Colorado SGP Calculation 2021

Growth & Pandemic Academic Impact

Damian Betebenner & Adam Van Iwaarden

Center for Assessment

Colorado Department of Education

TAP Meeting

September 30th, 2021

Catching Up with the Colorado TAP

- Nice to see you all again.
- What a difference two years make.
- USED granted testing waivers in 2020 due to the onset of the pandemic in March 2020.
  - CMAS testing cancelled in spring 2020.
  - WIDA-ACCESS testing occurred in 2020

Colorado Testing in 2021

- As you are all aware, the pandemic made for a complicated year of testing.
- Accountability waivers were granted for states but testing waivers were not universally granted.
- Colorado had reduced CMAS testing but had regular PSAT/SAT testing and regular WIDA-ACCESS testing.
  - CMAS ELA testing in grades 3, 5, 7
  - CMAS Mathematics testing in grades 4, 6, 8
  - PSAT Mathematics and ELA testing in grade 9 and 10
  - SAT Mathematics and ELA testing in grades 11
  - WIDA-ACCESS testing in grades

Colorado SGP Operational Considerations in 2021

- Colorado CMAS growth analyses are skip-year/gap year growth from 2019 to 2021:
  - ELA – Grade 3/2019 to Grade/5 2021; Grade 5/2019 to Grade 7/2021
  - Mathematics – Grade 4/2019 to Grade 6/2021; Grade 6/2019 to Grade 8/2021.
- Colorado WIDA-ACCESS growth analyses are consecutive year growth from 2020 to 2021.
- In 2021, two types of SGP analyses were conducted for Colorado data:
  - Cohort referenced analyses: These analyses create growth norms from the 2019 to 2021 students. These growth norms yield cohort means/median SGPs of 50 and obscure COVID-19 academic impact.
  - Baseline referenced analyses: Historical growth norms were created using 2019 and prior data. These growth growth analyses don't have mean/median SGPs of 50 and **can** be used to investigate overall COVID-19 academic impact.
- Projections (lagged and straight) were calculated for 2021 data using baseline growth norms in case they were needed.

SGP Summary of Cohort versus Baseline Referenced

- Cohort referenced SGPs
  - Growth norms created from the current cohort of students.
  - Typical way that SGPs are calculated.
  - Mean/median will be 50 for all grade x content area analyses.
  - Cannot investigate overall impact of the pandemic on students.
  - Can investigate differential impact of the pandemic on students.
  - Does not require scale stability across years.
- Baseline referenced SGPs.
  - Growth norms created from historical (i.e., pre-COVID) cohorts of students.
  - Mean/median 50 for historical cohort, not necessarily 50 for current cohort.
  - Extent of deviation from (usually below) 50 is indicative of less learning/pandemic related academic impact.
  - A mean SGP deviation of 5 corresponds to an effect size of 0.18.
  - Can investigate BOTH overall impact and differential impact
  - Requires scale stability across years.

SGP Calculations

- SGP calculations were performed using the SGP package and R software environment based upon data supplied by CDE.
- To facilitate the calculation of baseline SGPs across multiple states, we created an additional R package, SGPmatrices, that contains Colorado matrices, amongst others.
  - https://centerforassessment.github.io/SGPmatrices/
- Source code associated with calculations are available on GitHub:
  - Colorado: https://github.com/CenterForAssessment/Colorado

SGP Assumptions

- SGP analyses (cohort and baseline referenced) presume valid scale scores
  - Students forming the growth norm group are presumed to have taken the same (or equivalent) test at the same time. Scores across students in that instance should mean the same thing.
  - SGPs, however, are often used as a post-hoc check of scale score validity/comparability via *Reductio ad absurdum*
  - In 2021, check, for example, form comparability, mode-of-administration comparability, etc.
  - Colorado has extensive experience (e.g., PARCC paper/pencil versus online) investigating and accommodating mode-effects.
  - If scores are not comparable then we either attempt to make them comparable or run separate analyses.

Scale Score Validity

- Prior to the arrival of 2021 test results (across many states), there were pervasive concerns about ”comparability” of results to previous (e.g., 2019).
- The term comparability is often used in two different ways:
  - Comparability of individual scale scores: Does the scale mean the same thing that it used to.
  - Comparability of aggregate scores (e.g., percent proficient). Is percent proficient (e.g., for a school) comparable to prior years.
- Both issues are of relevance. Note that without comparability of individual scale scores, you lose comparability at the aggregate level.
- We focus initially on comparability of individual scale scores as that it critical for the calculation of baseline referenced SGPs.

Scale Score Validity

- Threats to comparability of individual level scores can occur due to numerous conditions including:
  - Different modes of administration (in person versus remote)
  - Differential motivation of test takers.
  - Equating anomalies (e.g., post-equating of results over the course of the pandemic)
  - Differential test taking conditions including altered/shortened test forms.
- We note that lower participation rates in and of themselves are not sufficient to undermine scale score validity unless these conditions can be shown to somehow impact the meaning of student scores.

Scale Score Validity

- Given plans to calculated baseline referenced SGPs in numerous states, the Center put in place several checks to ensure validity of scale scores.
- A priori Checks:
  - Reviewed results from states/organizations using baseline SGPs
  - Ran sensitivity analyses to examine to what extent perturbed data impact baseline SGP analyses.
  - Examined equating and psychometric reports (e.g., person fit statistics) for 2021 to detect anomalous data.
- Post-hoc Checks:
  - Compared results across states to see whether results conform to what is showing up in other states.
  - Inspected results within state looking for anomalous results which may indicate issues related to scale stability.

Baseline referenced SGPs

- Baseline referenced SGPs have been calculated in several states as well as with interim assessments over the last decade.
- The Georgia Department of Education calculated baseline SGPs and used them as part of their accountability system for 5 years in the early 2010s prior to the assessment transition in 2015.
  - Baseline SGPs demonstrated excellent stability for grade level assessments.
  - Goal of using baseline SGPs was to avoid zero-sum game associated with norm-referenced analyses each year.
- Renaissance Learning uses baseline matrices to calculate SGPs in real time.
- Quantile regression is equivariant to monotone transitions to scale so, in general, robust to minor perturbations of scale.

Baseline referenced SGPs

- We are not aware of any psychometric concerns regarding the validity of scale scores reported by CDE.
- Given that the state reported results to the public in early August, we presume that the psychometrics supported comparisons to 2019 results.
- Baseline growth results are consistent with what has been observed in the dozen states we have examined nationally.
- In the following slides we report the baseline referenced SGPs.
  - We also show 2019 SGPs (baseline/skip-year) to illustrate academic impact.

SGP Analyses and Academic Impact

- Understanding pandemic related impact necessarily requires looking back to pre-pandemic results.
- The impulse of stakeholders to want to compare 2021 results with 2019 results is the correct impulse at both the individual, school, district, and state level.
- Wanting to look at 2019 to 2021 change is an attempt to investigate impact.
- The challenge is making 2019 to 2021 comparisons apples-to-apples comparisons.
- Supporting these comparisons requires having a complete understanding is who constitutes the missing data.
  - This is the second comparability issue discussed earlier.

SGP Analyses and Academic Impact

- 2019 to 2021 Status and Growth comparisons get at the same thing: deceleration in student learning due to the COVID-19 pandemic “headwind”.
- One of the best ”unread” papers on accountability system design and indicators is by Dale Carlson where he discusses the relationship between changes in status (i.e., improvement) and changes in growth (acceleration). https://www.nciea.org/sites/default/files/publications/Dale020402.pdf
- We present (briefly) status results in addition to growth results.

CMAS/PSAT/SAT Attainment: 2019 to 2021 Percent Proficient

Grade	ELA			Mathematics		
	% Proficient 2019 (Count/Participation Rate)	% Proficient 2021 (Count/Participation Rate)	Difference 2021 to 2019	% Proficient 2019 (Count/Participation Rate)	% Proficient 2021 (Count/Participation Rate)	Difference 2021 to 2019
3	41.3% (60,796/95.3%)	39.1% (45,191/74.3%)	-2.2%	41.0% (62,560/95.4%)	----	----
4	48.0% (63,258/95.3%)	---	---	33.6% (64,474/95.4%)	28.5% (46,783/74.0%)	-5.1%
5	48.4% (65,757/94.9%)	47.2% (46,917/72.8%)	-1.2%	35.7% (65,917/95.1%)	---	---
6	43.6% (64,493/93.2%)	---	---	29.5% (64,650/93.4%)	24.1% (44,819/66.9%)	-5.4%
7	46.5% (62,645/91.1%)	42.6% (42,934/62.2%)	-3.9%	31.6% (62,790/91.3%)	---	---
8	46.9% (58,808/87.2%)	---	---	36.9% (58,863/87.3%)	29.5% (39,205/56.7%)	-7.4%
PSAT 9	66.5% (63,043/90.3%)	68.5% (49,811/70.6%)	2.0%	49.6% (63,043/90.3%)	43.0% (49,811/70.6%)	-6.6%
PSAT 10	64.9% (60,724/88.6%)	69.3% (48,413/70.8%)	4.4%	39.1% (60,724/88.6%)	40.1% (48,413/70.8%)	1.0%
SAT	58.6% (57,970/87.6%)	60.0% (51,211/76.0%)	1.4%	39.0% (57,970/87.6%)	36.4% (51,211/76.0%)	-2.6%

Points of Clarification for Status Results

- Decrease in attainment (as a decrease in percent proficient) between 2019 and 2021 is the most often reported statistic related to state testing in 2021.
- This analysis & statistic (percent proficient) are problematic.
- Using 2019 as a counterfactual for where students would have been had the pandemic not occurred is problematic for several reasons including:
  - Changes in student cohort composition is not accounted for.
  - Improvement between 2019 and 2021 that would have occurred is not accounted for.
  - Missing data in 2021 not accounted for.

Points of Clarification for Status Results

- Drops in percent proficient are fairly modest compared to what has been observed in other states (ELA high single digit decreases/mathematics low double-digit decreases).
- Based upon multiple imputation analyses, decreases observed in percent proficient almost certainly underestimate the true decrease in percent proficient had all students tested.
- Participation rates (derived from decreases in observed test scores between 2019 and 2021) vary by grade and state/organization but suggest participation between 60% and 80% (mostly toward the lower end)
- In general, current results based upon incomplete data likely represent a best-case scenario compared to what would appear given fuller participation.

CMAS/PSAT/SAT: 2019 to 2021 Median Baseline SGPs by Grade

Grade	ELA			Mathematics		
	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019
3	NA	NA	NA	NA	NA	NA
4	NA	NA	NA	NA	NA	NA
5	50.0 (57,083/82.4%)	46.0 (41,535/64.5%)	-4.0	50.0 (58,766/84.8%)	---	---
6	50/0 (56,629/81.8%)	---	---	50.0 (57,419/83.0%)	33.0 (40,775/60.9%)	-17.0
7	50.0 (55,057/80.1%)	40.0 (39,035/56.6%)	-10.0	50.0 (55,095/80.1%)	---	---
8	50.0 (51,616/76.5%)	---	---	50.0 (51,637/76.6%)	37.0 (35,647/51.6%)	-13.0
PSAT 9	50.0 (51,120/73.2%)	44.0 (42,889/59.1%)	-6.0	50.0 (48,727/69.8%)	36.0 (42,957/60.9%)	-14.0
PSAT 10	50.0 (47,463/69.2%)	47.0 (35,513/60.8%)	-3.0	50.0 (35,513/51.8%)	50.0 (40,434/59.1%)	0.0
SAT	---	---	---	---	---	---

Discussion

- Impacts in mathematics exceeded those in ELA.
  - Decreases in median SGP in mathematics were, in general, in the lower teens.
  - Decreases in median SGP in ELA were in, in general, the upper single digits. ELA grade 5 is anomalous for being somewhat small.
- The lack of impact in PSAT 10 math is likely due to the fact that the baseline matrices contained a subset of students (PSAT 10/2019 to Math/2017/Grade 8 excluded about 15,000 students who took Algebra I or Geometry in 2017 while in grade 8.
- The results are broadly consistent with those found in the dozen other states we've analyzed.
- Due to the fact that growth is uncorrelated with prior achievement, missing data impacts are much more muted with growth than with status.
- In other states we've classified academic impacts as:
  - > -5: Modest to None
  - -15 to -5: Moderate
  - < -15: Large

WIDA-ACCESS Growth: 2020 to 2021 Median Baseline SGPs

Grade	Composite		
	Median SGP 2020	Median SGP 2021	Difference 2021 to 2020
1	50.0 (8,668/85.8%)	32.0 (7,414/79.5%)	-18.0
2	50.0 (9,120/87.8%)	38.0 (7,613/80.0%)	-12.0
3	50.0 (9,086/89.1%)	32.0 (7,388/80.4%)	-18.0
4	50.0 (8,134/89.2%)	30.0 (6,590/78.3%)	-20.0
5	50.0 (6,975/87.9%)	28.0 (5,357/77.5%)	-22.0
6	50.0 (5,479/84.9%)	32.0 (3,988/69.7%)	-18.0

Grade	Composite		
	Median SGP 2020	Median SGP 2021	Difference 2021 to 2020
7	50.0 (5,219/83.9%)	37.0 (3,964/69.3%)	-13.0
8	50.0 (4,894/83.6%)	38.0 (3,838/70.5%)	-12.0
9	50.0 (4,184/72.7%)	49.0 (2,962/56.3%)	-1.0
10	50.0 (3,586/74.3%)	51.0 (2,424/56.6%)	1.0
11	50.0 (3,020/73.5%)	47.0 (2,007/53.9%)	-3.0
12	50.0 (2,430/63.6%)	51.0 (1,487/43.8%)	1.0

Discussion

- Impacts in lower grades much more severe than those in middle schools.
- Impacts in high school grades are negligible to non-existent.
- Same patterns seen in all other WIDA states we have looked at.
- Based upon a presentation by WIDA at NCSA/CCSSO in June, drops are largely attributable to decreases in scores in speaking.
- Lack of impact in high school grades likely indicates that typical, pre-pandemic growth represented impoverished rates of learning.

CMAS: 2019 to 2021 Median SGPs by Ethnicity

Ethnicity	ELA			Mathematics		
	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019
American Indian/Alaskan Native	45.0 (711)	37.0 (429)	-8.0	49.0 (697)	33.0 (377)	-16.0
Asian	62.0 (3,440)	53.0 (2,382)	-9.0	62.0 (3,535)	44.0 (2,335)	-18.0
Black/African American	44.0 (4,893)	40.0 (2,887)	-4.0	45.0 (4,781)	29.0 (2,508)	-16.0
Hispanic/Latino	46.0 (39,279)	38.0 (27,099)	-8.0	47.0 (39,720)	31.0 (26,766)	-16.0
Pacific Islander	49.0 (240)	41.0 (161)	-8.0	45.0 (260)	33.0 (152)	-12.0
White	52.0 (58,685)	45.0 (44,028)	-7.0	52.0 (55,464)	37.0 (41,139)	-15.0
Two or More Races	51.0 (4,882)	44.0 (3,581)	-7.0	49.0 (4,595)	36.0 (3,141)	-13.0

Discussion of Ethnicity Results

- Broken out by ethnicity, 2021 baseline SGP results show similar rank ordering (i.e., growth gaps) as seen in normal years.
- Student academic growth is lower in mathematics than in ELA for all ethnic subgroups.
- Growth impacts by ethnicity are fairly uniform.

