on the CAASPP site... Performance Levels... Can the test tell us if we should be...

Happy because we measure up? Sad because we don't? Carefree either way? We need to know more about SCALES.

Scaled Status

Reporting on the Dashboard... Status and Change

Status is determined by reference to a non-student referenced scale based on the number of points from level 3. The issue being raised around the methodology is that it lacks comparative value and is susceptible to measurement volatility as the test lacks development and conditions are refined. Our analysis will be based on comparison to the state as the larger and most stable data set and comparison to another district to provide some control over demographic variables. We will look at trends rather than status for relative cohort growth over time as a more reliable indicator.

Comparative Status

Math CAASPP

Demographic Group	Reference Group Mean (Subject Score)	CAASPP Full Scale Score (Student Footcandle)
African-American	68%	-47.5
White	75%	10.7
Hispanic	30%	-6.5
Asian-American	67%	15.3
English Learners	34%	-77.2
English Only	65%	-11.6
Reclassified ELLs	78%	-43.1
Socioeconomically disadvantaged	34%	-76.1
Learning disabled	55%	-32.5

Distance from "meeting standard" suggests that most groups are doing poorly compared to that scale as implied by negative scores. Data in information not readily available without downloading full data reports. Use and further analysis, the percentile rankings of the groups compared to their similar student groups statewide tends significantly different. For example, MHSLSU lowest performing group is actually performing in the top half of districts statewide. Our analysis will look at comparative results to identify areas of strength and challenge.

Imprecision

How are Performance Levels working with current test Precision

Director Level CLE Score: +/- 4 to 5px
School Level CLE Score: +/- 4 to 10 px
Student Level CLE Score: +/- 2 to 10 px

Director Level Math Score: +/- 3 to 5px
School Level Math Score: +/- 6 to 10 px
Student Level Math Score: +/- 2 to 5 px

Precision is not reported publicly but the standard error of measurement often overstates performance levels and dashboard cells. Our control for imprecision is to use scaled scores rather than performance levels and to validate with multiple measures.

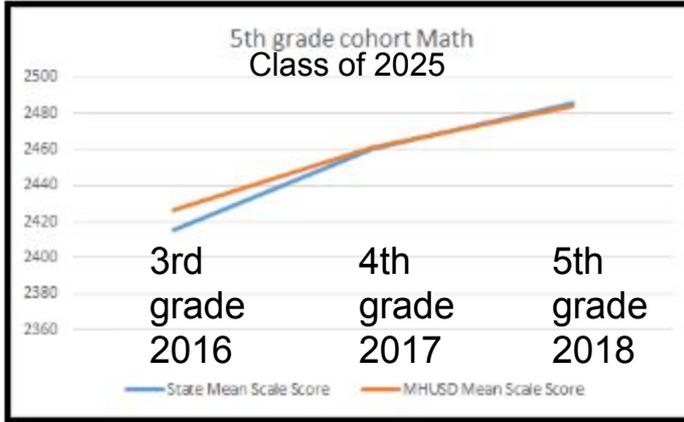
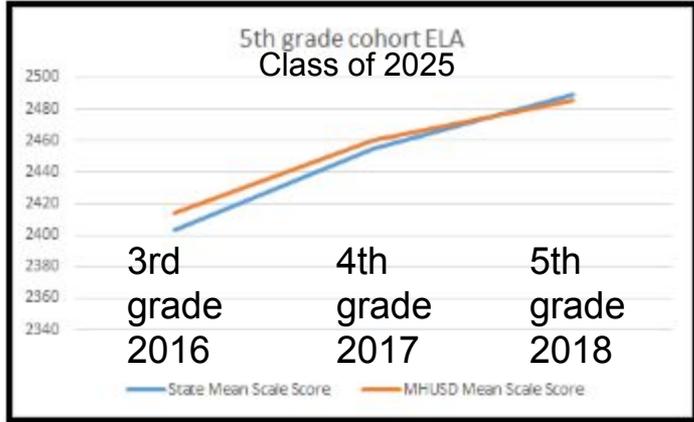
Comparative Change

Last year we saw that our grade level change scores had peaks and valleys that aligned with county and state, indicating that test functionality was an issue. Do we see any of that this year?

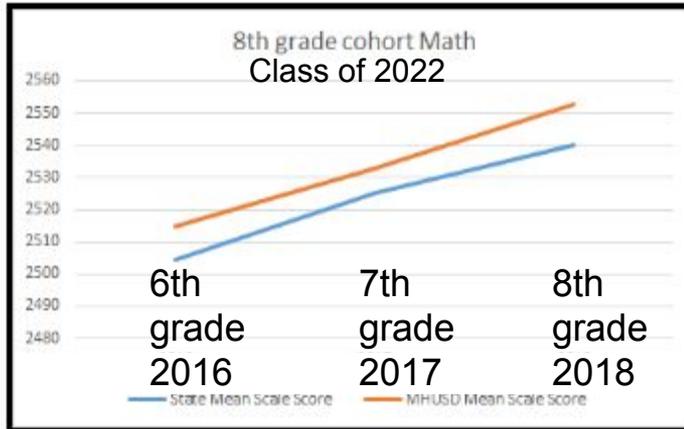
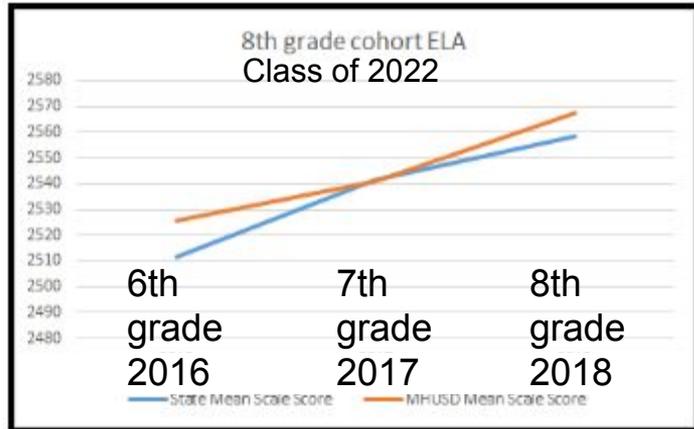
Similar change scores year to year indicate may suggest measurement that affects scores. For example, the largest gains in District CLE scores is 0.3 points at grade 5, but this gain returns the state and county gaps.

This data suggests we areas of need are with literacy and math in middle school (SLC). But is that a reliable conclusion? How data is year to year change data, so the student groups in each year are different. This can affect the district more than the state as the district is a much smaller group. This is an example of volatility due to changing test groups. What does Cohort growth data tell us?

Cohort Outcome Data for District and State as students pass through Elementary and Middle School. *The district is shown in orange and the state in blue. Growth from 3rd to 5th and 6th to 8th*

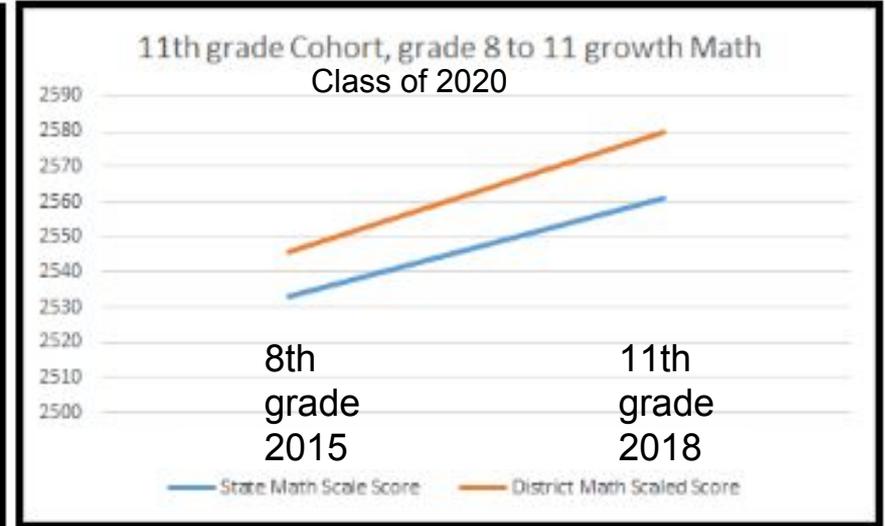
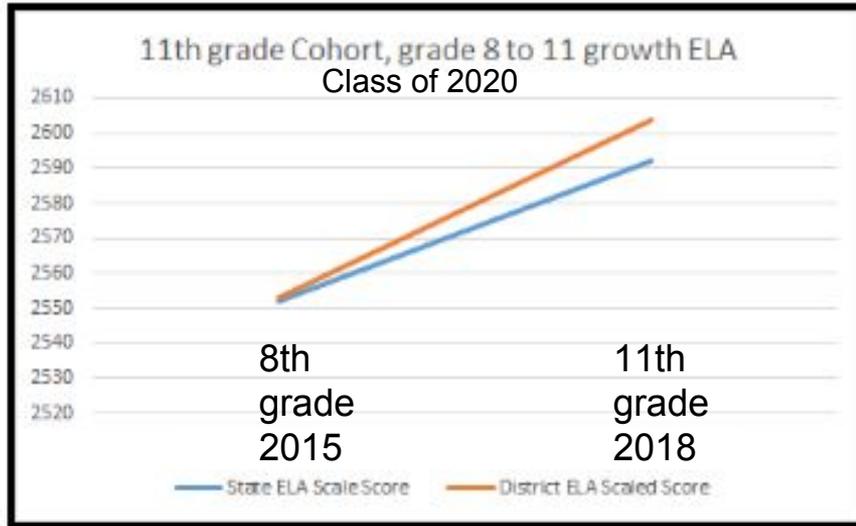


