Appendix C to the 2017-2018 Colorado CMAS/SAT Growth Model Report:
Investigation of Potential Ceiling and Floor Effects.

Adam R. VanIwaarden

Damian W. Betebenner

National Center for the Improvement of Educational Assessment (NCIEA)

October 2018

Abstract

DRAFT REPORT - DO NOT CITE!

1 Introduction

In the 2017-2018 academic year Colorado transitioned completely from the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium assessments to the Colorado Measures of Academic Success (CMAS) assessments in grades 3-8 in English Language Arts and mathematics and the PSAT9, PSAT10 and SAT assessments administered in 9th, 10th and 11th grades respectively. The transition included numerous changes to the assessment system including the incorporation of new performance standards. As other states have gone through similar assessment transitions, many have observed ceiling and floor effects in the new assessments (i.e. a relatively large proportion of students scoring at/near the scale extremes). These assessment ceilings/floors can telegraph onto the Student Growth Percentile (SGP) calculations.

With perfect data and model fit, the expectation is that the majority of SGPs for students scoring at or near the lowest obtainable scale score (LOSS) will be low (preferably less than 5 and not higher than 20), and that SGPs for students with the highest obtainable scale score (HOSS) will be high (higher than 95 and not less than 80). Ceiling effects in growth measures are somewhat more problematic than floor effects because students that consistently receive the highest scores are given lower than expected growth percentiles and are therefore negatively impacted. Conversely, the consistently lowest achieving students have higher estimated SGPs than expected. This could possibly conceal unacceptably low growth that might otherwise be identified and addressed.

In part, these problems are caused by the way in which a "percentile" is most typically defined to begin with, and the inability of the assessments (and therefore the SGP model) to make granular distinctions between kids who score at the extremes of the test year after year. As an example, if a group of students were given a relatively easy test and 20% of them received a perfect score, these students would be defined as being in the 80th percentile of achievement because they scored higher than 80% of their peers. This is somewhat misleading however, because their score was *equal to* or greater than *all* of their peers and so could be also described as achieving at the 99.9^{th} percentile under other equally valid definitions.

To extend this heuristic from achievement to growth, if 50% of those top students also scored perfectly on the next test, we might estimate that they had 50th percentile growth. Although there is nothing *technically* incorrect about this estimate since their growth is fairly typical for their academic peer group, it is an inadequate or unsatisfactory assessment of their growth because they have consistently attained the highest levels. Furthermore, if it is typical for their peers to maintain perfect scores then even small deviations from a perfect score could produce *low growth* SGP estimates.

Typically only a few students are impacted by ceiling and floor effects, making them difficult to detect using traditional SGP diagnostic tools. "Ceiling/Test Effects" indicators have been added to the SGP model goodness of fit plots and this Appendix to the annual technical reports provides even more rigorous diagnostic and descriptive analyses. This report includes:

1. Scatter plots of the current and the most recent prior year's test score distributions to indicate ceilings or floors in the data used in growth calculations.
2. Box plots showing the range and distribution of SGPs for the highest and lowest achieving students in the current year.

2 Prior- and Current-Year Score Distributions

The marginal and conditional distributions of test scores can serve as a preliminary indicator of potential ceiling or floor effects in the calculation of student growth percentiles. Some minor problems could occur if these characteristics are present in either the prior- or current-year scores, and are particularly likely when present in both.

The plots in the following sections depict distributions for the current year and the most recent prior year used in the SGP calculations for each content area and grade level. These plots start with a basic scatter plot of each student's scores to show their conditional (joint) score distributions, and each point is depicted as the estimated SGP value based on their scores[1]. On top of this is layered 1) **green contour lines** to provide a better sense of the score distribution density, 2) three **non-linear magenta lines** identifying the bivariate relationship between prior and current scores at the 5th, 50th and 95th percentiles[2], 3) **red** dotted lines that represent the cutoff for the highest and lowest 25+ current scale scores (corresponding with the first and last rows of the fit plot table), and 4) **rug plots** that depict the marginal distributions (prior scores shown in blue and current scores in red).

Ceiling or floor effects may be indicated by dark shaded SGP values in the extreme top-right or bottom-left corners of the plots. This suggests that staying at the extremes is common, which may lead to odd growth estimates for these high/low achieving students.

2.1 CMAS Content Areas

We see several areas of concern in both CMAS content areas and in nearly every grade. Both ceiling and floors are apparent in the 2017 and 2018 Colorado grade-level PARCC/CMAS data, which appear as clustered points in the scatter plots along the lower-left and upper-right corners of the plots. Dark lines at the extremes of the rug plots also confirm the high concentration of scores at the extremes. If a significant number of students have attained the HOSS in both prior years (2016 and 2017) we can expect to see SGPs that are much lower than 99. The results shown in Figures C.1 and C.2 require further investigation in section 3 to evaluate the extent of the test ceiling and floor effects may have had on the student growth estimates.

---

[1]Note that many SGPs are estimated using more than one prior score, and therefore plots may show SGP results from multiple analyses and/or varying SGPs for identical score combinations.

