Appendix C to the 2017-2018 Rhode Island RICAS/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DRAFT REPORT - DO NOT CITE!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 Rhode Island transitioned from the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium assessments, to the Rhode Island Comprehensive Assessment System (RICAS) in grades 3-8 in English Language Arts and mathematics and the PSAT10 and the SAT assessments administered in 10^{th} and 11^{th} grade. 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 80^{th} 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 50^{th} 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 5^{th}, 50^{th} and 95^{th} 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 RICAS Content Areas

We see very few issues in all content areas and grades in the 2017 and 2018 Rhode Island grade-level RICAS data. Where minor ceilings appear in either years' data, the opposite year score distributions for these students are well distributed, lessening the concern for a growth ceiling effect. In past years analyses, some students attaining the HOSS in consecutive years had slightly lower SGPs than 99 (very minor ceiling issues).

[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.

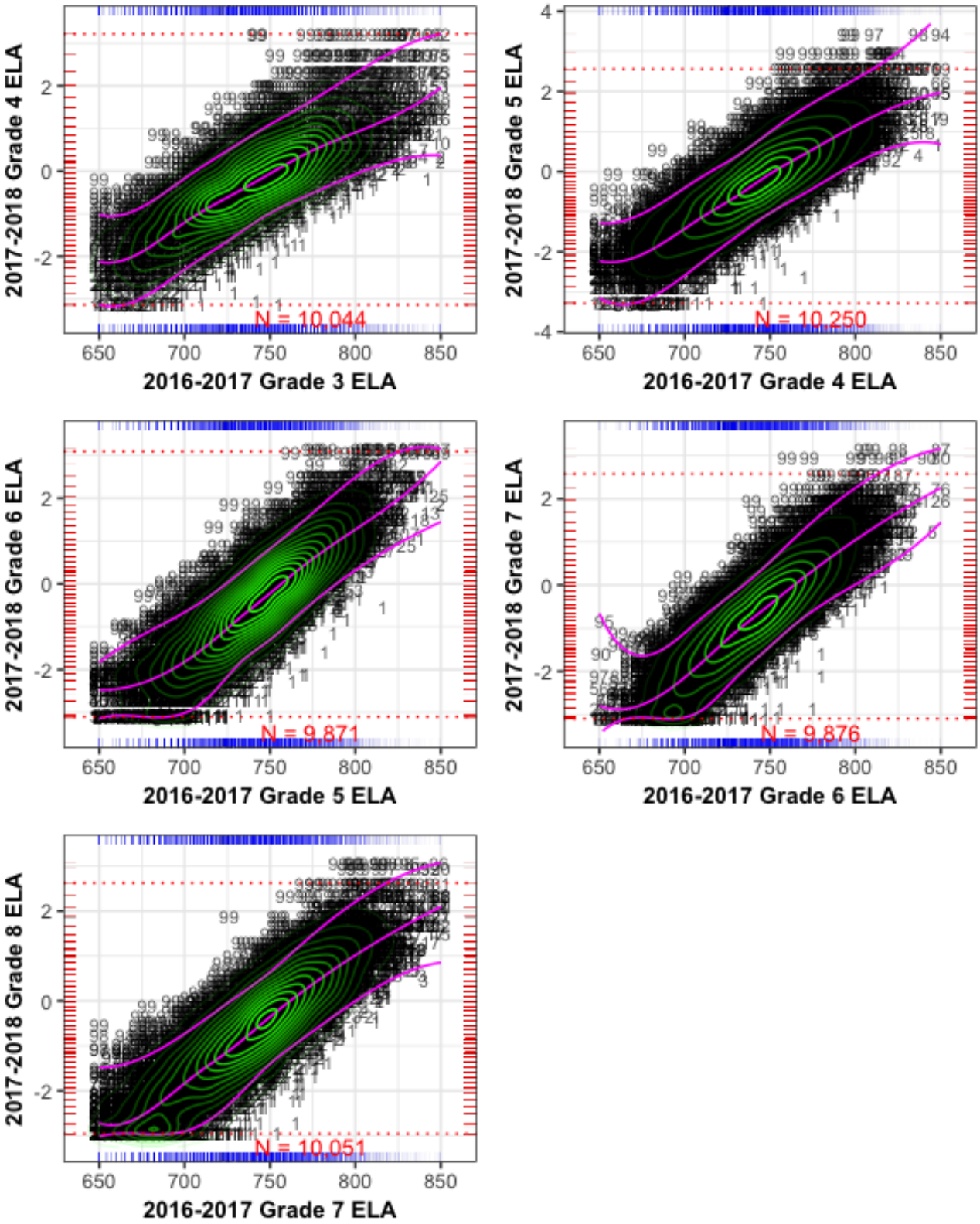
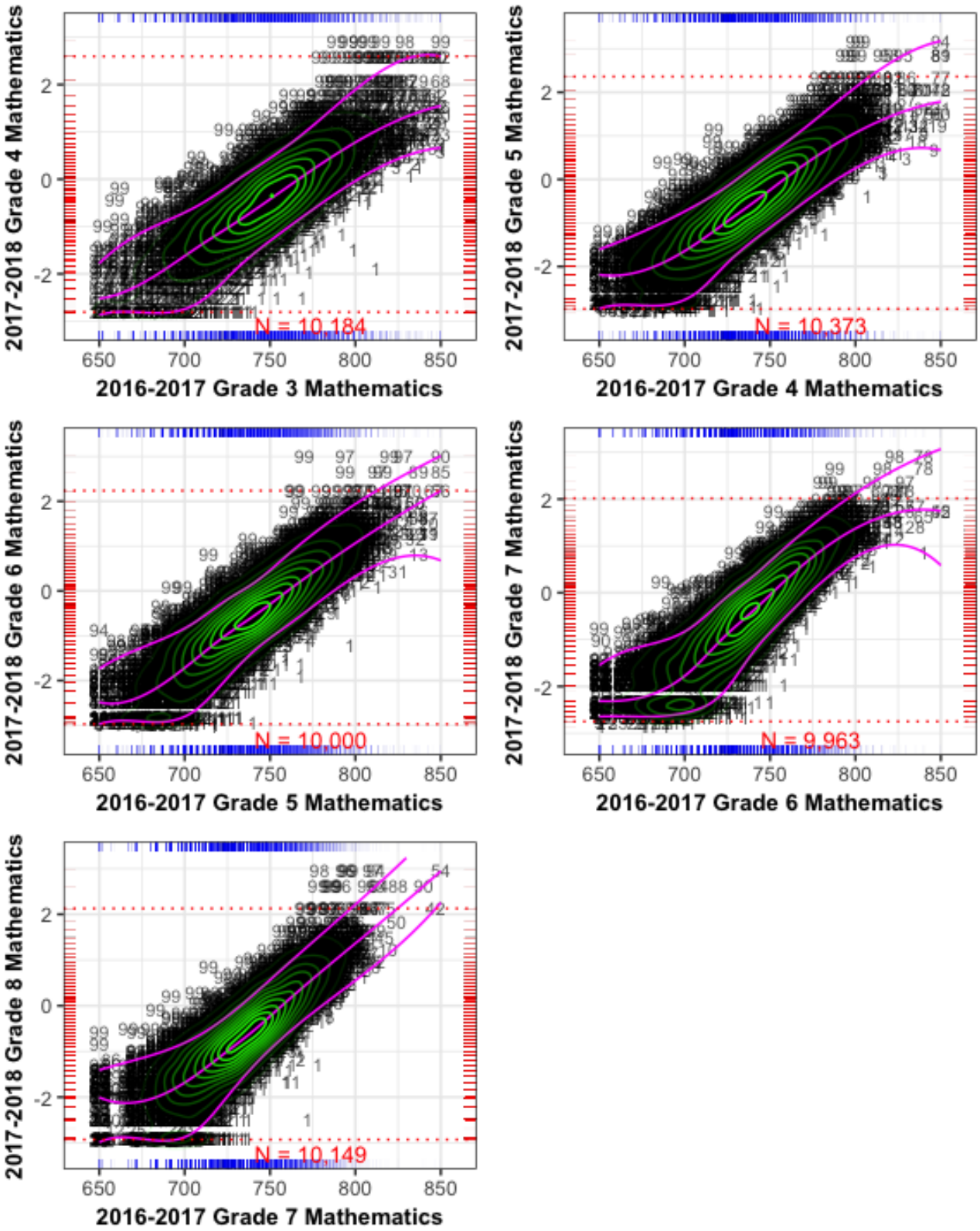


Fig. C.2: Conditional distributions of current and prior scale scores: Mathematics.



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. Some of these atypical norm group analyses present a handful of concerning estimates that require closer examination.

