COLORADO

Department of Education

September Technical Advisory Panel Meeting

September 13, 2019

Agenda

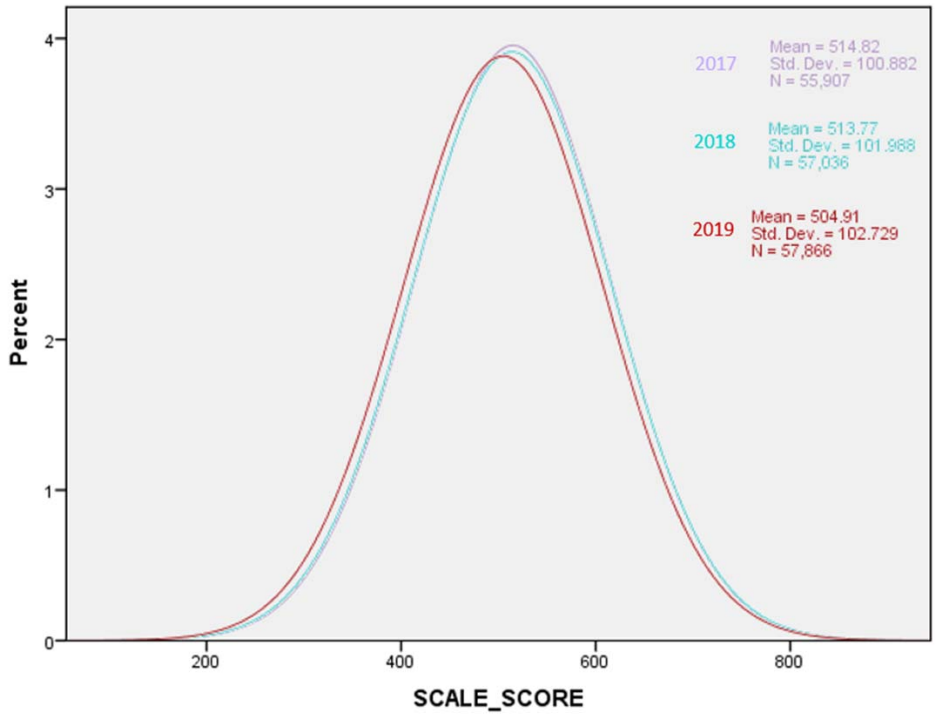
- Welcome and Introductions
- Membership Updates
- SAT Cut-scores Update - *Informational*
- 9th Grade EBRW Growth Update - *Informational*
- Potential SBE Framework Changes - *Seeking TAP Feedback*

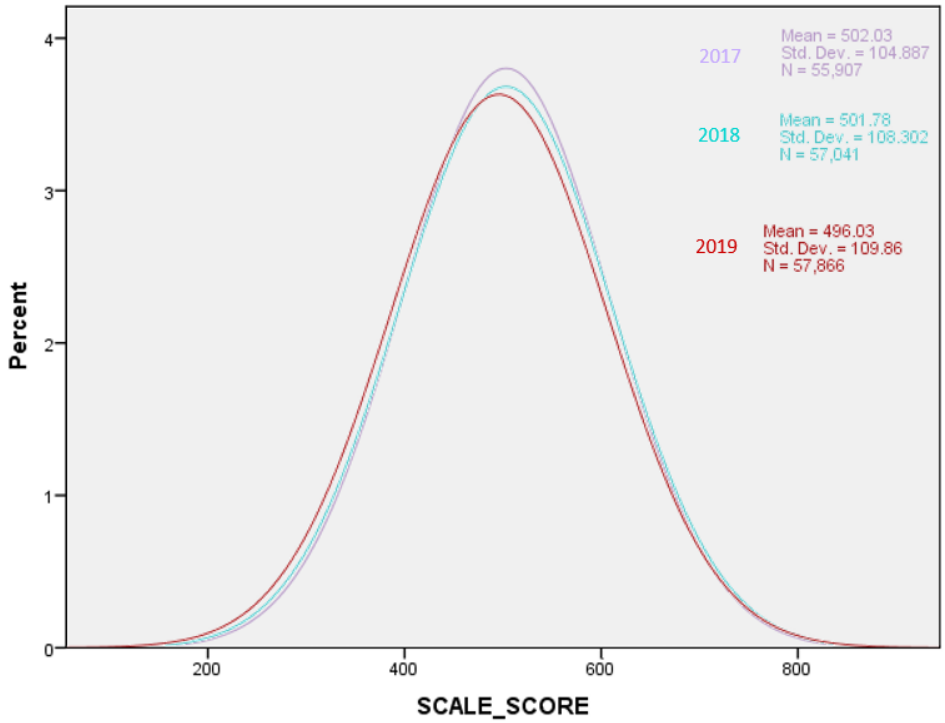
SAT Cut-Scores Update

Marie Huchton

2019 SAT Results

- State-level mean scale scores for g11 SAT were slightly lower than in 2017 and 2018, by 9 points in Evidence-based Reading and Writing (EBRW) and 6 points in Math





CDE and College Board Validation Procedures

- CDE engaged with the vendor, College Board, to verify that all of the technical and logistical processes and procedures were followed.
- Content:
  - Verified that no changes were made to the content specifications of the test
  - Verified that standard content and bias/sensitivity item review processes were followed
  - An analysis of differential item functioning (DIF) across major subgroups was completed using CO data. Out of 154 items analyzed, no items with high levels of DIF were identified
  - No irregularities were identified in content development. Established processes and procedures were followed.

CDE and College Board Validation Procedures

- Equating
  - Item usage and metadata were verified
  - Supplemental item analyses were completed based on Colorado students. When compared to the national equating sample, no aberrant performance patterns were identified
  - As part of the vendor's standard practices, initial equating is validated by the Center for the Advancement for Standards, Measurement and Assessment (CASMA)
    - In response to CDE request, the equating process and results were further reviewed by multiple psychometric experts.
  - No irregularities were identified in equating activities. Technically sound processes and procedures were followed.

CDE and College Board Validation Procedures

- Scoring
  - Item keys were verified and consistent with those from the equating administration
  - A review of scoring procedures was conducted. As standard practice, operational scoring is always completed through a dual process (two separate systems). College Board reviewed and confirmed that there were no discrepancies between scores originally generated through that standard practice (item, raw, and converted)
    - During the review process, a third set of independent programs were used to score all the student responses and the use of the correct raw score to scale score were verified. No discrepancies were found between any of the scores and the scores generated during production.
  - No irregularities were identified in scoring activities. Technically sound processes and procedures were followed.

- Cross Function Review
  - An end-to-end review was completed involving Psychometrics, Assessment Design and Development Team, IT, and Operations.
  - No irregularities were identified during the cross function review. Every hand-off was validated.

Implications for Student Scores

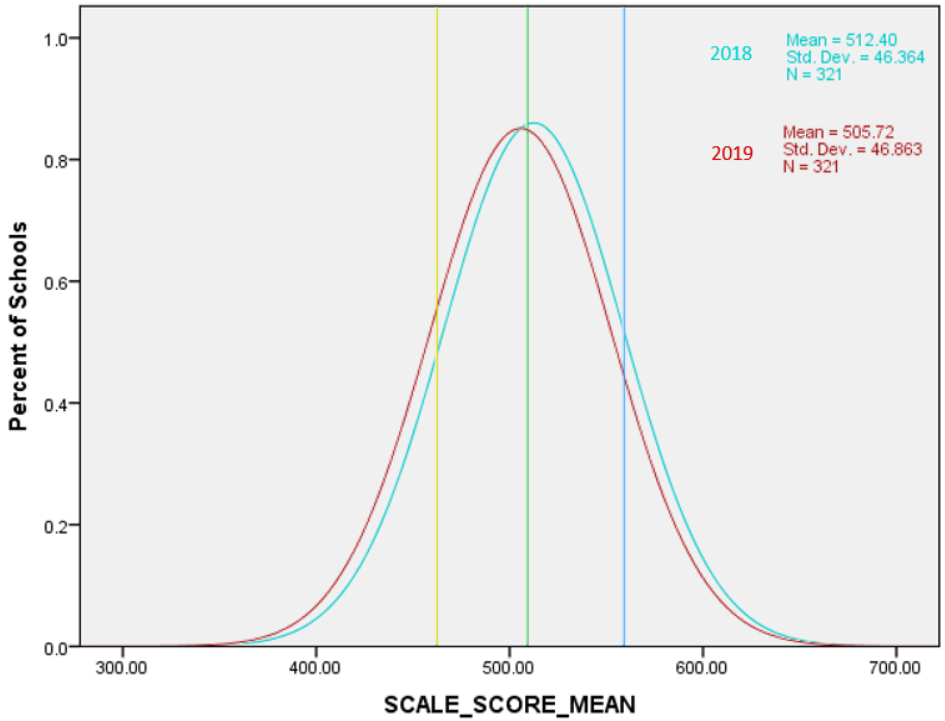
- The test reliabilities and relationships among scores for the spring 2019 SAT administration in Colorado fall within statistically reasonable ranges and may be used for making inferences about student performance, for inclusion in college and scholarship applications, and as a piece of information used by teachers for instruction.

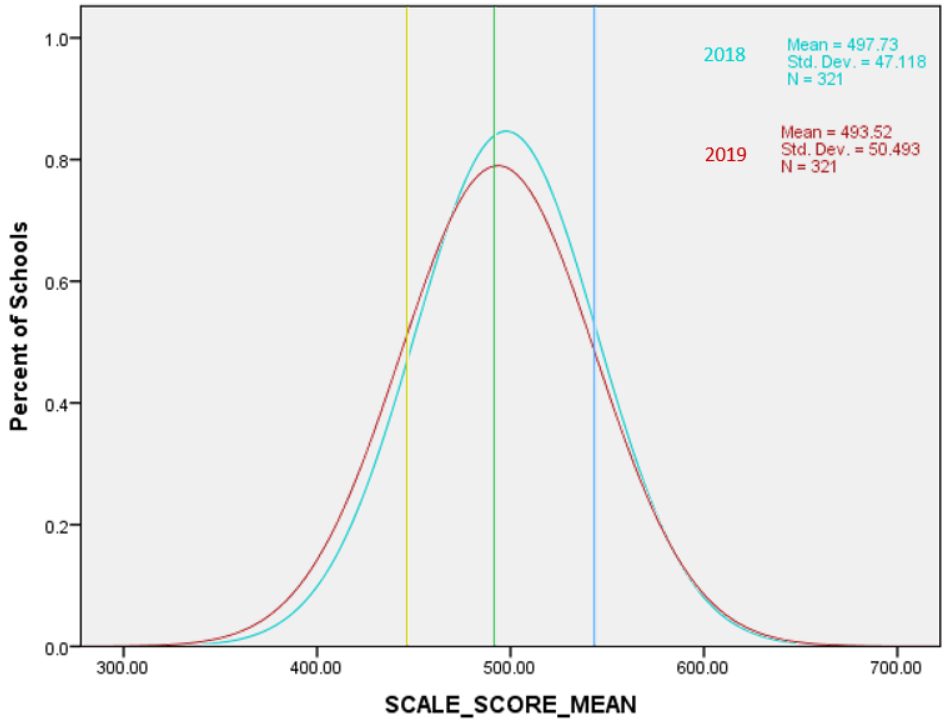
Implications for State Performance Frameworks Achievement Sub-indicators

- CDE's priority is to ensure fairness and consistency of state accountability rating interpretations across schools and districts over time.
- CDE annually reviews the Achievement sub-indicators (including SAT as a PWR Achievement measure) to see if the baselined cut-scores continue to meaningfully differentiate school ratings into roughly: 15% Does Not Meet, 35% Approaching, 35% Meets and 15% Exceeds.
- 2019 SAT results would have resulted in more schools and districts identified for Does Not Meet and Approaching ratings and fewer schools at Meets or Exceeds.
- To reduce this potential negative impact, the Department re-normed the achievement cut-scores for SAT at the 15th, 50th and 85th percentiles based on the 2019 state data.

Implications for State Performance Frameworks Achievement Sub-indicators

- Comparison school-level mean scale scores for g11 SAT between 2018 and 2019 (excluding AECs) showing 2018 cut-scores for Approaching, Meets, Exceeds





Implications for State Performance Frameworks Achievement Sub-indicators

- Counts of districts and high schools that would have received lower Achievement sub-indicator ratings if we had used the 2018 cut-scores:

School or District	Test & Content Area	Student Group	2018 Rating	Exceeds	Meets	Approaching	Does Not Meet	Total Changed
DIST	CO SAT - EBRW	All Students	Exceeds	6	0	0	0	20
			Meets	1	36	0	0	
			Approaching	0	14	79	0	
			Does Not Meet	0	0	5	33	
DIST	CO SAT - MATH	All Students	Exceeds	9	0	0	0	18
			Meets	0	42	0	0	
			Approaching	0	5	79	0	
			Does Not Meet	0	0	13	26	
SCH	CO SAT - EBRW	All Students	Exceeds	42	0	0	0	58
			Meets	9	102	0	0	
			Approaching	0	29	126	0	
			Does Not Meet	0	0	20	126	
SCH	CO SAT - MATH	All Students	Exceeds	51	0	0	0	35
			Meets	0	109	0	0	
			Approaching	0	17	129	0	
			Does Not Meet	0	0	18	130	

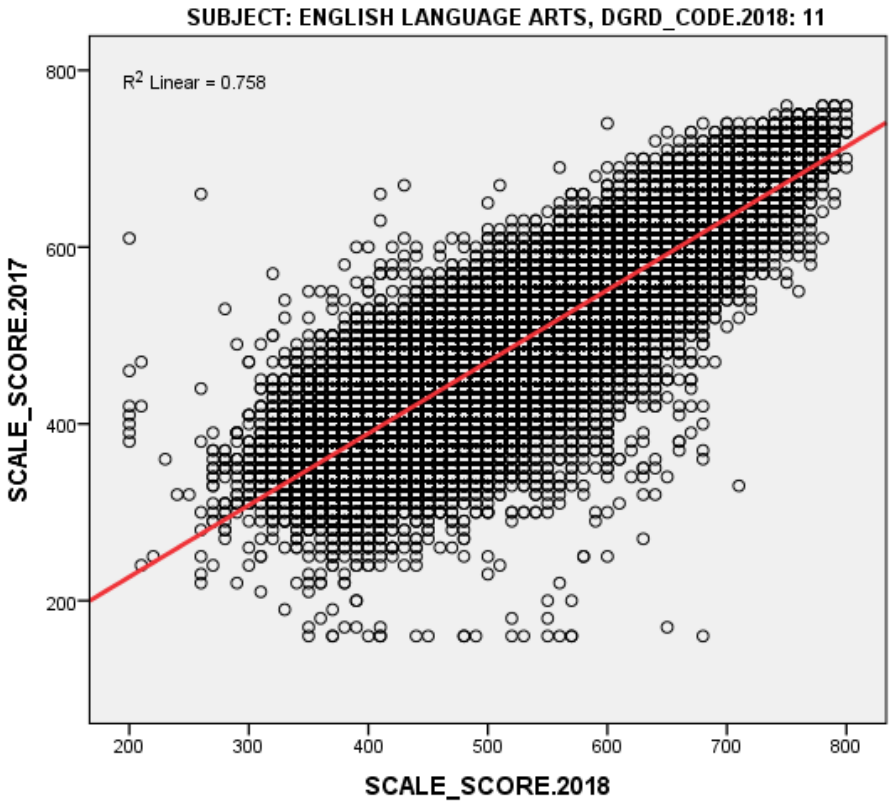
Implications for State Performance Frameworks Achievement Sub-indicators

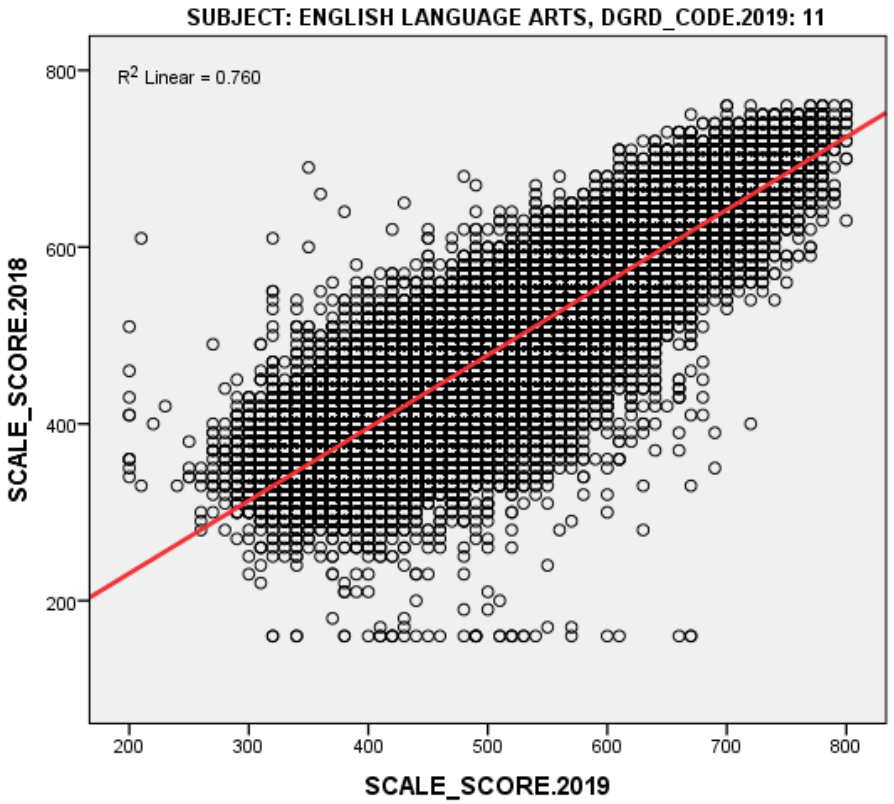
- Renorming of sub-indicator cut-scores has been done on occasions when the technical qualities of a metric have shifted unexpectedly and could potentially negatively impact overall rating results.
- CDE plans to return to the 2018 SAT baseline cut-scores for the 2020 frameworks.
- Adjustments to the cut-scores may again be warranted in 2020 if the observed student scores would continue to result in substantively higher-than-expected proportions of schools and districts identified for Does Not Meet or Approaching Achievement ratings.

Implications for State Performance Frameworks Growth Sub-indicators

- CDE also investigated potential impacts of the SAT score shift on the growth model calculations for 2019, and found that the slight score drop was consistent enough across the state to maintain the relative ranking of students and therefore did not adversely affect the calculation outcomes.
- The aggregated 2019 median growth percentiles for schools and districts were comparable to those from 2018 and displayed enough consistency to warrant continued inclusion in the performance frameworks.