Last year's 5th grade students experienced similar scores and growth compared to statewide scores. There is convergence of scores that **suggests K-5 ELA and Math as areas for attention.**



Last year's 8th grade students experienced higher scores and higher rates of growth than their statewide counterparts indicating a strong middle grade program with **grade 7 ELA warranting attention.**

Cohort Outcome Data for Grade 11 is now available as last year's Juniors tested in grade 8
The district is shown in orange and the state in blue.

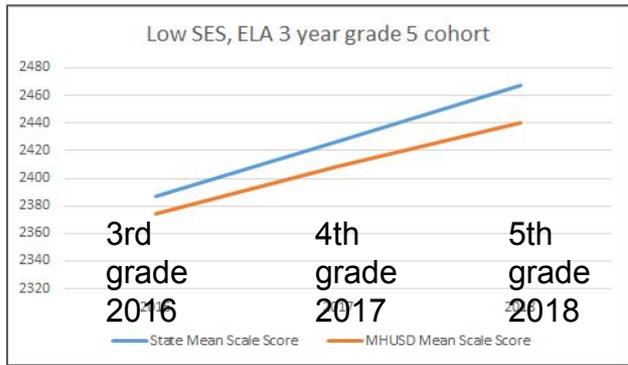


Last year's 11th grade students experienced higher scores and a higher growth rate (slope) compared to their statewide counterparts. This positive divergence is an indicator of a **strong high school program, direction, and final K-12 outcomes for our students.**

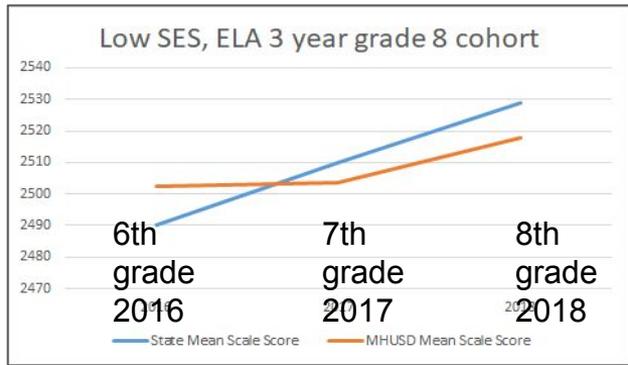
Cohort Outcome Data for Socioeconomically Disadvantaged Students

The district is shown in orange and the state in blue.

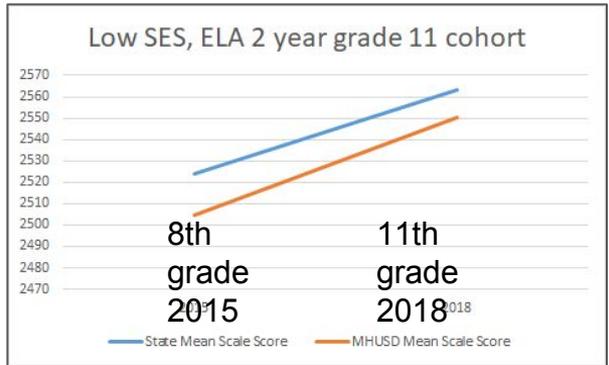
Grade 5 cohort 3 year trend



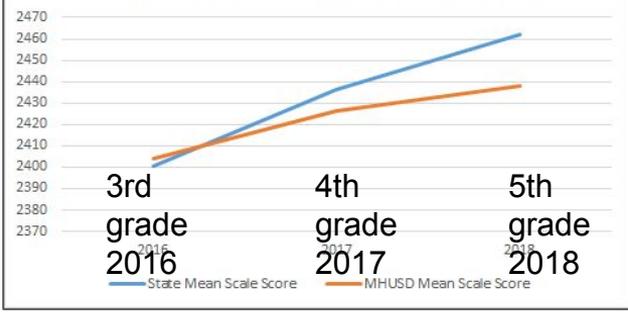
Grade 8 cohort 3 year trend



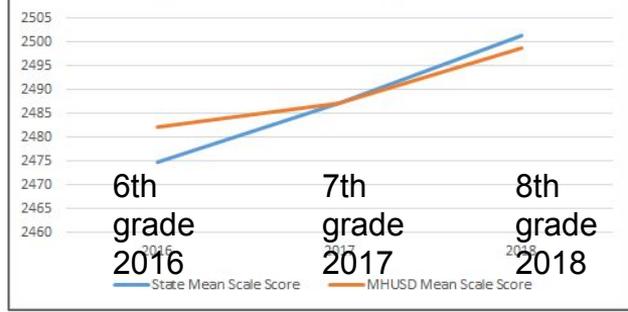
Grade 11 cohort 2 year trend



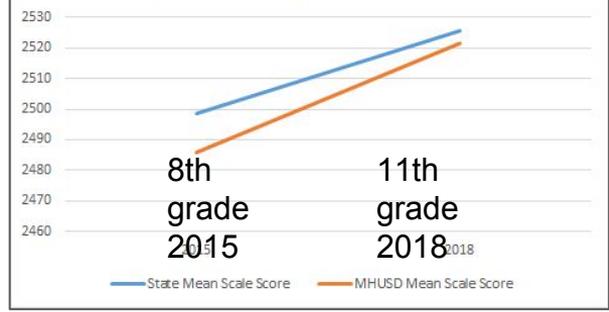
Low SES, Math 3 year grade 5 cohort



Low SES, Math 3 year grade 5 cohort



Low SES, Math 2 year grade 11 cohort



Low SES testing data shows areas of need in early and middle grades (markedly in 7th) with recovery through high school. Comparison groups do not account for language or mobility variables in cohort due to limits in state data disaggregation

Cohort Outcome Data for Ever English Language Learners

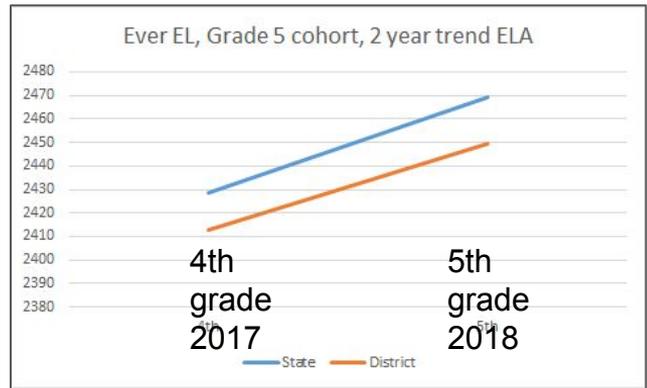
The district is shown in orange and the state in blue.

Ever EL's is a new metric now being reported by the state as an attempt to improve the meaning of data for a student group which is constantly changing by mobility, by program design, and by a variety of reclassification practices.