These analyses can have problems due to sparsity of data points in the high achievement regions, which leads the SGP model to struggle to produce adequate and sensible growth estimates.

Fig. C.3: Conditional distributions of current and prior scale scores: ELA PSAT 10.

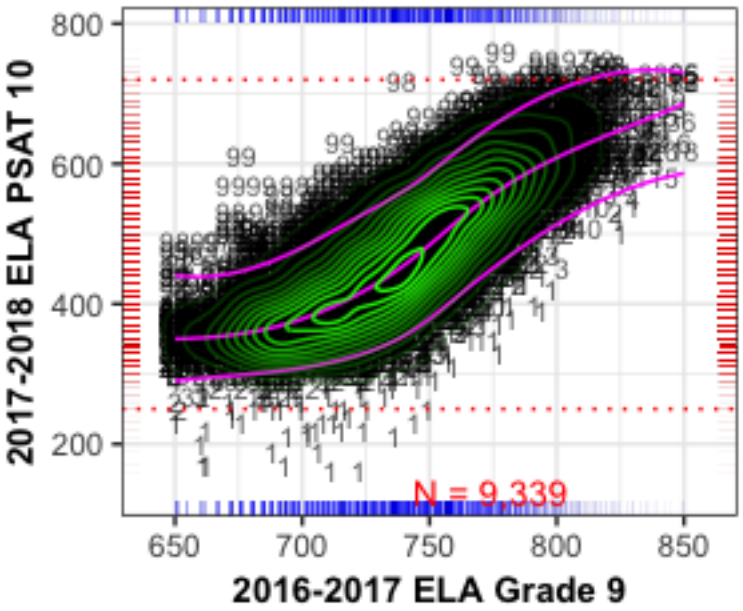


Fig. C.4: Conditional distributions of current and prior scale scores: ELA SAT.

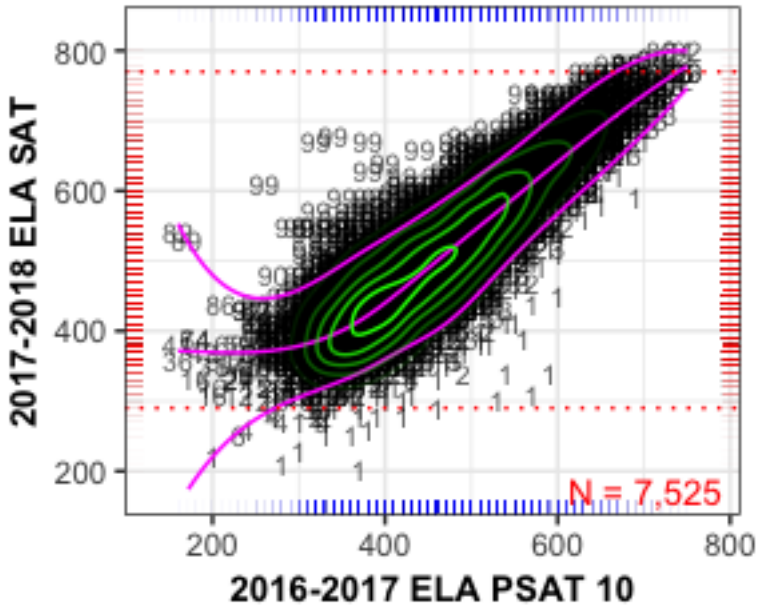


Fig. C.5: Conditional distributions of current and prior scale scores: Mathematics PSAT 10.