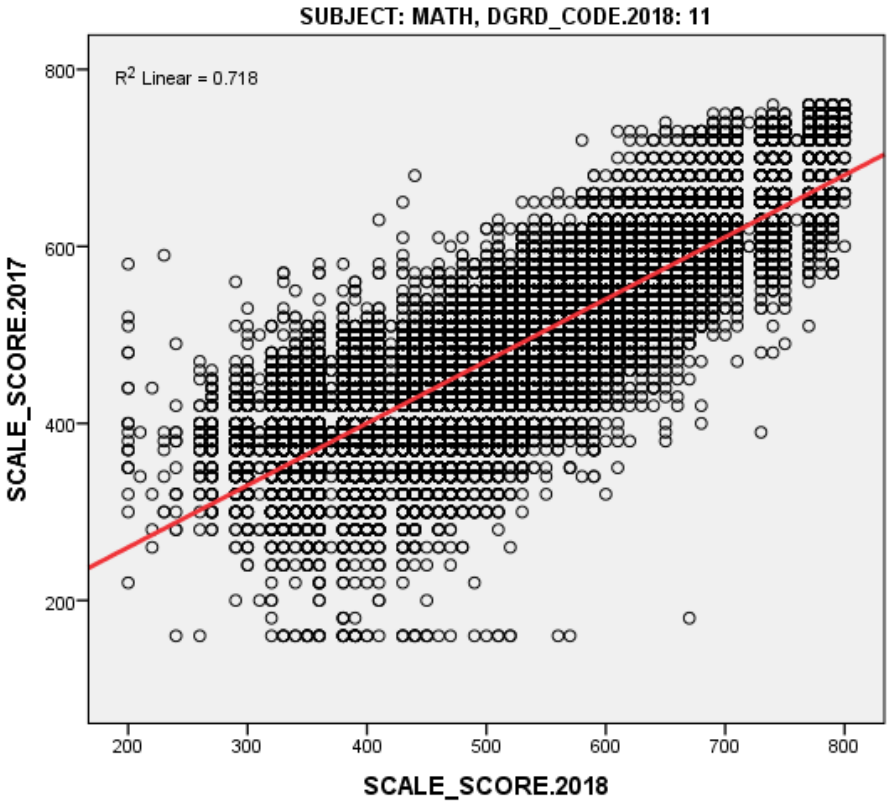
Implications for State Performance Frameworks Growth Sub-indicators

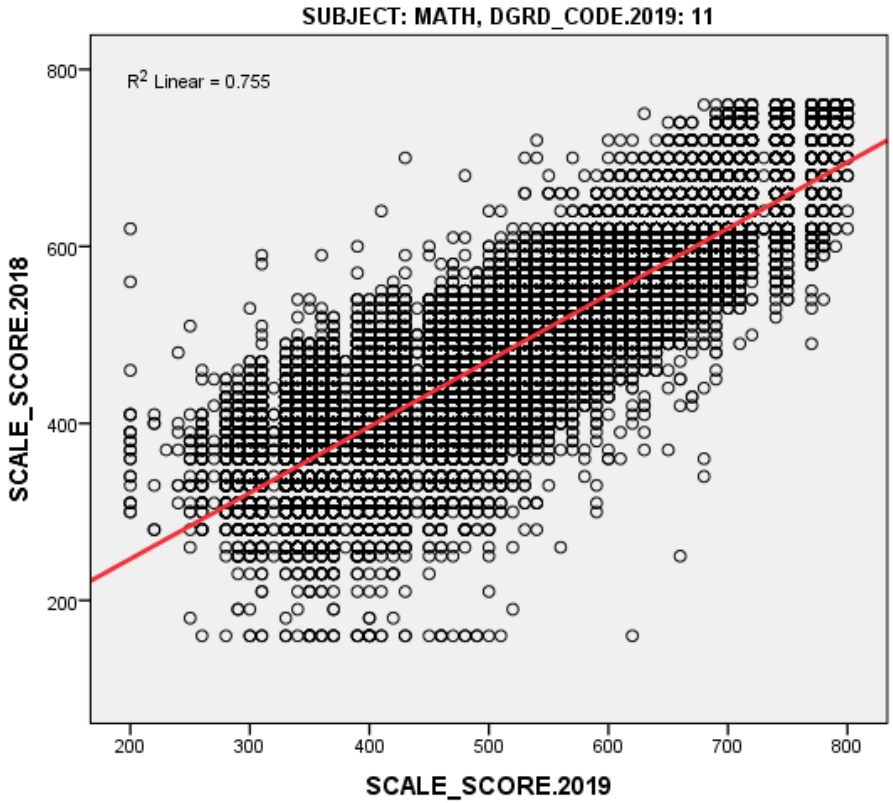




The correlation from 2018 to 2019 for ELA was 0.872 which is nearly identical to the 0.872 correlation from 2017 to 2018, despite the slight overall score decrease.

Implications for State Performance Frameworks Growth Sub-indicators

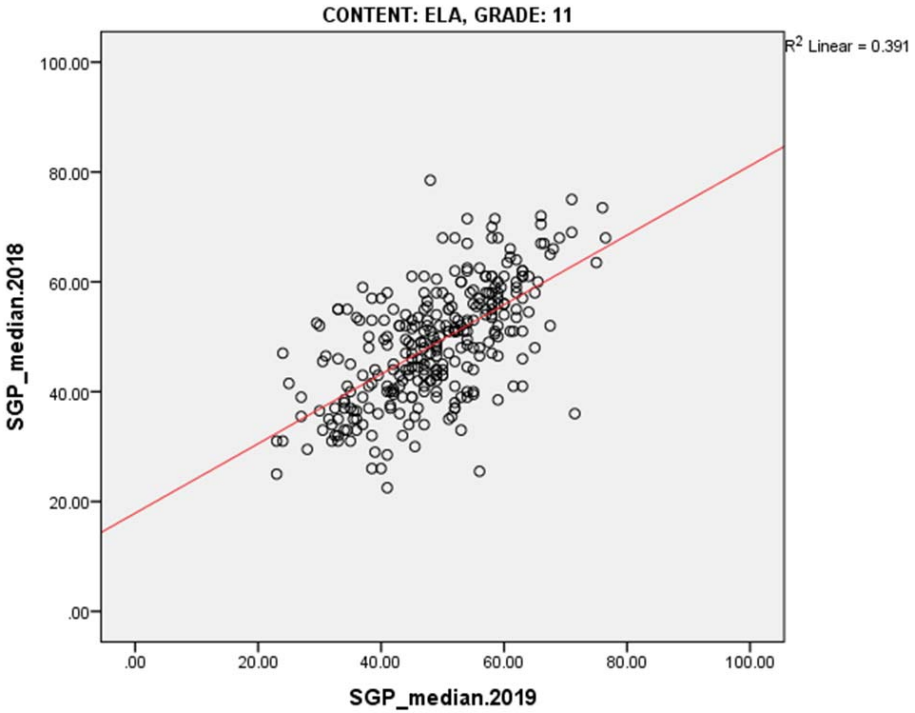


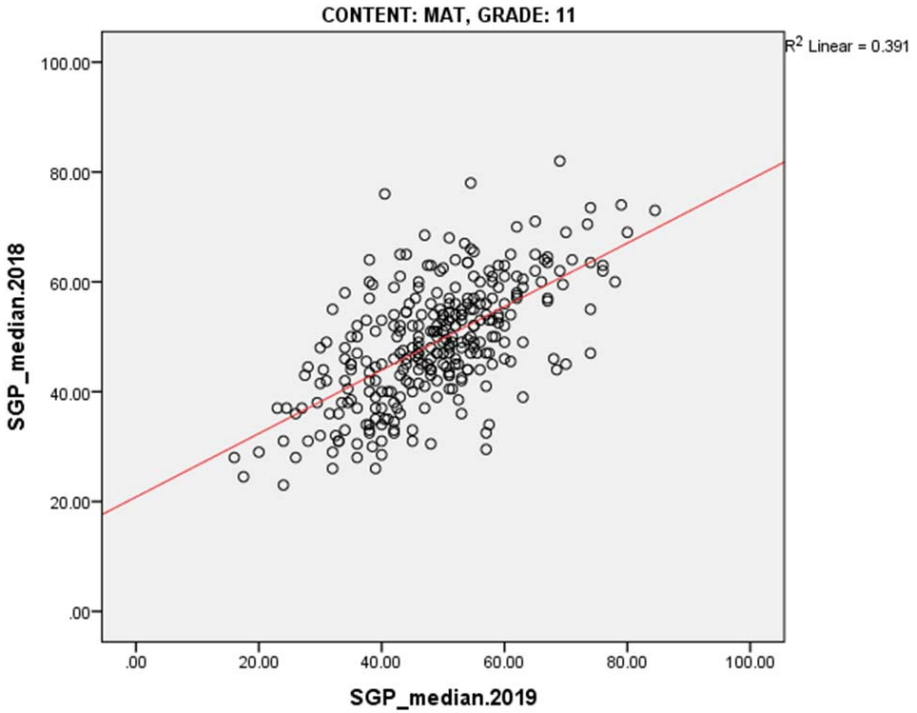


The correlation from 2018 to 2019 for Math was 0.869, which is slightly higher than the 0.847 correlation from 2017 to 2018, despite the slight overall score decrease.

Implications for State Performance Frameworks Growth Sub-indicators

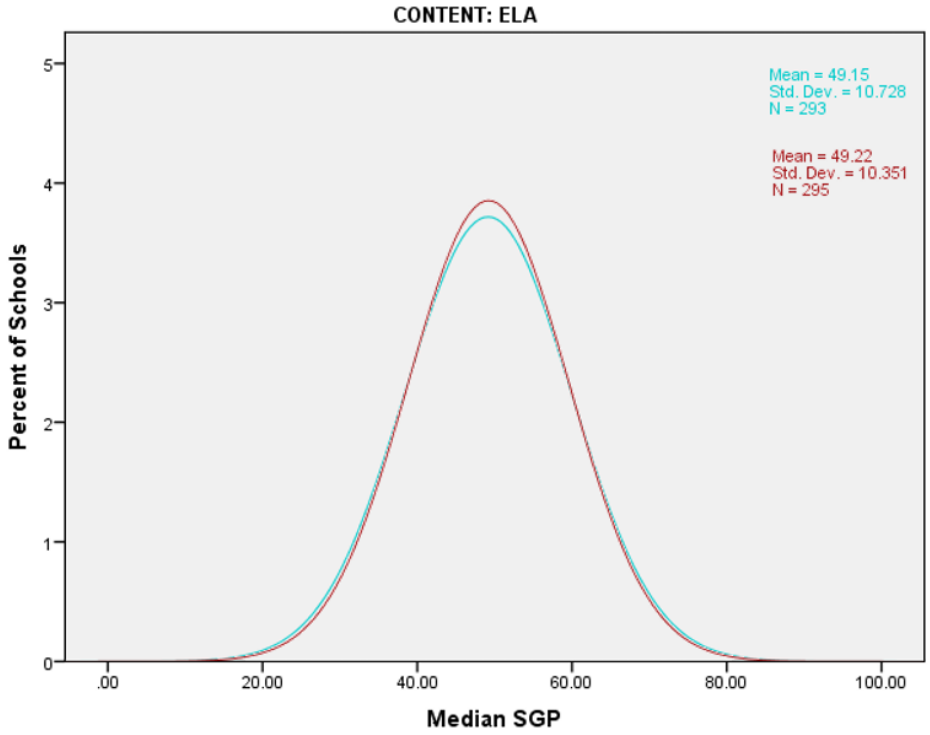
- School-level MGPs remained fairly consistent between 2018 and 2019, the correlation for ELA was 0.625 and for Math 0.626. They are consistent with previous results and slightly higher than the correlations for CMAS between 2018 and 2019.

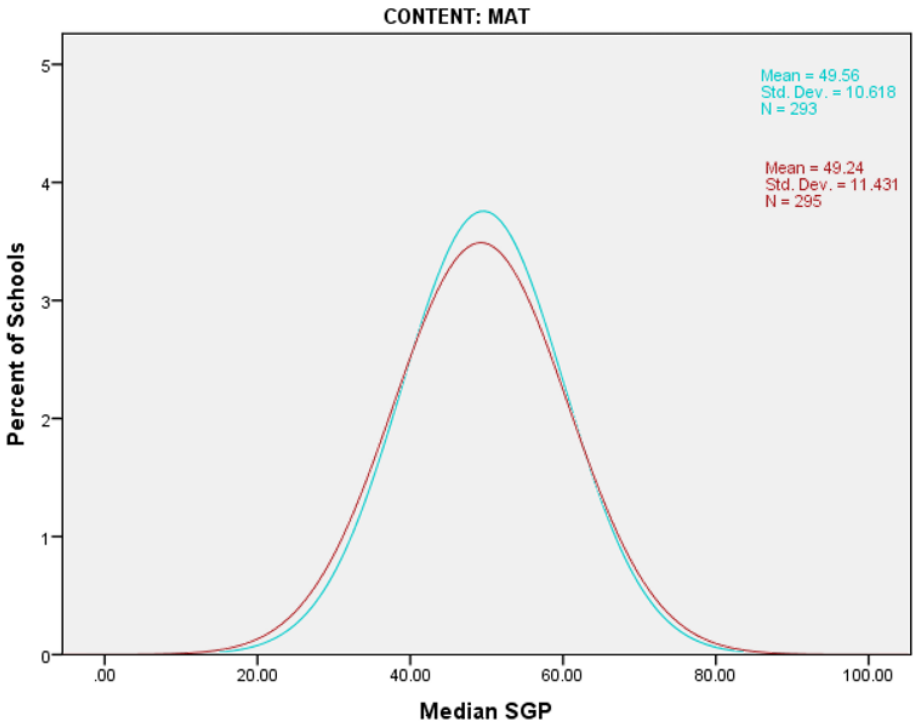




Implications for State Performance Frameworks Growth Sub-indicators

- The school-level MGP distributions also remained consistent, indicating the frameworks are identifying about the same number of schools in each sub-indicator rating category.





Implications for Federal Accountability Identifications

- The US Department of Education required Colorado submit specific SAT sub-indicator cut-score values in order to approve our ESSA state plan.
- Applying these historic cuts to the 2019 SAT results in more high schools receiving Does Not Meet and Approaching Achievement ratings than in past years.

9th Grade EBRW Growth Update

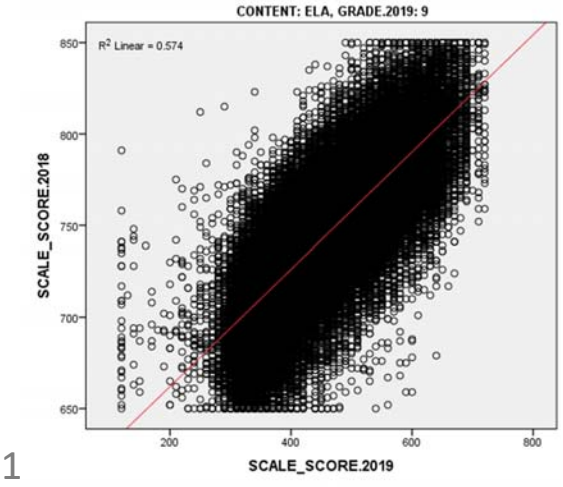
Marie Huchton

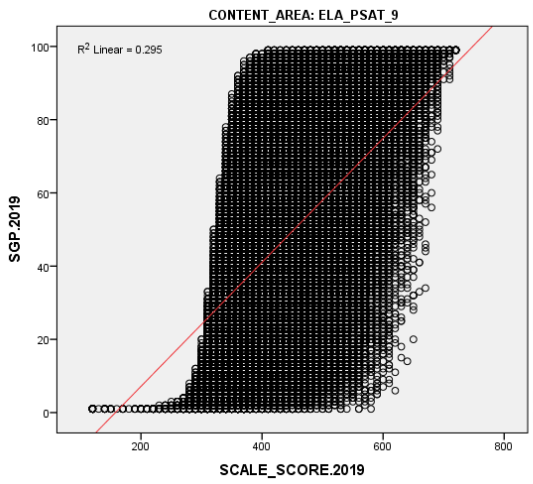
Context for PSAT/SAT Transition

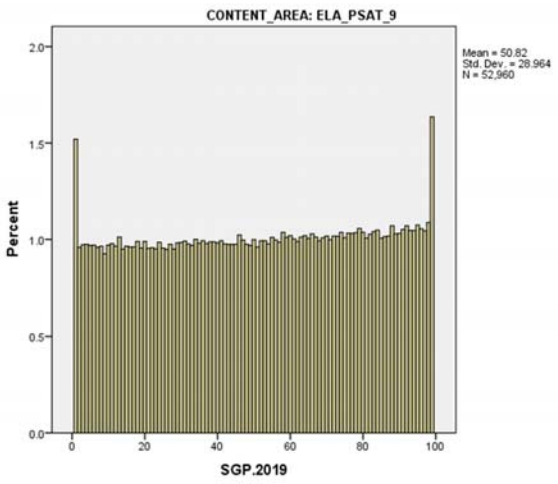
- Legislative requirement to use a nationally recognized college entrance examination for grade 11, and the precursor assessments for grades 9 and 10.
- Transition to the PSAT/SAT assessments was largely intended to encourage high school student participation in state assessments following several years of high parent excusal rates.
- Unfortunately it resulted in a reduction in the alignment of the high school assessments to our state content standards.
- We assumed that the assessment constructs underlying both CMAS and PSAT/SAT would be similar enough to run growth calculations.

Introduction of PSAT g9 in 2018

- 2018 was the first year the PSAT g9 Assessment was utilized for census testing in Colorado.
- The Accountability team received the student results data at the end of July and rushed to incorporate it into the Achievement and Growth indicators on the 2018 school and district performance frameworks.
- Our standard validation steps all checked out and it appeared that the data growth were consistent with previous tests.







Differing Assessment Constructs

- In preparing for this year's growth calculations, CDE dug a bit deeper into the growth results and saw lower between-year correlations (related to gender differences in the student performance trends from CMAS ELA to PSAT EBRW) as compared to the other grade and content areas.
- The primary difference between these assessments is that the PSAT and SAT do not include substantial writing components (selected response items only) while the CMAS ELA assessments include explicit text-based writing tasks (constructed response items in addition to selected response items).