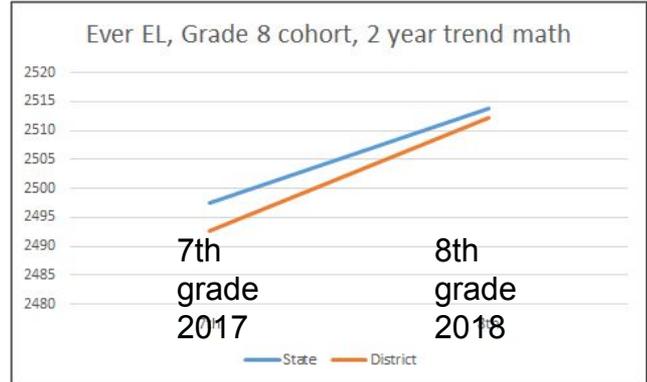
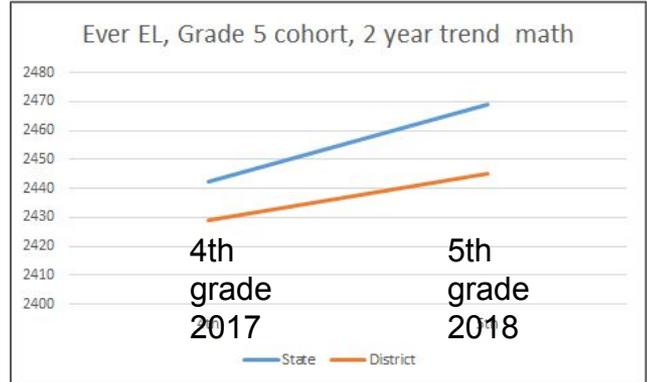
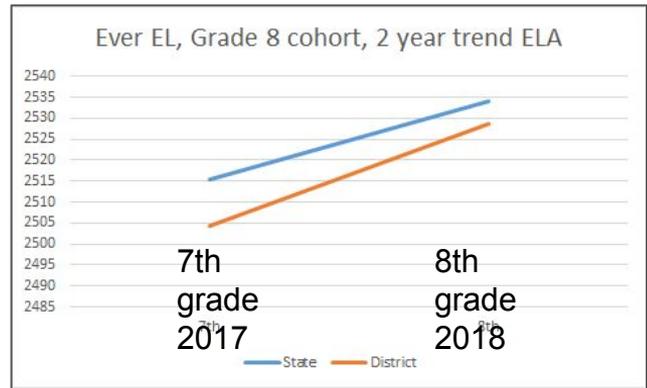
The Ever EL data set cannot be further disaggregated to determine the effects of other variables such as Initial fluency level, fluency in primary language, socioeconomic status, migrant status, parent education level etc.

Most important are relative slopes or growth rates which indicate strong programs with the highest area of **need in elementary math literacy**

Grade 5 cohort 2 year trend



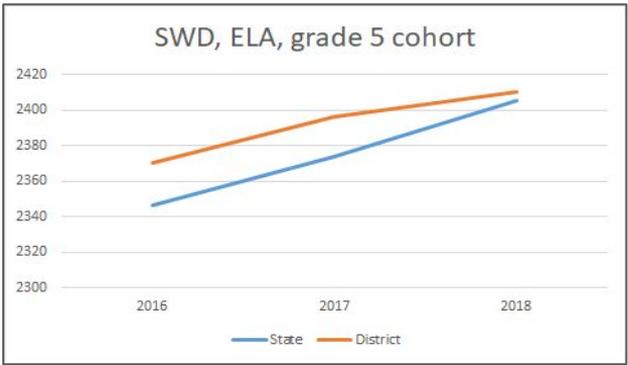
Grade 8 cohort 2 year trend



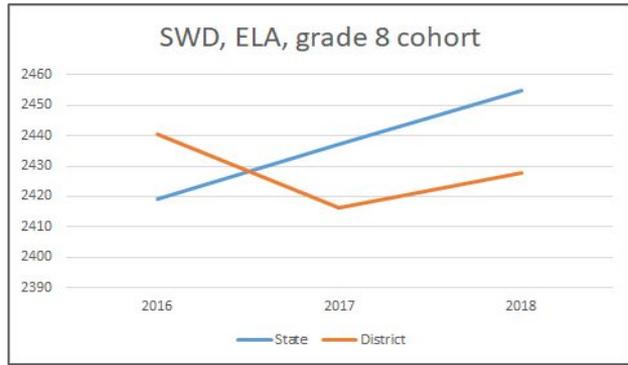
Cohort Outcome Data for **Students with Disabilities**

The district is shown in orange and the state in blue.

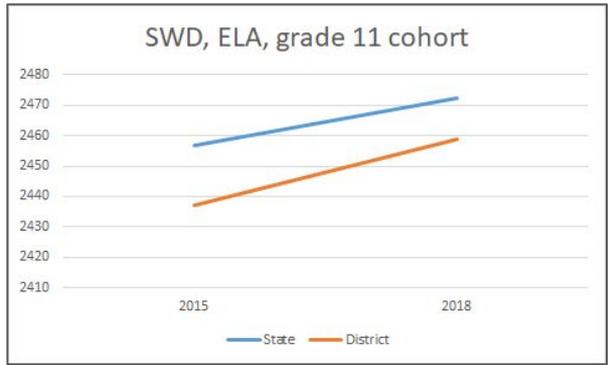
Grade 5 cohort 3 year trend



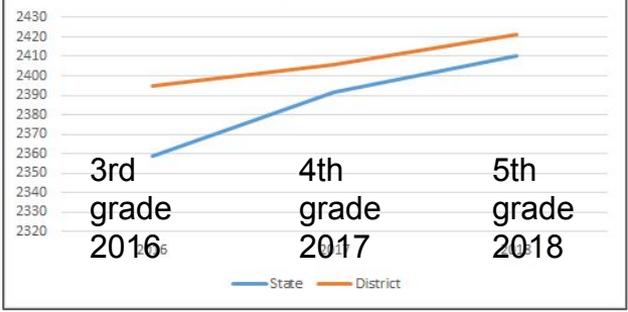
Grade 8 cohort 3 year trend



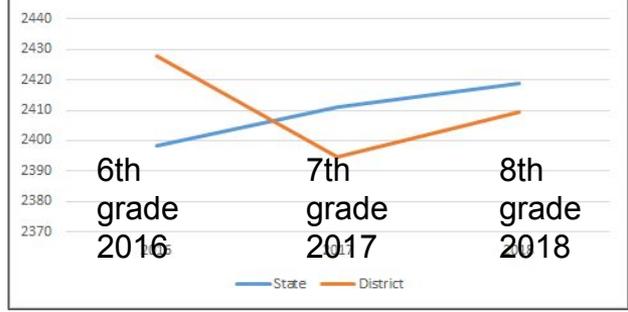
Grade 11 cohort 2 year trend



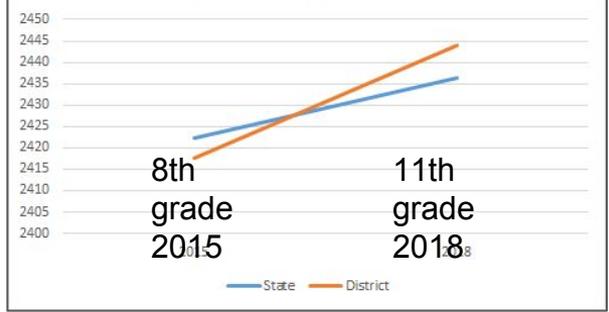
SWD, Math, grade 5 cohort



SWD, Math, grade 8 cohort



SWD, Math, grade 11 cohort

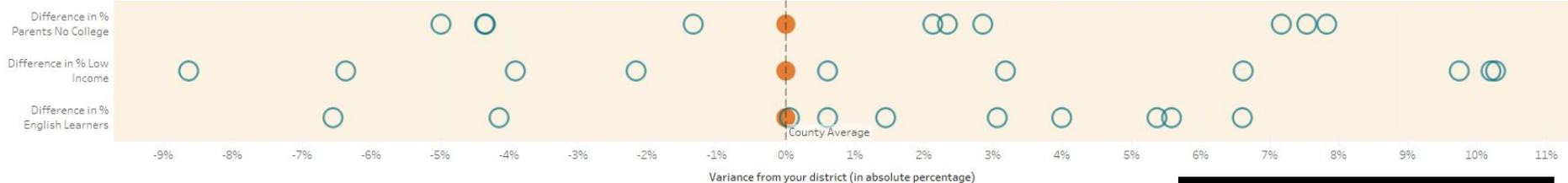


SWD testing data does not account for disability types or severity in the testing group. The middle school pattern suggests a possible programmatic issue and need to **examine special education instructional rigor and inclusion practices at the secondary grades.**

Similar Districts:

*How similar are they and what does the distribution suggest about relative position?
And what does relative position tell us about program implementation?*