[2]Produced using quantile regression similar to, but not the same as, that used in calculation of the SGPs.
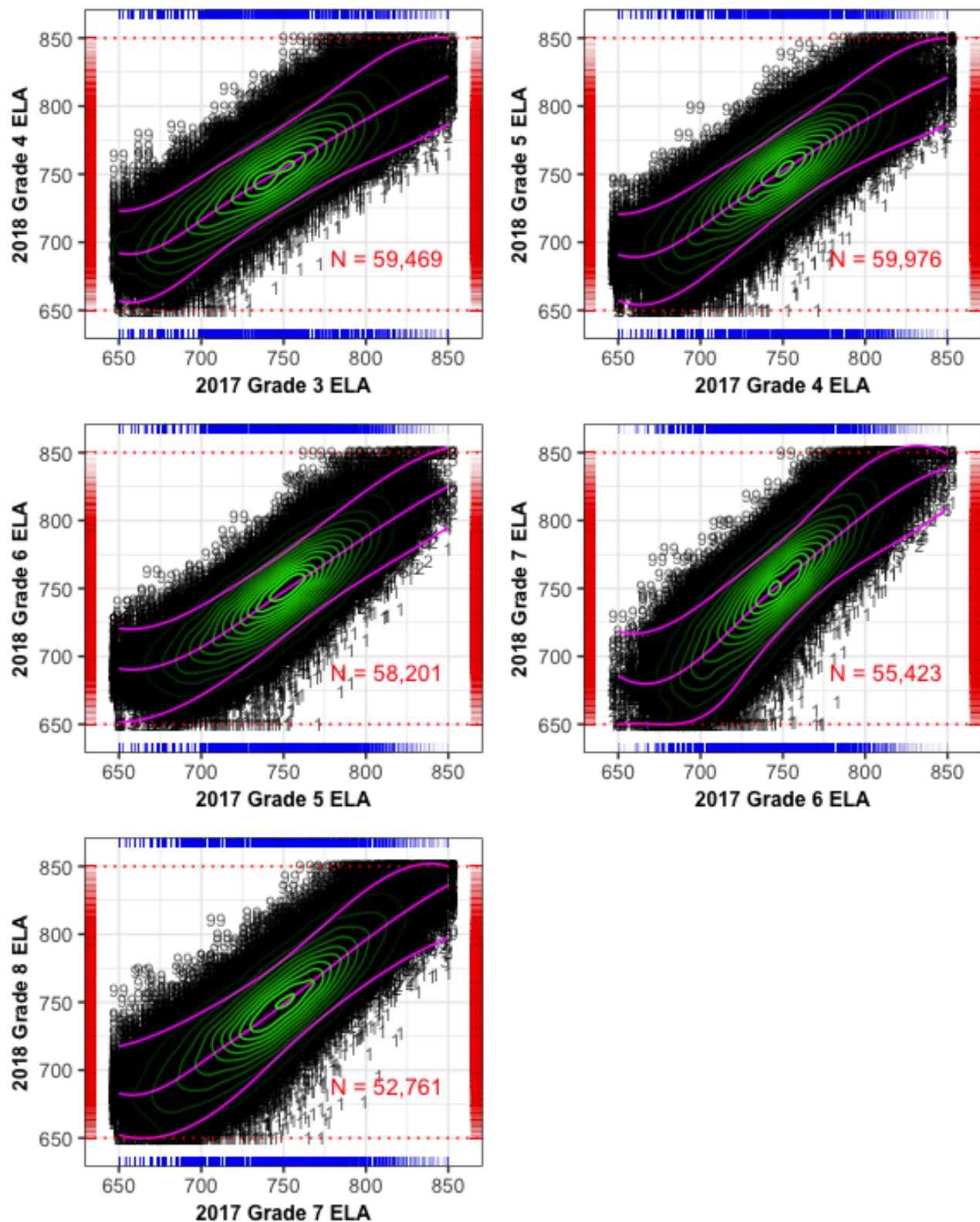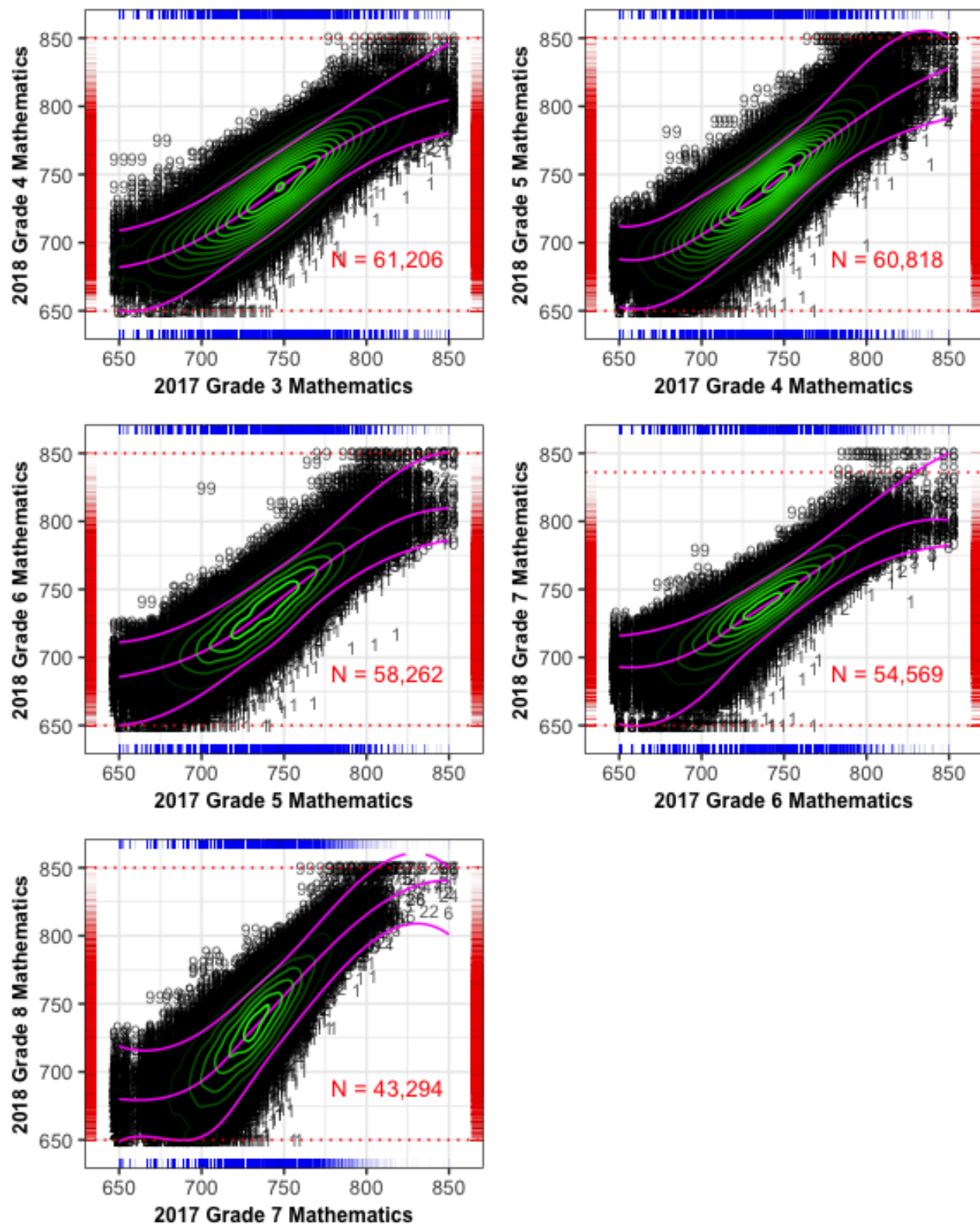
Fig. C.1: Conditional distributions of current and prior scale scores: ELA.

Fig. C.2: Conditional distributions of current and prior scale scores: Mathematics.

2.2   PSAT/SAT Subjects

The conditional density plots for the PSAT/SAT subjects are displayed below. The most recent prior is used in each plot to provide insight on the specific academic peer group on which each analysis is based. There are multiple such cohort norm groups for the Mathematics PSAT analyses.