Student-level Between-Year Scale Score Correlations

		2017 to 2018		2018 to 2019		
		N	Corr	N	Corr	Corr Diff
ELA	CMAS g3 to CMAS g4	59463	0.833	58726	0.827	-0.006
	CMAS g4 to CMAS g5	59974	0.815	60687	0.822	0.007
	CMAS g5 to CMAS g6	58195	0.811	60091	0.804	-0.007
	CMAS g6 to CMAS g7	55418	0.833	58033	0.824	-0.009
	CMAS g7 to CMAS g8	52759	0.834	54081	0.821	-0.013
	CMAS g8 to PSAT g9	50793	0.769	52960	0.758	-0.011
	CMAS/PSAT g9 to PSAT g10	44447	0.776	55790	0.872	0.096
	PSAT g10 to SAT g11	51840	0.871	52781	0.872	0.001
Math	CMAS g3 to CMAS g4	61120	0.85	60316	0.85	0
	CMAS g4 to CMAS g5	60775	0.851	61823	0.852	0.001
	CMAS g5 to CMAS g6	58219	0.848	60216	0.837	-0.011
	CMAS g6 to CMAS g7	55479	0.854	58087	0.855	0.001
	CMAS g7 to CMAS g8	52668	0.84	54089	0.848	0.008
	CMAS g8 to PSAT g9	50727	0.814	52973	0.823	0.009
	CMAS/PSAT g9 to PSAT g10	44296	0.79	55790	0.843	0.053
	PSAT g10 to SAT g11	51840	0.847	52781	0.869	0.022

2017 to 2018 growth for g9 CMAS to g10 PSAT highlighted in red

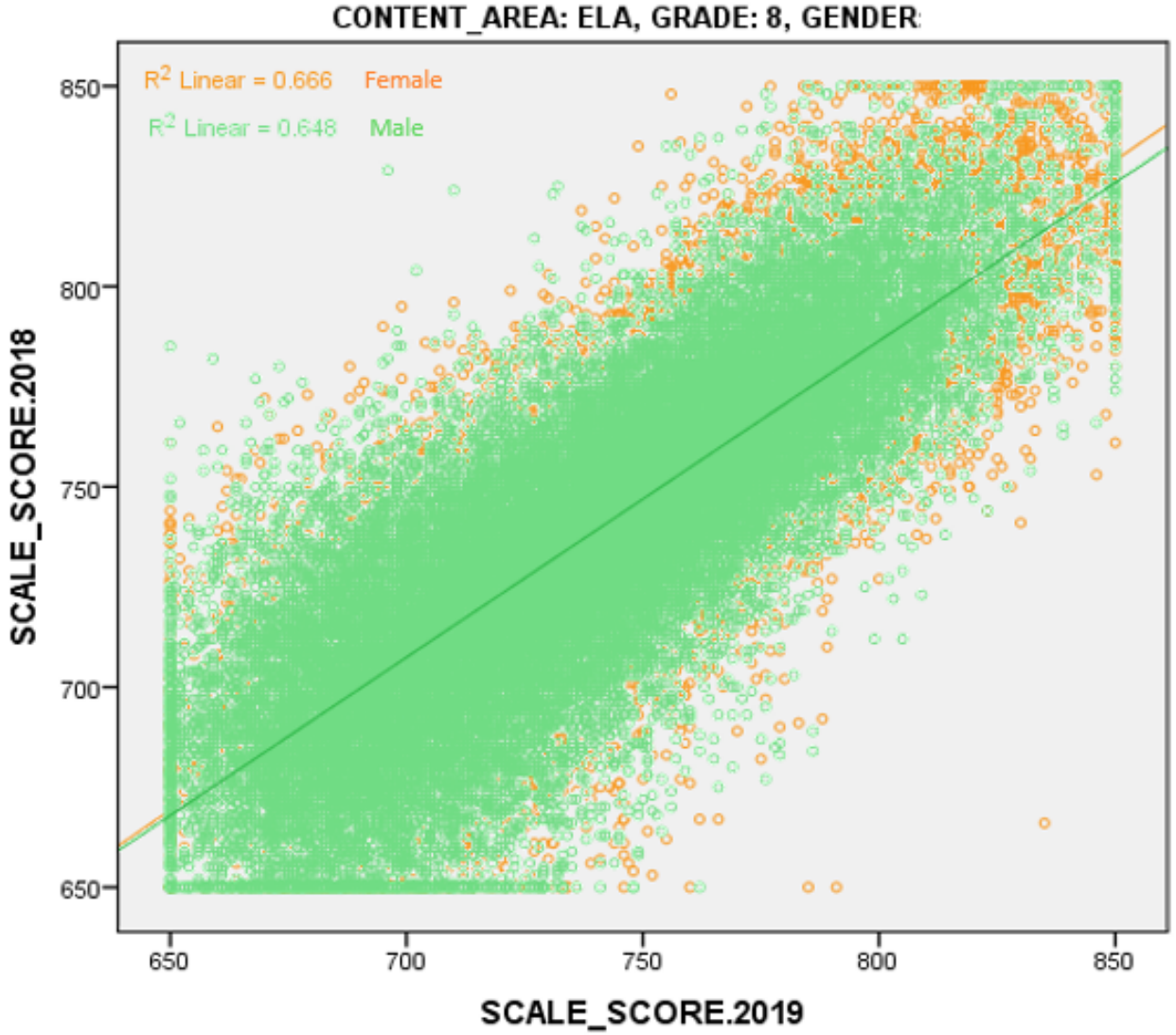
Differing Assessment Constructs

- The CMAS ELA assessment was designed to measure Colorado's academic content standards, including a student's ability to produce original written work.
- The PSAT EBRW construct does not align as closely with Colorado's academic content standards, including not requiring original student writing tasks, which was an acknowledged part of the trade-off in moving to PSAT/SAT.
- Achievement and growth results for girls taking Colorado's ELA assessments have historically been higher than for boys, particularly for CSAP/TCAP Writing and CMAS ELA, likely due to the inclusion of open-ended constructed response items.

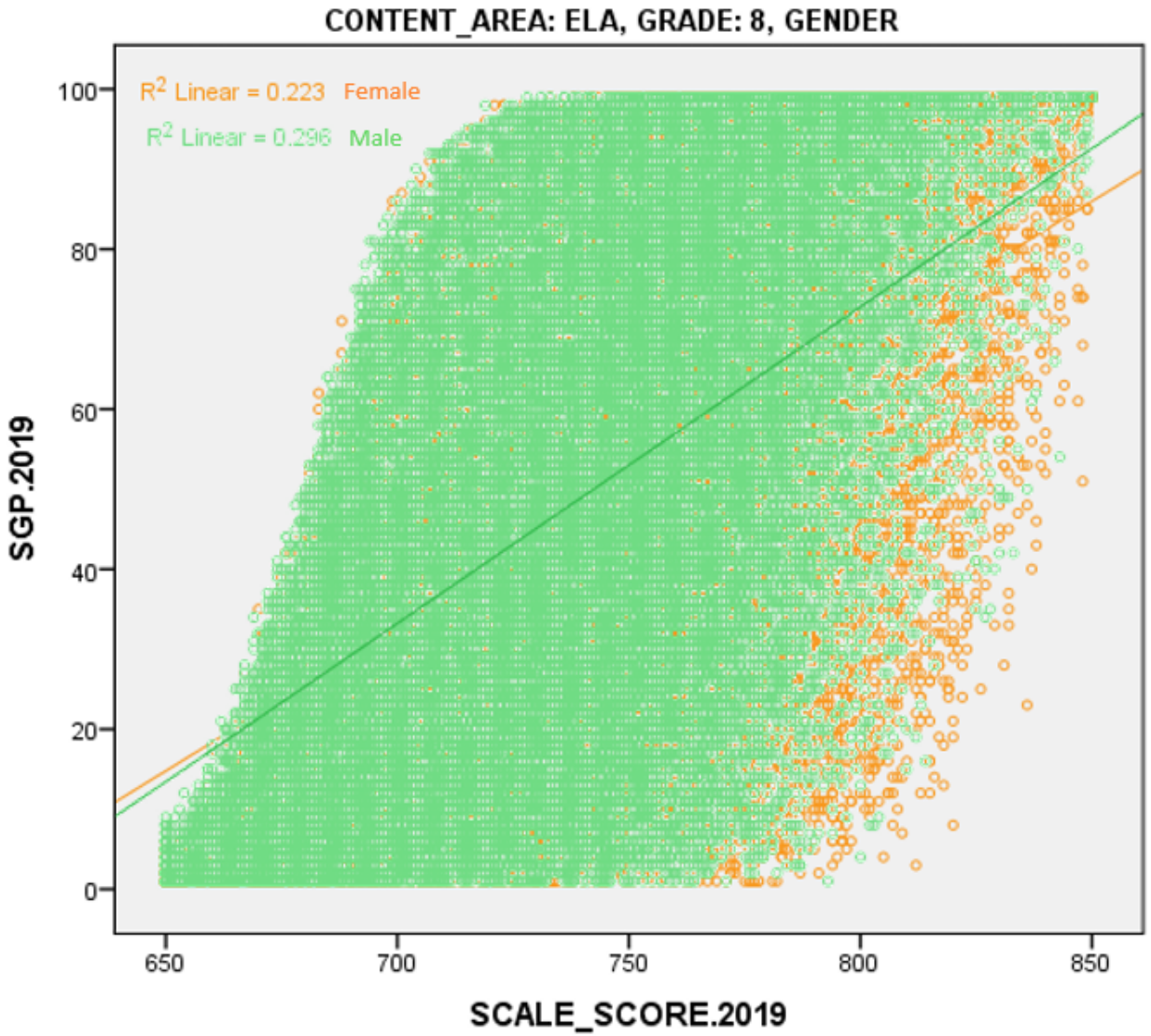
Differing Assessment Constructs

- As students transition from CMAS ELA to PSAT EBRW, the relative student rankings shift in different ways for boys and girls due to the different assessment constructs.
- An unfortunate side effect of Colorado's normative growth model calculation is that any systematic differences in score rankings between years are magnified in the student growth percentile results.
- Comparing the scale score and growth results from CMAS g8 ELA to PSAT g9 EBRW for girls and boys against the other grades/content areas shows this limitation.

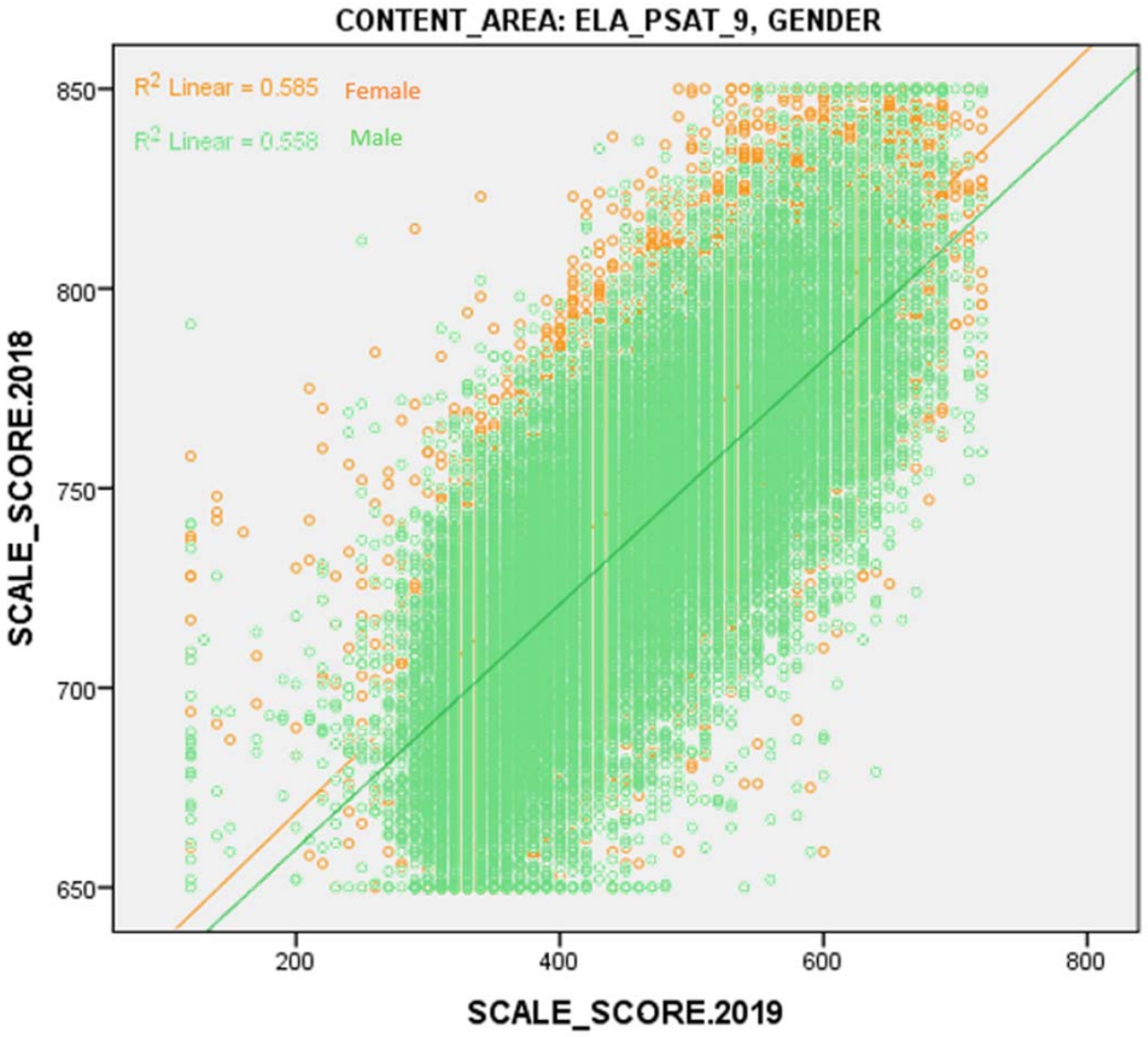
Scale Score Scatter plots for 2018 to 2019: CMAS g7 to CMAS g8 ELA by Gender



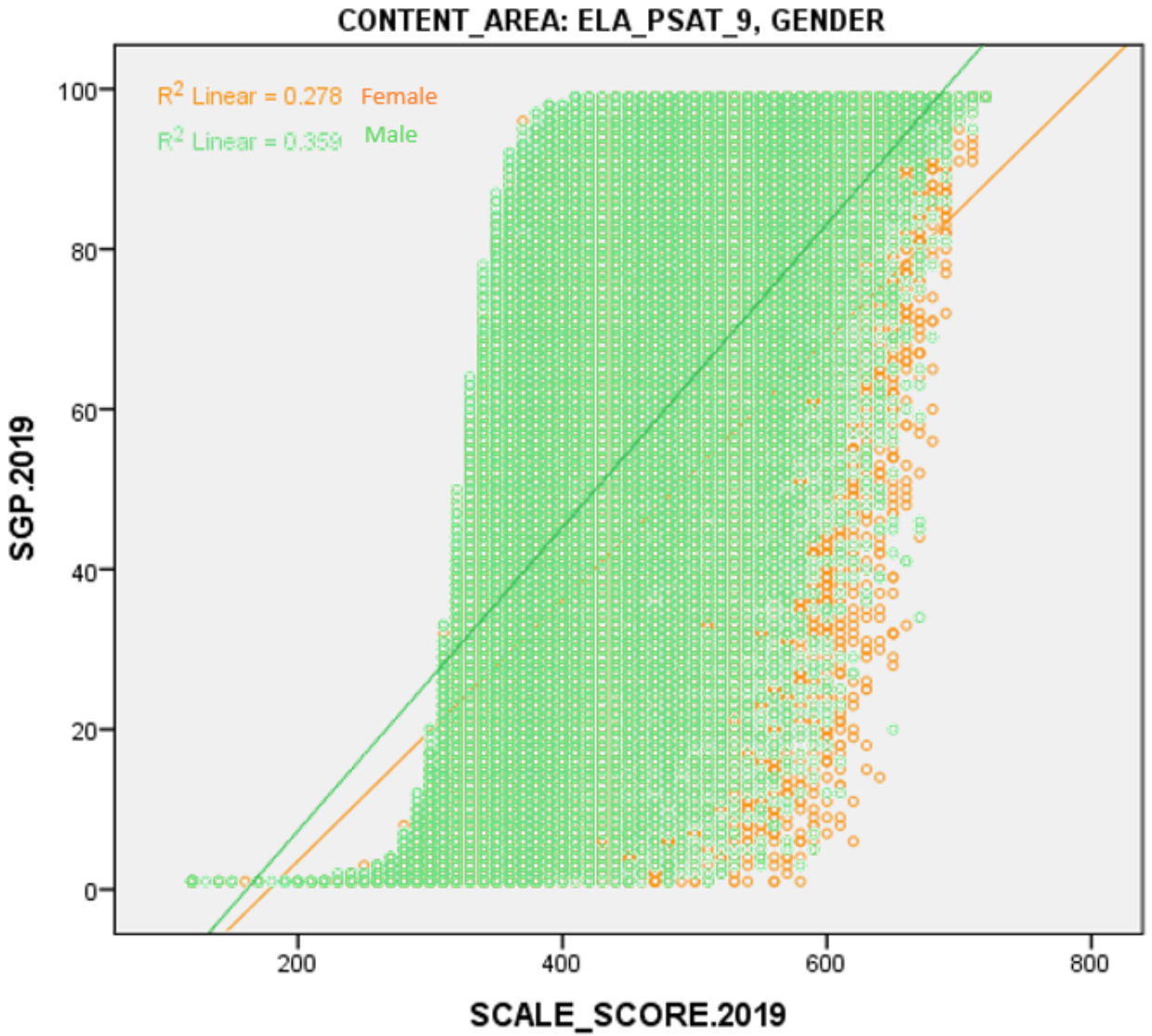
SGP to Scale Score Scatter plot for 2019: CMAS g8 ELA by Gender