County	District	City	Grade Range	Most Common Language	Ethnic Diversity and Largest Ethnic Group	# Schools	Enrollment	Parents No College	Low Income	English learners
Santa Clara County	Morgan Hill USD	Morgan Hill	K-12	Spanish	Hispanic (52%)	13	8,501	31%	41%	18%
Alameda County	New Haven USD	Union City	K-12	Spanish	Hispanic (38%)	13	12,150	34%	45%	24%
	Newark USD	Newark	K-12	Spanish	Hispanic (53%)	12	6,008	38%	48%	23%
Marin County	Novato USD	Novato	K-12	Spanish	White (50%)	15	7,658	26%	35%	18%
San Bernardino Co..	Upland USD	Upland	K-12	Spanish	Hispanic (55%)	14	11,138	26%	52%	11%
San Luis Obispo County	Lucia Mar USD	Arroyo Grande	K-12	Spanish	White (50%)	18	10,704	26%	52%	14%
	Paso Robles Jt. USD	Paso Robles	K-12	Spanish	Hispanic (52%)	12	6,710	39%	51%	22%
San Mateo County	South San Francisco USD	South San Francisco	K-12	Spanish	Hispanic (48%)	16	8,924	33%	39%	23%
Sonoma County	Cotati-Rohnert Park USD	Rohnert Park	K-12	Spanish	White (45%)	12	5,855	33%	42%	19%
	Windsor USD	Windsor	K-12	Spanish	Hispanic (50%)	8	5,243	38%	37%	21%
Ventura County	Moorpark USD	Moorpark	K-12	Spanish	Hispanic (46%)	11	6,513	29%	33%	18%



Similar District Data: how do we compare against the growth and status scores for the most similar districts in the state? (similarity based on socioeconomic status, English Learners and parent education level).

ELA

Relative Position by grade level, ELA	2018	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade	11th Grade
Novato Unified	2018 ELA	2436.7	2468.6	2507.4	2508.5	2564.2	2571.9	2623.9
New Haven Unified	2018 ELA	2429.1	2464.6	2504	2504.9	2531.3	2563.1	2619.7
Newark Unified	2018 ELA	2429.1	2464.6	2504	2504.9	2531.3	2563.1	2619.7
Morgan Hill Unified	2018 ELA	2416.7	2458.8	2490.6	2510.3	2534.3	2560.1	2604.6
Lucia Mar Unified	2018 ELA	2432.1	2472.3	2502.8	2536.9	2551.9	2566.8	2600.9
Upland Unified	2018 ELA	2444.6	2458.7	2501.9	2532.6	2553.4	2567.9	2596.2
Windsor Unified	2018 ELA	2420.7	2441.3	2487.2	2507.8	2531.3	2552.5	2594.8
Moorpark Unified	2018 ELA	2455.5	2508.9	2535.5	2537.3	2572	2594.1	2594.5
Paso Robles Unified	2018 ELA	2420.4	2451.7	2481.1	2515.6	2540.2	2559.8	2590.3
So. San Fran Unified	2018 ELA	2429.4	2459.1	2495.2	2506.2	2551.2	2564.4	2585.9
Cotati-Rhonert Park Unified	2018 ELA	2418.9	2461.7	2482.6	2519.8	2527.3	2546.4	2587.7
Average	2018 ELA	2430.291	2464.573	2499.3	2516.8	2544.4	2564.555	2601.2
Relative Position 1 through 11	2018 ELA	11	8	8	6	7	8	4
MHUSD Cohort Growth 2017-2018	2017-18 ELA	N/A	40.1	33.3	25.0	12.2	21.0	36.9
Group Average Cohort Growth	2017-18 ELA	N/A	45.6	43.7	23.0	22.9	17.2	38.2
District Standard Error of Measurement	2018 ELA	+/- 4	+/- 4	+/- 4	+/- 4	+/- 4	+/- 4	+/- 5

Initial relative position within the comparison group improves as students move up through the grade levels. This is a significant trend among a group of similar districts who are all engaged in their own improvement efforts.

MHUSD Cohort growth rates among fifth and seventh grade are somewhat lower than the comparison group average.

Conclusions from this data is that MHUSD has a strong overall K-12 program with areas of challenge being grades 3, 5 and 7 ELA.

Similar District Data: how do we compare against the growth and status scores for the most similar districts in the state? (similarity based on socioeconomic status, English Learners and parent education level).

Math

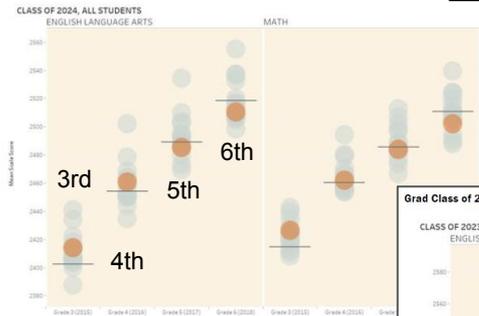
Relative Position by grade level, Math	2018	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade	11th Grade
Novato Unified	2018 Math	2431.8	2472.7	2502.1	2510.9	2572	2569.2	2598.5
Moorpark Unified	2018 Math	2463.9	2498.5	2518.8	2515	2555.3	2574.5	2581.9
Morgan Hill Unified	2018 Math	2425.7	2463.9	2486.7	2502.1	2529.3	2548.6	2580.1
New Haven Unified	2018 Math	2442.1	2467.5	2497.6	2504.8	2522.5	2551.5	2578.7
Newark Unified	2018 Math	2442.1	2467.5	2497.6	2504.8	2522.5	2551.5	2578.7
Cotati-Rhonert Park Unified	2018 Math	2421	2458.9	2472.8	2487.7	2498.4	2512.2	2565.1
Paso Robles Unified	2018 Math	2428.6	2453.2	2483.6	2506.1	2499	2549.8	2563.1
Lucia Mar Unified	2018 Math	2435.2	2472.2	2492.3	2524.1	2530.6	2542.9	2558.8
Windsor Unified	2018 Math	2425.3	2449.1	2471.2	2490.4	2500.3	2531.7	2556.9
So. San Fran Unified	2018 Math	2440.8	2480.4	2507.4	2519.8	2537.7	2558.5	2555.4
Upland Unified	2018 Math	2440.5	2462.2	2487.4	2524.4	2517.1	2561.7	2547.5
Average	2018 Math	2436.1	2467.8	2492.5	2508.2	2525.9	2550.2	2569.5
Relative Position 1 through 11	2018 Math	9	7	8	9	5	8	3
MHUSD Cohort Growth 2017-2018	2017-18 Math	N/A	35.6	18.8	18.3	17.3	19.9	27.2
Group Average Cohort Growth	2017-18 Math	N/A	37.2	27.8	19.2	15.3	18.3	20.8
District Standard Error of Measurement	2018 Math	+/- 4	+/- 3	+/- 4	+/- 4	+/- 4	+/- 5	+/- 5

Initial relative position within the comparison group improves as students move up through the grade levels. This is a significant trend among a group of similar districts who are all engaged in their own improvement efforts.

MHUSD Cohort growth rates are strong among secondary grades and similar at elementary grades with grade 5 math having lower growth

Conclusions from this data is that MHUSD has a strong overall K-12 program with an area for focus being grade 5 Math based on cohort growth

Grad class of 2024



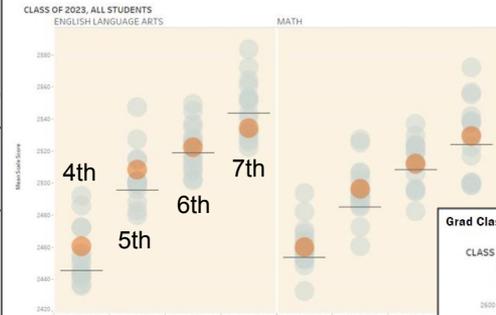
Average scale score for all districts is indicated by the horizontal black line. Morgan Hill appears in orange.

7th graders

Cohort Similar Districts Position

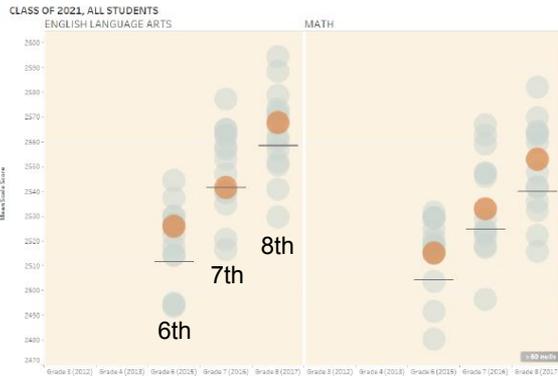
8th graders

Grad Class of 2023



Average scale score for all districts is indicated by the horizontal black line. Similar districts appear in light blue. Morgan Hill appears in orange.

Grad class 2021



Average scale score for all districts is indicated by the horizontal black line. Similar districts appear in light blue. Morgan Hill appears in orange.