CMAS Growth: 2019 to 2021 Median SGPs by Disadvantaged Status

Free/Reduced Lunch Status	ELA			Mathematics		
	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019
No	53.0 (64,111)	46.0 (50,336)	-7.0	53.0 (62,510)	38.0 (48,114)	-15.0
Yes	46.0 (48,008)	38.0 (30,215)	-8.0	46.0 (46,526)	30.0 (28,285)	-16.0

Discussion of disadvantaged status

- Baseline SGP results show similar growth gaps in 2021 (students indicated as receiving free/reduced lunch have lower SGPs than students indicated as not receiving free/reduced lunch.
- Academic impacts are uniform across the subgroups.

CMAS Growth: 2019 to 2021 Median SGPs by Gender

Gender	ELA			Mathematics		
	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019
Female	55.0 (54,606)	45.0 (38,864)	-10.0	52.0 (52,913)	36.0 (36,362)	-16.0
Male	46.0 (57,534)	41.0 (41,706)	-5.0	49.0 (56,143)	34.0 (40,060)	-15.0

Discussion of Gender Results

- Growth gender gaps historically exist (favoring females) in both ELA and mathematics.
- Growth gender gaps are historically larger in ELA than mathematics.
- Gaps are similarly large in math, but smaller for males in ELA than for females.

CMAS Growth: 2019 to 2021 Median SGPs by IEP Status

IEP Status	ELA			Mathematics		
	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019	Median SGP 2019	Median SGP 2021	Difference 2021 to 2019
IEP No	51.0 (100,268)	43.0 (71,454)	-8.0	51.0 (98,272)	35.0 (68,240)	-16.0
IEP Yes	41.0 (11,872)	39.0 (9,116)	-2.0	43.0 (10,784)	34.0 (8,182)	-9.0

Discussion of IEP Status Results

- IEP status gaps historically exist and favor non-IEP students in both ELA and mathematics.
- Gaps in 2021 were substantially larger for non-IEP students than for IEP students.
- Potential explanation include:
  - Concerted effort to educate IEP students during the pandemic thus minimizing the academic impact to that group of students.
  - Impoverished growth rates pre-pandemic such that pandemic impacts didn’t substantially change rates of learning.

Academic Impact by Prior Achievement

- In examining results associated with SGP fit, we discovered that growth and associated impact varied considerably based upon the prior achievement of the student.
- The following slides illustrate these differences.
- The appendix contains additional figures looking at growth by prior score decile by ethnicity.

Colorado Median/Mean Skip-Year SGPs by Prior Achievement: Grade 5 ELA

Full Population

Pre–Covid (2019) Results

Median SGP: 50
Mean Scale Score: 748
Percent Proficient: 49.7%
SGP Count: 57083

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,684)	9.46	10.09	9.96	10.24	9.87	10.02	9.89	9.94	9.89	10.65	9.4%/50
2nd/[684,703)	9.34	9.90	10.70	9.59	10.60	10.22	10.19	9.64	10.10	9.71	10.3%/49
3rd/[703,717)	9.62	9.82	9.44	9.91	9.24	10.29	10.12	10.18	10.27	11.11	10.2%/51
4th/[717,729)	9.60	10.34	9.81	10.33	10.42	9.47	10.21	9.70	9.79	10.33	9.2%/49
5th/[729,740)	9.33	9.97	10.49	10.02	9.93	10.10	9.85	10.07	9.93	10.32	10.5%/50
6th/[740,751)	9.82	10.46	9.72	9.57	10.17	10.39	9.84	9.55	9.89	10.58	10.2%/50
7th/[751,761)	9.29	9.60	10.21	10.14	10.01	9.36	10.25	10.56	10.09	10.48	9.7%/50
8th/[761,774)	9.65	10.15	9.86	10.09	10.16	10.04	9.79	9.93	10.04	10.29	9.8%/50
9th/[774,791)	9.90	9.89	10.29	9.56	9.97	10.13	9.71	9.97	9.84	10.75	10.8%/50
10th/[791,850]	9.25	9.97	9.88	10.39	9.97	9.70	10.14	10.19	10.09	10.42	10%/50

*Prior score deciles can be uneven depending upon the prior score distribution

Covid (2021) Results

Median SGP: 46
Mean Scale Score: 747
Percent Proficient: 48.2%
SGP Count: 41535

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,688)	5.77	8.25	8.79	9.13	10.08	11.78	11.90	12.34	11.73	10.22	9.9%/57
2nd/[688,706)	9.50	9.78	11.61	10.38	11.10	12.16	9.98	8.47	9.04	7.99	10%/48
3rd/[706,719)	11.38	11.17	12.09	10.51	11.15	10.43	8.95	8.67	7.91	7.73	9.4%/44
4th/[719,731)	11.23	11.39	11.11	9.66	10.11	10.39	9.90	9.71	7.94	8.56	10.1%/46
5th/[731,742)	12.32	11.15	10.60	9.93	9.54	9.93	9.84	9.11	8.74	8.85	10.5%/46
6th/[742,752)	12.65	11.03	10.93	9.87	9.36	9.00	9.77	9.49	9.33	8.56	9.4%/46
7th/[752,763)	12.23	10.57	11.38	9.61	10.45	10.09	8.94	8.70	9.25	8.78	10.1%/45
8th/[763,776)	12.73	10.95	10.47	9.91	11.23	10.97	9.52	9.77	7.59	6.85	10.4%/45
9th/[776,793)	13.47	11.36	9.95	11.36	10.74	9.58	9.95	9.48	7.77	6.33	9.7%/43
10th/[793,850]	15.48	13.86	11.52	11.81	10.76	8.98	7.60	8.20	6.77	5.02	10.5%/37

*Prior score deciles can be uneven depending upon the prior score distribution

Covid Impact (2021 – 2019) Results

Median SGP Difference: –4
Mean Scale Score Difference: –1
Percent Proficient Difference: –1.5%

Median SGP Difference by Prior Score Decile

Prior Score Decile	Median SGP Difference
1st/[650,688)	7
2nd/[688,706)	–1
3rd/[706,719)	–7
4th/[719,731)	–3
5th/[731,742)	–4
6th/[742,752)	–4
7th/[752,763)	–5
8th/[763,776)	–5
9th/[776,793)	–7
10th/[793,850]	–13

Colorado Median/Mean Skip-Year SGPs by Prior Achievement: Grade 7 ELA

Full Population

Pre–Covid (2019) Results

Median SGP: 50
Mean Scale Score: 746
Percent Proficient: 47.1%
SGP Count: 55057

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,699)	9.58	10.00	9.54	10.38	9.92	9.82	10.20	9.94	9.96	10.68	9.1%/50
2nd/[699,716)	9.48	10.24	10.23	9.64	9.88	10.07	9.57	10.12	9.79	10.99	10.5%/50
3rd/[716,728)	9.86	10.07	9.65	9.82	10.29	9.41	10.54	10.16	9.97	10.24	9.7%/50
4th/[728,738)	9.01	9.78	9.72	9.97	10.31	10.76	9.47	9.96	10.06	10.95	10.2%/51
5th/[738,747)	9.51	10.11	10.36	10.33	9.38	9.34	10.20	10.11	10.20	10.46	10%/50
6th/[747,755)	9.86	9.59	9.70	9.88	9.90	10.48	10.17	10.46	9.99	9.99	9.9%/51
7th/[755,764)	9.03	9.97	10.35	10.09	10.40	9.55	10.16	9.73	9.71	11.01	10%/50
8th/[764,774)	9.28	10.10	9.99	9.70	9.88	10.38	9.61	9.84	10.45	10.78	9.9%/51
9th/[774,788)	9.71	9.95	9.91	9.93	10.14	9.69	10.16	10.03	10.16	10.33	10.5%/50
10th/[788,850]	9.53	9.97	10.04	10.19	9.85	10.22	9.94	10.01	9.69	10.56	10.2%/50

*Prior score deciles can be uneven depending upon the prior score distribution

Covid (2021) Results

Median SGP: 40
Mean Scale Score: 743
Percent Proficient: 43.2%
SGP Count: 39035

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,702)	10.67	11.40	11.69	10.61	10.86	9.69	9.29	10.05	8.83	6.91	9.5%/44
2nd/[702,718)	10.97	11.36	10.92	11.96	10.79	10.84	10.82	9.00	8.12	5.21	9.9%/44
3rd/[718,730)	10.30	11.98	12.50	12.50	11.70	10.14	9.96	7.99	7.37	5.55	9.9%/42
4th/[730,740)	9.93	13.72	12.86	12.34	11.46	10.04	9.47	8.12	6.64	5.42	9.9%/40
5th/[740,749)	10.71	14.97	12.93	12.68	10.84	9.26	8.95	7.04	7.12	5.51	10%/38
6th/[749,758)	12.67	14.08	13.63	11.86	10.47	9.87	8.07	7.72	6.96	4.67	10.2%/37
7th/[758,767)	13.32	13.47	12.95	10.19	10.26	10.31	8.93	8.53	7.10	4.94	10.2%/40
8th/[767,777)	13.68	13.42	12.78	10.55	10.45	10.75	8.06	7.62	6.88	5.80	10%/39
9th/[777,791)	15.79	12.05	12.49	10.65	10.04	9.70	7.43	9.26	7.11	5.47	10.5%/38
10th/[791,850]	17.35	15.02	11.18	9.98	9.90	9.01	9.11	6.83	6.53	5.09	10%/36

*Prior score deciles can be uneven depending upon the prior score distribution

Covid Impact (2021 – 2019) Results

Median SGP Difference: –10
Mean Scale Score Difference: –3
Percent Proficient Difference: –3.9%

Median SGP Difference by Prior Score Decile
–6
–6
–8
–11
–12
–14
–10
–12
–12
–14

Colorado Median/Mean Skip-Year SGPs by Prior Achievement: Grade 6 Mathematics

Full Population

Pre–Covid (2019) Results

Median SGP: 50
Mean Scale Score: 733
Percent Proficient: 29.9%
SGP Count: 57419

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,693)	9.73	9.80	10.17	10.19	10.06	9.65	10.13	9.76	9.80	10.72	9.5%/50
2nd/[693,706)	9.61	10.02	9.94	10.27	9.45	10.60	9.69	10.09	10.09	10.25	9%/50
3rd/[706,718)	9.15	10.02	10.04	9.58	10.22	9.59	10.53	10.15	9.99	10.73	11.4%/50
4th/[718,727)	9.49	9.83	9.81	9.64	10.27	10.42	9.62	9.96	10.18	10.77	9.4%/50
5th/[727,736)	9.49	10.03	10.21	10.41	9.53	9.82	10.34	9.91	10.03	10.23	9.6%/50
6th/[736,745)	9.83	10.18	9.73	10.40	10.19	9.67	9.83	10.02	9.52	10.64	11%/49
7th/[745,754)	9.10	9.46	10.20	9.85	10.47	10.03	10.14	9.96	9.85	10.94	9.7%/50
8th/[754,765)	9.31	9.96	10.07	9.63	9.29	10.49	9.86	10.33	10.84	10.22	10.4%/51
9th/[765,779)	10.05	10.50	9.42	9.72	10.52	9.77	10.15	9.75	9.56	10.57	10%/49
10th/[779,850]	9.19	9.89	10.20	10.32	9.84	9.86	10.04	10.13	10.16	10.37	10.1%/50

*Prior score deciles can be uneven depending upon the prior score distribution

Covid (2021) Results

Median SGP: 33
Mean Scale Score: 727
Percent Proficient: 24.4%
SGP Count: 40775

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,695)	21.57	13.91	11.33	10.05	9.84	9.14	7.53	6.50	5.12	5.00	9.8%/33
2nd/[695,708)	23.89	14.80	11.16	10.39	8.59	7.17	6.69	6.45	5.68	5.17	10.2%/30
3rd/[708,718)	21.00	15.26	10.86	9.88	8.78	7.77	8.01	6.59	6.07	5.77	9.5%/33
4th/[718,727)	19.99	13.17	11.79	9.74	8.53	8.61	8.28	7.07	7.10	5.72	9.6%/35
5th/[727,736)	18.82	12.73	11.74	10.15	9.39	9.66	8.49	6.88	6.29	5.84	9.9%/36
6th/[736,746)	18.93	13.32	11.54	10.29	9.62	8.46	8.53	7.57	6.28	5.46	11%/35
7th/[746,755)	19.18	13.88	12.07	10.11	9.44	9.00	6.77	7.34	5.89	6.31	9.5%/34
8th/[755,765)	21.66	15.18	11.90	10.49	8.35	8.80	7.06	6.18	4.99	5.37	9.7%/31
9th/[765,779)	21.32	15.04	11.20	9.64	8.39	7.88	7.36	6.51	6.06	6.60	10.4%/32
10th/[779,850]	22.50	13.98	11.32	9.09	8.45	7.34	7.37	6.21	7.51	6.21	10.4%/32

*Prior score deciles can be uneven depending upon the prior score distribution

Covid Impact (2021 – 2019) Results

Median SGP Difference: –17
Mean Scale Score Difference: –6
Percent Proficient Difference: –5.5%

Median SGP Difference by Prior Score Decile

Prior Score Decile
–17
–20
–17
–15
–14
–14
–16
–20
–17
–18

Colorado Median/Mean Skip-Year SGPs by Prior Achievement: Grade 8 Mathematics

Full Population

Pre–Covid (2019) Results

Median SGP: 50
Mean Scale Score: 737
Percent Proficient: 37.7%
SGP Count: 51637

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,692)	10.12	9.55	9.91	10.21	10.06	9.62	10.15	9.85	9.81	10.72	9.2%/50
2nd/[692,707)	9.27	9.98	9.95	9.63	10.26	10.51	9.81	10.06	10.25	10.28	10.3%/50
3rd/[707,717)	9.35	10.09	10.17	10.07	9.72	9.59	10.13	10.23	9.72	10.93	9.9%/50
4th/[717,726)	9.45	9.54	10.21	9.75	10.43	9.86	9.71	9.69	10.64	10.71	10.4%/50
5th/[726,734)	9.44	10.63	9.90	10.17	9.50	10.38	10.23	10.07	9.38	10.30	10.1%/50
6th/[734,743)	9.83	9.15	9.59	10.01	10.19	9.43	9.89	10.69	10.57	10.65	9.7%/51
7th/[743,752)	9.27	9.61	10.10	10.12	10.29	10.38	10.04	9.48	9.89	10.82	10.3%/50
8th/[752,761)	9.63	11.03	10.13	9.89	9.78	9.76	9.22	9.98	9.74	10.82	9%/49
9th/[761,776)	9.48	9.61	9.86	10.23	9.77	9.88	10.27	10.30	10.40	10.19	10.6%/51
10th/[776,850]	9.52	9.74	10.02	9.93	10.09	10.15	10.07	10.13	9.83	10.52	10.5%/50