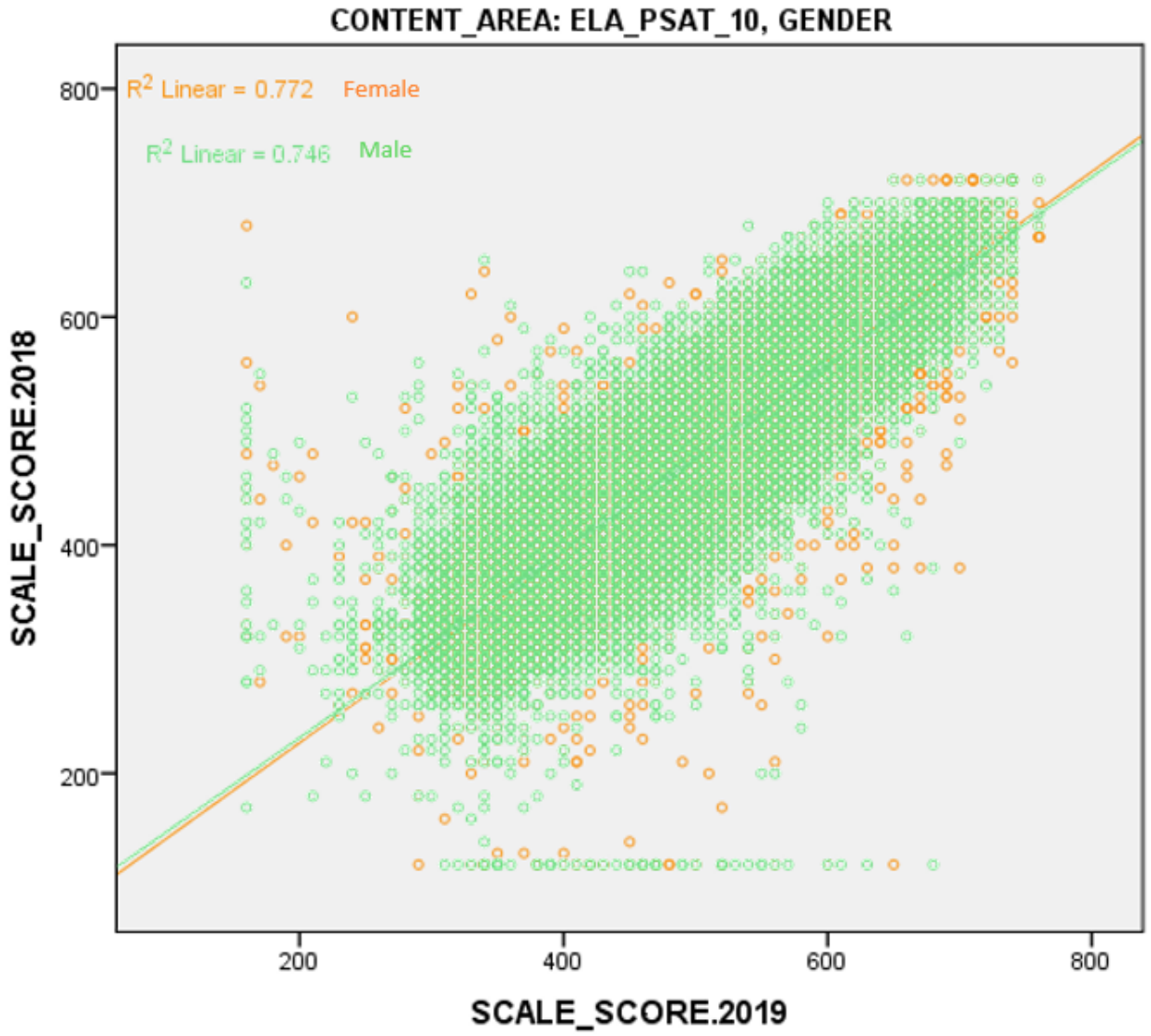
Scale Score Scatter plots for 2018 to 2019: CMAS g8 ELA to PSAT g9 EBRW by Gender



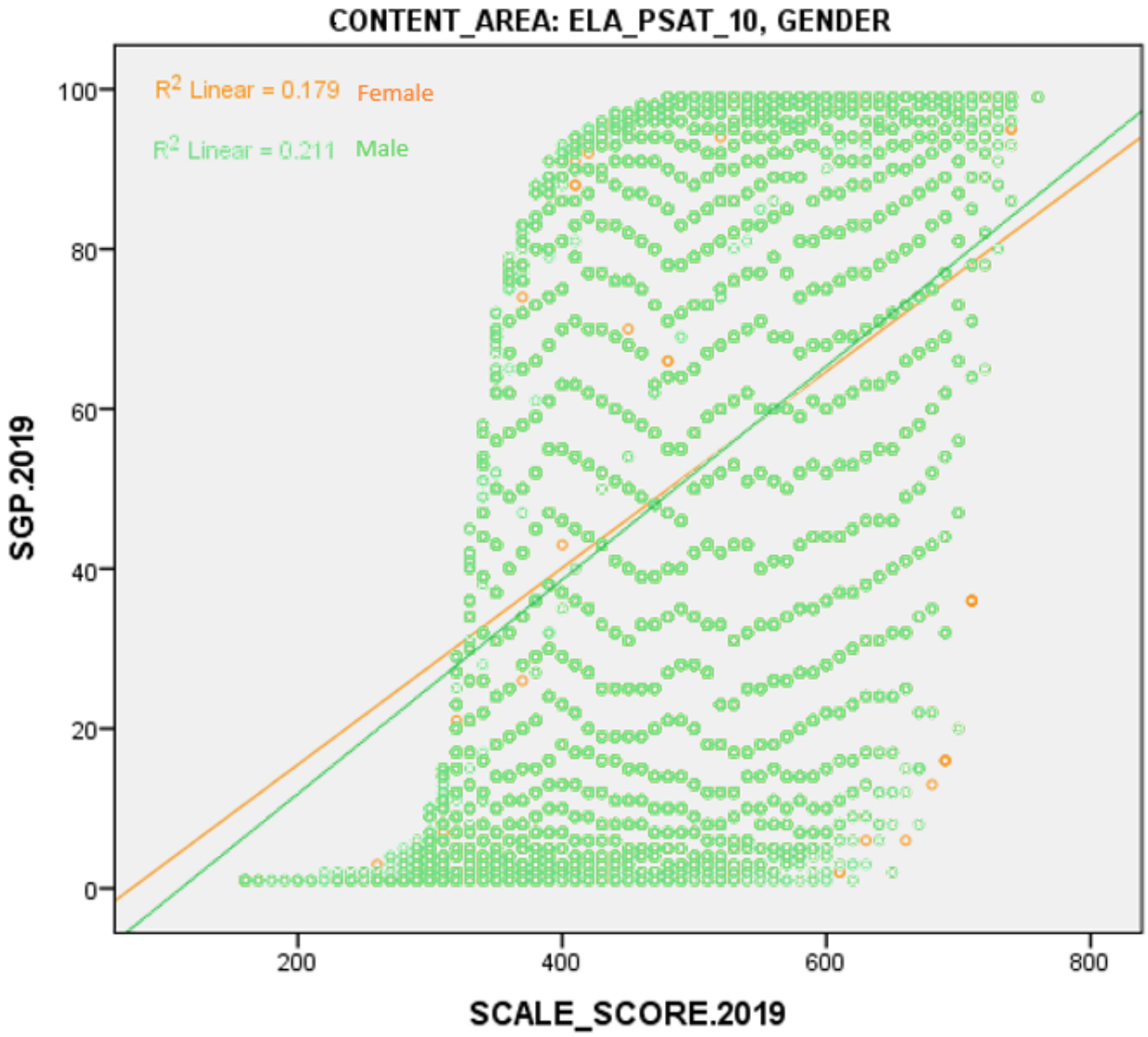
SGP to Scale Score Scatter plot for 2019: PSAT g9 EBRW by Gender



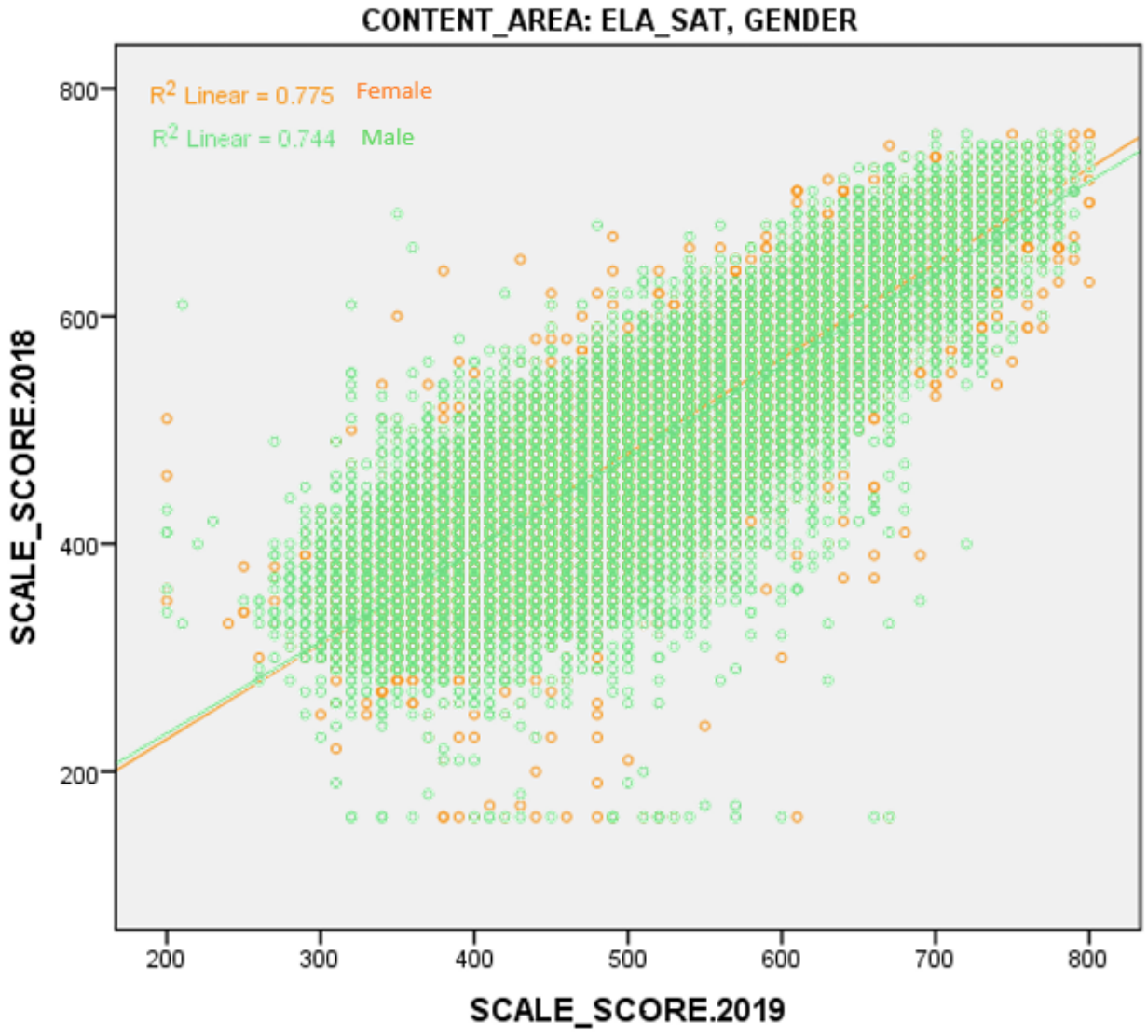
Scale Score Scatter plots for 2018 to 2019: PSAT g9 to PSAT g10 EBRW by Gender



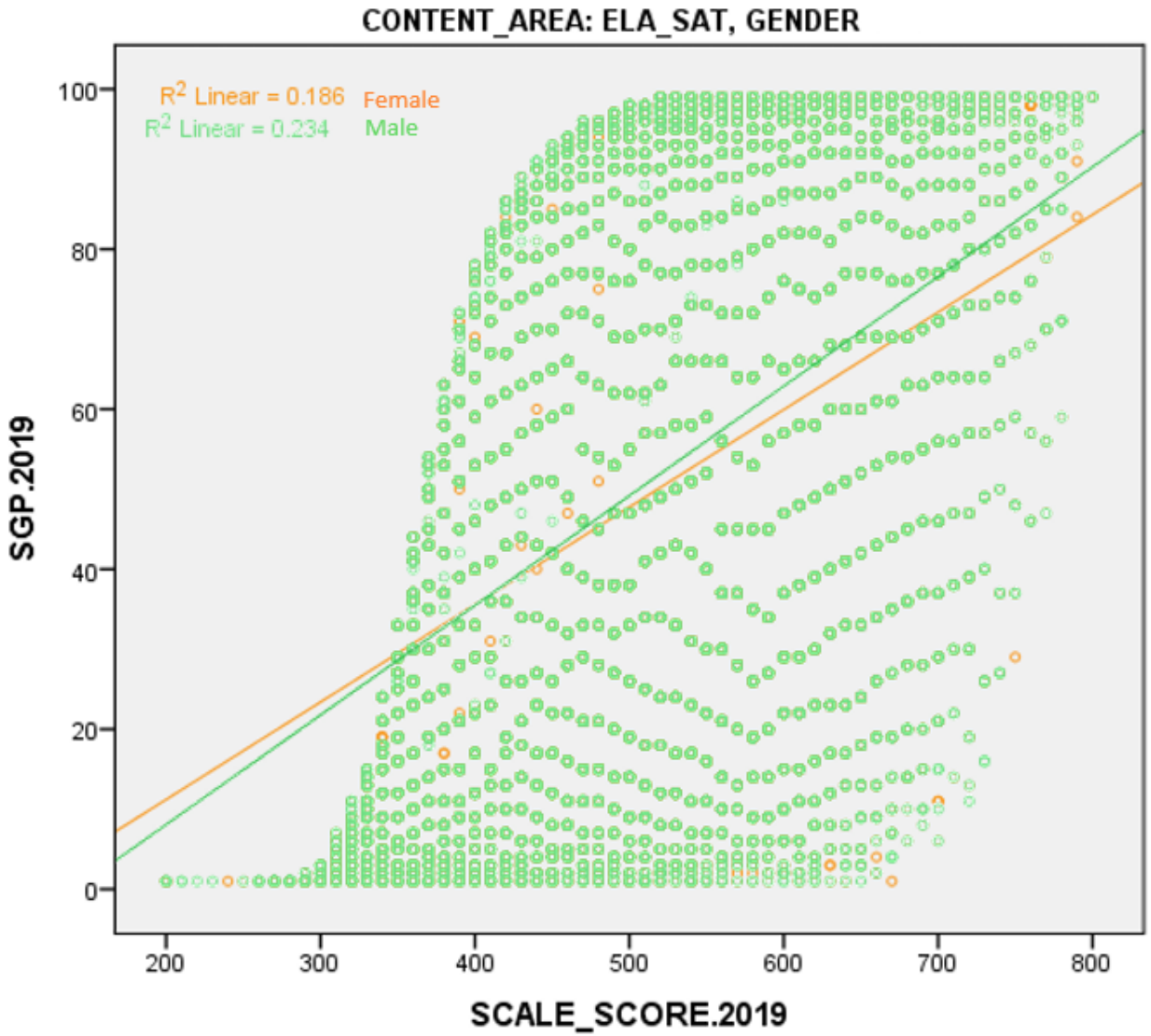
SGP to Scale Score Scatter plot for 2019: PSAT g10 EBRW by Gender



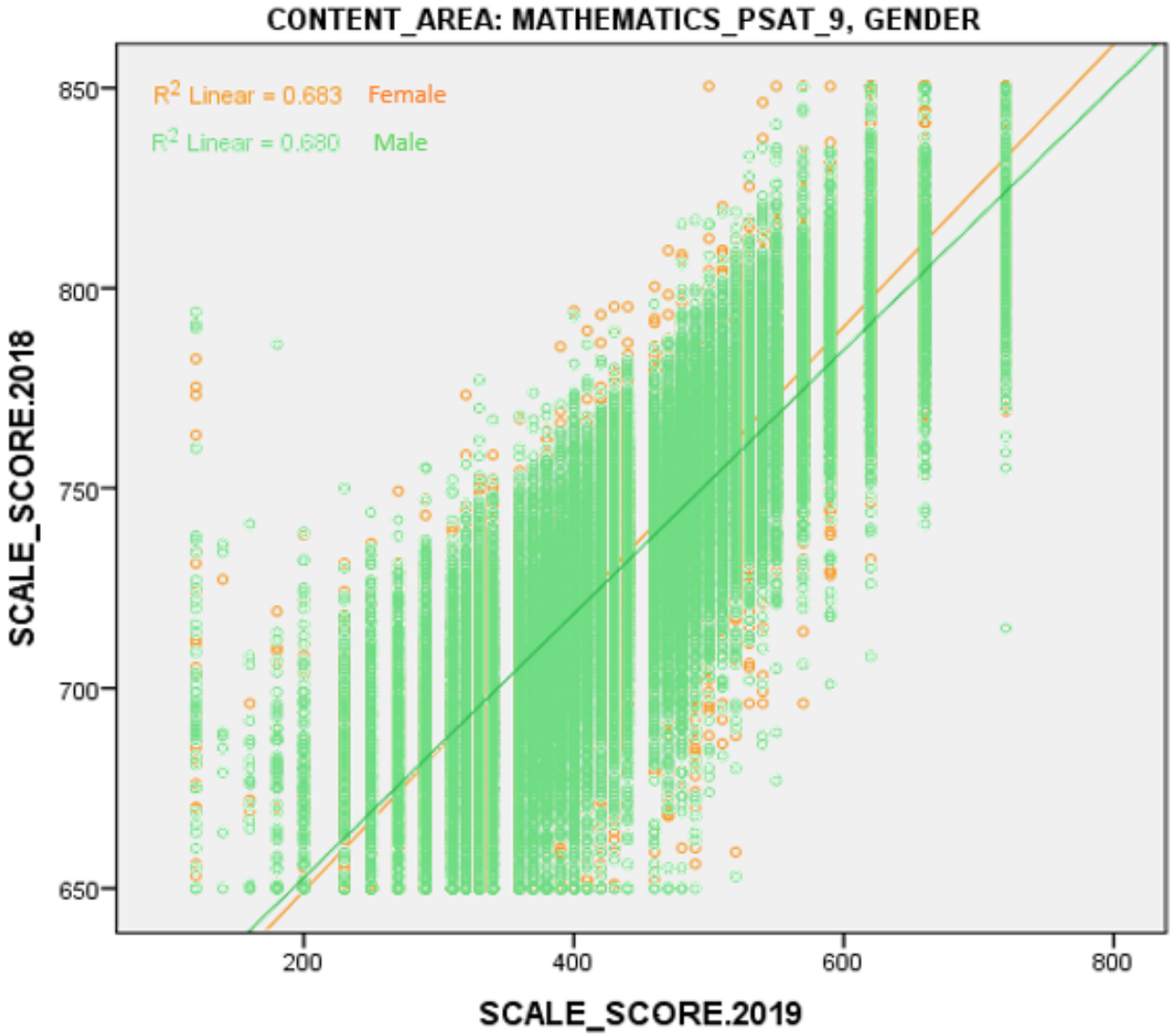
Scale Score Scatter plots for 2018 to 2019: PSAT g10 to SAT 11 EBRW by Gender