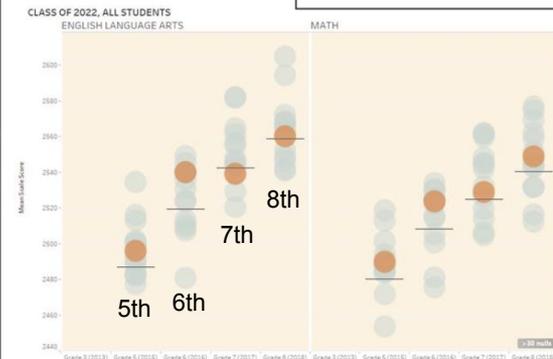


Sophomores

Four year **time sequence** for class cohorts that span the testing grades

This visualization reveals **grade 7 ELA** as a **relative area for attention** with an **overall strong progression of relative standing as cohorts move up through the grades**

Grad Class 2022



Average scale score for all districts is indicated by the horizontal black line. Similar districts appear in light blue. Morgan Hill appears in orange.



Freshman



Schoolwise Press,
K-12 Measures

Summary of Areas of Strengths and Challenges by Grade as indicated by CAASPP Data

Data Lens	ELA Challenges	ELA Strengths	Math Challenges	Math Strengths
From County & State grade level change comparison (non cohort)	Grade 3,4,6,7,8	Grade 11 (includes progress made during grades 9-11)	Grade 3,4,6	Grade 11 (includes progress made during grades 9-11)
From State Cohort growth for grade band outcomes	Elementary program and grade 7	Secondary program: grade 8 and 11 outcomes	Elementary program: grade 5 outcomes	Secondary program: grade 8 and 11 outcomes
From Most Similar District comparison Status	Grade 3, 4, 5	Grade 6 and 11	Grade 3, 5, 6	Grade 7 and 11
From Most Similar District comparison Cohort Growth over time	Grade 3, 5 and 7	Grade 6 and 8	Grade 5	Grade 7, 8 and 11
Priority areas of strength/challenge	Grades 3,4,5,7 All, SED, EL, SWD	Grades 6,8, (9-10)11	Grades 3,5,6 All, SED, EL, SWD	Grades 7,8, (9-10)11

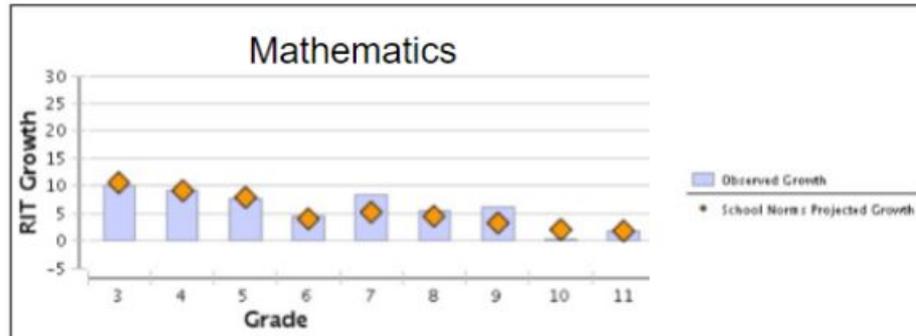
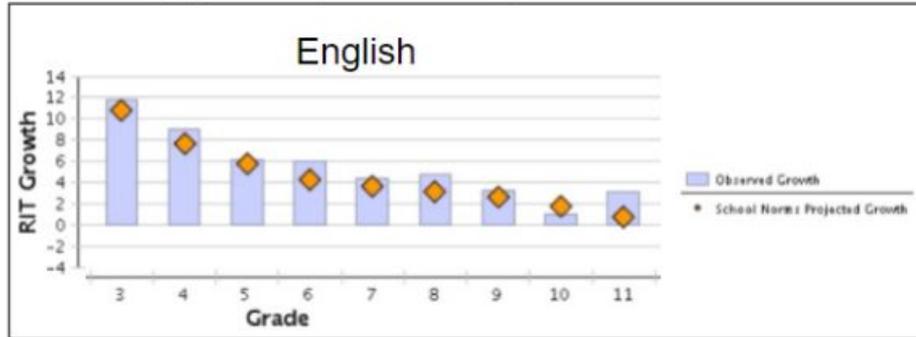
Overall, CAASPP data based on scales and change indicate more areas of challenge than CAASPP data analysed by cohort growth and similar comparison groups.

How do these areas compare with multiple measures? Let's see...

Validation of areas of strengths and challenges by comparison to NWEA MAP results:

Priority areas of strength/challenge	Grades 3,4,5,7 All, SED, EL, SWD	Grades 6,8,11	Grades 3,5,6 All, SED, EL, SWD	Grades 7,8,11
--------------------------------------	-------------------------------------	---------------	-----------------------------------	---------------

Diamond represents national normed growth expectation for similar districts and bar represents MHUSD growth (Fall to Fall cohort growth scores)



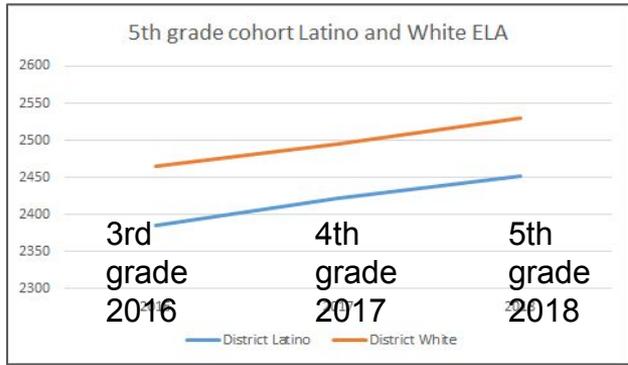
MAP tests verify that grades 6,8, and 11 are outpacing national English growth by the largest margins. Grades 3 and 4 also show high growth, with grades 5 and 7 the most similar to national growth. This indicates an overall strong program with **5th and 7th grade ELA marginally indicated as areas for attention.**

MAP tests show that grades 3,4,5, and 6 are borderline to national Math norms with 7th grade having stronger results. **This validates elementary math as an area for attention.**

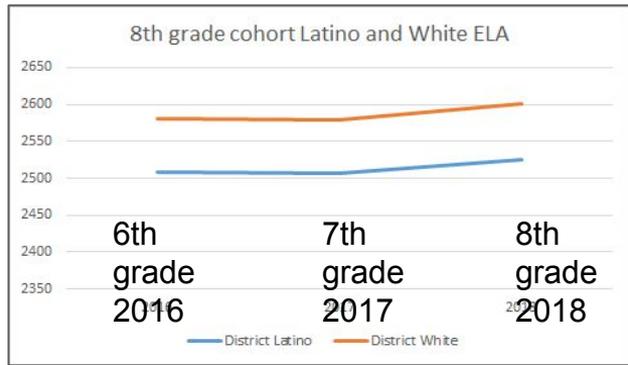
Latino-White achievement gaps by cohorts

The district is shown in orange and the state in blue.