*Prior score deciles can be uneven depending upon the prior score distribution

Covid (2021) Results

Median SGP: 37
Mean Scale Score: 731
Percent Proficient: 30%
SGP Count: 35647

Student Growth Percentile Range

Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	Marginal Proportion/Median SGP
1st/[650,695)	7.20	8.06	11.81	10.95	12.19	10.25	11.54	11.28	9.21	7.50	9.5%/49
2nd/[695,707)	7.49	10.28	12.53	12.45	12.56	10.50	11.42	9.03	7.88	5.85	10.1%/46
3rd/[707,716)	9.85	11.28	13.94	12.22	12.76	11.10	10.27	7.99	5.98	4.62	9.5%/42
4th/[716,725)	11.97	13.44	15.23	12.48	12.24	10.66	7.35	7.00	5.29	4.33	10.5%/37
5th/[725,733)	13.37	16.45	14.40	11.84	11.39	9.19	7.23	7.02	5.21	3.92	9.3%/34
6th/[733,742)	16.94	16.41	14.09	11.18	10.19	9.47	6.70	6.50	4.29	4.23	10.1%/31
7th/[742,751)	18.83	18.50	14.30	10.30	9.66	7.04	6.90	5.57	4.73	4.17	10%/29
8th/[751,762)	20.46	15.74	12.91	10.77	8.93	7.31	6.89	6.73	4.86	5.39	10.6%/30
9th/[762,775)	18.97	15.06	12.26	11.46	9.99	8.46	7.13	6.19	5.44	5.05	10.1%/32
10th/[775,850]	19.17	15.47	10.02	10.95	8.80	8.22	7.57	7.60	6.54	5.66	10.3%/34

*Prior score deciles can be uneven depending upon the prior score distribution

Covid Impact (2021 – 2019) Results

Median SGP Difference: –13
Mean Scale Score Difference: –6
Percent Proficient Difference: –7.7%

Median SGP Difference by Prior Score Decile

Prior Scale Score Decile/Range*	Median SGP Difference
1st/[650,695)	–1
2nd/[695,707)	–4
3rd/[707,716)	–8
4th/[716,725)	–13
5th/[725,733)	–16
6th/[733,742)	–20
7th/[742,751)	–21
8th/[751,762)	–19
9th/[762,775)	–19
10th/[775,850]	–16

Academic Impact by School and District

- Here we present some preliminary finding when examining academic impact at the school and district level.
- Because of the reduced testing in Colorado, results are presented at the school x grade and district x grade level.
- Growth is less impacted by missing data than status, but caution should be exercised in interpretations especially with substantial missing data for the school or district.
- We are finalizing multiple imputation results that can be used to better understand the impact of missing data for schools and districts.

Participation Rates and Interpretation

- An issue impacting the interpretation of growth (and status) results in 2021 is participation in the state assessment.
- Low participation doesn't impact calculation of cohort referenced growth norms nor use of baseline referenced growth norms.
- Low participation does impact comparisons of group level results with historical (e.g., 2019) results based upon higher participation rates.
- For example, how should one compare growth and status results for school in 2019 and 2021 with very different rates of participation.

Multiple Imputation

- Given the rich historical data available for students missing in 2021, and the observed performance of comparable students from 2019 to 2021, it is possible to simulate what we think their performance might have been.
- Combining multiple simulations can inform us about potential bias in what we observed at state/district/school levels in 2021, as well as inform us about levels of uncertainty around them.
- Multiple imputation (MI) of missing data is widely adopted and well-developed statistical analysis technique used in many fields to deal with missing data problems.

Multiple Imputation

- MI uses information from the observed data to generate parameter estimates in three steps:
  - **Imputation**: A prediction model generates a set of plausible values for the missing observations, resulting in *M* imputed data sets.
  - **Analysis**: The analysis (e.g., baseline student growth percentiles) is conducted on each of the *M* data sets.
  - **Pooling**: Parameter and (more importantly) uncertainty estimates are constructed by pooling across the *M* analyses.

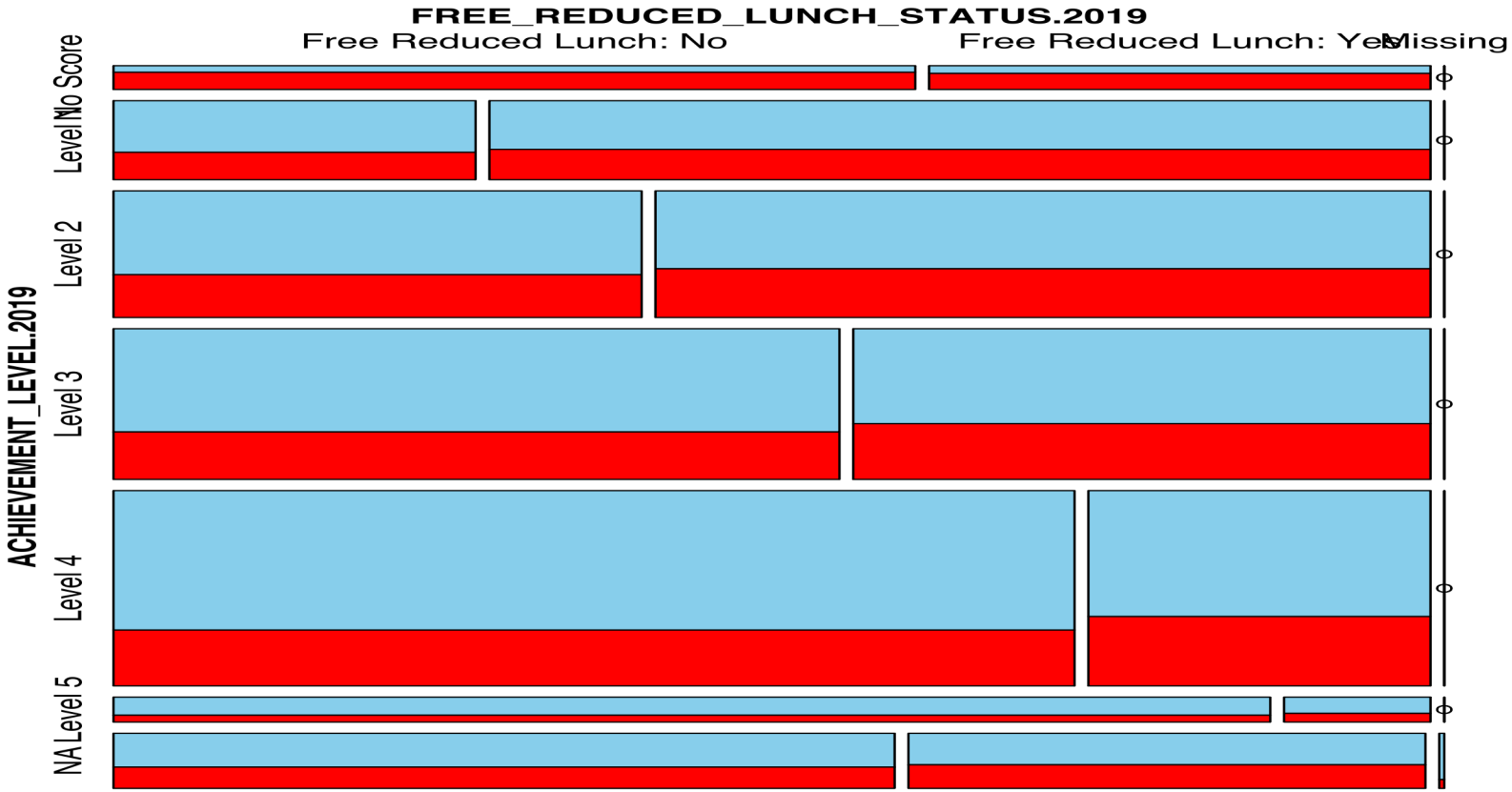
Multiple Imputation

- In the context of COVID impact analyses, we use the multiple (plausible) parameter estimates to inform us about the uncertainty around what was observed. That is, we are more interested in using MI to determine whether we trust the statistics calculated from the observed (with missing) data than to produce "corrected" parameter estimates.
- Statistics of particular interest:
  - Mean/median scale scores
  - Mean/median baseline SGPs
  - Effect size estimates
  - Percent Proficient

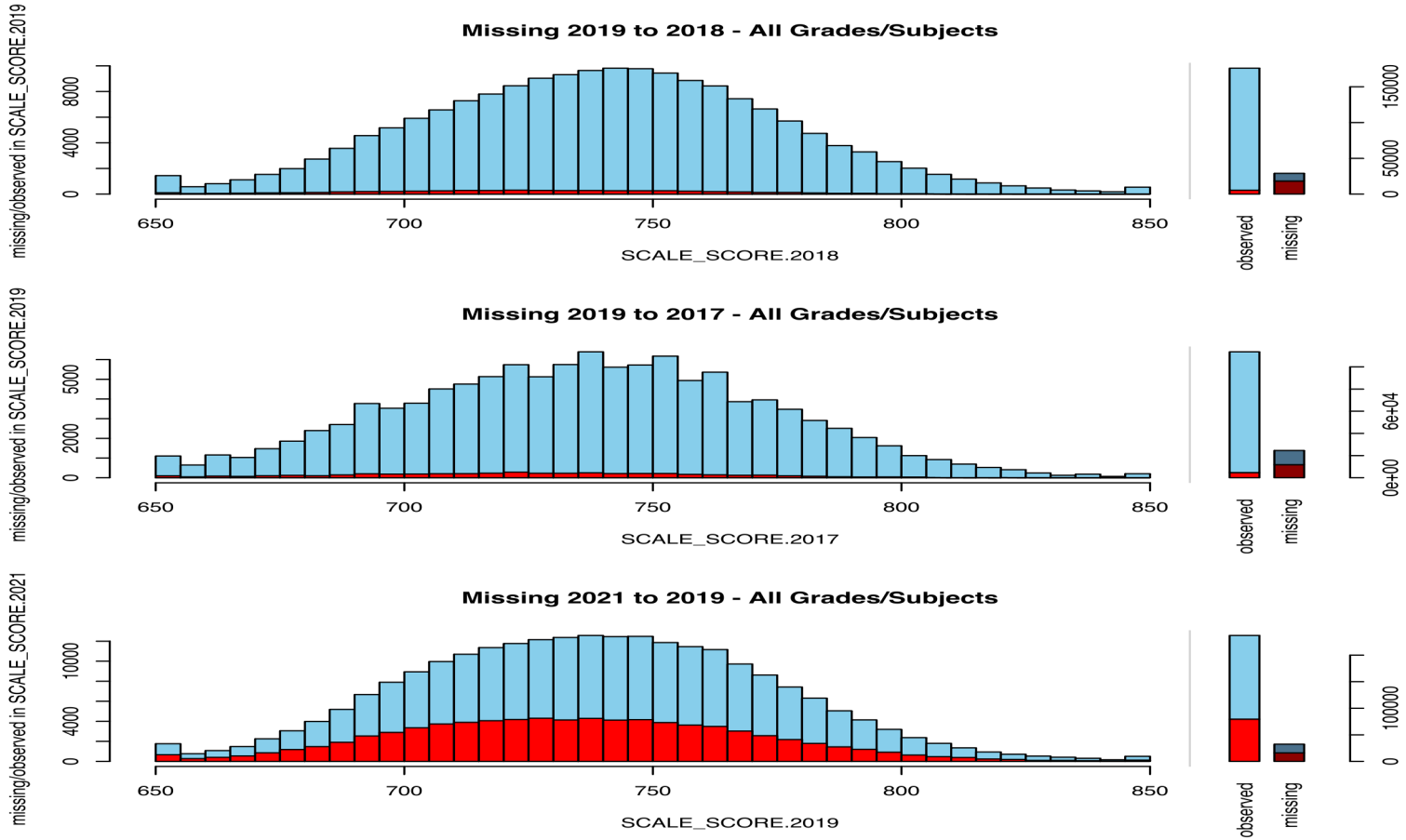
Multiple Imputation

- Before performing a MI analysis with Colorado data, missing data patterns were investigated.
- Patterns suggest some relationship to factors such economic status and ethnicity.
  - The relationship with prior achievement is relatively small.
- These relationships suggest that missingness in Colorado is non-ignorable, but does not preclude the use of MI.
  - Data can be considered “Missing-at-random” – relationships exist between missingness and the observed data (and so can not be considered “missing completely-at-random”) but the relationships are not systematic (e.g. all students below proficiency in 2019 were missing in 2021).