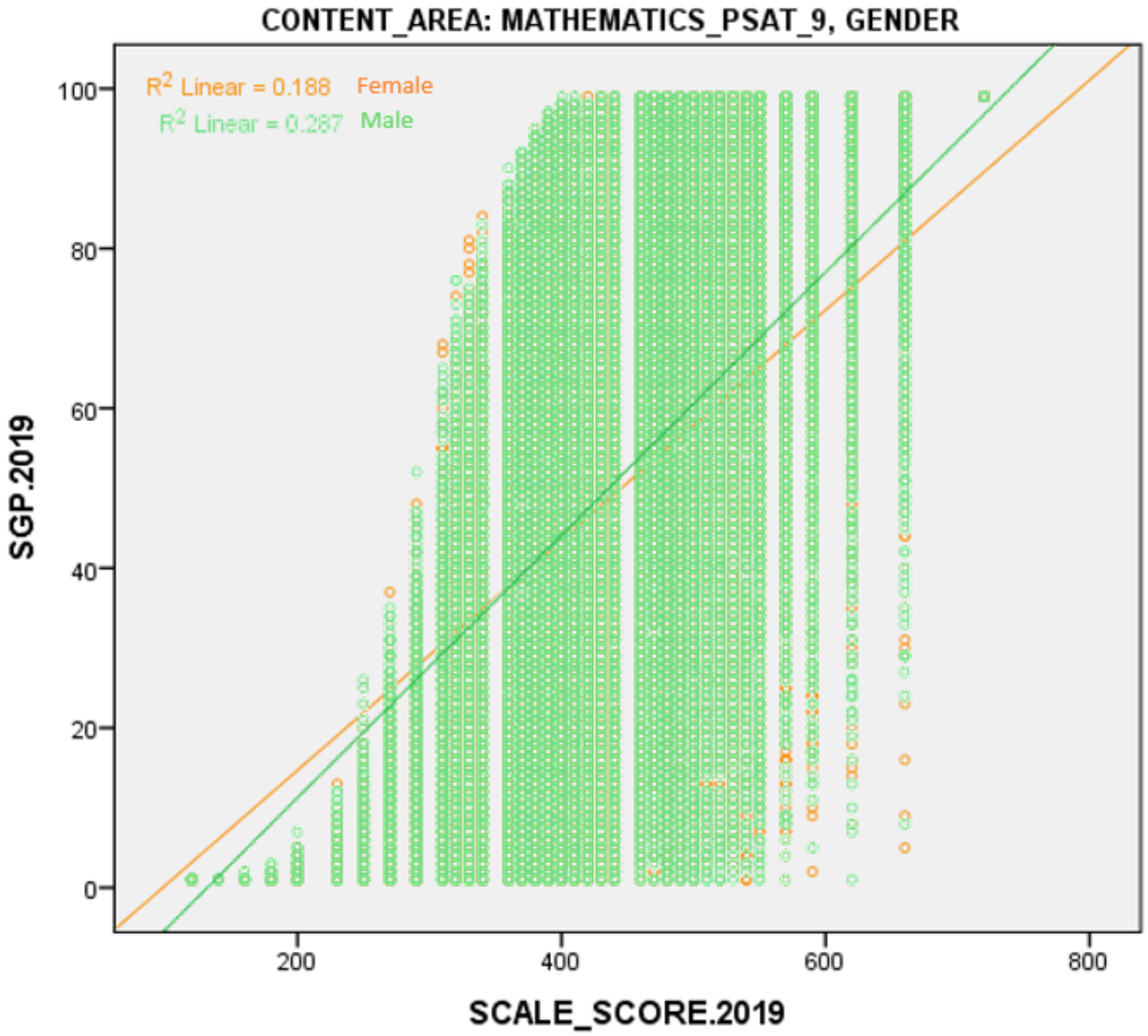
SGP to Scale Score Scatter plot for 2019: PSAT g11 EBRW by Gender



Scale Score Scatter plots for 2018 to 2019: CMAS g8 ELA to PSAT g9 Math by Gender

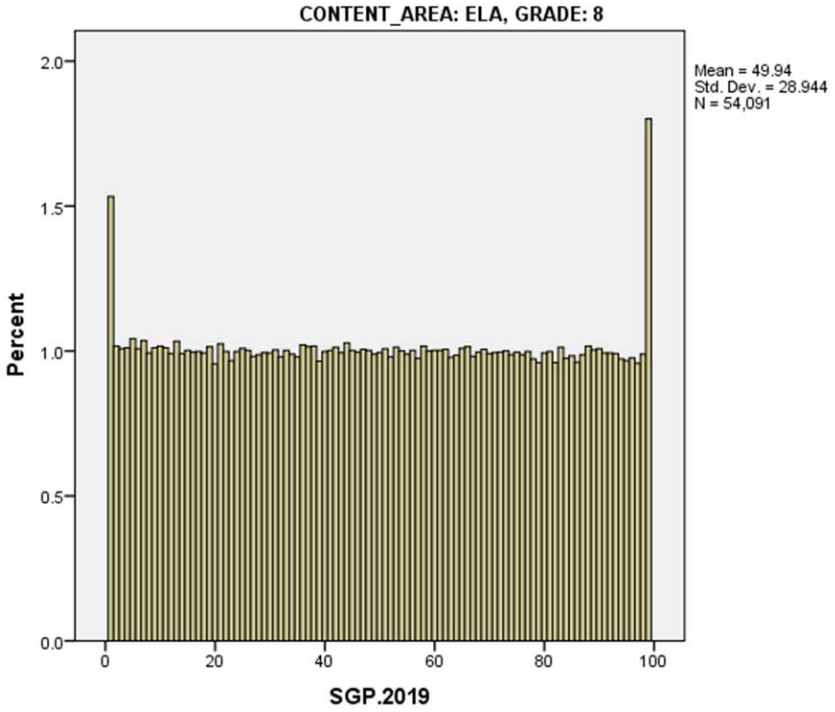


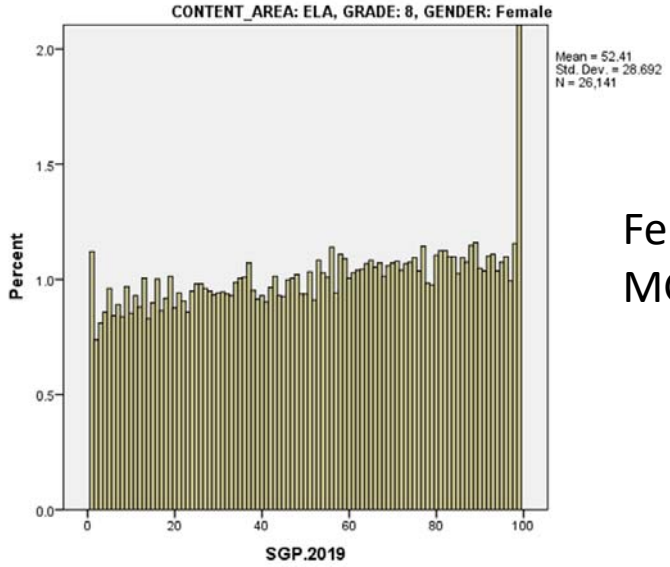
SGP to Scale Score Scatter plot for 2019: PSAT g9 Math by Gender



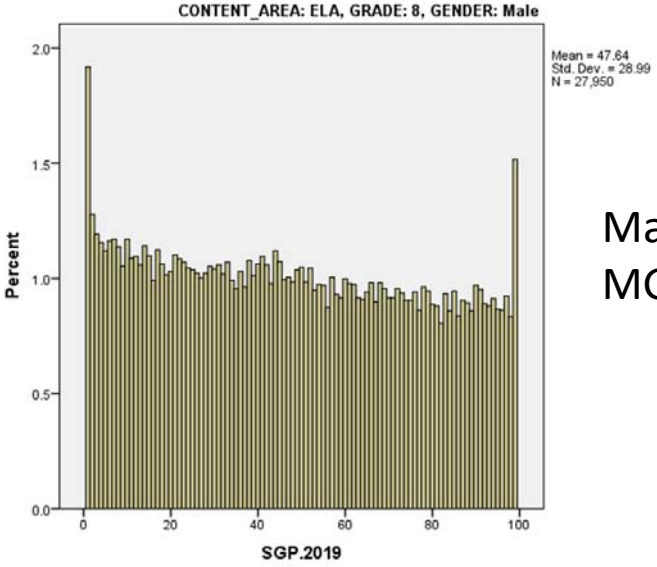
Student Growth Percentile Histograms- 2018 to 2019 CMAS g7 to CMAS g8 ELA

All Students
MGP = 50.0





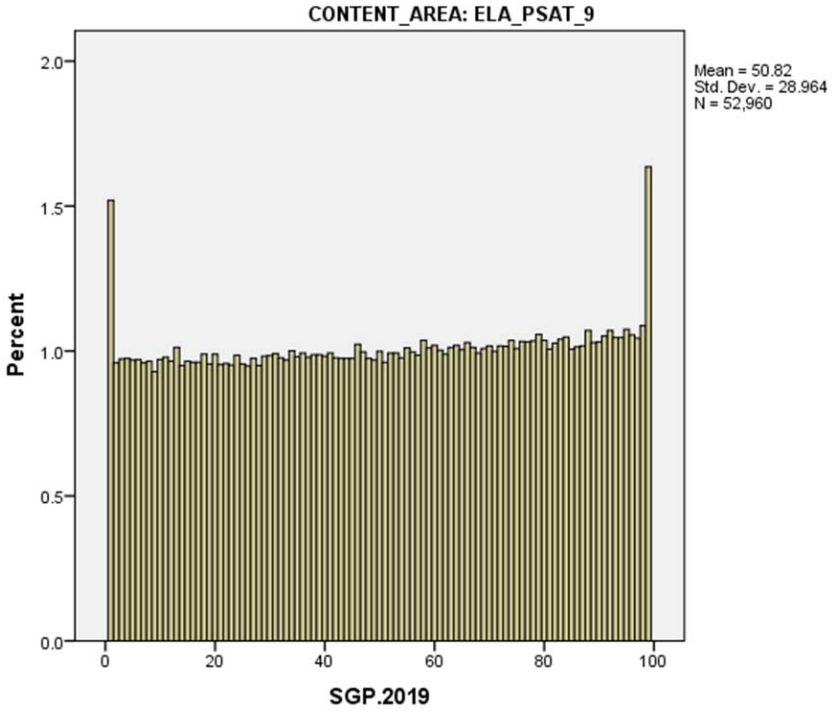
Female
MGP = 54.0

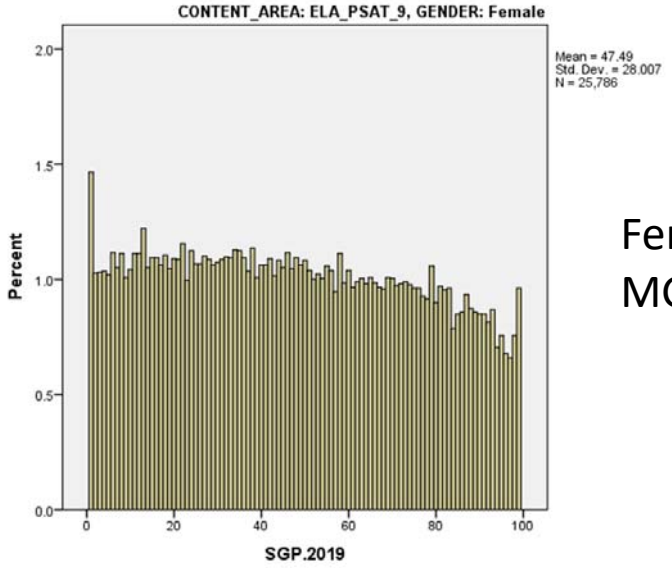


Male
MGP = 46.0

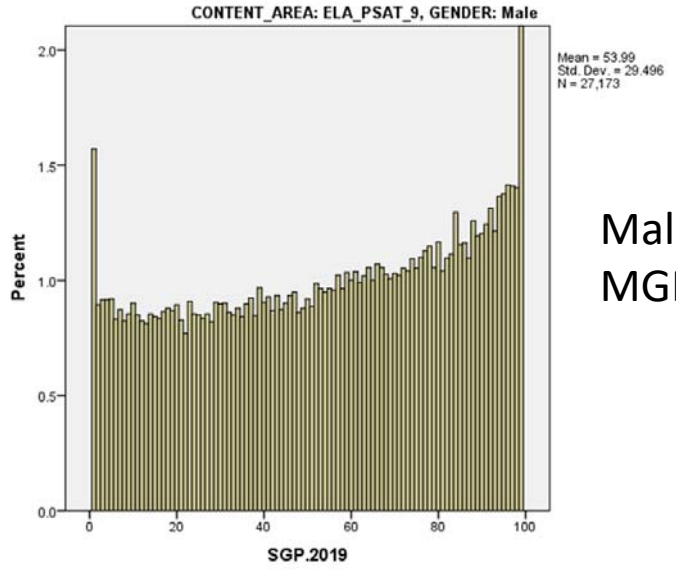
Student Growth Percentile Histograms- 2018 to 2019 CMAS g8 ELA to PSAT g9 EBRW

All Students
MGP = 51.0





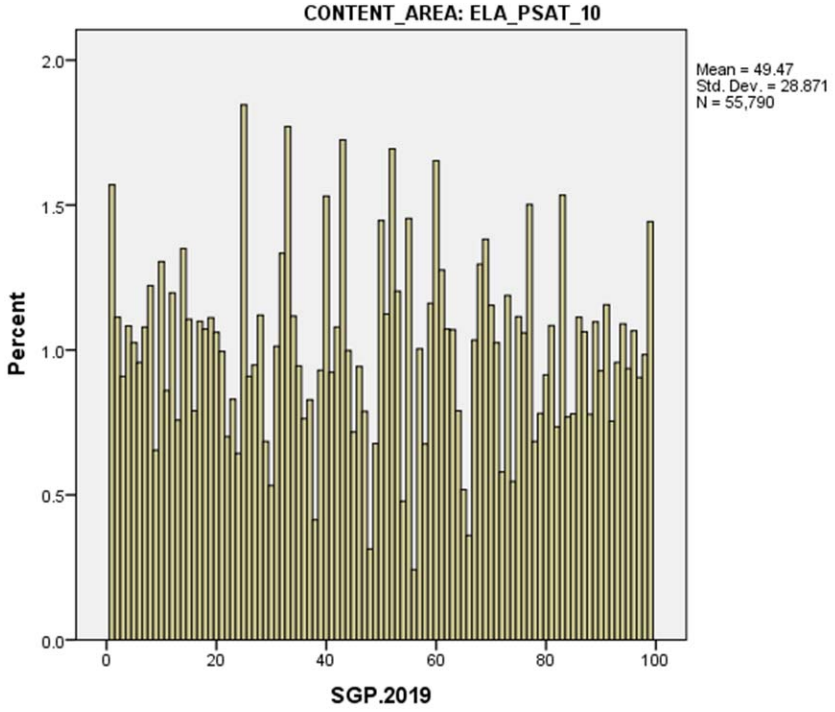
Female
MGP = 46.0

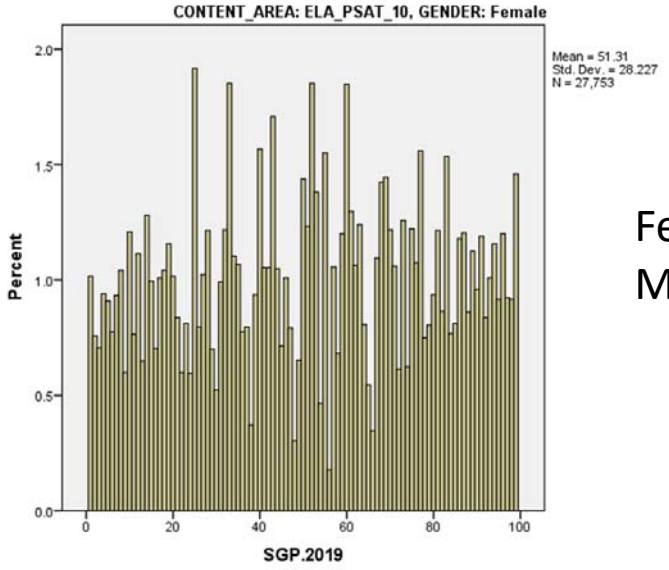


Male
MGP = 56.0

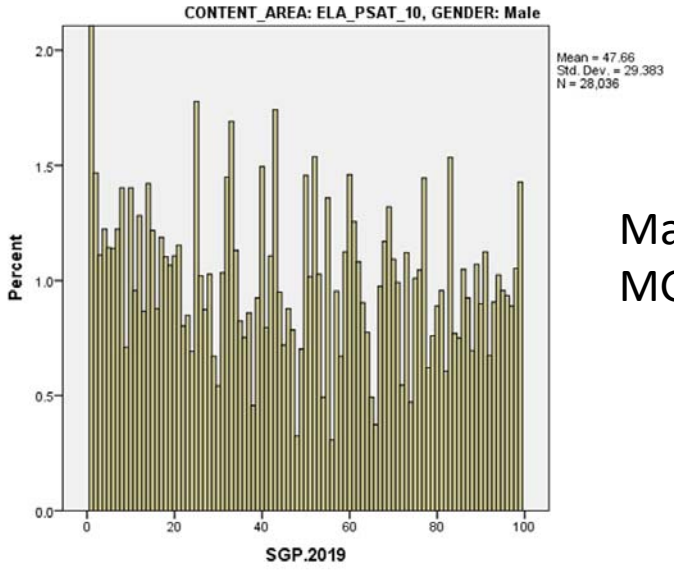
Student Growth Percentile Histograms- 2018 to 2019 PSAT g9 to PSAT g10 EBRW