As with the CMAS analyses, we again see cause for concern with the PSAT/SAT analyses, particularly with potential ceiling effects evident by the clustering of scores and convergence of the quantile regression lines in the upper-right corner of the scatter plots. The Mathematics PSAT/SAT analyses appear to be particularly affected. As with the CMAS scatter plots, the results shown in Figures C.3 through C.8 suggest the need for further investigation in section 3 to evaluate the extent to which test ceilings may have affected growth estimates.

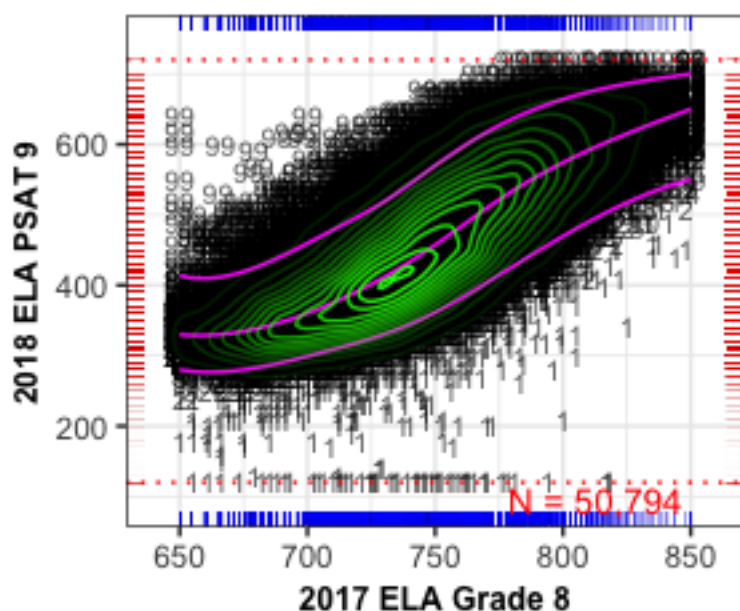**Fig. C.3:**   Conditional distributions of current and prior scale scores: ELA PSAT 9

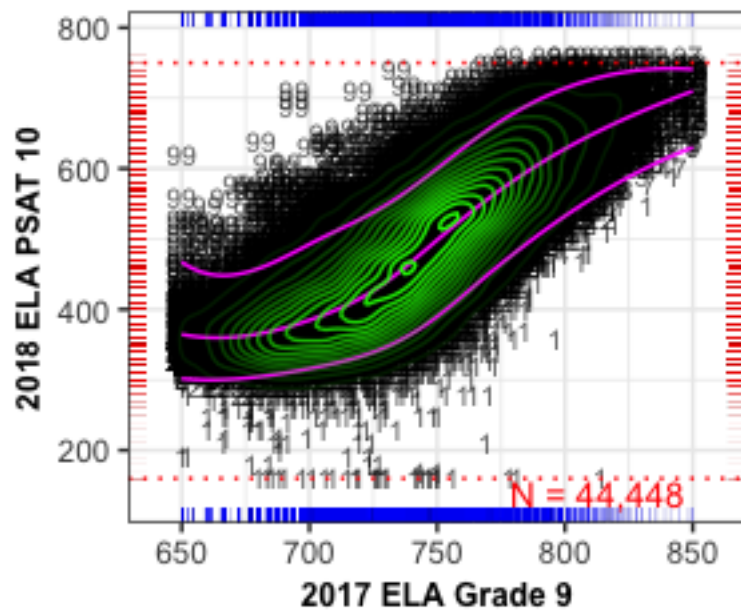**Fig. C.4:** Conditional distributions of current and prior scale scores: ELA PSAT 10.

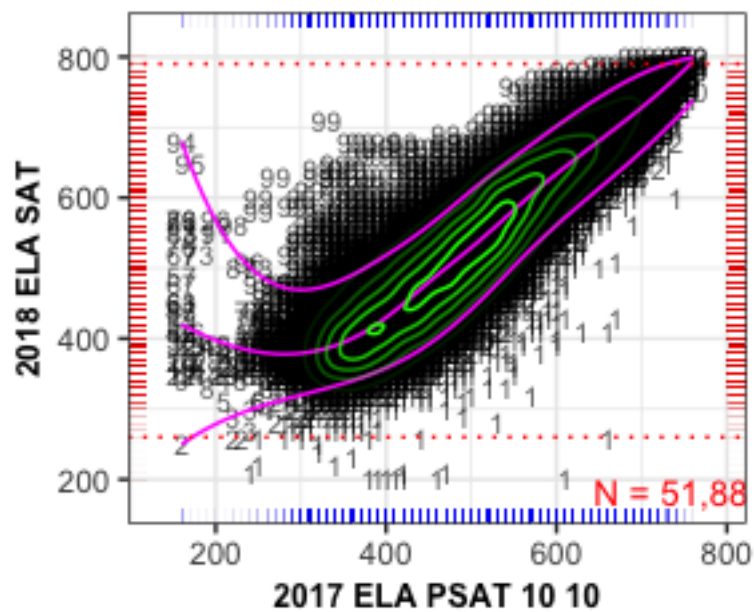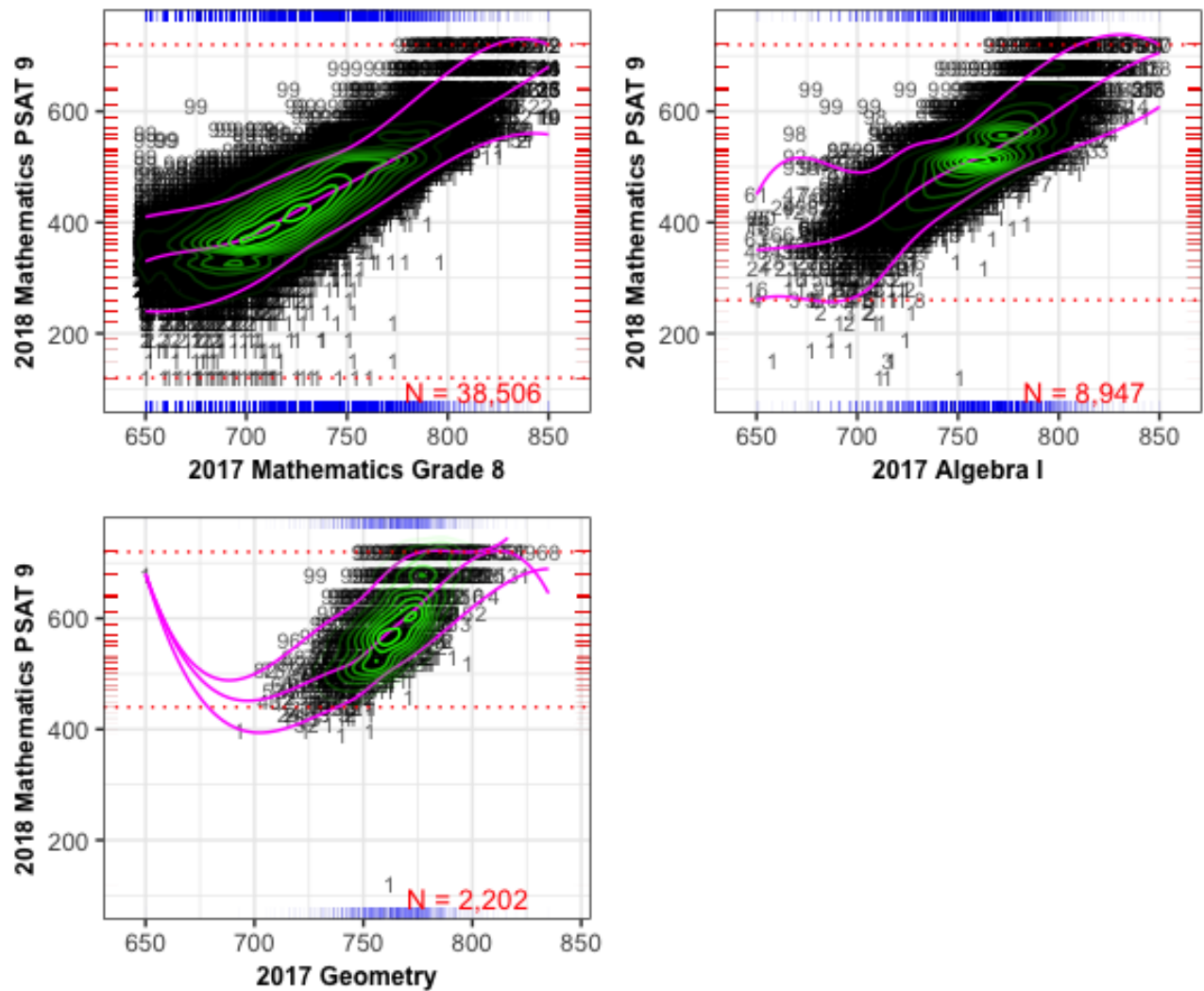**Fig. C.5:** Conditional distributions of current and prior scale scores: ELA SAT.