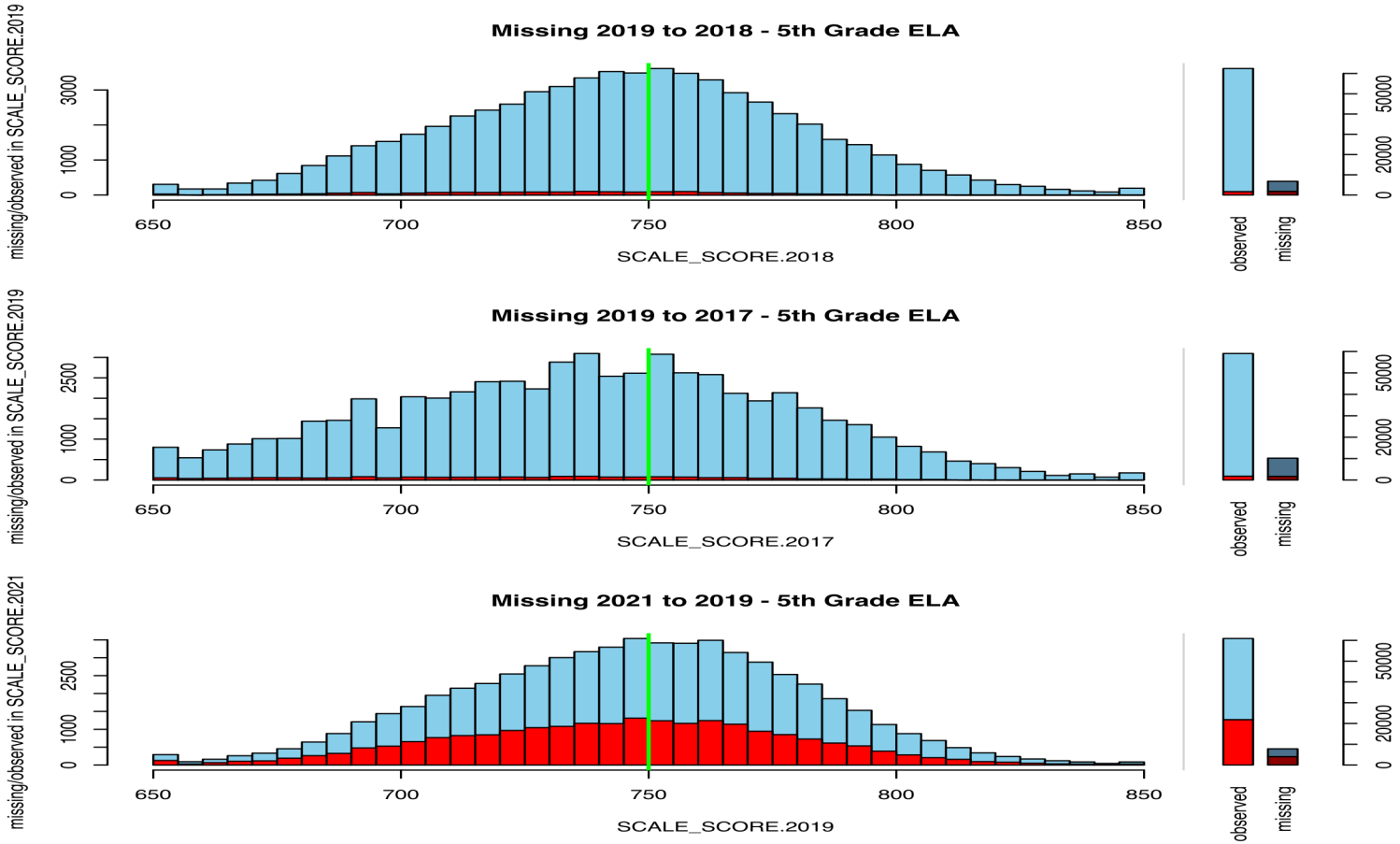
Missing Data Patterns



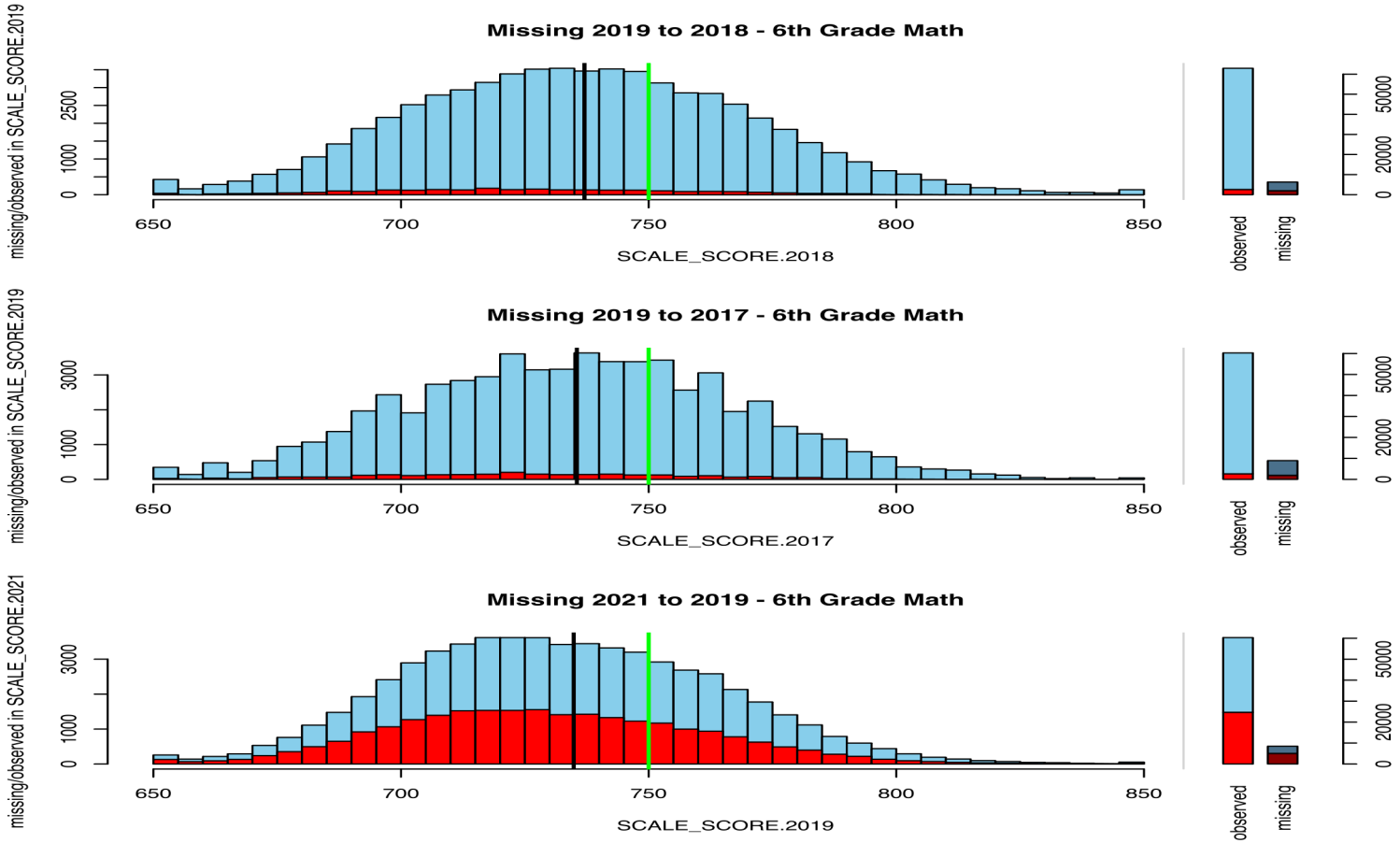
Missing Data Patterns



Missing Data Patterns



Missing Data Patterns



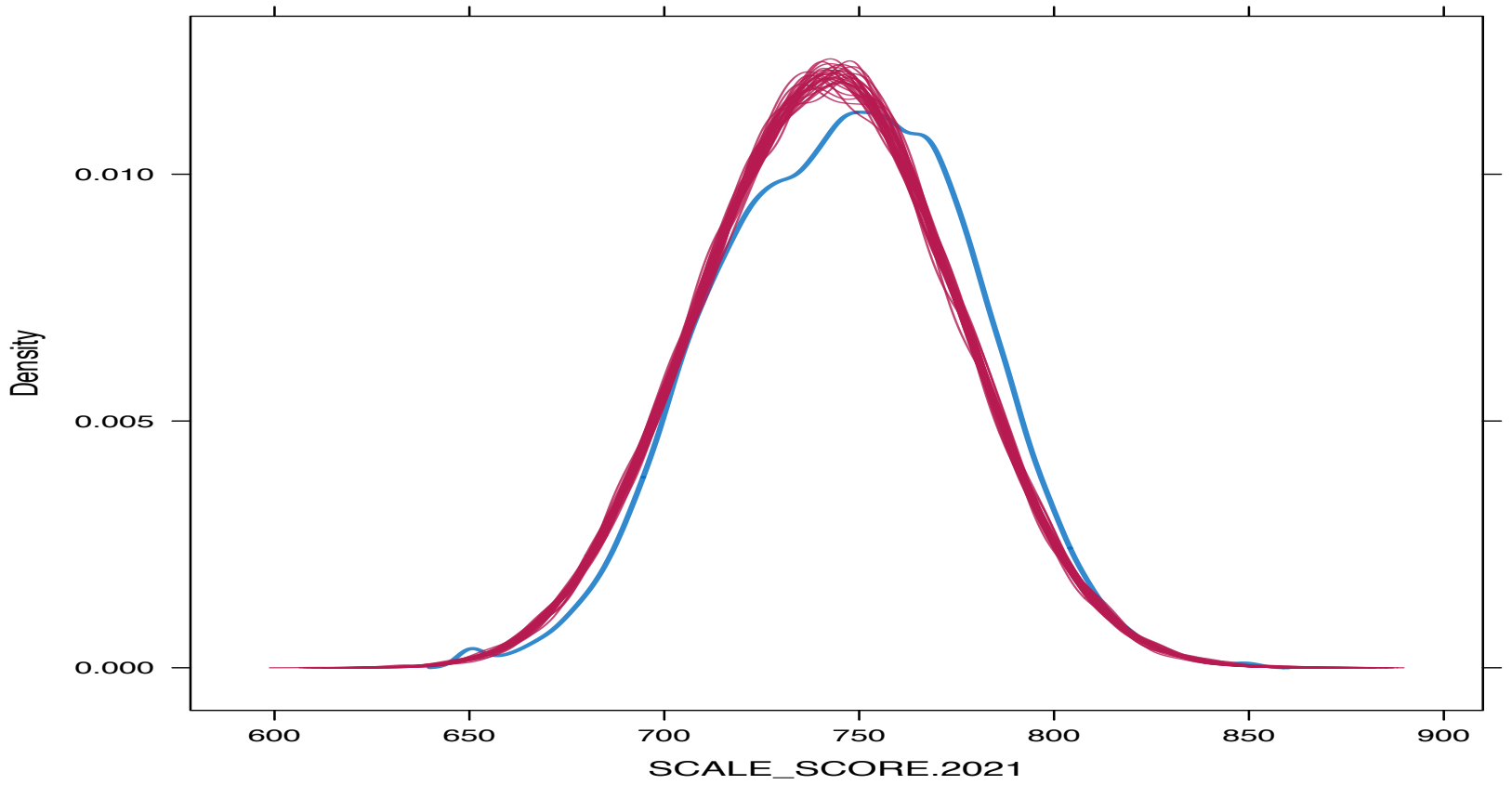
Multiple Imputation

- Given the 2021 missingness patterns and the good coverage of observed scores for students from all demographic groups and prior achievement levels, MI is a reasonable approach to dealing with CMAS missing data.
- Imputation models included:
  - Student prior achievement
  - Student demographics
    - FRL Status
    - Disability status
    - English language proficiency status
    - Gender
    - Ethnicity dummy variables for White, African American, Hispanic and Asian students
  - School level clustering
- 30 imputations (*M* = 30 with a maximum of 10 iterations per imputation) were conducted.

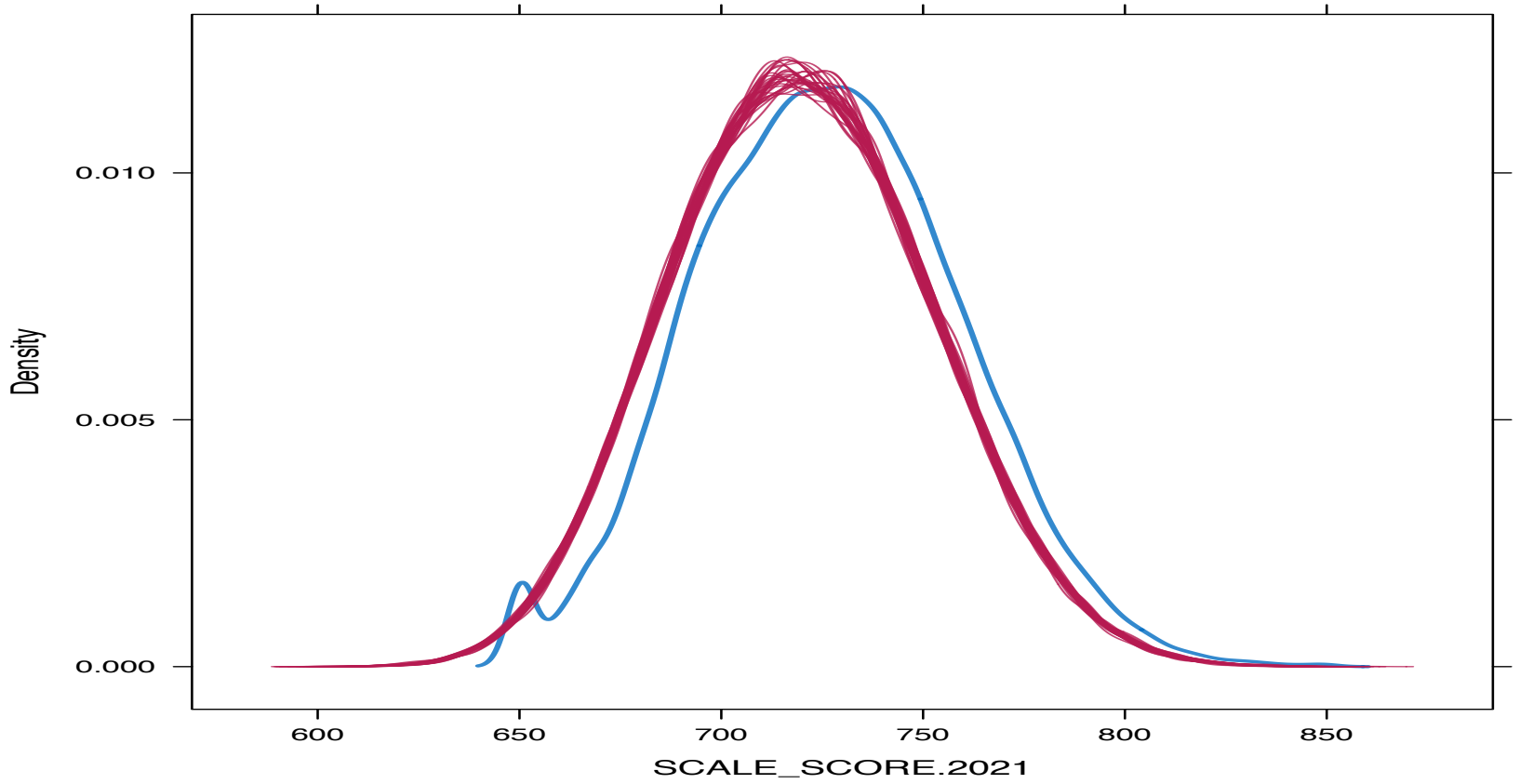
Multiple Imputation

- Imputation models were restricted to include only students who were enrolled, but not tested.
  - Knowing school/district affiliation is critical for assessing the impact of missing data for subsequent analyses at these levels.
  - Excluding students who “disappeared” from the system may not give us a full account of the impact of missing data at the state level. However, a comparison of the enrolled vs. non-enrolled missing students does not lead us to believe their inclusion would meaningfully change inferences at the state level.