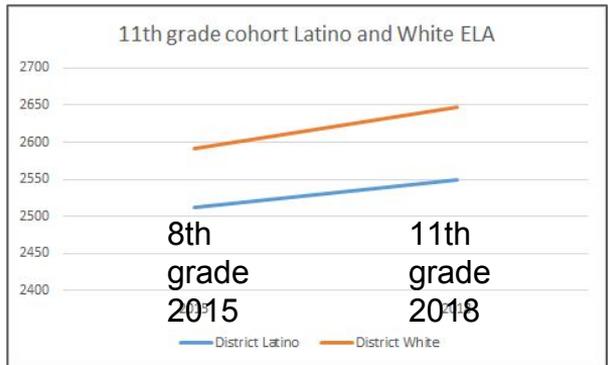
Grade 5 cohort 3 year trend



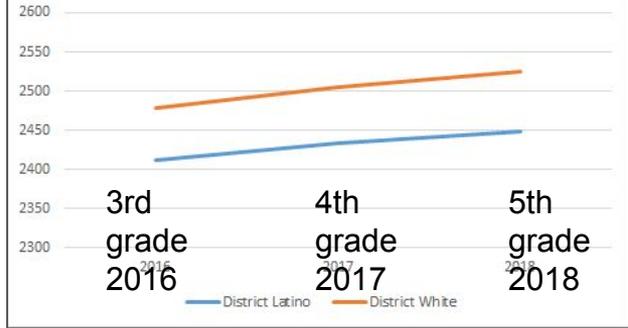
Grade 8 cohort 3 year trend



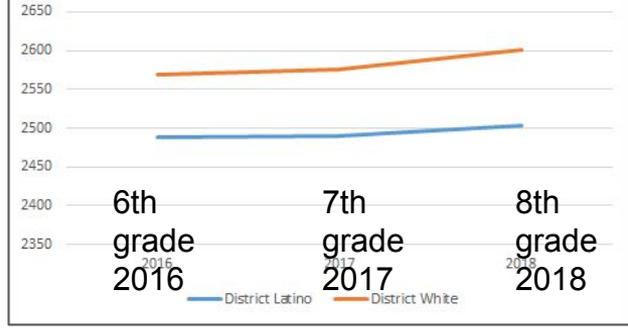
Grade 11 cohort 2 year trend



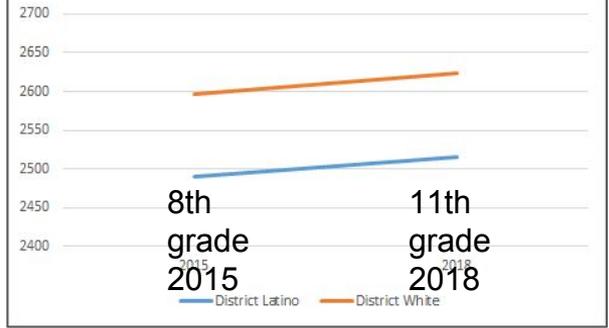
5th grade cohort Latino and White Math



8th grade cohort Latino and White Math



11th grade cohort Latino and White Math



The growth rates of both ethnic groups is similar; however, the gap is persistent. Comparison groups do not control major variables such as language fluency, socioeconomic status, parent education level, or measurement bias to isolate causality.

Strengths and challenges indicated by student group gap data

Improvement in the EL gap with other gaps relatively unchanged.

ELA	White to Other	Latino to Other	EL to Other	SpEd to Other	Low SES to Other
2017 ELA Gap	37.6	-35.4	-131.7	-93.8	-46.9
2018 ELA Gap	39.7	-37.7	-118.8	-100.3	-47.7

Math	White to Other	Latino to Other	EL to Other	SpEd to Other	Low SES to Other
2017 Math Gap	40.0	-39.9	-121.9	-99.8	-40.9
2018 Math Gap	41.3	-41.2	-111.3	-96.1	-42.1

Note: "All" means "All other than the group identified", gaps are based on average gaps of all tested grade levels.

Causality Variables

2018 Gaps	White & Latino Gap	White Latino Non English Learner Gap	White Latino Non Disadvantaged Gap	White Latino Similar Parent Ed. Gap	All 3 controlled, remaining GAP
ELA	79	55	47	46	19
Math	85	66	56	55	21
Effect		33%	28%	28%	11%

Additional strengths or challenges by school site data

Similar students comparison, grades 3-5 schools

Non-SED, Non EL, White			Non-SED, Non EL, Latino			Low SES, Non EL, Latino		
	ELA	Math		ELA	Math		ELA	Math
Barrett	2475	2484	Barrett	2441	2447	Barrett	2377	2428
El Toro	2471	2476	El Toro	2424	2422	El Toro	2403	2422
Jackson	2510	2516	Jackson	2498	2485	Jackson	2459	2468
Los Paseos	2498	2479	Los Paseos	2468	2458	Los Paseos	2427	2418
Nordstrom	2525	2511	Nordstrom	2501	2488	Nordstrom	2414	2431
PA Walsh	***	***	PA Walsh	2469	2456	PA Walsh	2408	2406
Paradise	2533	2540	Paradise	2479	2497	Paradise	***	***
San Martin Gwinn	2498	2511	San Martin Gwinn	2483	2469	San Martin Gwinn	2424	2423

Scaled scores for 2018 CAASPP tests by site, *** indicates insufficient sample size for score (n<25)

Shifting relative positions and close range of scores (recall standard error) indicate that the instructional program at all schools is producing similar results. Test scores do not provide a credible method for ranking schools or predicting student outcomes by school.

Additional strengths or challenges by school site data

Middle Schools

Non-SED, Non EL, White		
	ELA	Math
Britton	2562	2572
Murphy	2588	2582
Jackson	2606	2596

Non-SED, Non EL, Latino		
	ELA	Math
Britton	2537	2534
Murphy	2554	2541
Jackson	2548	2531

Low SES, Non EL, Latino		
	ELA	Math
Britton	2506	2493
Murphy	2511	2458
Jackson	2525	2514

High Schools

Non-SED, Non EL, White		
	ELA	Math
Ann Sobrato	2669	2669
Live Oak	2655	2613

Non-SED, Non EL, Latino		
	ELA	Math
Ann Sobrato	2617	2578
Live Oak	2623	2557

Low SES, Non EL, Latino		
	ELA	Math
Ann Sobrato	2536	2532
Live Oak	2584	2516

Scaled scores for 2018 CAASPP tests by site

As with elementary, shifting relative positions and close range of scores (recall standard error) indicate that the instructional program at all schools is producing similar results.

Test scores do not provide a credible method for ranking schools or predicting student outcomes by school.

Additional strengths or challenges by claim data

ELA/Literacy	Grade 5	Grade 8	Grade 11
Reading	33 42 25	31 46 23	22 42 36
Writing	33 38 29	28 46 26	24 36 41
Listening	24 60 16	19 66 15	16 58 27
Research/Inquiry	29 44 27	24 43 33	21 35 43

ELA Claim data shows a **steady decrease in the lowest performance level as measured in grades 5, 8, and 11 last spring** with the highest achievement levels in high school

Although there are a relatively high number of advanced writers by grade 11, writing has the highest number of students in the lowest performance level...**suggests a program need for expository writing based on evidence drawn from reading passages.**

Mathematics	Grade 5	Grade 8	Grade 11
Concepts and Procedures	49 27 24	45 31 25	48 24 28
Problem Solving and Modeling & Data Analysis	38 43 20	30 44 26	35 44 21
Communicating Reasoning	37 47 16	29 48 24	25 50 24

Math Claim data shows that **concepts and procedures** is a high area of need in the lowest performance levels indicating a focus area for interventions. **Communicating higher order problem solving** for all students is also an area for attention.

How **initiatives correlate** to our priority areas of **focus**: *Elementary ELA and Math outcomes*

MTSS: Refining metrics, building regular data reports, and data analysis training.

- Deepening our use of **multiple measures** and data to connect students with **strategic supports** via the **MTSS** initiative. (SED, EL, low Parent Ed, SpEd)

Elementary Math Literacy: SVCF and Project Light Grant initiatives

- Developing **math literacy for expressing thinking and problem solving for higher order performance tasks** similar to the those on the CAASPP with the goal of increasing measurable achievement as students complete K-5

Early Literacy: Tiered interventions and Writing Initiative.

- Elementary schools are now staffed with reading specialists, MTSS para-professionals and instructional materials for **leveled literacy intervention**.
- We are **developing a district-wide expository K-5 writing** program with rubrics, prompts, and calibrating exemplars to support interdisciplinary writing.

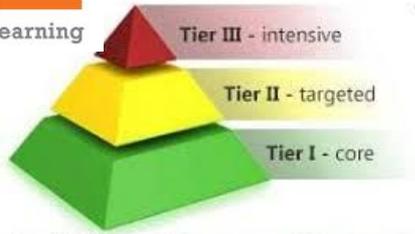


OUR OBJECTIVES

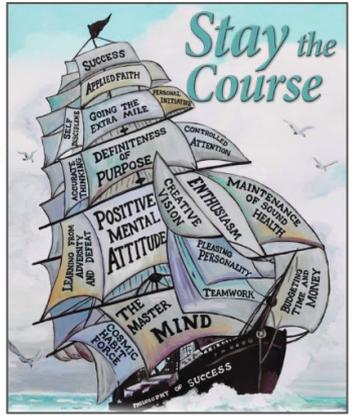
- Gain perspective of the current status of the SBAC test and how to use the data in the most credible and reliable manner.
- Understand our areas of strengths and challenges as indicated by our CAASPP scores and validated with multiple measures.
- Understand how our current actions correlate to challenges as revealed by data.



Use Multiple Lenses



Multi-Tiered System of Supports



Reference Slides

A ship in harbor is safe, but that is not what ships are built for.
- John A. Sheehy, 1928



2018 Accountability Update

2018 CAASPP Test Results,
Report to the Governing Board
November 13, 2018




Glen Webb, Director of Curriculum, Instruction and Assessment

The Shift to a Growth Model

Meeting all students where they are and supporting their growth...It's a statewide, multi-year journey.