Fig. C.6: Conditional distributions of current and prior scale scores: Mathematics PSAT 9.

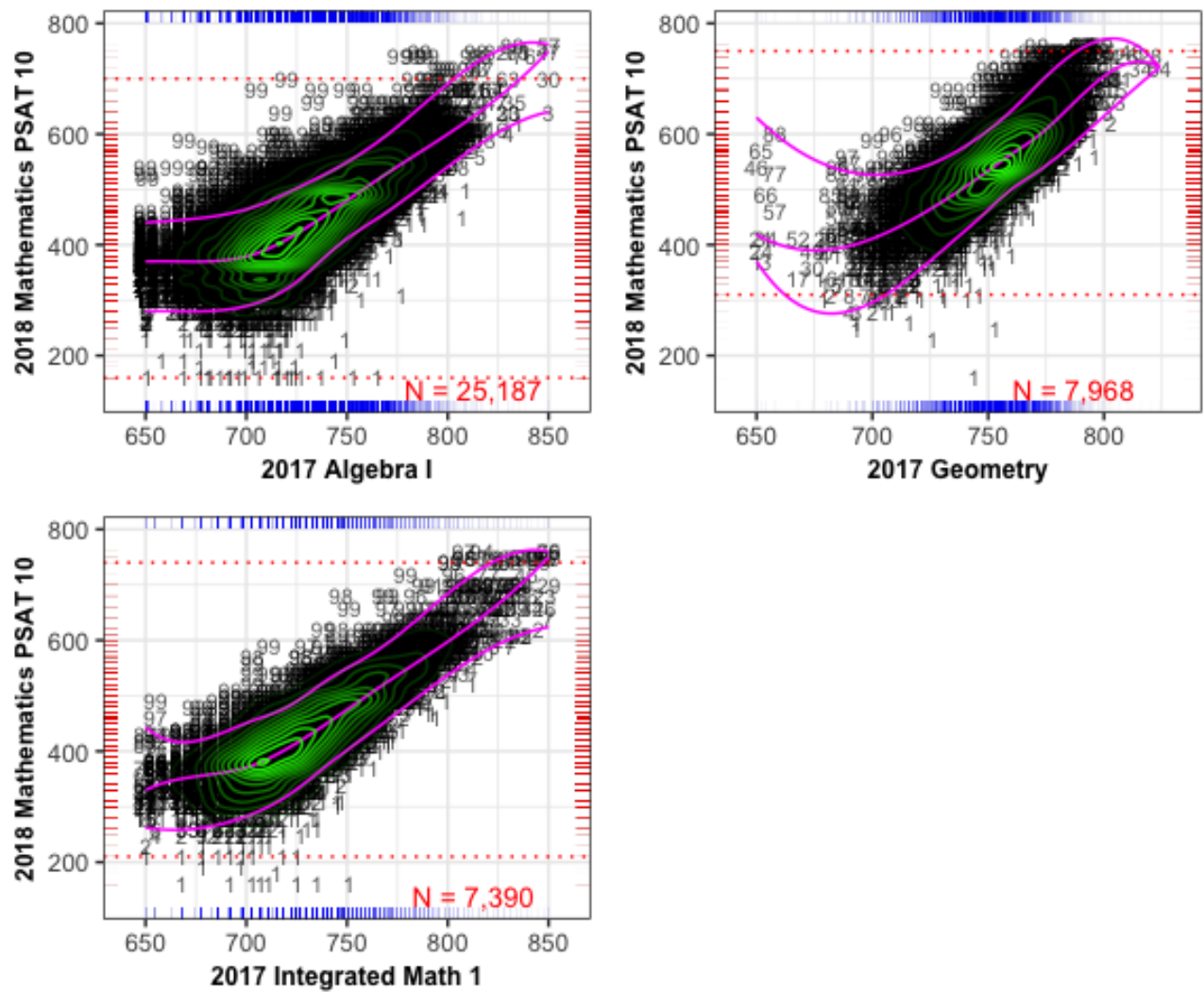**Fig. C.7:** Conditional distributions of current and prior scale scores: Mathematics PSAT 10.
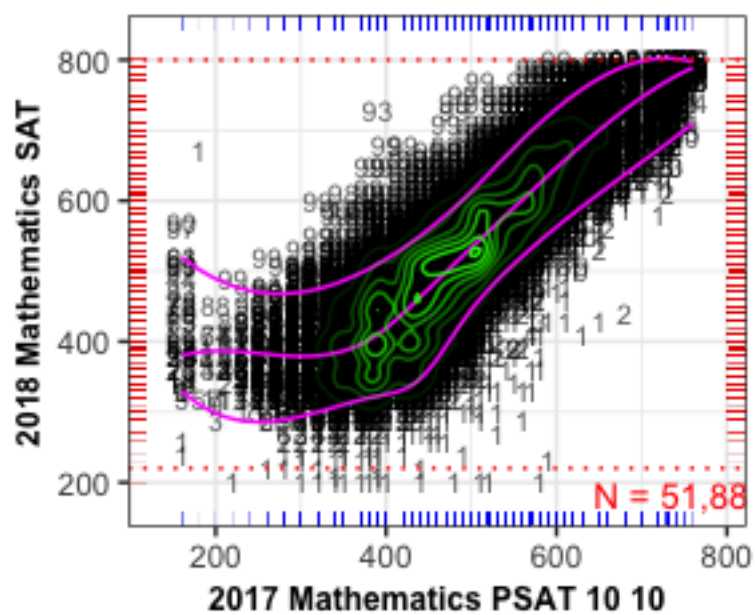
Fig. C.8: Conditional distributions of current and prior scale scores: Mathematics SAT.