All Students
MGP = 50.0





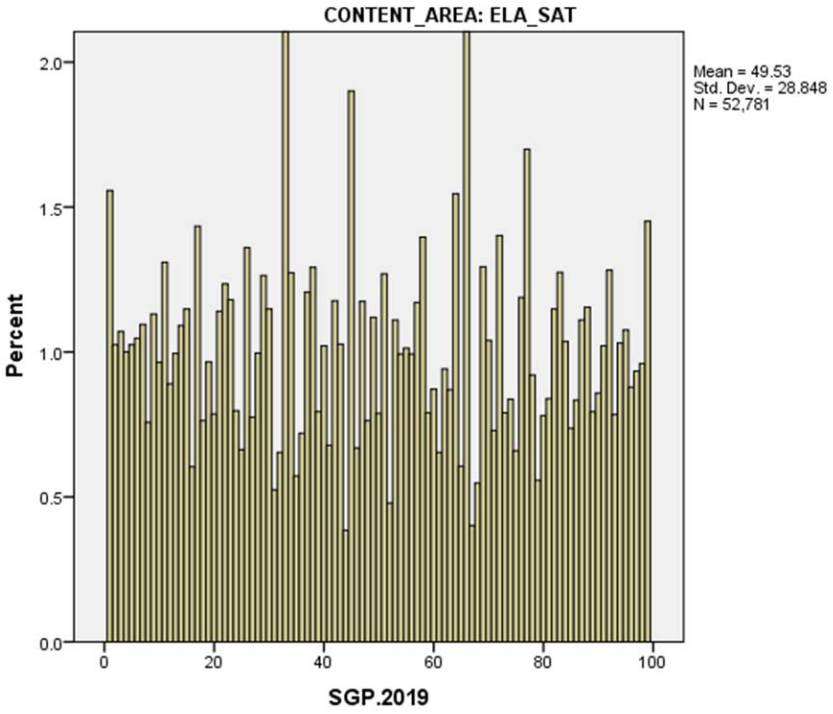
Female
MGP 52.0

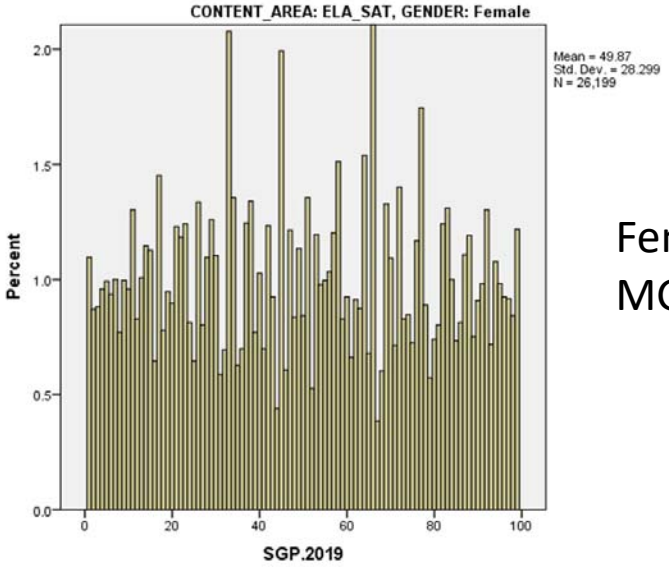


Male
MGP = 46.0

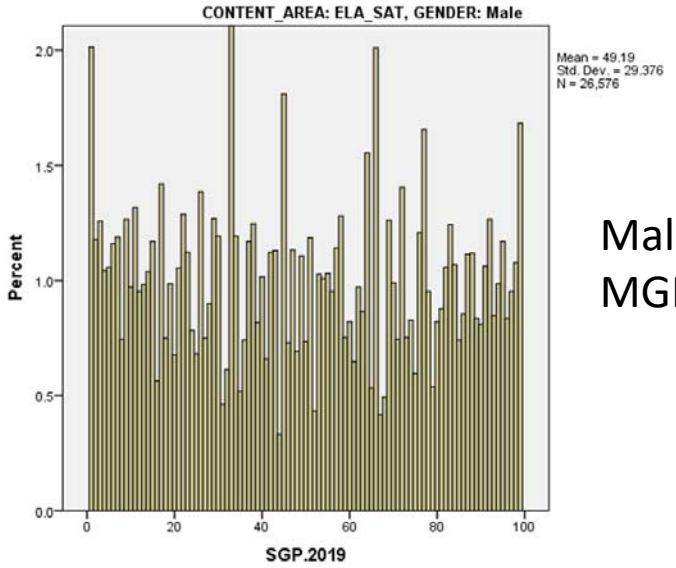
Student Growth Percentile Histograms- 2018 to 2019 PSAT g10 to SAT g11 EBRW

All Students
MGP = 49.0





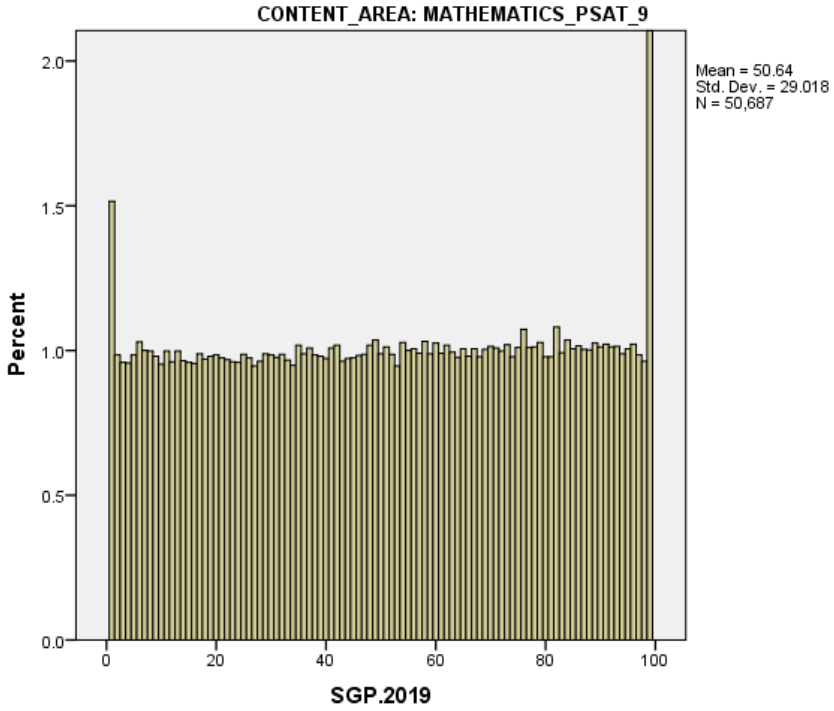
Female
MGP = 50.0

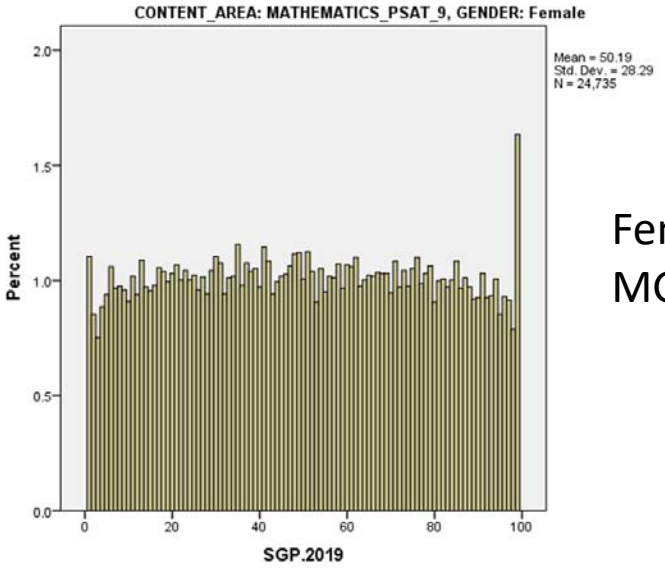


Male
MGP = 49.0

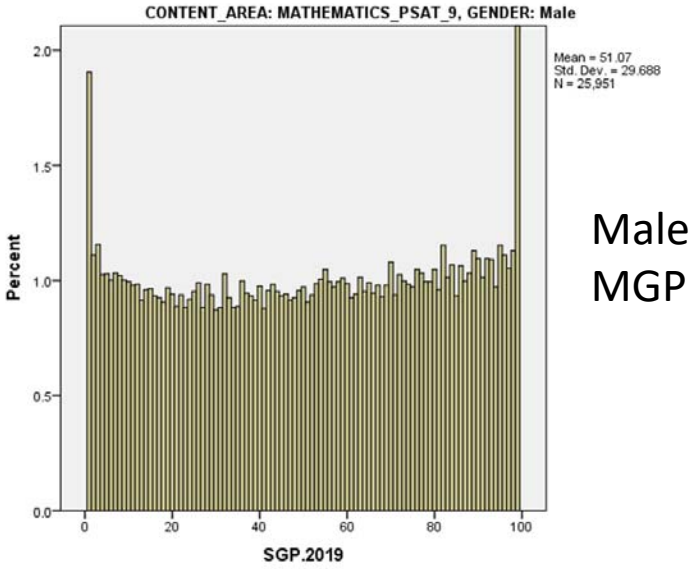
Student Growth Percentile Histograms- 2018 to 2019 CMAS g8 ELA to PSAT g9 Math

All Students
MGP = 51.0





Female
MGP = 50.0



Male
MGP = 52.0

Differing Assessment Constructs

- In reviewing these data, CDE staff has determined that the constructs underlying CMAS ELA and PSAT Evidence-based Reading and Writing (EBRW) do not align sufficiently to provide meaningful inferences about student growth for continued use between grades 8 and grades 9.

- Note that after the May TAP meeting, CDE did look into running growth from the CMAS Reading sub-claim score to PSAT EBRW, however it did not prove viable given the technical constraints of the growth model programming and the limited timeframe available to conduct such analyses.

Impact on State Accountability

- As the grade 9 EBRW student growth will likely result in less accurate inferences of aggregate school and district performance, the state has determined that these data are not appropriate for continued use in accountability calculations.
- For this reason, starting in 2019 and moving forward, grade 9 EBRW growth data will not be included on the school and district performance frameworks.
- Based on the recommendation from the TAP and analysis ensuring the construct comparability in math between CMAS and PSAT, CDE has decided to continue including grade 9 Math growth on the performance frameworks.

Request to Reconsider Using g9 EBRW Growth

- Recognizing the removal of g9 EBRW growth as a change from the previous year and following historical practice, CDE has identified 4 districts and 12 schools that would have received a higher rating from the inclusion of these data in 2019 and is offering an expedited request to reconsider process to remove any negative impact on the final rating for these systems.

- For districts submitting a request to reconsider, CDE will consider compound requests using the g9 EBRW growth impact in addition to other criteria that can be incorporated into the R2R scoring rubric. (Note that participation requests are not eligible for combination).

Availability of Data

- Student level g9 EBRW growth data will not provided to districts or schools in 2019 and moving forward.
- CDE will be releasing upon request a data file containing 2019 school and district-level g9-11 PSAT/SAT EBRW median growth percentiles. This file will also provide for each school or district the rating and total percent of points that would have been earned on the performance frameworks if these data had been included in 2019 accountability calculations.
- Note this file will not be available in future years as CDE will no longer run g9 EBRW growth calculations.

BREAK

Potential SBE Framework Changes

Marie Huchton

Colorado's system of school and district accountability is primarily designed to provide valid and actionable information regarding the progress of all students toward meeting the state academic standards and prioritize support for schools and districts identified for academic improvements.

EVALUATE

Local	*State*
Schools and districts analyze state and local data.	**CDE** creates **School & District Performance Frameworks**. **CDE** recognizes areas of success and identifies schools and districts for additional support based on student academic outcomes.

ASSESS NEEDS AND PLAN

Local	*State*
Schools and districts work with **local communities** to assess needs and select strategies to support continuous improvement.	**CDE** supports the **Unified Improvement Planning** process for all school and districts informed by state and local data.

IMPLEMENT

Local	*State*
Schools and districts apply for additional resources and implement selected strategies for improvement.	**CDE** allocates **resources and supports** in alignment with school and district needs identified in Unified Improvement Plans.

INTERVENE

Local	*State*
Low performing schools and districts take more rigorous action if student performance remains consistently low.	**The State Board of Education** works with CDE and districts to determine more rigorous action steps for persistently low performing systems through **Performance Watch**.

Considerations

Timing:

- Flexibility on year of implementation, but no sooner than 2020 for informational purposes and 2021 for points

Policy:

- State statute specifies the minimum required school and district rating categories
- State Board of Education sets cut points between performance levels

Considerations For Accountability Metrics: Commissioner's Thoughts

- Consider that different schools have vastly different starting points (in some systems, many students that come into the school can be years behind). Thus, metrics that show how well the system is *growing* students is an important consideration.
- Accountability systems should consider how much adults are contributing to the learning of students and catching them up to grade-level achievement.
- An effective school performance framework strives to measure progress and outcomes, rather than inputs.

Feedback from Fall 2018 Stakeholder Meetings

Timeline for Adoption

Allow the appropriate amount of time to have thoughtful accountability conversations

- ✓ Gather authentic stakeholder feedback on all areas that will be addressed during the rulemaking process
- ✓ Allow appropriate lead time for any adjustments
- ✓ Develop and evaluate models to determine the impact of any potential adjustments

Colorado Values

Ensure that Colorado's values are reflected in accountability policy

- ✓ Define and adhere to the purpose of school and district accountability
- ✓ There is a value reflected in the inclusion of growth measures on performance frameworks

Feedback from Fall 2019 Stakeholder Outreach

Distinction Schools

Stakeholders conveyed varying opinions about the need for and value of an additional Distinction category for schools

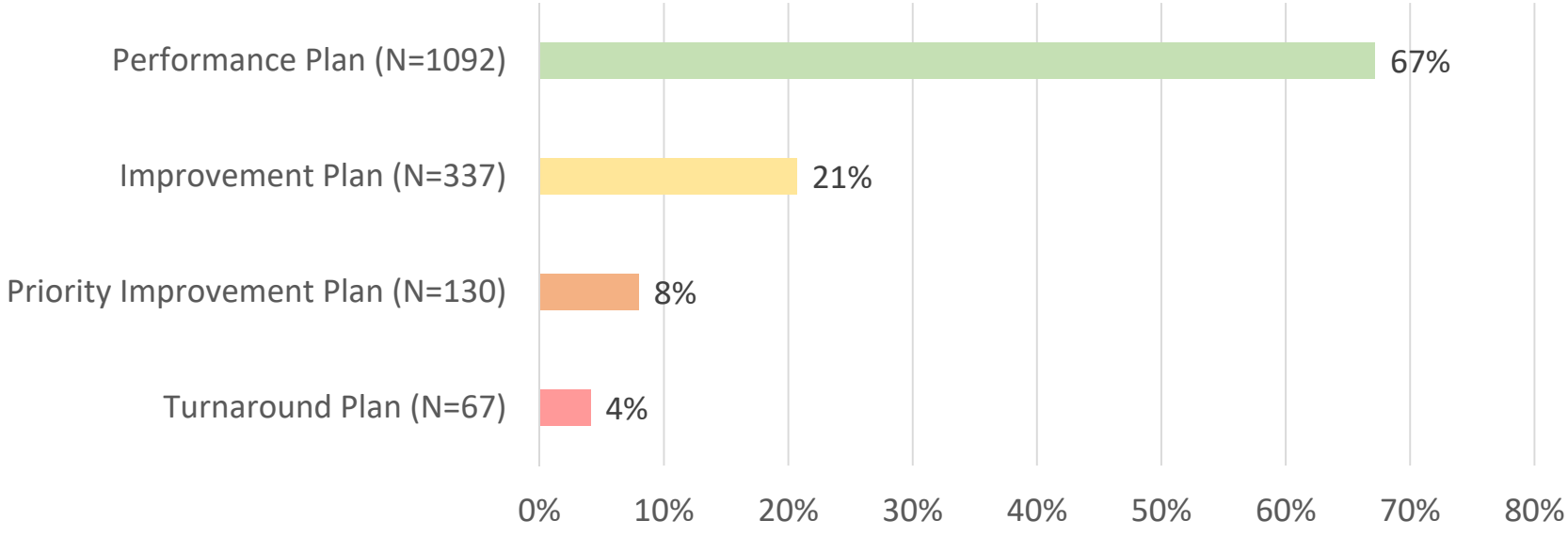
- ✓ Would provide positive feedback to schools
- ✓ Districts already have the authority to accredit schools with distinction based on local values
- ✓ Concerns around creating a negative competitive environment among schools within a district

Refined Possible Models

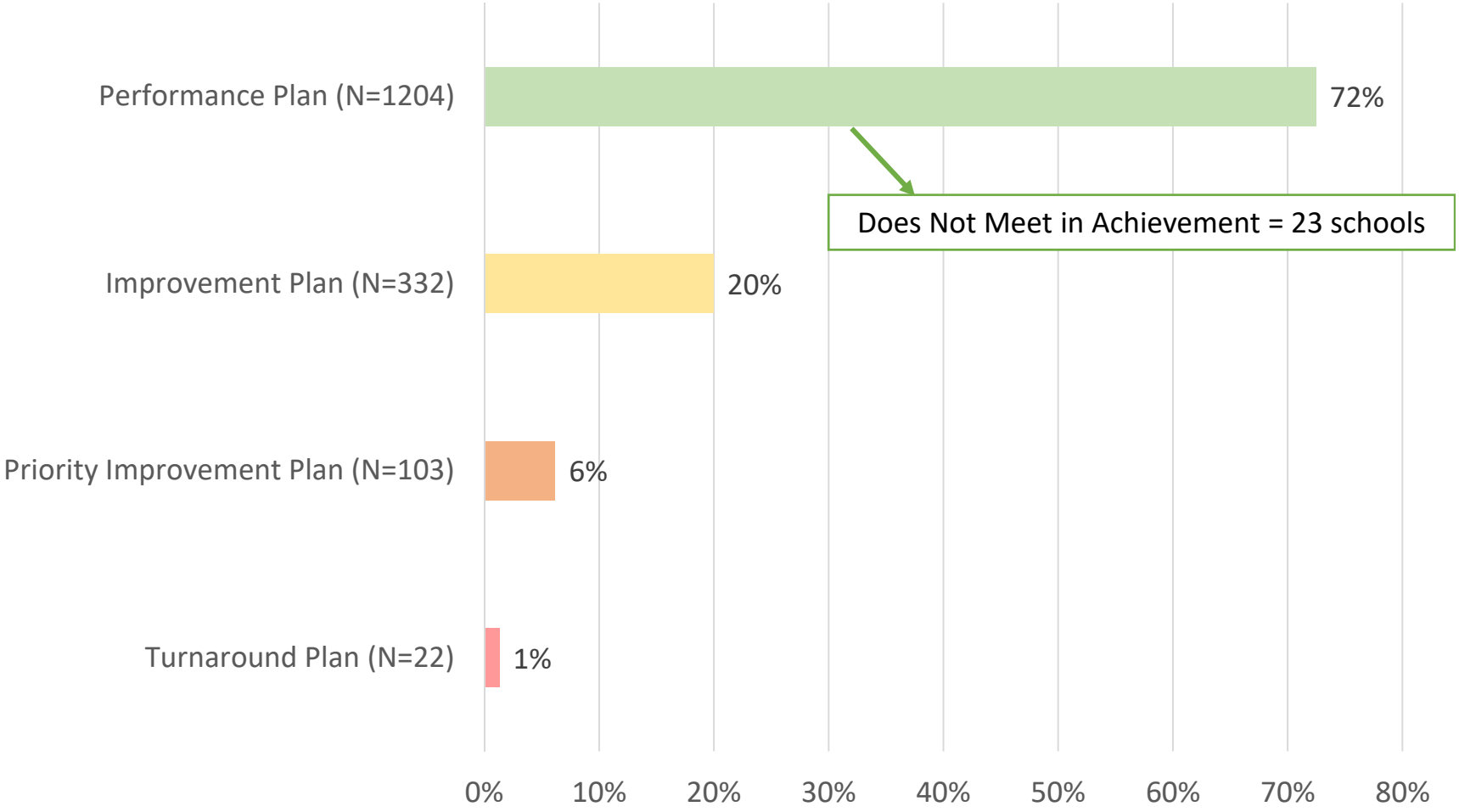
2010 School Ratings (All EMH combinations)

The state board has discretion over how to identify schools within each plan type or accreditation rating but must include the performance indicators in statute.