Multiple Imputation



Multiple Imputation



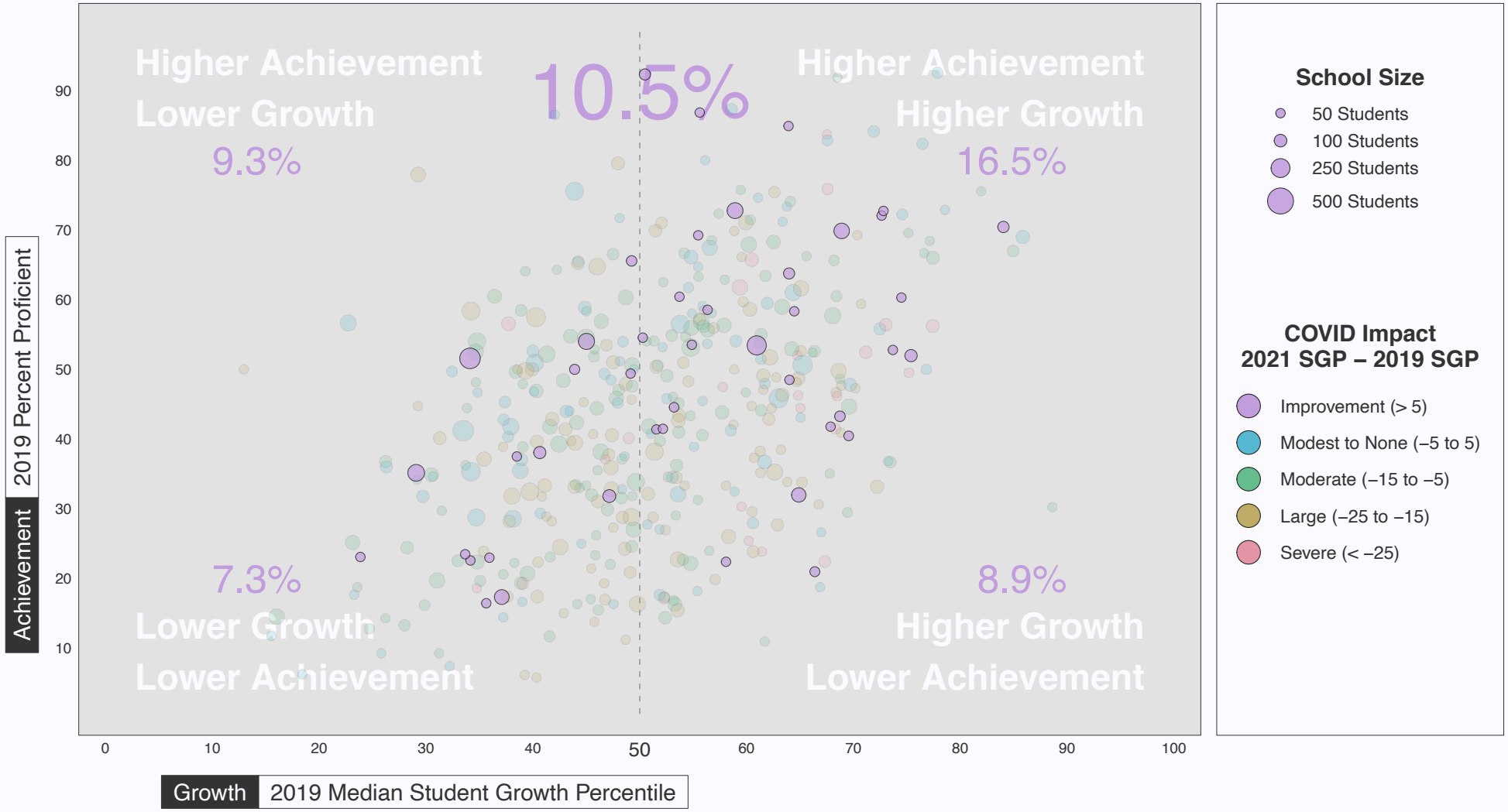
District & School Level Results

- Due to larger fluctuations in participation rates at the district and school level, growth comparisons between 2019 and 2021 must be approached more cautiously.
- In general, in other states we have had good success via multiple imputation that comparisons of status and growth results from 2019 to 2021.
- Multiple imputation indicates both when inference is defensible based upon observed data and when such an inference appears problematic.
- We have found that whether a school's results support cross-year comparison depends more on the size of the school than on the participation rate/enrollment change.
- With coming enrollment data to determine missingness, we will update results to include district and school level impact.

CMAS: Grade 7 ELA

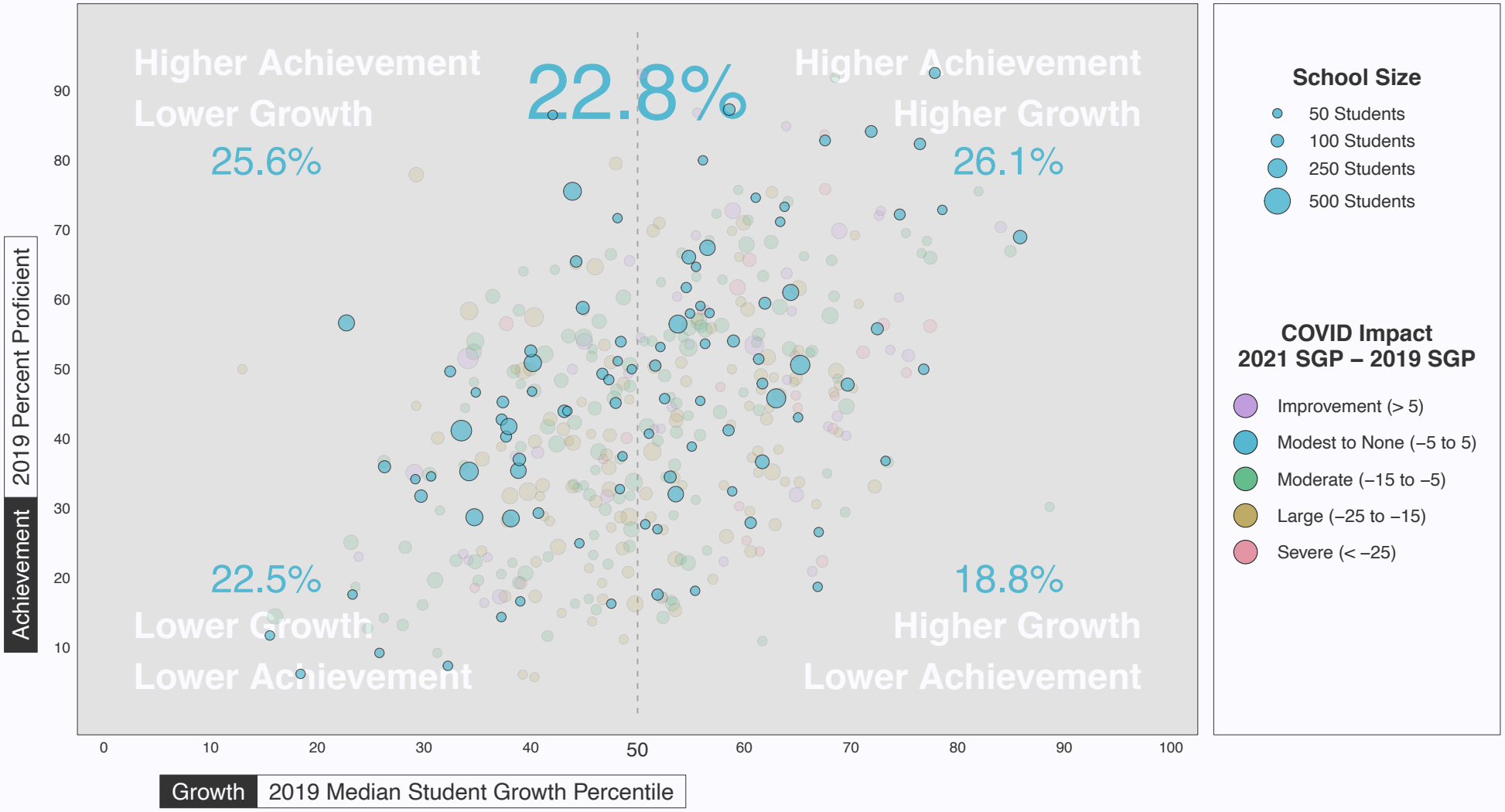
Grade by School Level COVID Academic Impact