3 SGP Ranges for the Highest- and Lowest-Achieving Students

In order to isolate the impact of assessment ceilings/floors on student growth percentile (SGP) calculations, the following section provides box plots of SGP distributions for the highest and lowest achieving students. We are specifically interested in the growth percentiles for students scoring at the highest/lowest obtainable scale score (HOSS/LOSS - i.e. the test ceiling/floor) on the current year test. However, in order to assure that an adequate number of students are included, the first set of plots uses, at a *minimum*, the highest/lowest 25 scores. These plots are provided as a starting point since this roughly corresponds to the number of students in the top and bottom rows of the table included in the SGP model "Goodness of Fit" plots. All students with a score in these students' range of scores are included. Consequently, the number of students in each box plot may be greater than 25 (the exact number is shown at the margins in red text).

The second set of box plots isolate **only** those students scoring the HOSS/LOSS. These plots may then incorporate a varying number of students depending on the prevalence of a ceiling/floor in the current year assessments.

The box plots provide several descriptive statistics. The dark line within the box marks the *median* SGP, while the ends ("hinges") of the boxes correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the highest value that is within $1.5 \times$ IQR of the hinge, where IQR is the interquartile range, or distance between the first and third quartiles. The lower whisker extends from the hinge to the lowest value within $1.5 \times$ IQR of the hinge. Data beyond the end of the whiskers are outliers and plotted as individual points. Evidence of a *lack* of either a ceiling or floor effect would be to have all high achieving students with SGPs near 99 and all low achieving students with SGPs near 1. That is, the desired visual evidence is a solid line at SGP = 99/1.

3.1 CMAS Content Areas

The scatter plots in the previous section showed that many students did score the LOSS and HOSS on both ELA and Mathematics assessments in the current and most recent prior years. Out of the tens of thousands of students, a small number are concerning. Grade 8 is impacted the most in both ELA and Mathematics. In ELA 212 students have scored the HOSS, of which 58 students have an SGP that is less than 90 (ranging from 61 to 89). In mathematics there are 42 students who attained the HOSS, 16 of which have an SGP less than 90. Here we see one student with a SGP estimate of 42. This student had perfect scores (850) on their 2016 and 2018 tests. A slight dip in their 2017 score (838) was enough to drop their SGP estimate significantly. Although all other students scoring the HOSS received a SGP estimate greater than 60, these estimates may still cause concern for parents and/or educators.

An even larger number of students in all grades scored the LOSS. Although the SGPs for these students are inflated as expected, the floor effect is not as significant as the ceiling effects. All LOSS scoring students SGPs are lower than 20, and the vast majority are less than 5. Grade 8 Mathematics is impacted the most by the floor effect with 232 students scoring the LOSS and receiving a SGP greater than 5 (max 18).

Overall the impacted kids represent a very small fraction of the grade-level cohorts: HOSS

students with SGPs less than 90 in 8th grade account for 0.1% of their ELA cohort (58 out of 52,761), and 0.04% in mathematics. Still, these small number of cases can, and often do, draw out-sized attention and understandable concern from students, parents, educators and others.