California's progress towards a new accountability system

2016-17
 ✓ First year Smarter Balanced assessment scores used to calculate status and pre-calculate into the evaluation matrix by January 2017
 ✓ Second year Smarter Balanced assessment scores available (7 as 2016)
 ✓ First year of California Alternate Assessment scores available (7 as 2016)
 ✓ SDE approve a set of criteria to initiate the selection of a growth model methodology
 ✓ SDE will use analysis from technical experts will explore growth model methodologies that meet the state's operational needs

2017-18
 ✓ Third year Smarter Balanced assessment scores available (7 as 2017)
 ✓ Second year of California Alternate Assessment scores available (Fall 2017)
 ✓ SDE approve a growth model methodology
 ✓ CCC with assistance from technical experts will perform growth model calculations for the purpose of reporting to the SDE and tracking business rules and requirements
 ✓ It's a Journey

2018-19
 ✓ CCC with assistance from technical experts will perform growth model calculations for the purpose of public reporting

2019-2020
 Growth model available for accountability and continuous improvement

OUR OBJECTIVES

- Gain perspective of the current status of the SBAC test and how to use the data in the most credible and reliable manner.
- Understand our areas of strengths and challenges as indicated by our CAASPP scores and validated with multiple measures.
- Understand how our current actions correlate to challenges as revealed by data.

1

2

3

4

How is the Test Functioning?

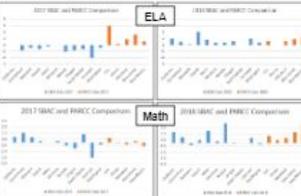
STATUS **VOLATILITY**
BIAS **Change** **GROWTH**
CERTAINTY Accuracy Vs Precision





How is the test looking for year to year Volatility?

Gain scores between the SBAC and PRRC systems appear more consistent than they were last year. However, gain scores as reported on the dashboard will use last year's very volatile results and will take time to stabilize. Our analysis will use comparative growth as a control for volatility.



Reporting on the Dashboard... Status and Change

Assign by Five reference charts are used to design an indicator score that will be displayed on the state dashboard. This example for English Language Arts categorizes the change in score from the previous year along the top axis and status measured as distance from level 3 along the vertical axis.

The boxes being raised around this methodology is that "change" is not "growth" as it uses different graduating years of students for year to year comparisons rather than the same cohort to determine growth. Our analysis today will be based on cohorts to help control measurement volatility.



Reporting Status on the CAASPP site...

Performance Levels...
Can the test tell us if we should be...

Happy because we measure up? Sad because we don't? Carefree either way? We need to know more about SCALES



5

6

7

8

Reporting on the Dashboard... Status and Change

Status is determined by reference to a non-stem referenced scale based on the number of points from level 3. The boxes being raised around this methodology is that it lacks comparative value and is unacceptable to measurement validity as the test bank develops and configurations are refined. Our analysis today will be based on comparison to the state as the largest and most stable data set and comparison to similar districts to provide some control over demographic variables. We will look at trends rather than status for relative cohort growth over time as a more reliable indicator.



Math CAASPP	Percent of total test takers in each category	CAASPP score minus distance from level 3
African-American	68%	-47.5
White	72%	10.7
Hispanic	30%	-6.5
Asian-American	67%	55.1
English Learners	34%	-77.2
English Only	66%	-11.6
Reclassified ELLs	38%	-43.1
Socioeconomically disadvantaged	24%	-76.1
Learning disabled	52%	-121.5

Distance from "meeting standard" suggests that most groups are doing poorly compared to that scale as implied by negative scores. But, in information not readily available without downloading big data, research is less and further analysis, the percentile rankings of the groups compared to their similar student groups statewide looks significantly off-peak. For example, MUSD's lowest performing group is actually performing in the top half of districts statewide. Our analysis will look at comparative results to identify areas of strength and challenge.

How are Performance Levels working with current test Precision

Clara of Standard Errors of Measurement

Director Level CLA Score: +/- 4 to 5 pts
 School Level CLA Score: +/- 4 to 10 pts
 Student Level CLA Score: +/- 21 to 47 pts

Director Level Math Score: +/- 3 to 5 pts
 School Level Math Score: +/- 4 to 10 pts
 Student Level Math Score: +/- 21 to 70 pts

Precision is not reported publicly but the standard error of measurement often overlap adjacent performance levels and dashboard cells. Our control for imprecision is to use scaled scores rather than performance levels and to validate with multiple measures.



Last year we saw that our grade level change scores had peaks and valleys that aligned with county and state, indicating that test functionality was an issue. Do we see any of that this year?

Grade Level	San Joaquin District Change, 2017 to 2018					
	1	2	3	4	5	6
Math Score	2	4	1	0	7	8
Change	0	1	0	0	1	0
State Change	4	1	0	0	1	0
County Change	4	1	0	0	1	0
District Standard Error	10	10	10	10	10	10
State Standard Error	10	10	10	10	10	10
County Standard Error	10	10	10	10	10	10

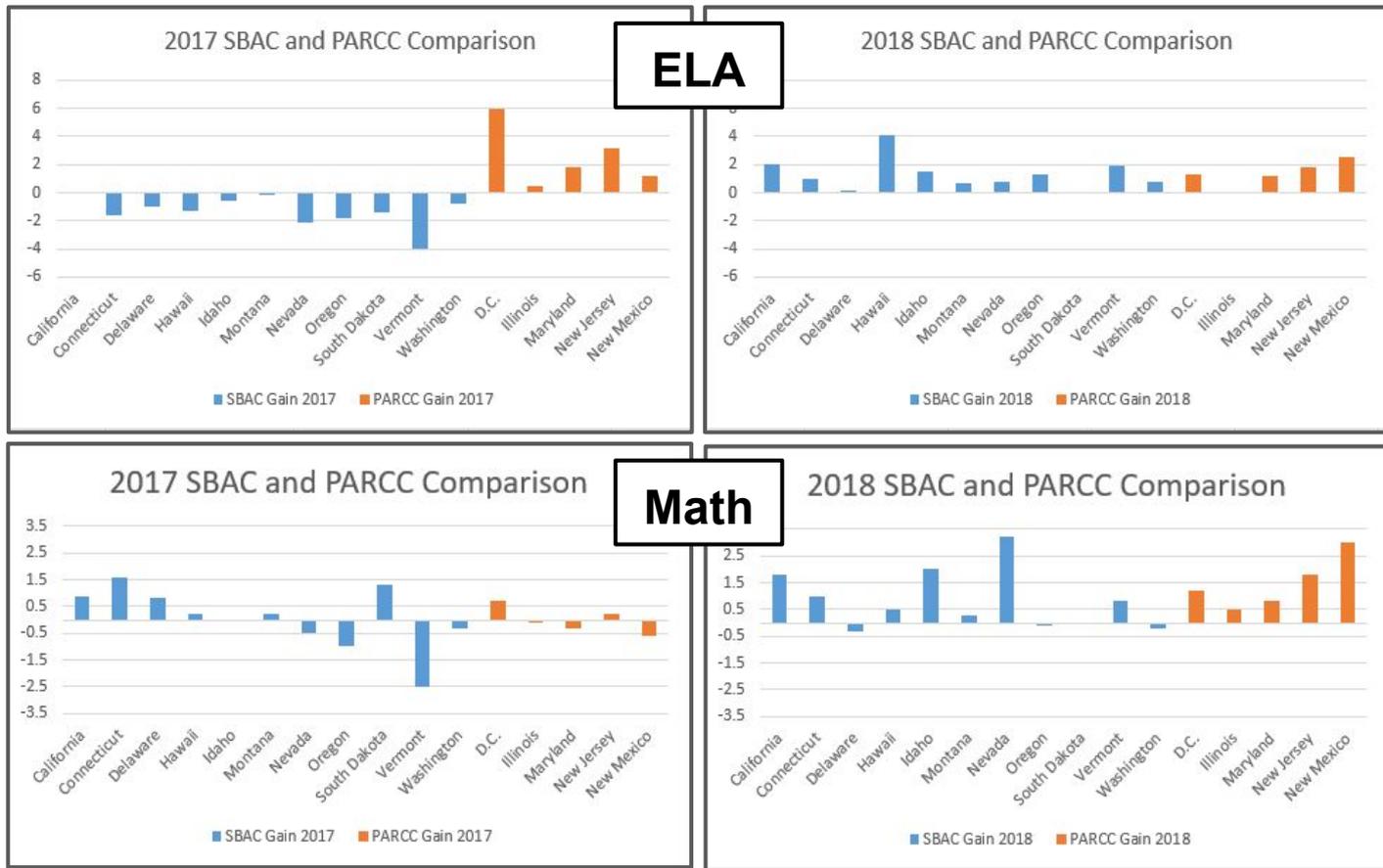
Similar change scores year to year indicate they ongoing recognition that affects scores. For example, the largest gain in District LLA scores is 5.3 points at grade 5...but the gain returns the state and county gains.