2019 Growth by Achievement



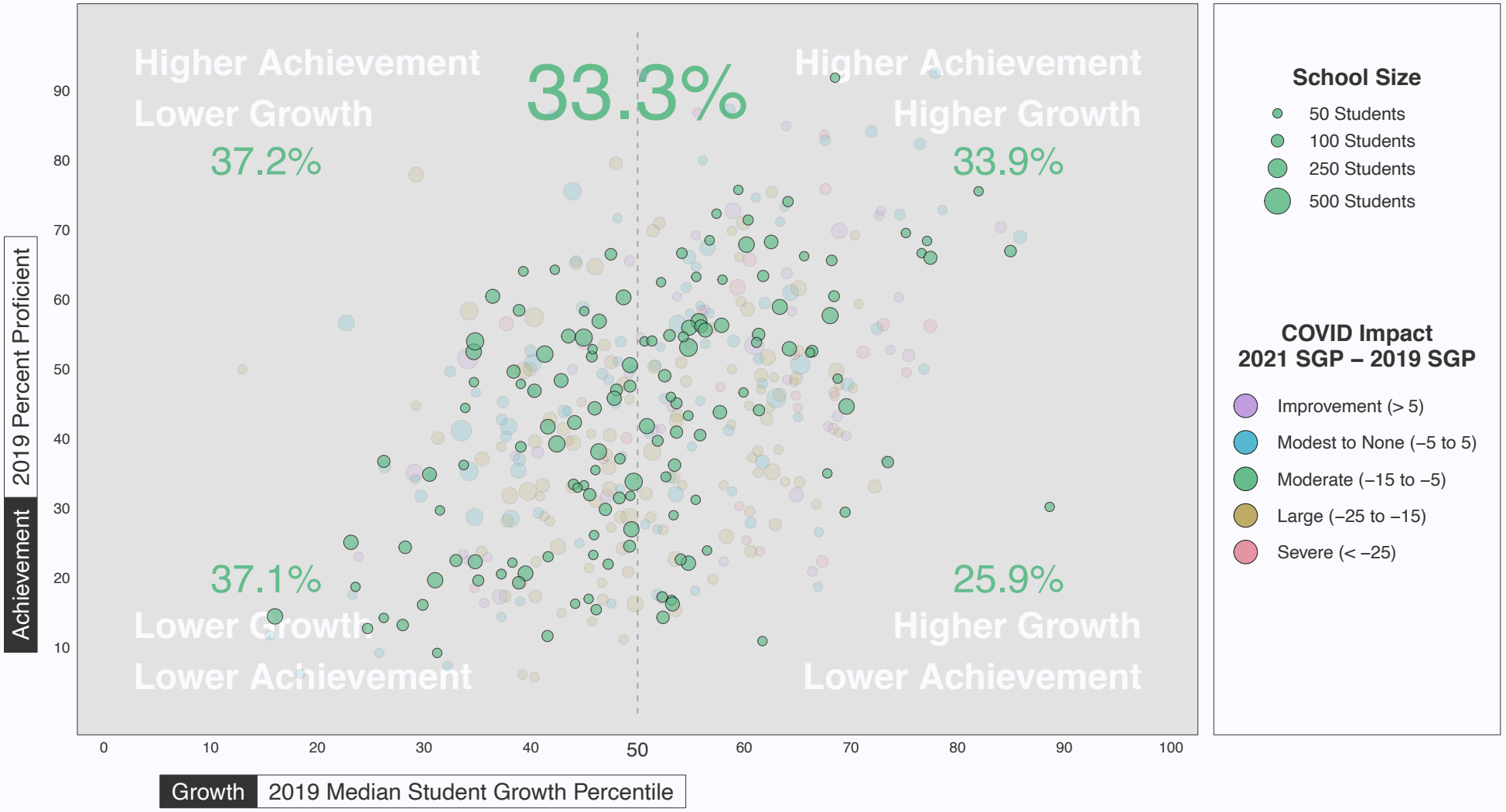
CMAS: Grade 7 ELA

Grade by School Level COVID Academic Impact
2019 Growth by Achievement



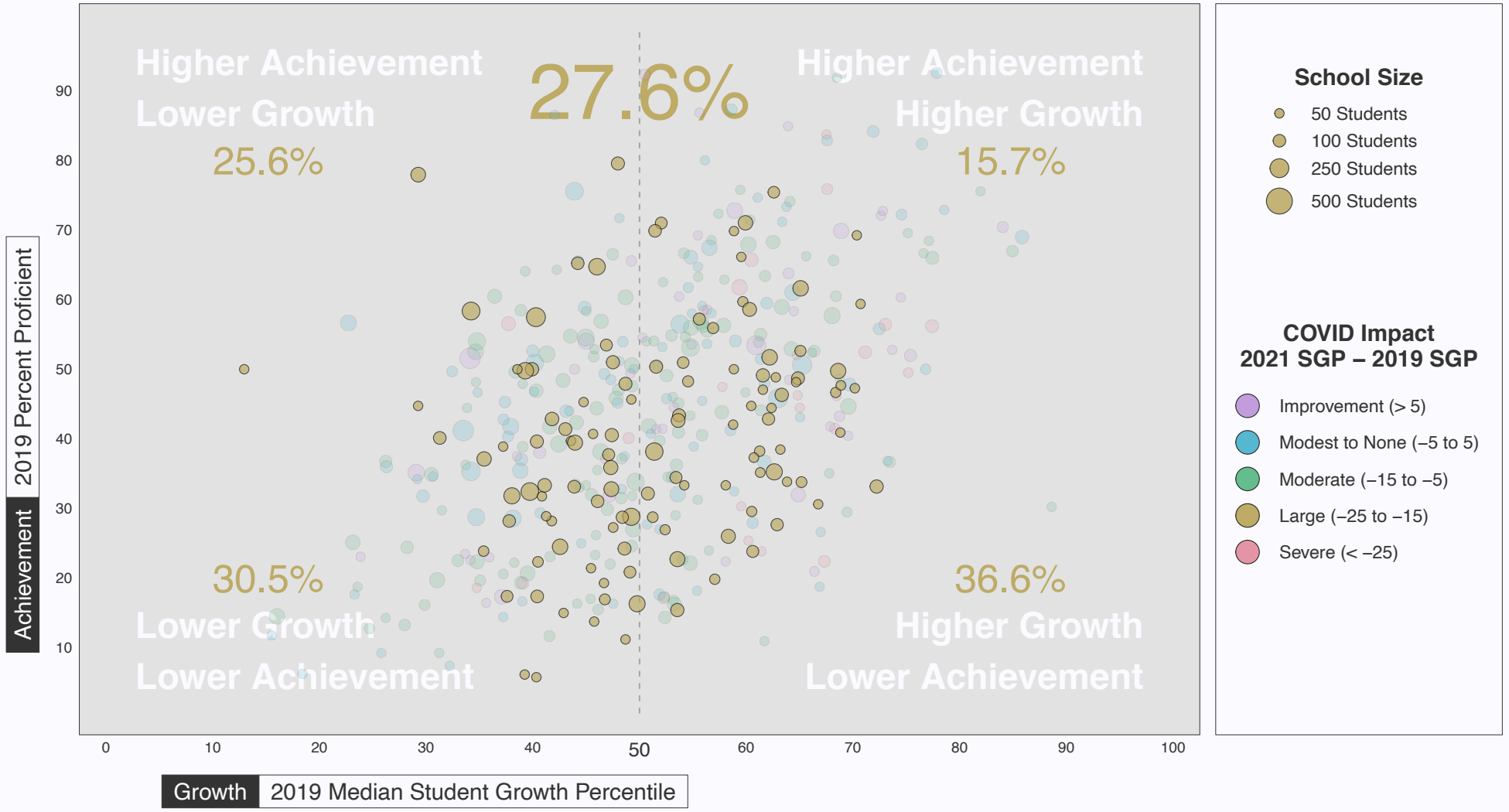
CMAS: Grade 7 ELA

Grade by School Level COVID Academic Impact
2019 Growth by Achievement



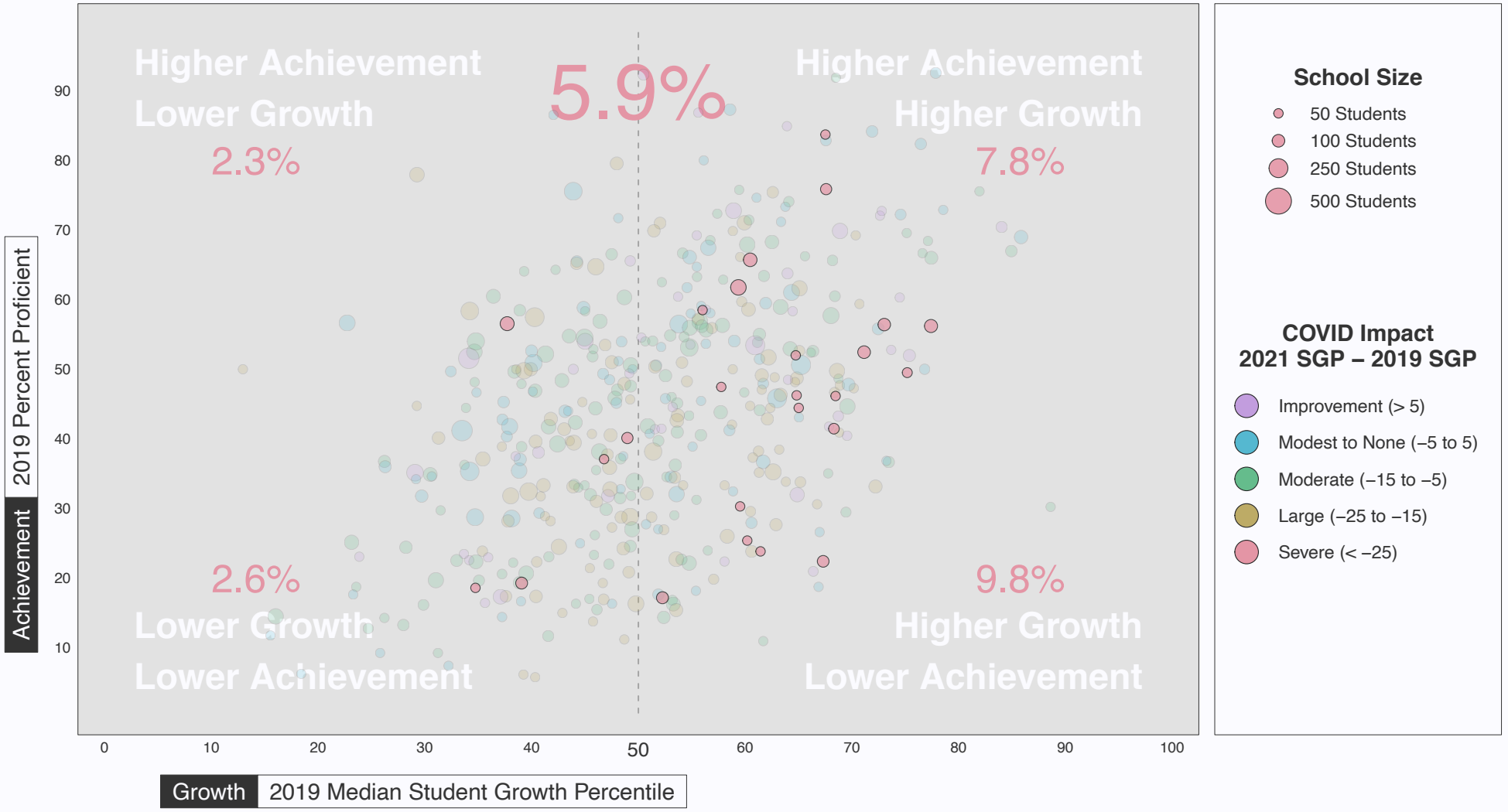
CMAS: Grade 7 ELA

Grade by School Level COVID Academic Impact
2019 Growth by Achievement



CMAS: Grade 7 ELA

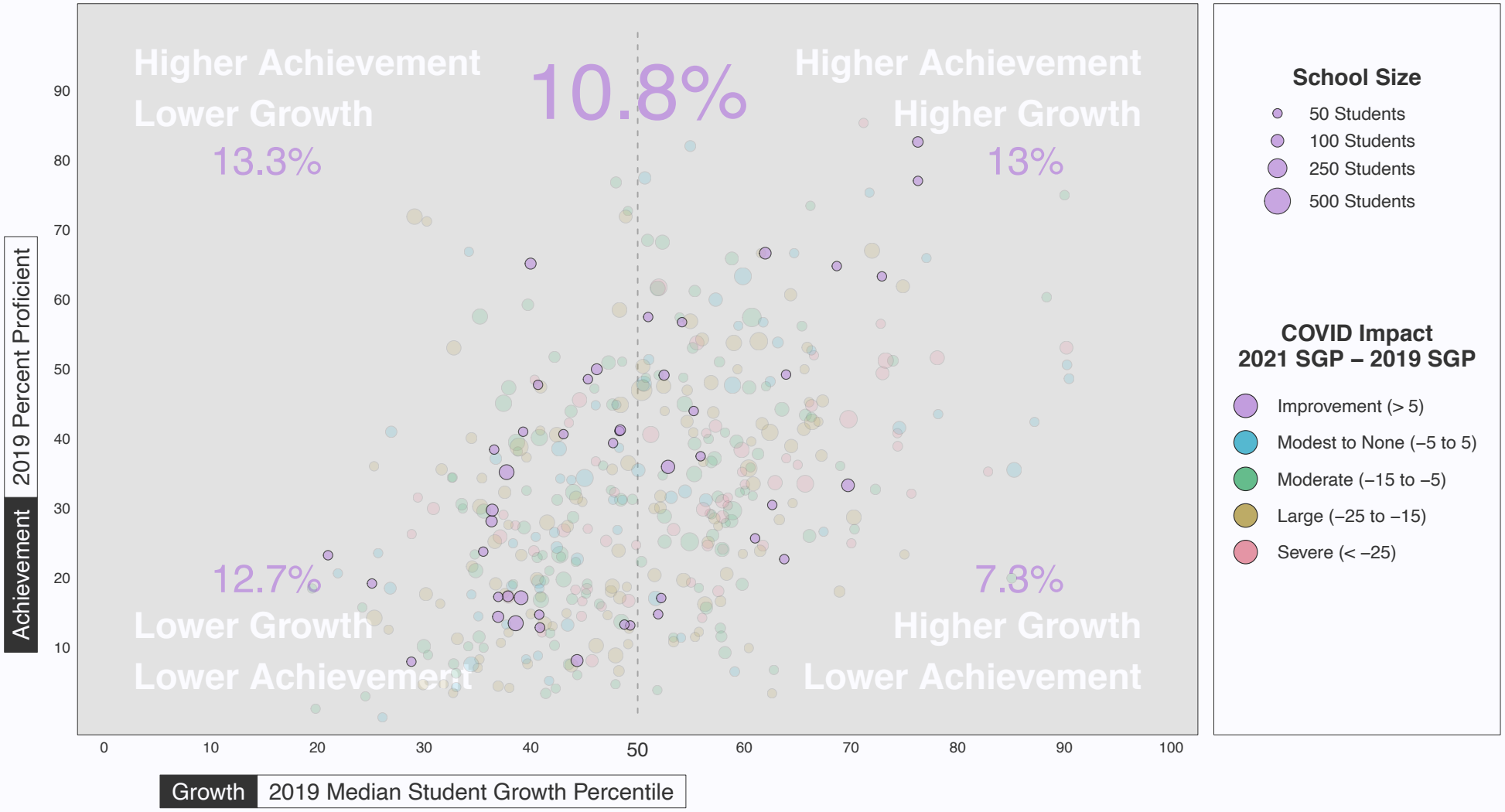
Grade by School Level COVID Academic Impact
2019 Growth by Achievement