Rating cut points were set in 2010 and the ratings reflected:



2019 School Performance Framework Ratings Using Current Calculation Methodology (All EMH combinations)

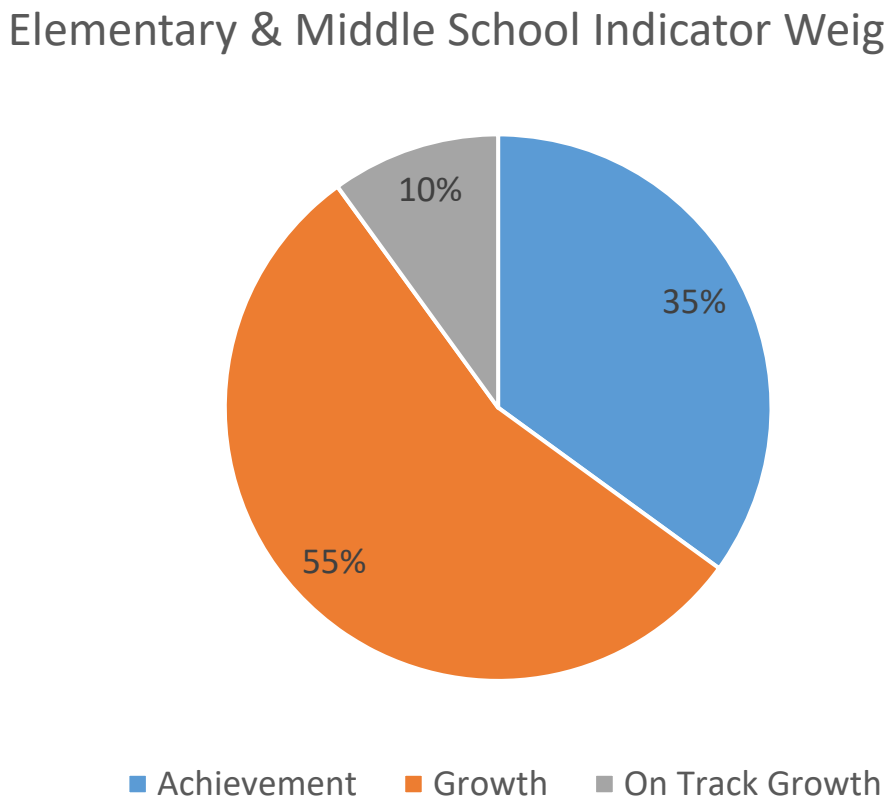


Refined Options for Shifting Performance Framework Calculations

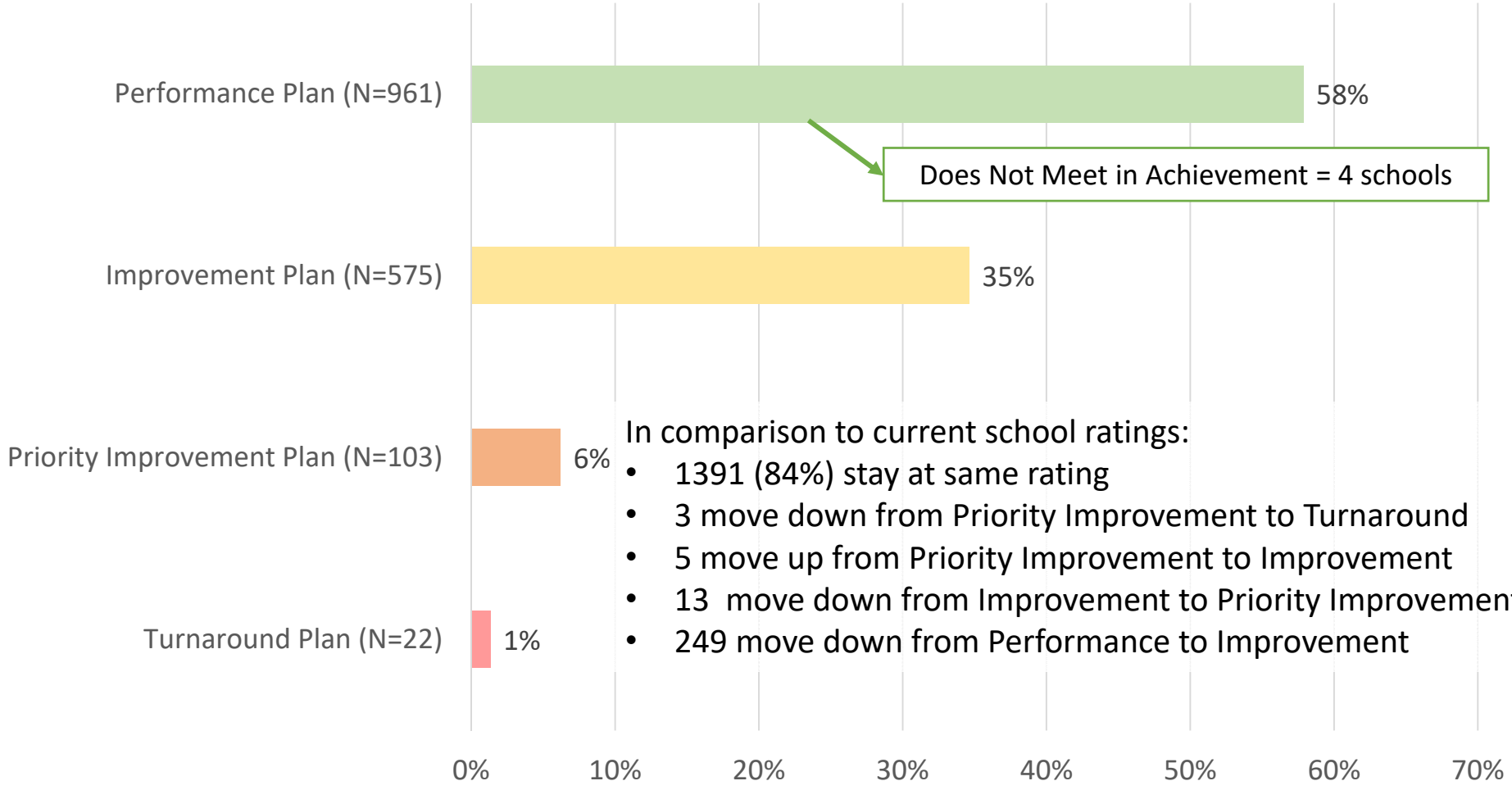
At the request of State Board members, CDE refined the scenarios presented at the June board meeting: keeping the 2 year Catch Up metric, shifting the % of points earned cut-score between Improvement and Performance ratings, and adjusting the proportion of schools identified for Distinction (see table below). Impact results from each of these scenarios will be shared in the following slides.

	Catch Up Timeframe for On Track Growth	Increase Cut-score for Performance Rating	Distinction Category?
Scenario A	2 years	Up 6%	No
Scenario B	2 years	Up 6%	Yes- 10%
Scenario C	2 years	Up 8%	No
Scenario D	2 years	Up 8%	Yes- 10%

Indicator Weighting with new Elementary & Middle On Track Growth Metric



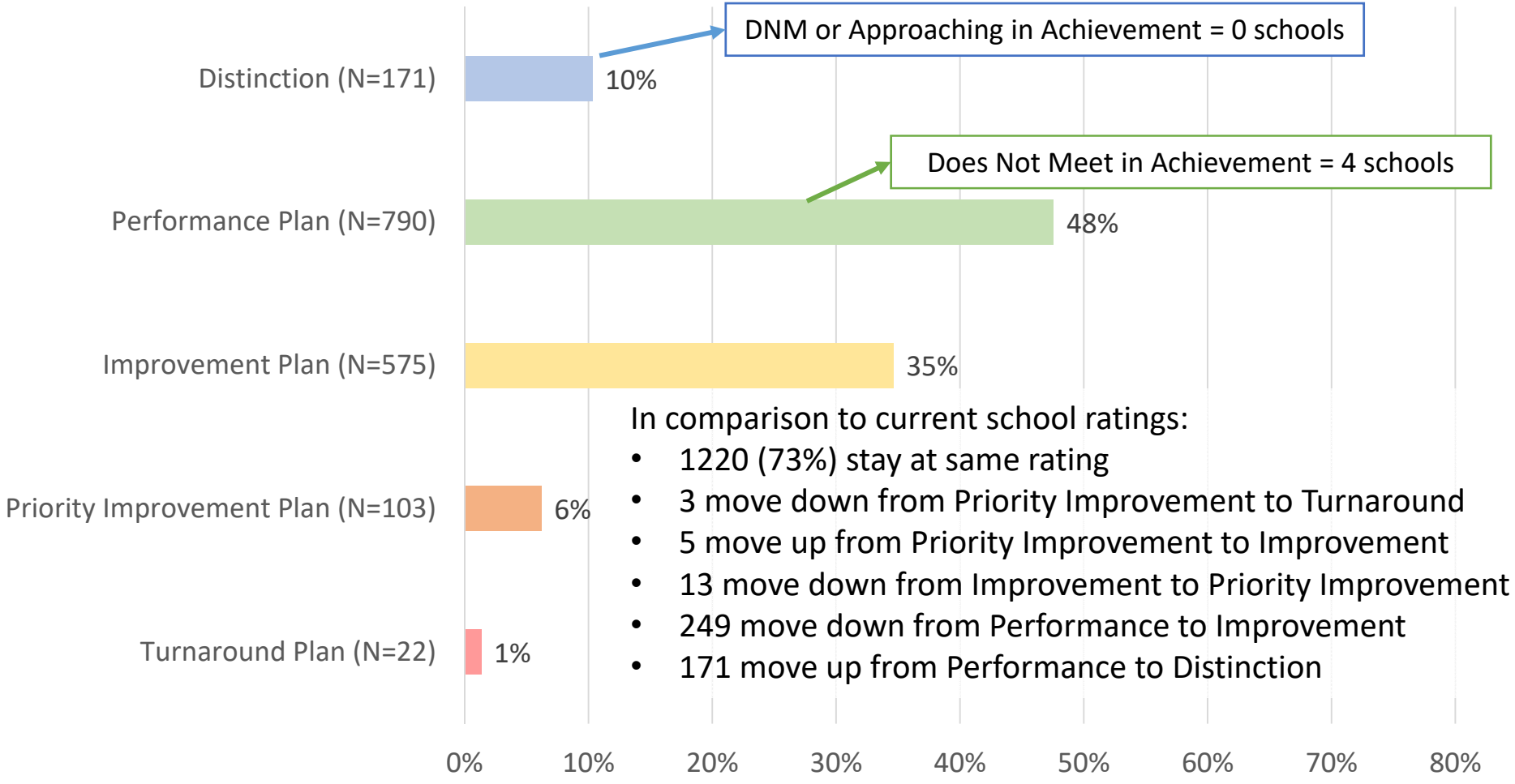
Scenario A: 2 years to Catch Up, Performance Cut Increases by 6%, No Distinction Category



In comparison to current school ratings:

- 1391 (84%) stay at same rating
- 3 move down from Priority Improvement to Turnaround
- 5 move up from Priority Improvement to Improvement
- 13 move down from Improvement to Priority Improvement
- 249 move down from Performance to Improvement

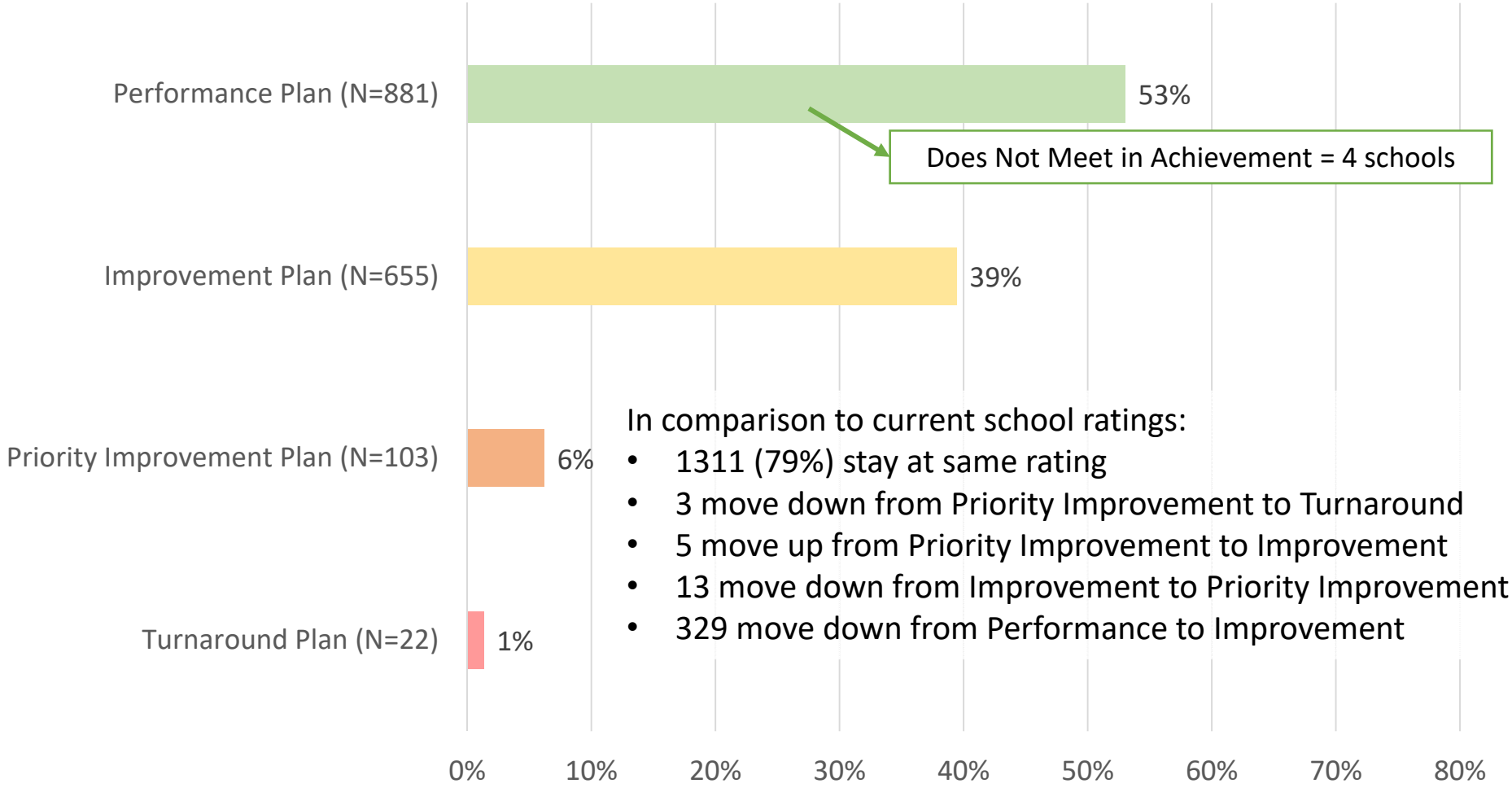
Scenario B: 2 years to Catch Up, Performance Cut Increases by 6%, Adds Distinction Category at 10%



In comparison to current school ratings:

- 1220 (73%) stay at same rating
- 3 move down from Priority Improvement to Turnaround
- 5 move up from Priority Improvement to Improvement
- 13 move down from Improvement to Priority Improvement
- 249 move down from Performance to Improvement
- 171 move up from Performance to Distinction