This data suggests our areas of need are early literacy and math and middle school LLA. But is that a reliable conclusion? This data is year to year change data, so the student groups in each year are different. This can reflect the district more than the state as the district is a much smaller group...this is an example of volatility due to changing test groups. What does cohort growth data tell us?

How is the test looking for year to year **Volatility?**

Gain scores between the **SBAC and PARCC systems appear more consistent** than they were last year; however, gain scores as reported on the dashboard will use last year's very volatile results and will take time to stabilize.

Our analysis will use **comparative growth as a control for volatility**



Reporting on the Dashboard... Status and Change

Five by Five reference charts are used to assign an indicator color that will be displayed on the state dashboard. This example for English Language Arts categorizes the change in score from the previous year along the top axis and status measured as distance from level 3 along the vertical axis.

The issues being raised around this methodology is that **“change” is not “growth”** as it uses different graduating years of students for year to year comparisons rather than the same cohort to determine growth. **Our analysis today will be based on cohorts to help control measurement volatility.**

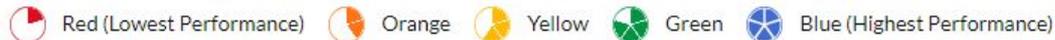
STATUS



Academic Indicator Five-by-Five Color Tables
English Language Arts/Literacy Assessment Five-by-Five Color Table

Level	Declined Significantly by more than 15 points (Change)	Declined by 3 to 15 points (Change)	Maintained Declined by less than 3 points or Increased by less than 3 points (Change)	Increased by 3 to less than 15 points (Change)	Increased Significantly by 15 points or more (Change)
Very High 45 points or Higher (Status)	Green	Green	Blue	Blue	Blue
High 10 to 44.9 points (Status)	Green	Green	Green	Green	Blue
Medium -5 points to +9.9 points (Status)	Yellow	Yellow	Yellow	Green	Green
Low -5.1 to -70 points (Status)	Orange	Orange	Orange	Yellow	Yellow
Very Low -70.1 points or lower (Status)	Red	Red	Red	Orange	Orange

Performance Levels:



Reporting **Status** on the **CAASPP** site....



Happy because we measure up?

Performance Levels...

Can the test tell us if we should be....

Sad because we don't

Carefree either way

We need to know more about **SCALES**



Reporting on the Dashboard...

Status and Change

Status is determined by reference to a non-norm referenced scale based on the number of points from level 3. The issues being raised around this methodology is that it lacks comparative value and is susceptible to measurement volatility as the test bank develops and calibrations are refined.

Our analysis today will be based on **comparison to the state** as the largest and most stable data set and **comparison to similar districts** to provide some control over **demographic variables**. We will look at **trends rather than status** for **relative cohort growth over time** as a more reliable indicator.

Little comparative value

S
T
A
T
U
S

← Change →

Academic Indicator Five-by-Five Color Tables
English Language Arts/Literacy Assessment Five-by-Five Color Table

Level	Declined Significantly by more than 15 points (Change)	Declined by 3 to 15 points (Change)	Maintained Declined by less than 3 points or Increased by less than 3 points (Change)	Increased by 3 to less than 15 points (Change)	Increased Significantly by 15 points or more (Change)
Very High 45 points or higher (Status)	Green	Green	Blue	Blue	Blue
High 10 to 44.9 points (Status)	Green	Green	Green	Green	Blue
Medium -5 points to +9.9 points (Status)	Yellow	Yellow	Yellow	Green	Green
Low -5.1 to -70 points (Status)	Orange	Orange	Orange	Yellow	Yellow
Very Low -70.1 points or lower (Status)	Red	Red	Red	Orange	Orange

Performance Levels:

-  Red (Lowest Performance)
-  Orange
-  Yellow
-  Green
-  Blue (Highest Performance)

Math CAASPP	Percentile Among Peer Subgroups in Calif	CAASPP Scale Score Distance from Level 3
African-American	68%	-47.5
White	75%	10.7
Hispanic	36%	-66.5
Asian-American	67%	55.3
English Learners	34%	-77.2
English Only	66%	-11.6
Reclassified ELL's	28%	-43.1
Socioeconomically disadvantaged	24%	-76.1
Learning disabled	55%	-112.5

Distance from “meeting standard” suggests that most groups are doing poorly compared to that scale as implied by negative scores

But, in information not readily available without downloading big data research files and further analysis, the percentile rankings of the groups compared to their similar student groups statewide looks significantly different. For example, MHUSD's lowest performing group is actually performing in the top half of districts statewide.

Our analysis will look at **comparative results** to identify areas of **strength and challenge**.

How are Performance Levels working with current test Precision

Chart of Standard Errors of Measurement

District Level ELA Score: +/- 4 to 5 pts	District Level Math Score: +/- : 3 to 5 pts
School Level ELA Score: +/- 6 to 12 pts	School Level Math Score: +/-: 6 to 14 pts
Student level ELA Score: +/- 21 to 47 pts	Student Level Math Score: +/-: 21 to 76 pts

Precision is not reported publicly but the standard error of measurement often overlaps adjacent performance levels and dashboard cells. Our control for imprecision is to use scaled scores rather than performance levels and to validate with multiple measures.

Grade	Not Met	Nearly Met	Met	Exceeded
3	2189-2380	2381-2435	2436-2600	2601-2621
4	2204-2410	2411-2484	2485-2548	2549-2659
5	2219-2454	2455-2527	2528-2578	2579-2700
6	2235-2472	2473-2551	2552-2609	2610-2748
7	2250-2483	2484-2566	2567-2634	2635-2778
8	2265-2503	2504-2585	2586-2652	2653-2802
11	2280-2542	2543-2627	2628-2717	2718-2862

2265 ±76
2643 ±23
2576 ±27
2698 ±21
2511 ±31
2503 ±32
2733 ±22
2445 ±37
2514 ±31
2486 ±33
2288 ±64
2385 ±43
30
38
40
34
2365 ±45

LEVEL	Declined Significantly by more than 15 points	Declined by 3 to 15 points	Maintained Declined by less than 3 point or increased by less than 3 points	Increased by 3 to less than 15 points	Increased Significantly by 15 points or more
Medium 25 points below to zero	Yellow (None)	Yellow (None)	Yellow (None)	Green All Students (School District)	Green (None)

Name	Number of Students	Average Scale Score
Morgan Hill Unified (43695830000000)	616	2553 ±5
Jackson Academy of Music and Math (JAMM) (43695836098263)	61	2582 ±12

Example: Jackson Math shows an increase of 7.4 points in math on the dashboard with a standard error of +/- 12 points (reported on ORS). Within the limits of measurement certainly, they fall anywhere within 4 adjacent columns--from declined to increased significantly.

Last year we saw that our **grade level change scores had peaks and valleys** that aligned with county and state, indicating that **test functionality was an issue**. Do we see any of that this year?

State, County, District Change, 2017 to 2018							
Grade Tested	3	4	5	6	7	8	11
State Change	9	7.8	6.8	0	1.5	0.3	-10.1
County Change	6.9	7.2	5.3	-2.8	-0.7	1.2	-10.8
District Change	-2	1.5	5.3	-11.8	-4.8	-7.6	-5.7
District Standard Error	+/- 4	+/- 4	+/- 4	+/- 4	+/- 4	+/- 4	+/- 5
within standard error?	no	no	yes	no	no	no	yes

Similar change scores year to year indicate may **ongoing recalibration** that **affects scores**. For example, the largest gain in District ELA scores is 5.3 points at grade 5...but this gain mirrors the state and county gains.

State, County, District Change, 2017 to 2018							
Grade Tested	3	4	5	6	7	8	11
State Change	3	4.8	4.6	2.5	-0.6	0.3	-3.6
County Change	4.1	3.9	4.7	-2	-2.1	3.8	-1.7
District Change	-2.6	-4	2.9	-9.9	0.6	-4.3	-4.7
District Standard Error	+/- 4	+/- 3	+/- 4	+/- 4	+/- 4	+/- 5	+/- 5
within standard error?	no	no	yes	no	yes	yes	yes

This data suggests our **areas of need** are **early literacy and math** and **middle school ELA**. But is that a reliable conclusion? **This data is year to year change data**, so the student groups in each year are different. This can affect the district more than the state as the district is a much smaller group...

...this is an example of **volatility** due to changing test groups. **What does Cohort growth data** tell us?