CMAS: Grade 8 Mathematics

Grade by School Level COVID Academic Impact

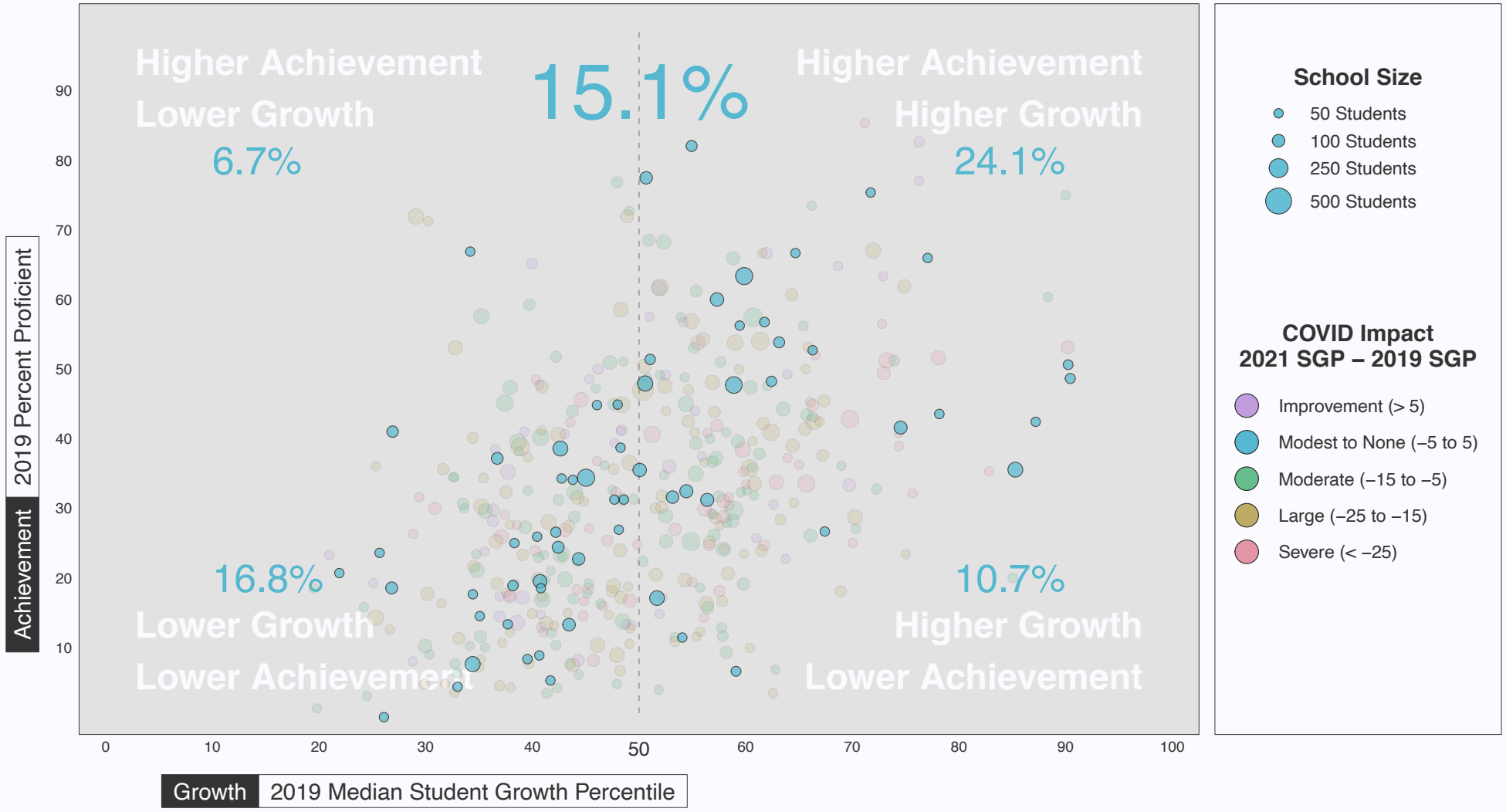
2019 Growth by Achievement



CMAS: Grade 8 Mathematics

Grade by School Level COVID Academic Impact

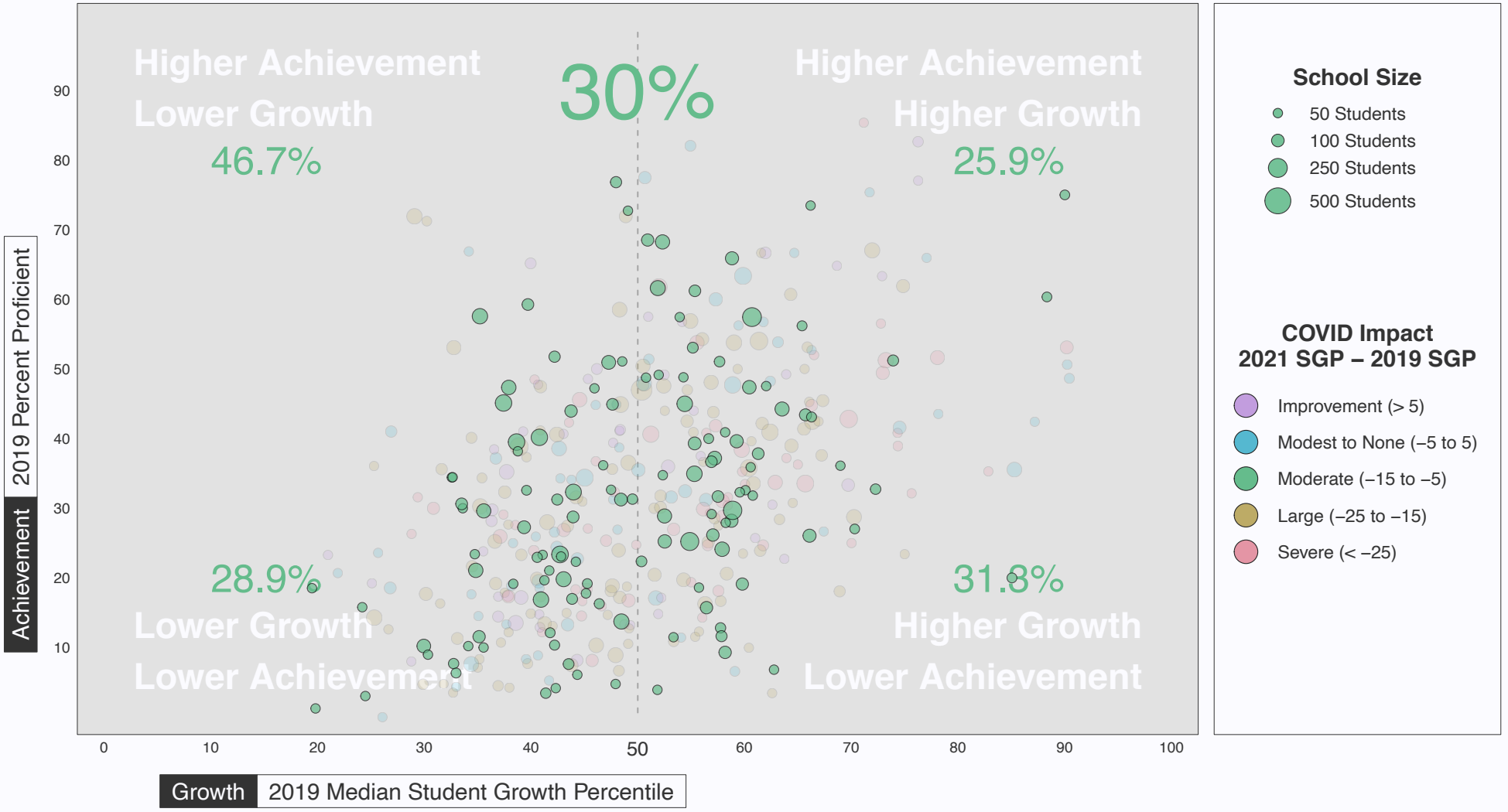
2019 Growth by Achievement



CMAS: Grade 8 Mathematics

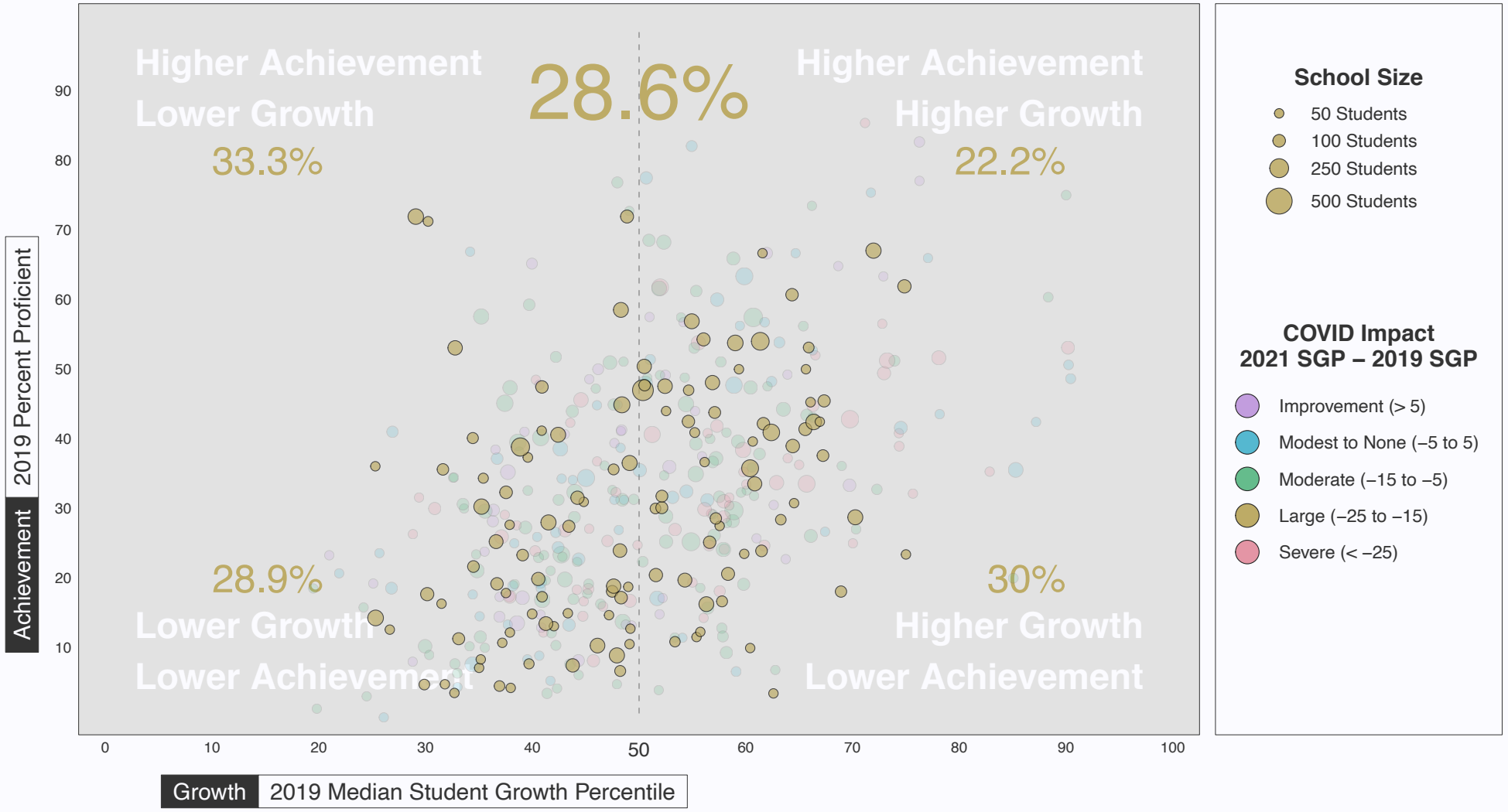
Grade by School Level COVID Academic Impact

2019 Growth by Achievement



CMAS: Grade 8 Mathematics

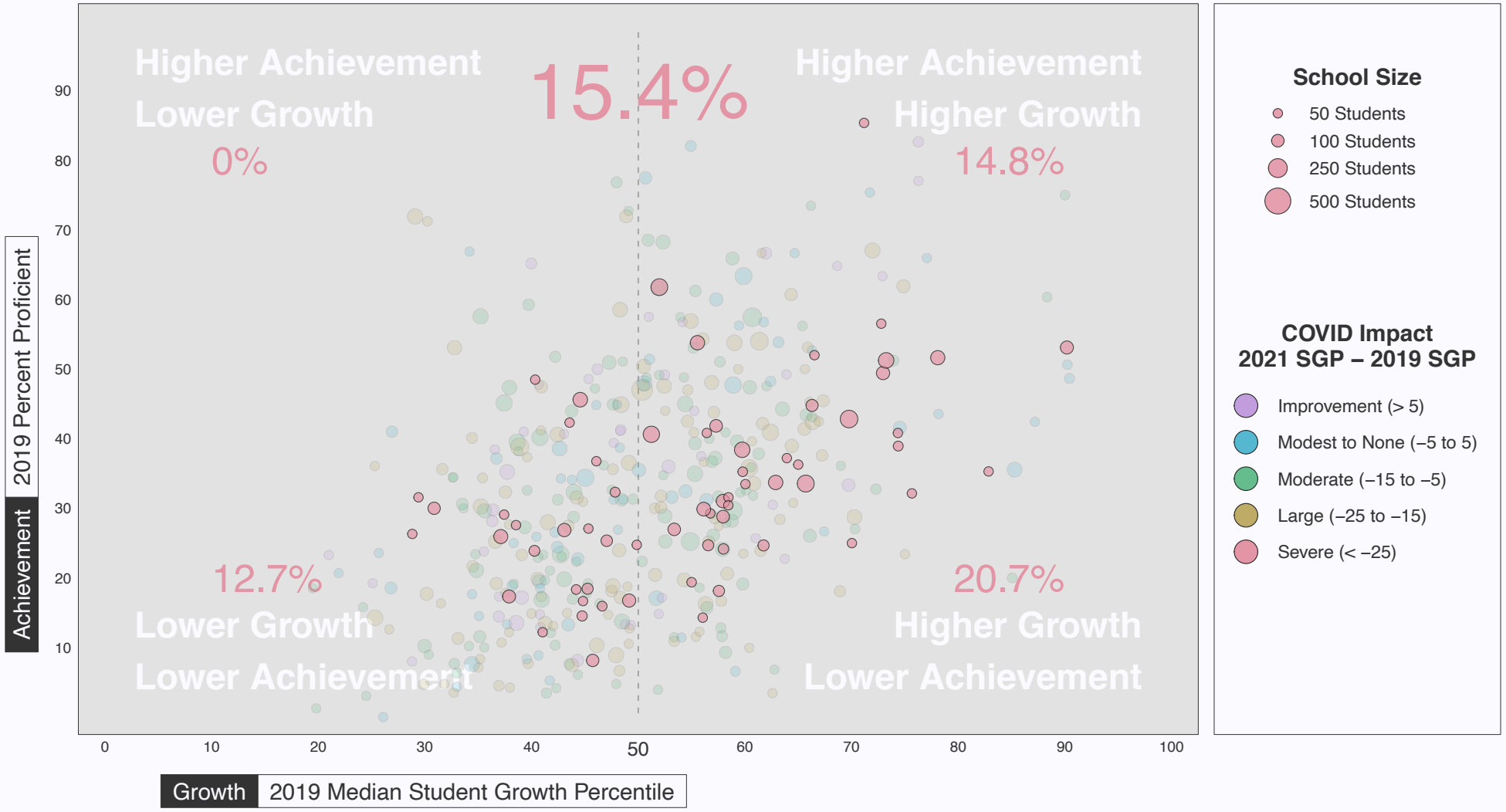
Grade by School Level COVID Academic Impact
2019 Growth by Achievement



CMAS: Grade 8 Mathematics

Grade by School Level COVID Academic Impact

2019 Growth by Achievement



Communication of Results in 2021

- Technical challenges to the calculation of growth in 2021 are likely minor compared to communication challenges.
- State reporting is, in general, highly aligned with accountability.
- Even though analyses outlined herein deal with status/growth, they don't necessarily align with accountability narratives.
- It is critical for leadership (at the highest level) to map out a communication plan and then build analytics and reporting to support that plan.
- Without a well considered, disciplined and coordinated communication plan, accountability narratives are likely to take over.

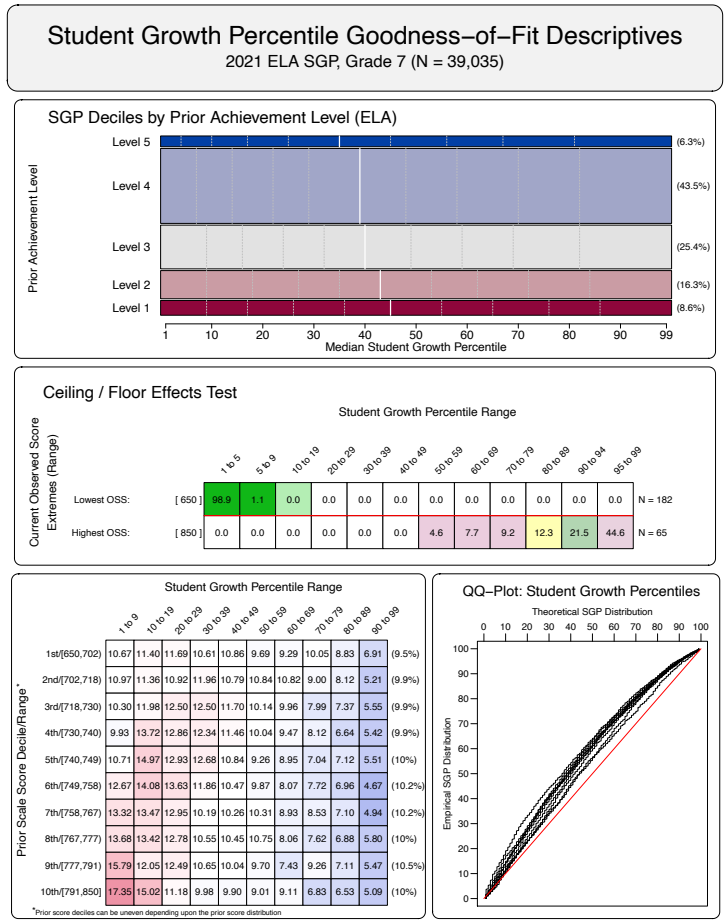
SGP Analysis Plans post 2021 (2022 and beyond)

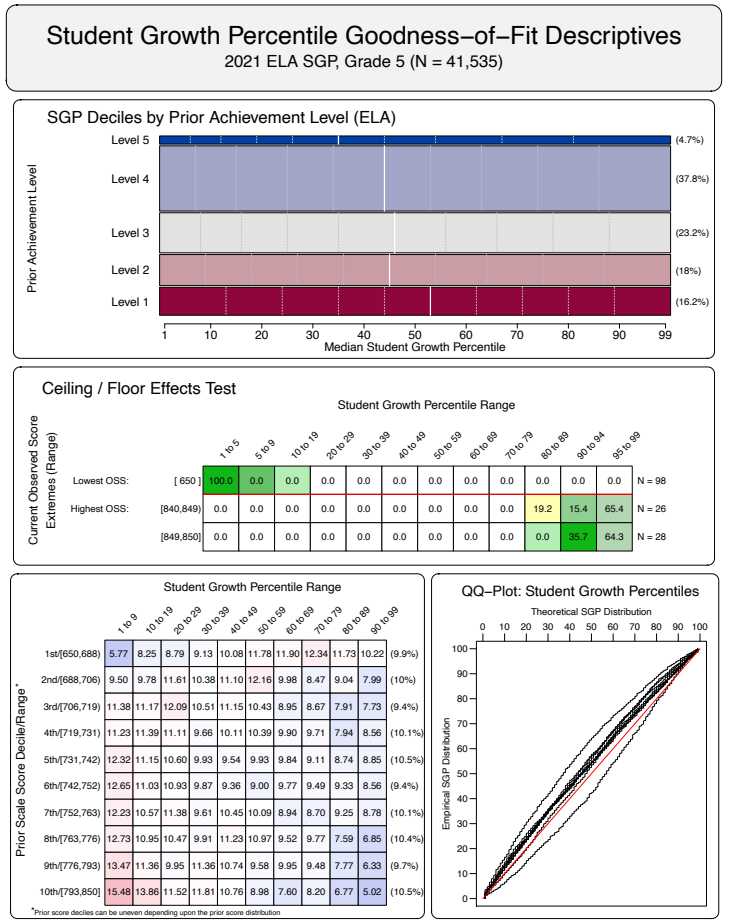
- Spring 2021 represented an ideal time to get a snapshot of student attainment as we are presumably/hopefully at or near the point of maximal impact on student attainment.
- That is, recovery has not really begun for most students in earnest.
- It's impossible to talk about recovery sensibly without understanding how big the impact is.
- Analytics in 2022 will be keenly interested about whether recovery is occurring: Who, what, how much.
- Because of CMAS content x grade coverage in 2021, growth will be limited in 2022.
- Monitoring recovery will be at least as critical as monitoring the impact of the pandemic on student achievement.
- Beware the K-shaped recovery!

Summary of Findings

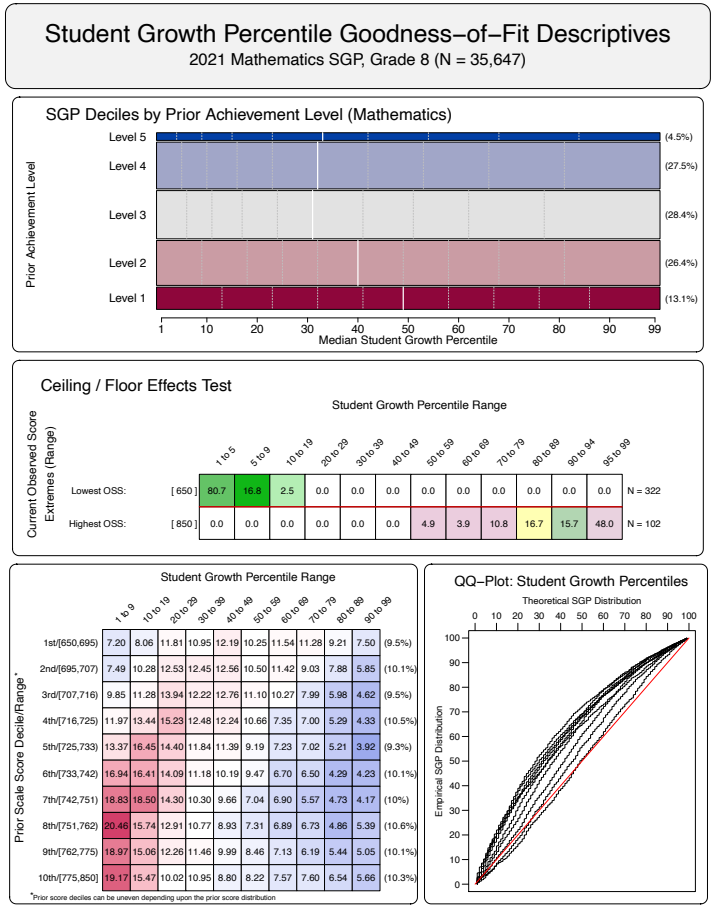
- SGP analyses were successfully conducted for Colorado CMAS/PSAT/SAT/WIDA-ACCESS testing in spring 2021
- For the CMAS, results of the analyses suggest moderate academic impacts in ELA and large academic impacts in mathematics due to the pandemic.
- For WIDA-ACCESS, academic impact was large in elementary grades, moderate in middle school grades, and no academic impact was indicated in the high school grades (likely due to the fact that pre-pandemic growth was already impoverished for high school students).
- Though CMAS testing is half of what it usually is, results suggest uniform impactby content area across elementary, middle and high school.
- Results show variation across academic subgroups (e.g., high versus low achievers in some content areas).
- Recovery from impacts this large will likely take multiple years and require additional supports beyond that provided by normal schooling.