Fig. C.9: Grade Level SGP distributions for highest and lowest 25+ scale scores by content area and grade.
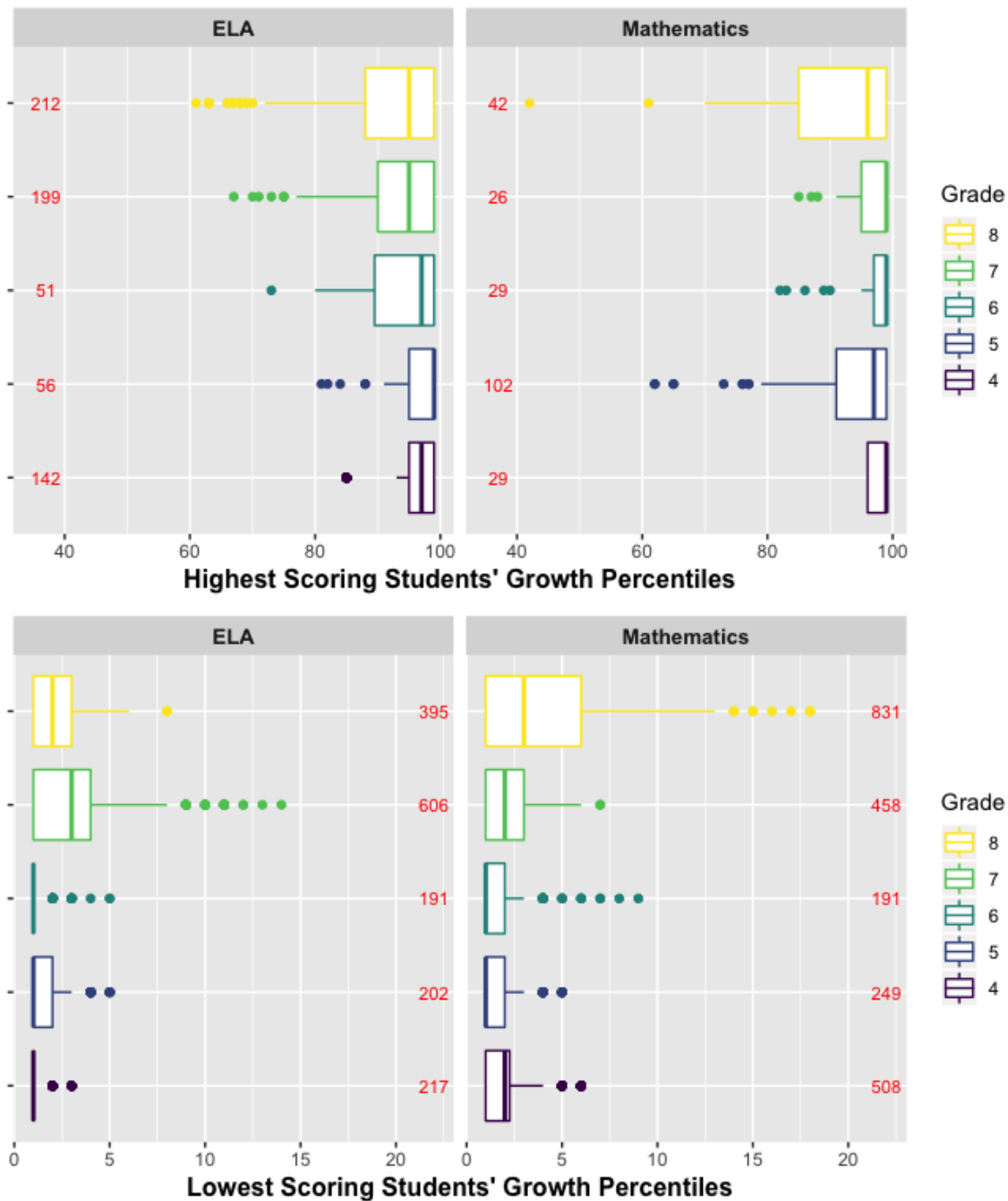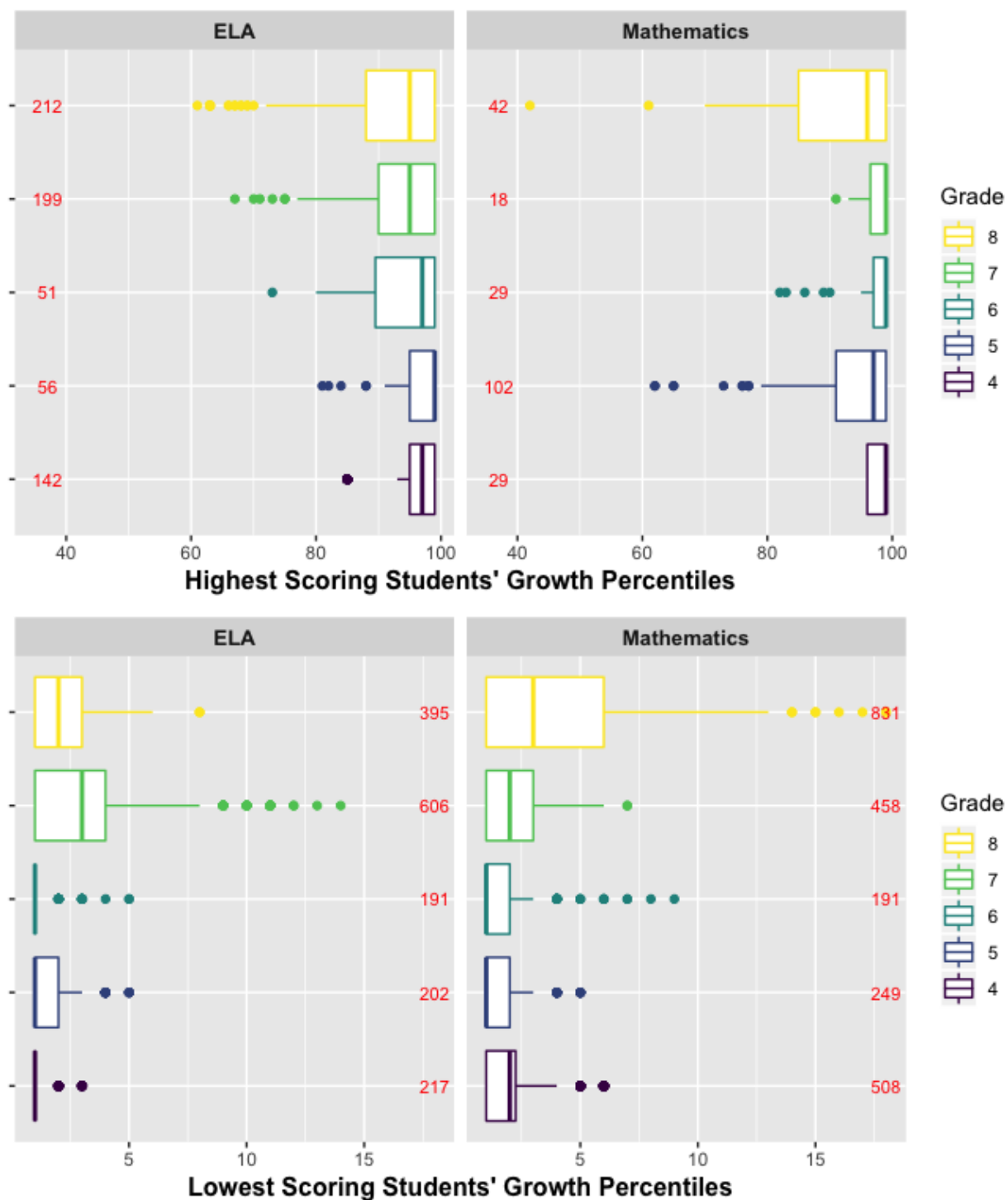
Fig. C.10: Grade Level SGP distributions for the HOSS and LOSS scores by content area and grade.

3.2 PSAT/SAT Subjects

The PSAT/SAT subject results are shown here *only for students scoring exactly the HOSS and LOSS* respectively. Ceiling effects are evident in the Mathematics PSAT/SAT analyses and a couple ELA SAT students.