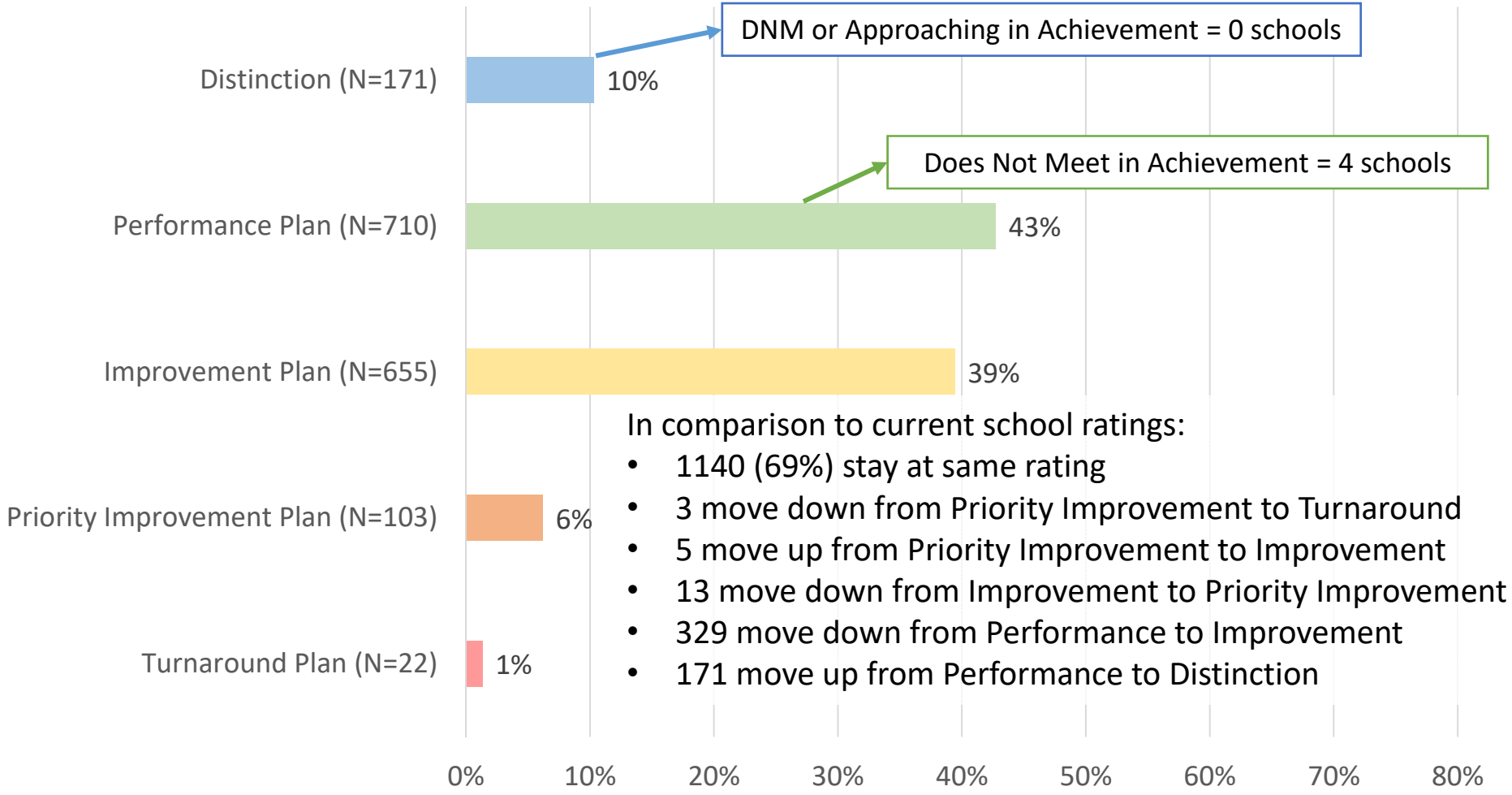
Scenario C: 2 years to Catch Up, Performance Cut Increases by 8%, No Distinction Category



In comparison to current school ratings:

- 1311 (79%) stay at same rating
- 3 move down from Priority Improvement to Turnaround
- 5 move up from Priority Improvement to Improvement
- 13 move down from Improvement to Priority Improvement
- 329 move down from Performance to Improvement

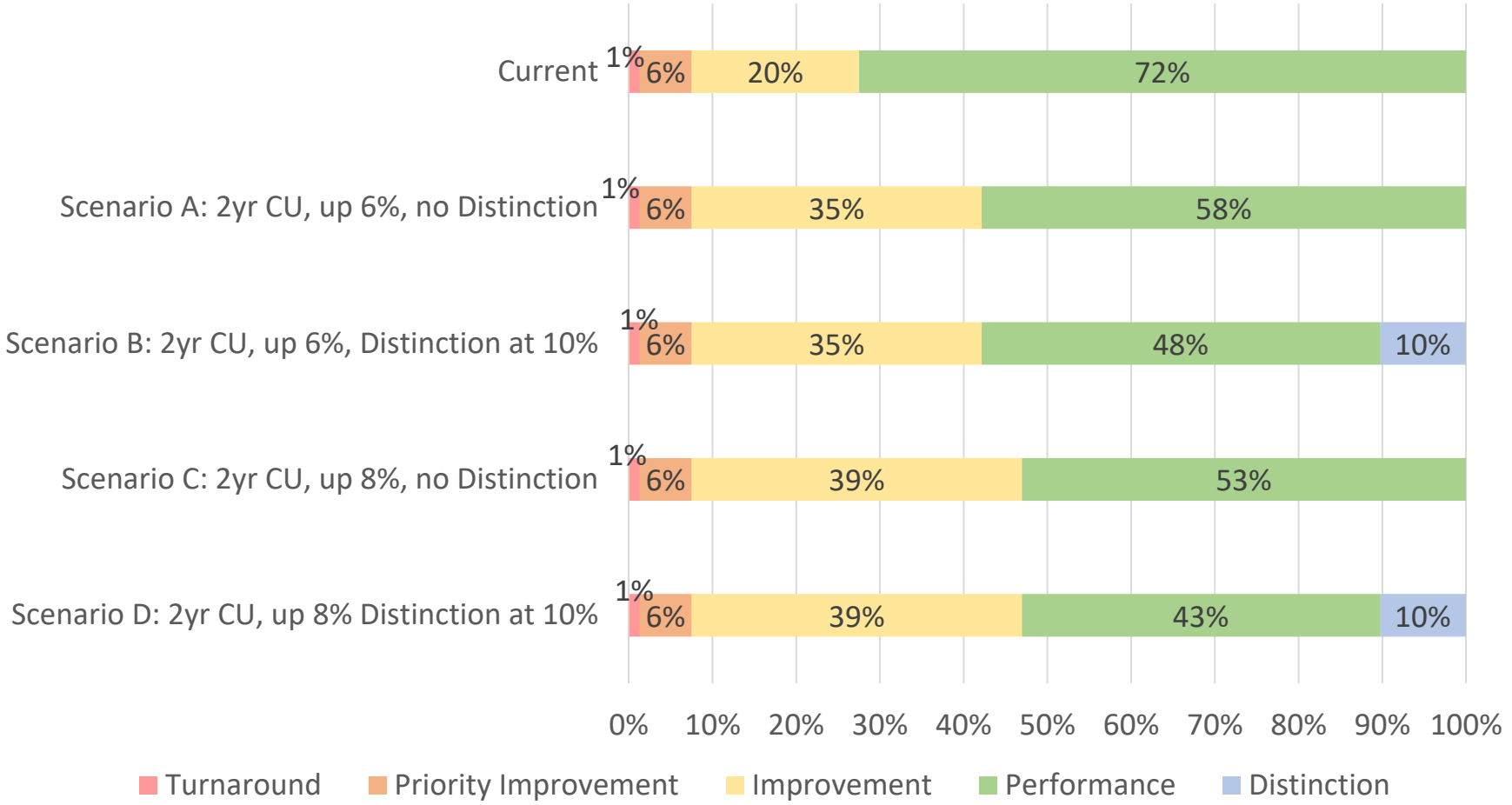
Scenario D: 2 years to Catch Up, Performance Cut Increases by 8%, Adds Distinction Category at 10%



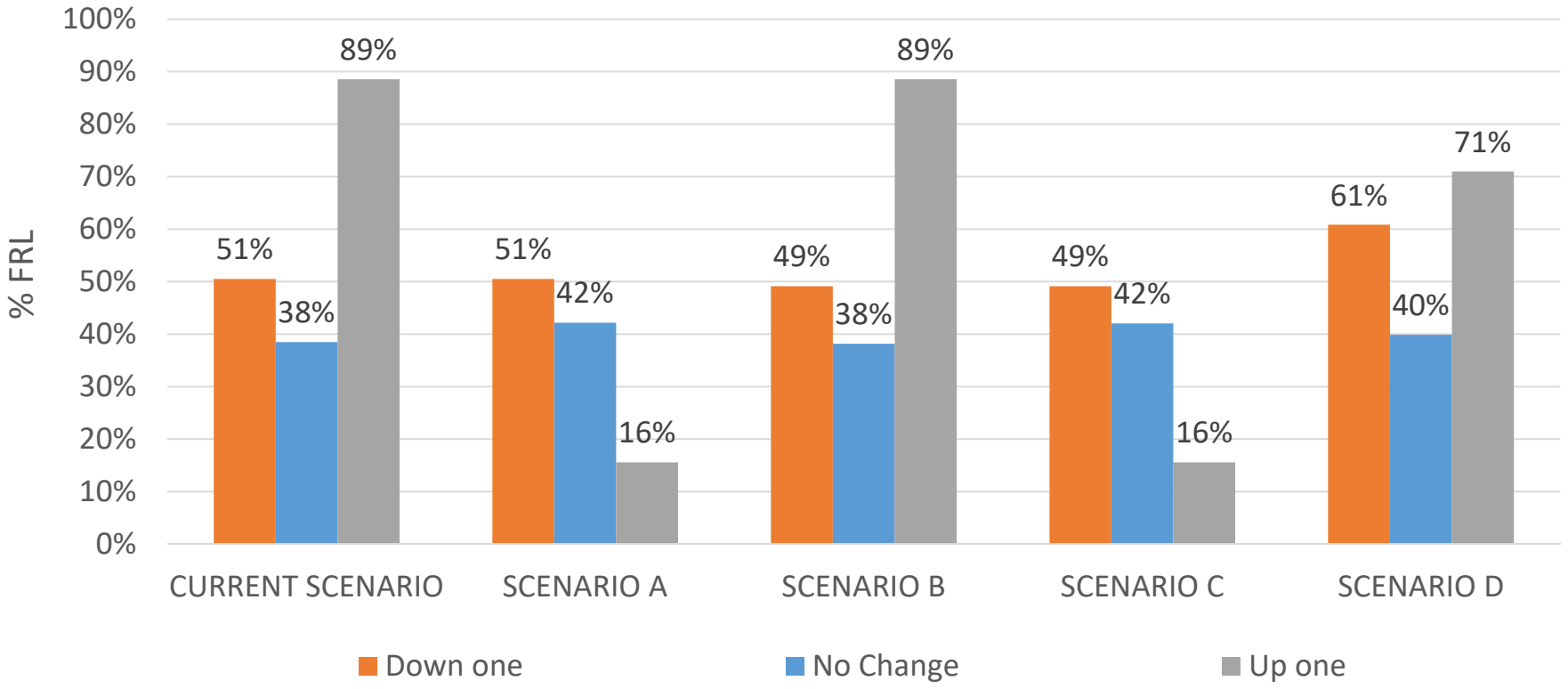
In comparison to current school ratings:

- 1140 (69%) stay at same rating
- 3 move down from Priority Improvement to Turnaround
- 5 move up from Priority Improvement to Improvement
- 13 move down from Improvement to Priority Improvement
- 329 move down from Performance to Improvement
- 171 move up from Performance to Distinction

Scenario Comparison: All Scenarios

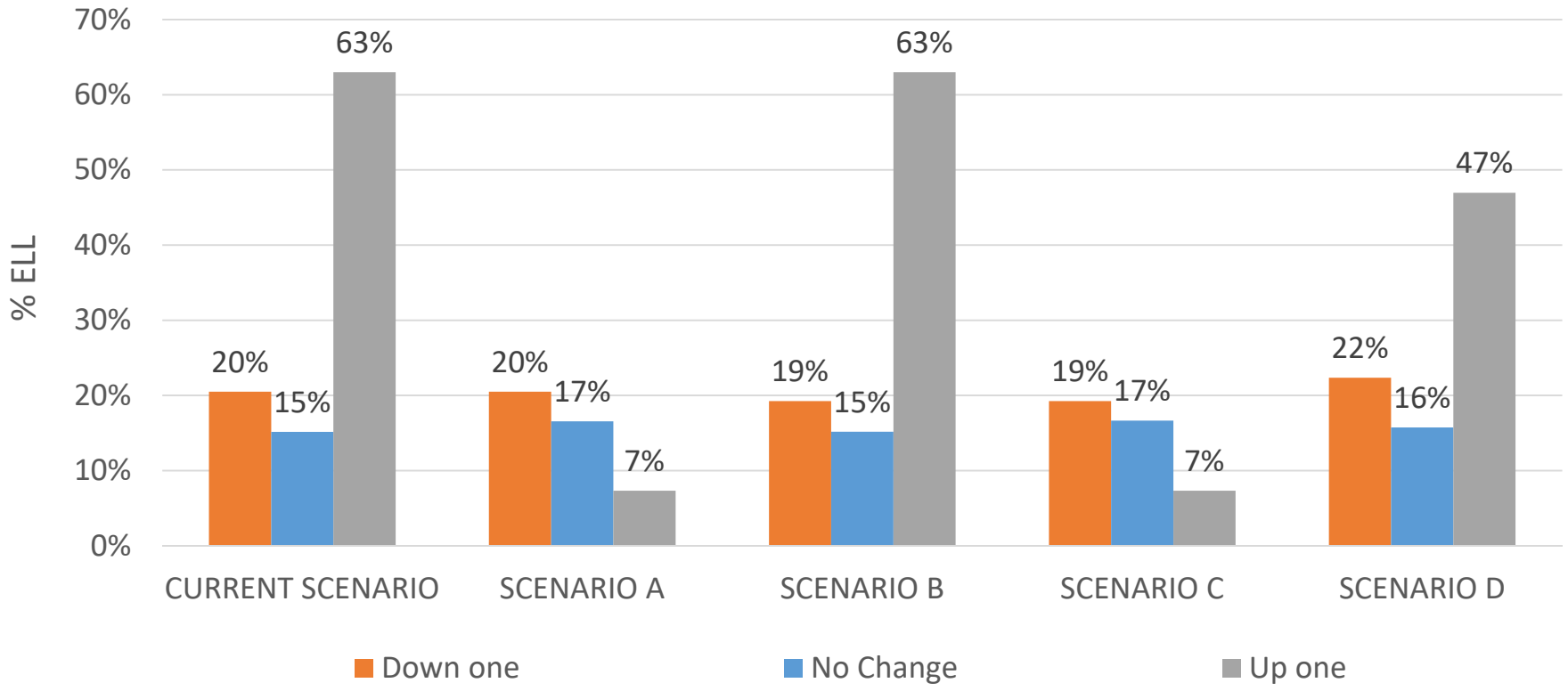


Scenario Impacts by School Demographic- Percent Eligible for Free- or Reduced-Price Lunch Programs



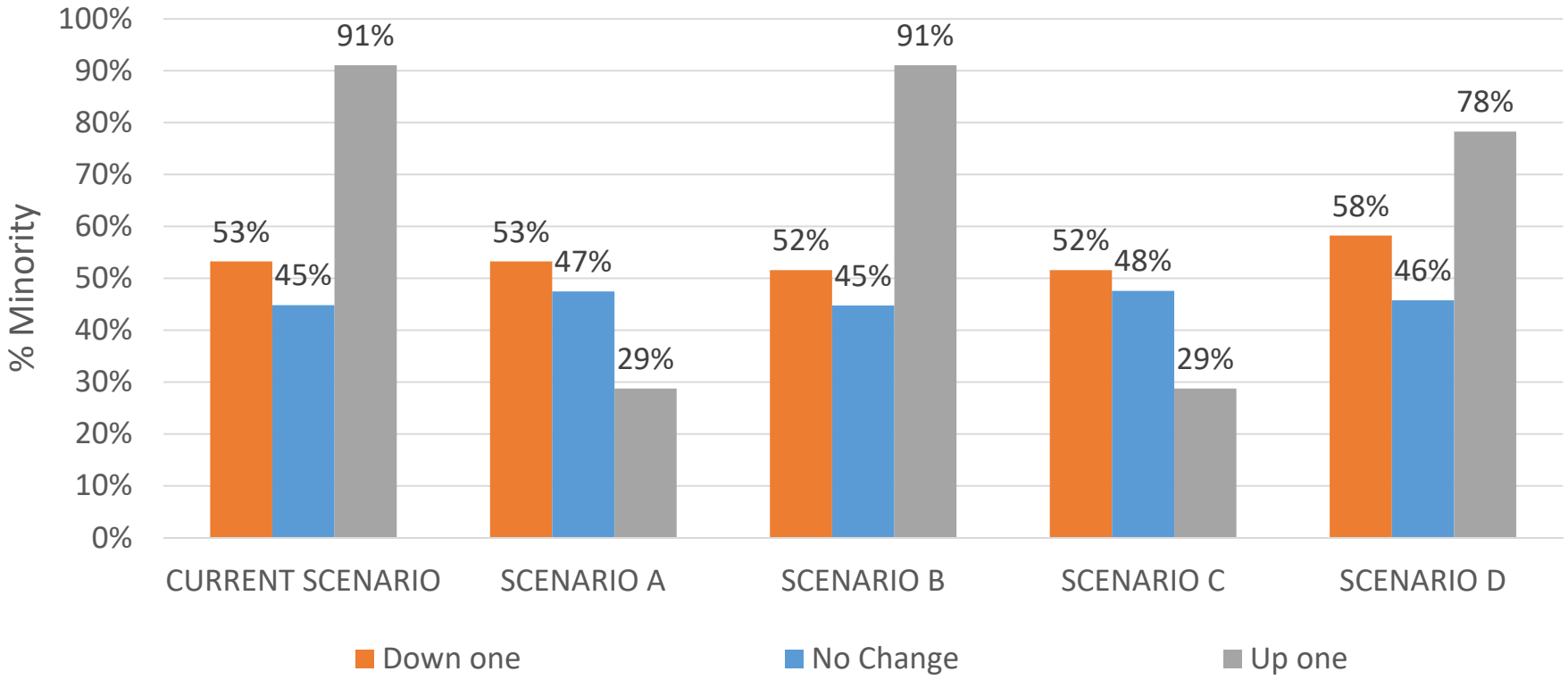
Rating Change	School counts				
	CURRENT	SCENARIO A	SCENARIO B	SCENARIO C	SCENARIO D
Down one	31	265	265	345	345
No Change	1616	1391	1220	1311	1140
Up one	14	5	176	5	176

Scenario Impacts by School Demographic- Percent English Learners



Rating Change	School counts				
	CURRENT	SCENARIO A	SCENARIO B	SCENARIO C	SCENARIO D
Down one	31	265	265	345	345
No Change	1616	1391	1220	1311	1140
Up one	14	5	176	5	176

Scenario Impacts by School Demographic- Percent Minority



Rating Change	School counts				
	CURRENT	SCENARIO A	SCENARIO B	SCENARIO C	SCENARIO D
Down one	31	265	265	345	345
No Change	1616	1391	1220	1311	1140
Up one	14	5	176	5	176

TAP Feedback

- What feedback would the TAP like to provide the State Board of Education in regards to the potential Framework Rating changes scenarios?
- Which scenario (if any) does the TAP recommended the State Board adopt?

Technical Advisory Panel

- Meeting Summary:
  - Suggested future analysis
  - TAP recommendations from this meeting
- Public Comment
- Close Meeting
  - Next Scheduled Meeting, Friday, October 25th, 9-noon (CDE)

What questions do you have?