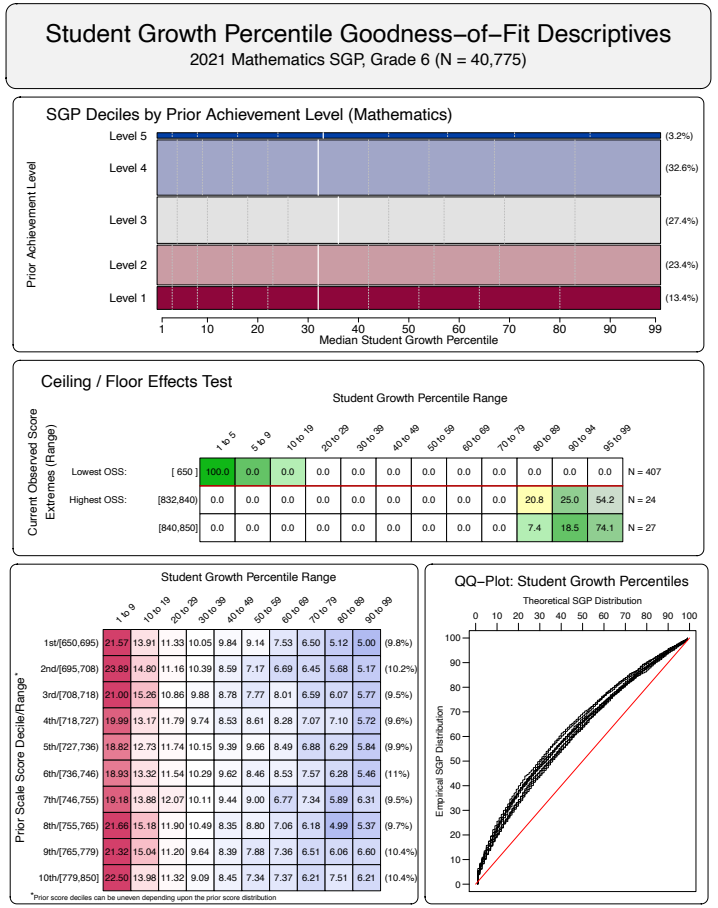
SGP Goodness of Fit Plots: Colorado ELA



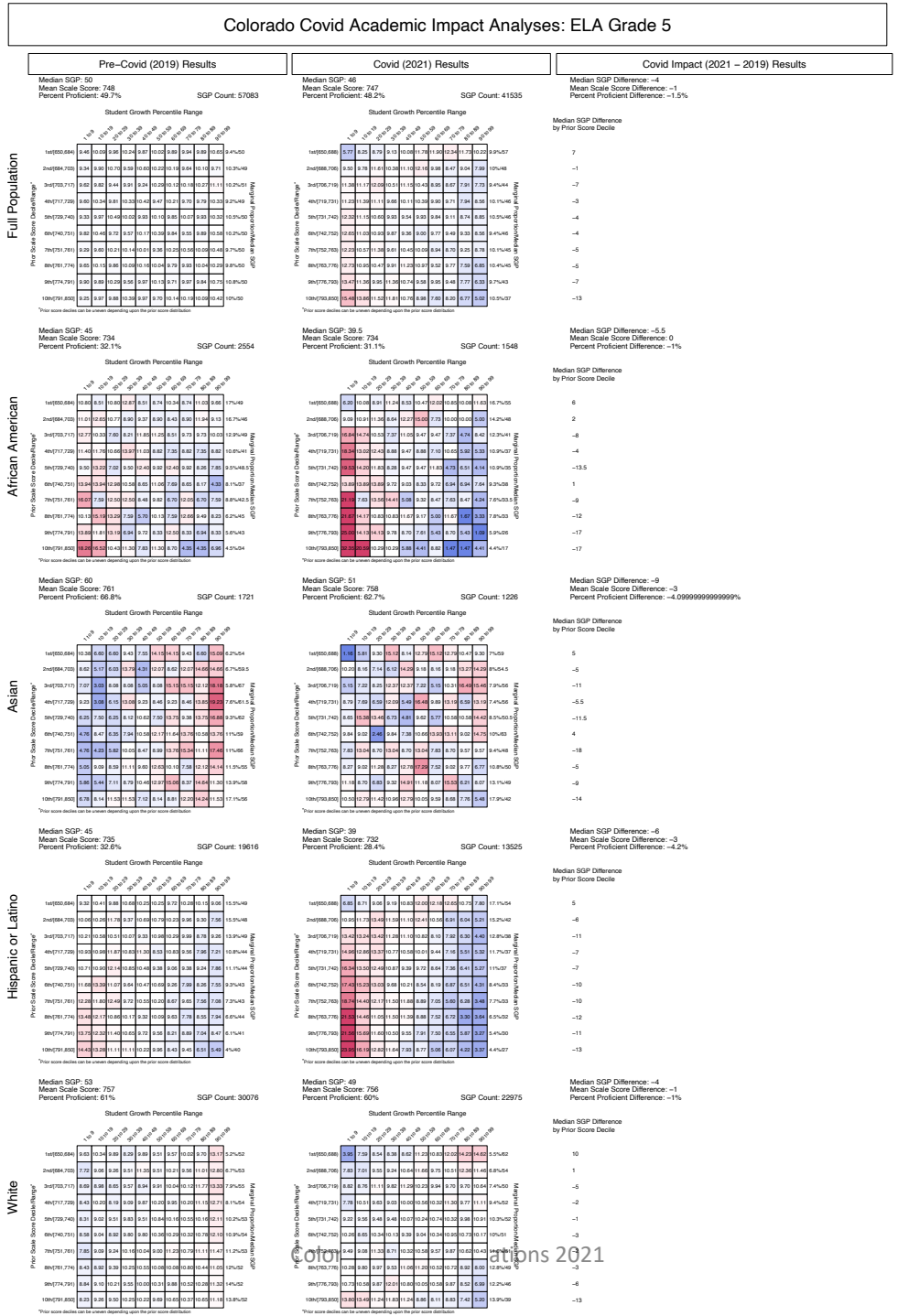


SGP Goodness of Fit Plots: Colorado Mathematics

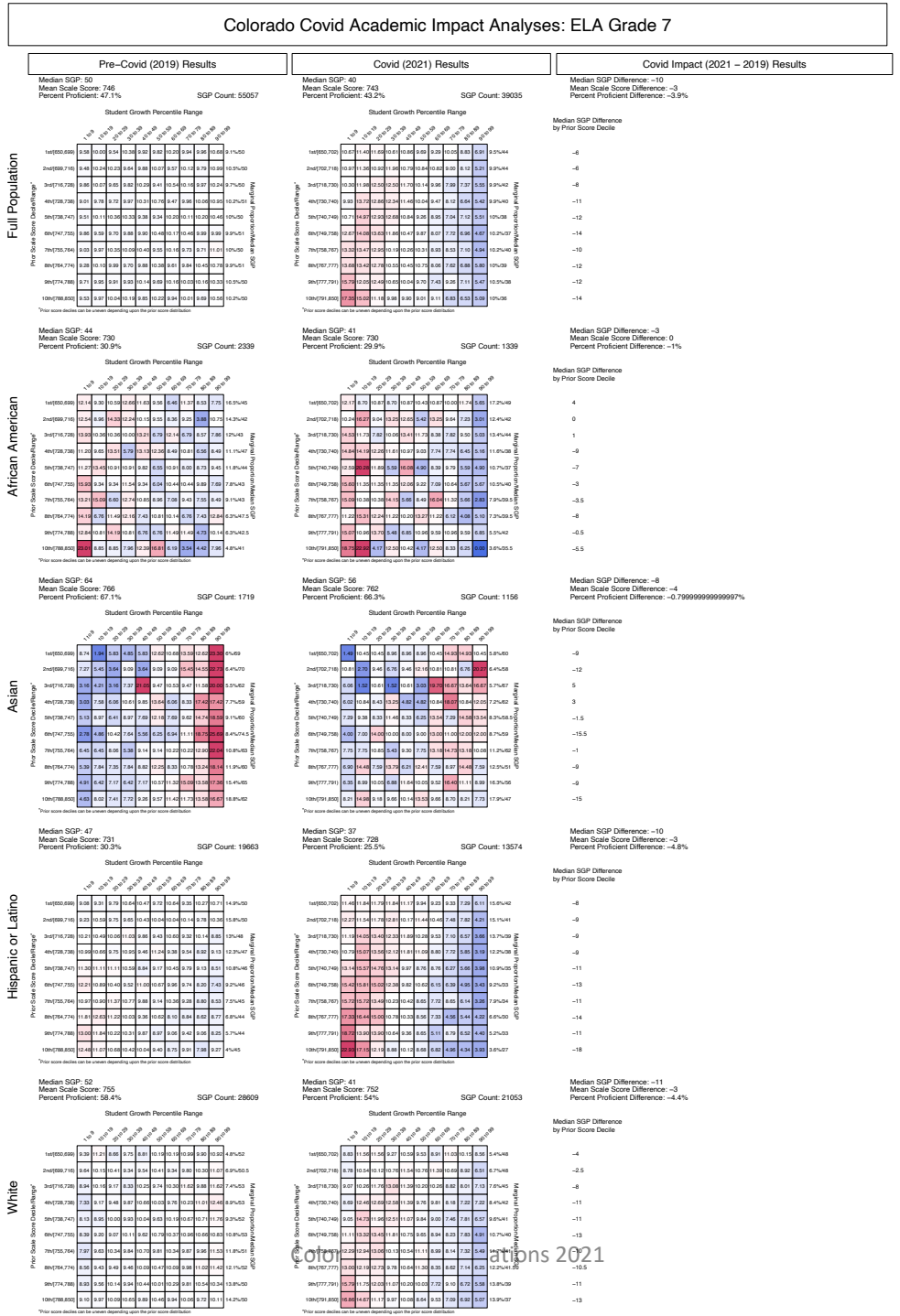




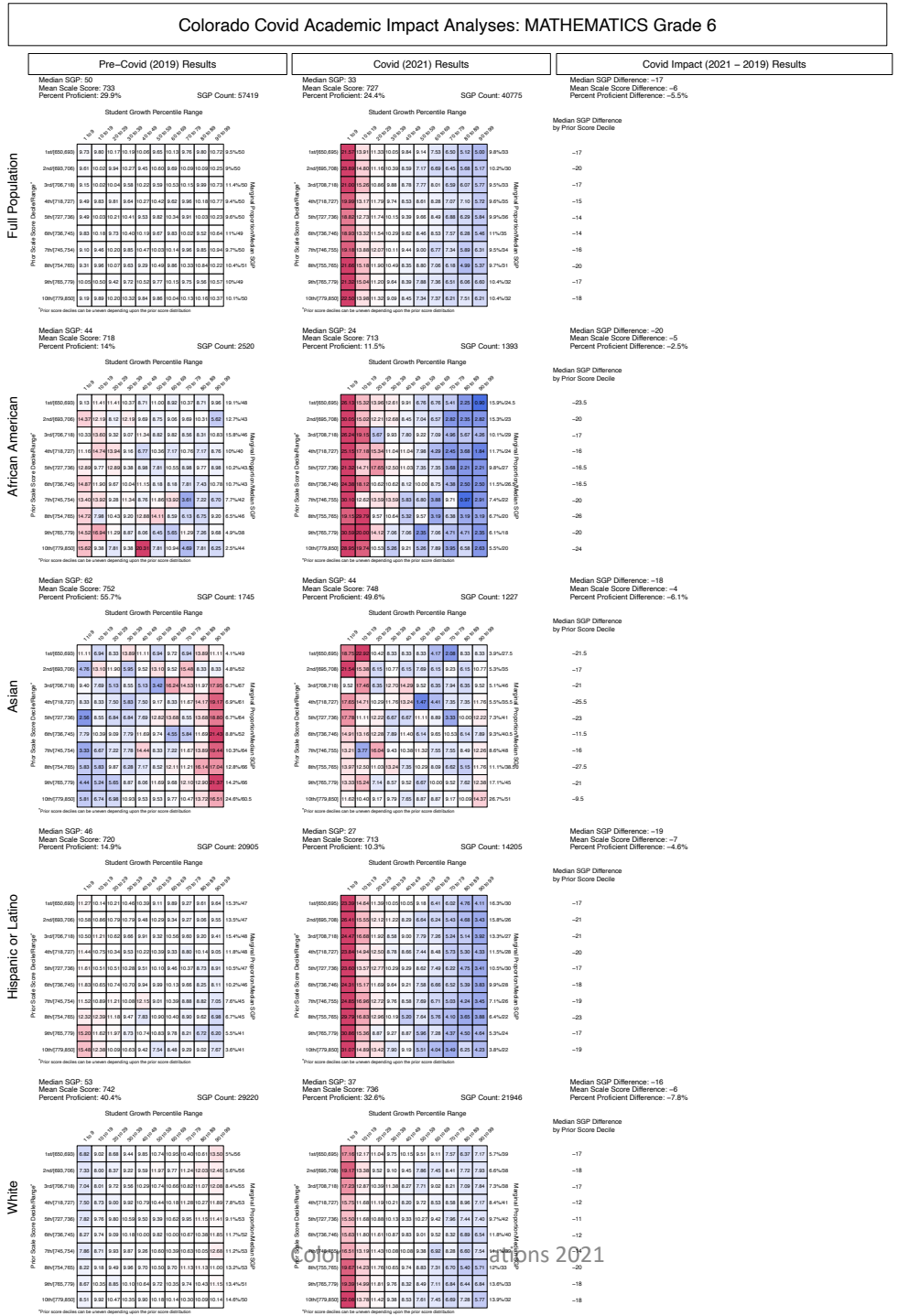
Growth by Prior Score Decile & Ethnicity: ELA - Grade 5



Growth by Prior Score Decile & Ethnicity: ELA - Grade 7



Growth by Prior Score Decile & Ethnicity: Mathematics - Grade 6



Growth by Prior Score Decile & Ethnicity: Mathematics - Grade 8