Fig. C.11: PSAT/SAT SGP distributions for the HOSS and LOSS scores by content area.
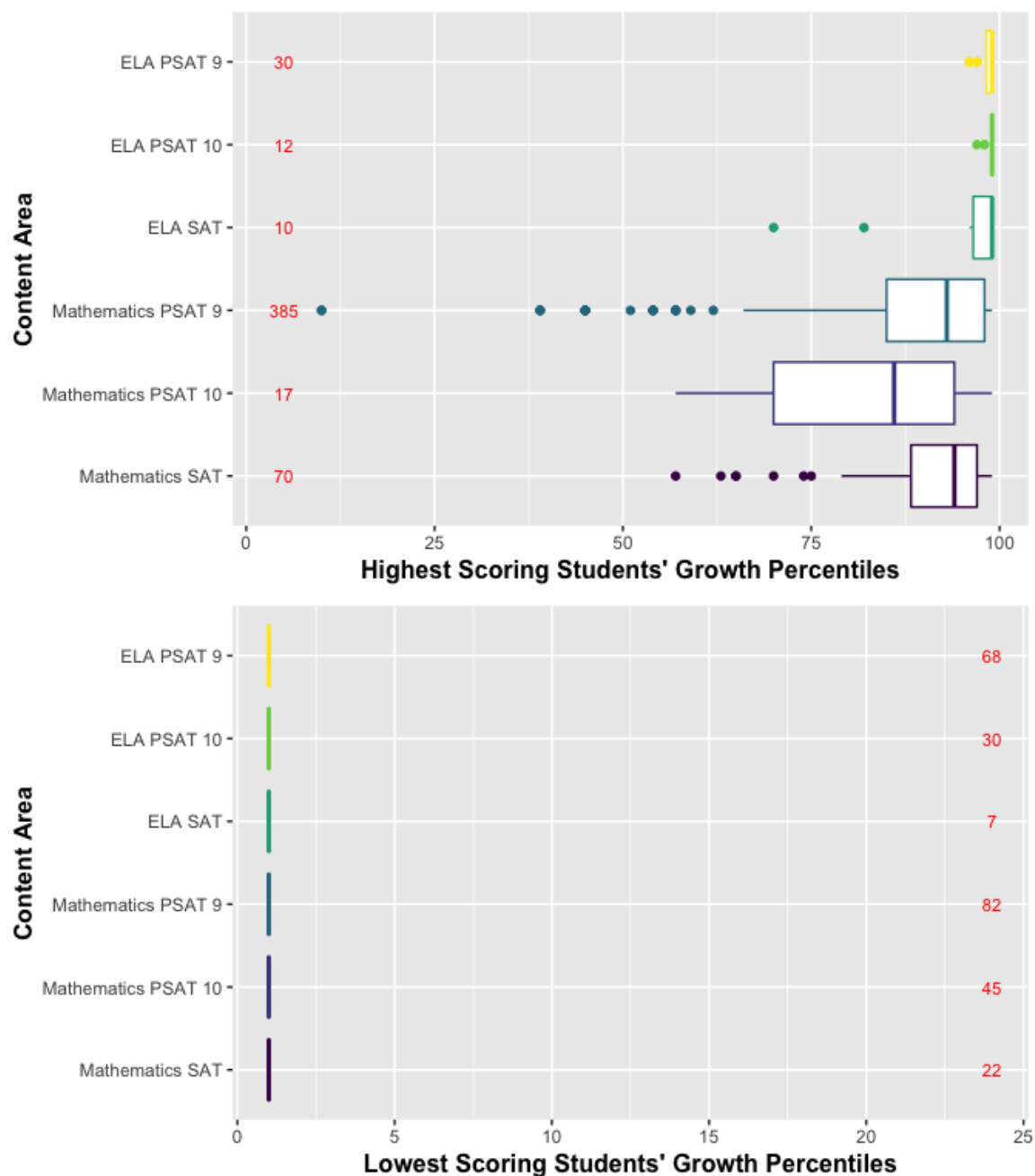
The following box plots disaggregate the Mathematics PSAT 9 analyses by the most recent prior to reflect their constituent norm groups more closely. All three cohorts exhibit growth ceiling effects, and re-examining the plots in Figure C.6 it appears as though the main cause of ceiling effect in growth is the high incidence of HOSS scores in the PSAT test (i.e. there does not appear to be ceiling issues in the prior tests).

The Algebra I cohort has the most extreme cases, where two students who also had perfect scores in Algebra I and Grade 7 Math tests received a SGP estimate of 10. This cohort has 28 HOSS students with SGPs below 90 (0.3% of the cohort). The Geometry cohort has the most extensive growth ceiling effect with 93 HOSS students (4.2%) given SGPs below 90. Although the Grade 8 Math test shows highest incidence of test ceilings (see Figure C.6), it has the most minor growth ceiling issues, with 20 students with SGPs below 90 (0.05% of the total cohort).

Fig. C.12: PSAT/SAT SGP distributions for the HOSS and LOSS scores by norm group: Mathematics PSAT 9.
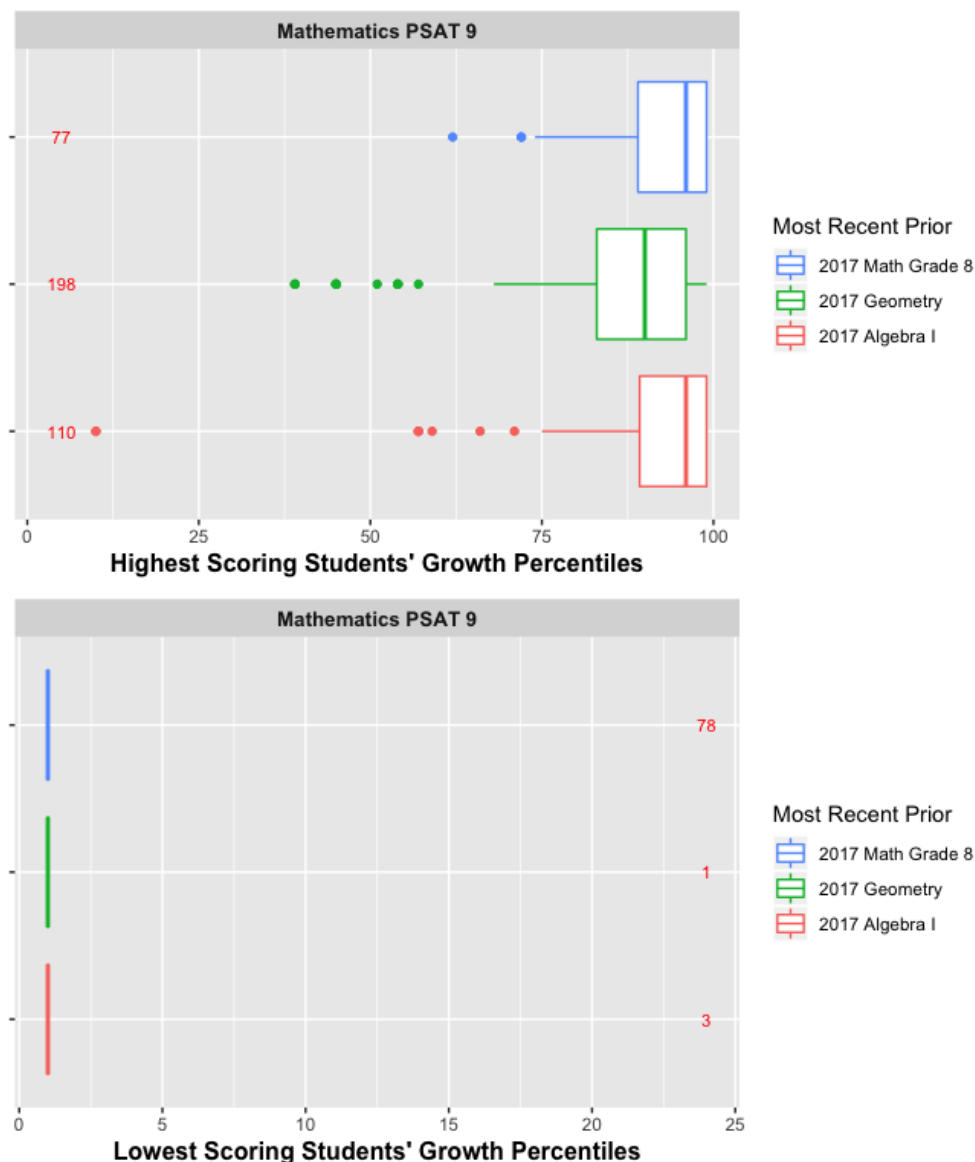
The final box plots presented in this report disaggregate the Mathematics PSAT 10 analyses by the most recent prior. Although the plots indicate that all three cohorts exhibit growth ceiling effects, only 17 students in total scored the HOSS in this exam in 2018.

Re-examining the plots in Figure C.7 shows that the cause of ceiling effect in growth for Algebra I and Integrated Math 1 priors is different than the causes in the CMAS and other PSAT/SAT analyses. The scatter plots for these subjects indicate a sparsity of data near the high achievement regions, which may lead the SGP model to struggle to produce adequate and sensible growth estimates for a handful of kids. The top-right scatter plot in Figure C.7 indicates a minor test ceiling for PSAT 10 Math, but these kids' prior Geometry scores were not at the HOSS, which mitigated any growth ceiling potential (although some of their SGPs were still lower than expected - 5 were less than 90.)

Fig. C.13: PSAT/SAT SGP distributions for the HOSS and LOSS scores by norm group: Mathematics PSAT 10.
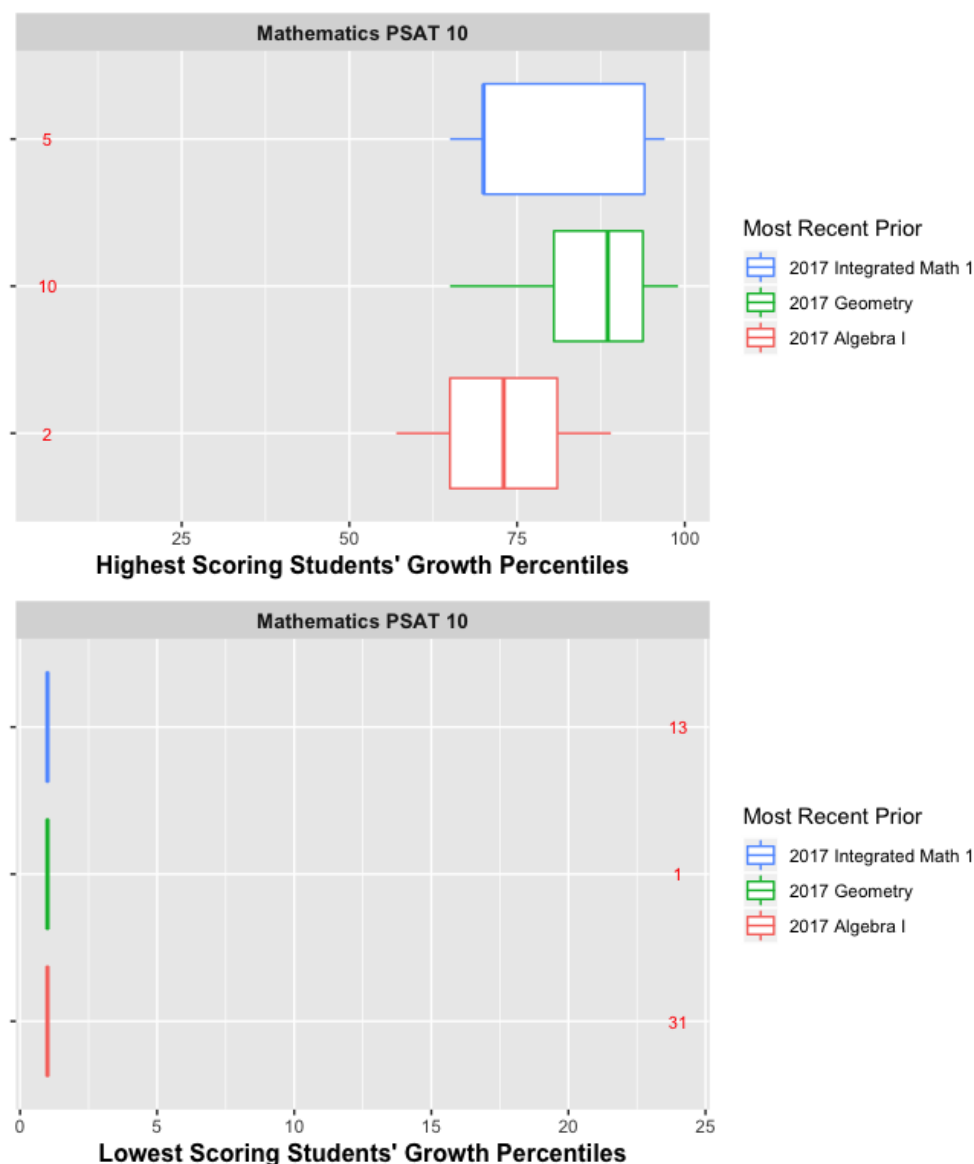
4   Discussion

Overall there is evidence of both floor and ceiling effects in the 2017-2018 Colorado CMAS/SAT SGP analyses. When ceiling or floor effects are encountered, there are several ways in which they can be "corrected" manually or analytically. These include (but are not limited to):

1. Convert all students scoring at the HOSS (LOSS) to 99 (1).
2. Run SGP analyses with more granular scores. For example, many tests that use Item Response Theory (IRT) to analyse test results provide scaled scores that enforce an artificial ceiling (floor), but also have more granular achievement scores available (IRT θ estimates).
3. Leave the results without a correction.