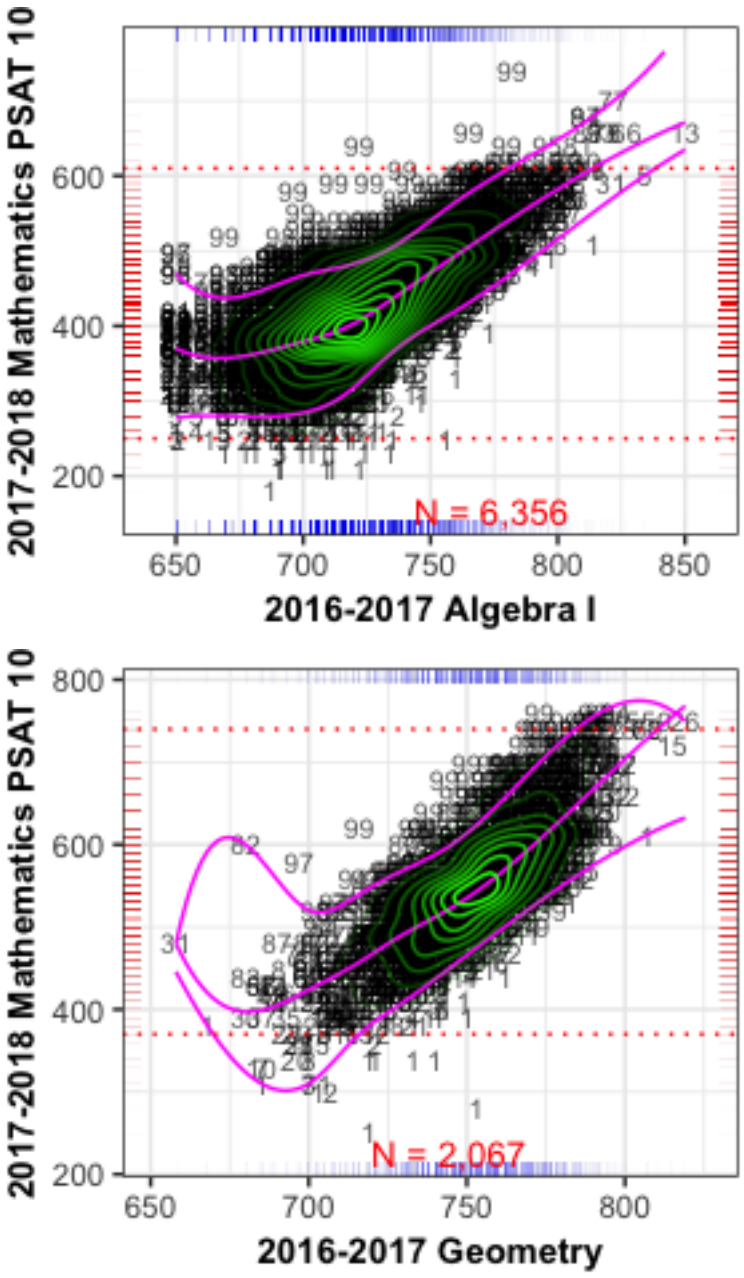
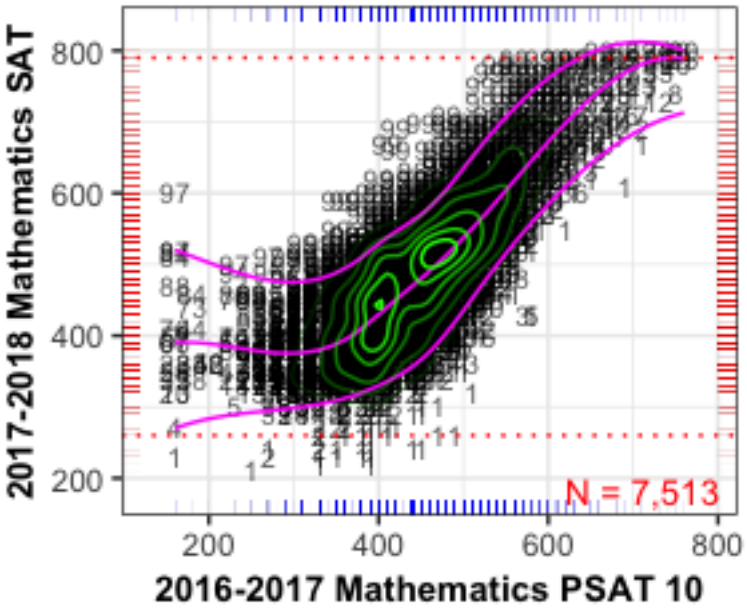


Fig. C.6: Conditional distributions of current and prior scale scores: Mathematics SAT.



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 25^{th} and 75^{th} 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 RICAS Content Areas

The scatter plots in the previous section showed that, although uncommon, students do score the HOSS on both ELA and Mathematics assessments. However, out of the tens of thousands of students, only a handful appear to be concerning. There is one student in Grade 6 ELA who has scored the HOSS and has an estimated SGP of 37, and another in Grade 8 Mathematics with a SGP estimate of 54. However, a majority of these students have SGPs in the 90s. Similarly, the students scoring the LOSS have a few SGP estimates with minor floor effects indicated, particularly in Grade 8 ELA.

Fig. C.7: Grade Level SGP distributions for highest and lowest 25+ scale scores by content area and grade.

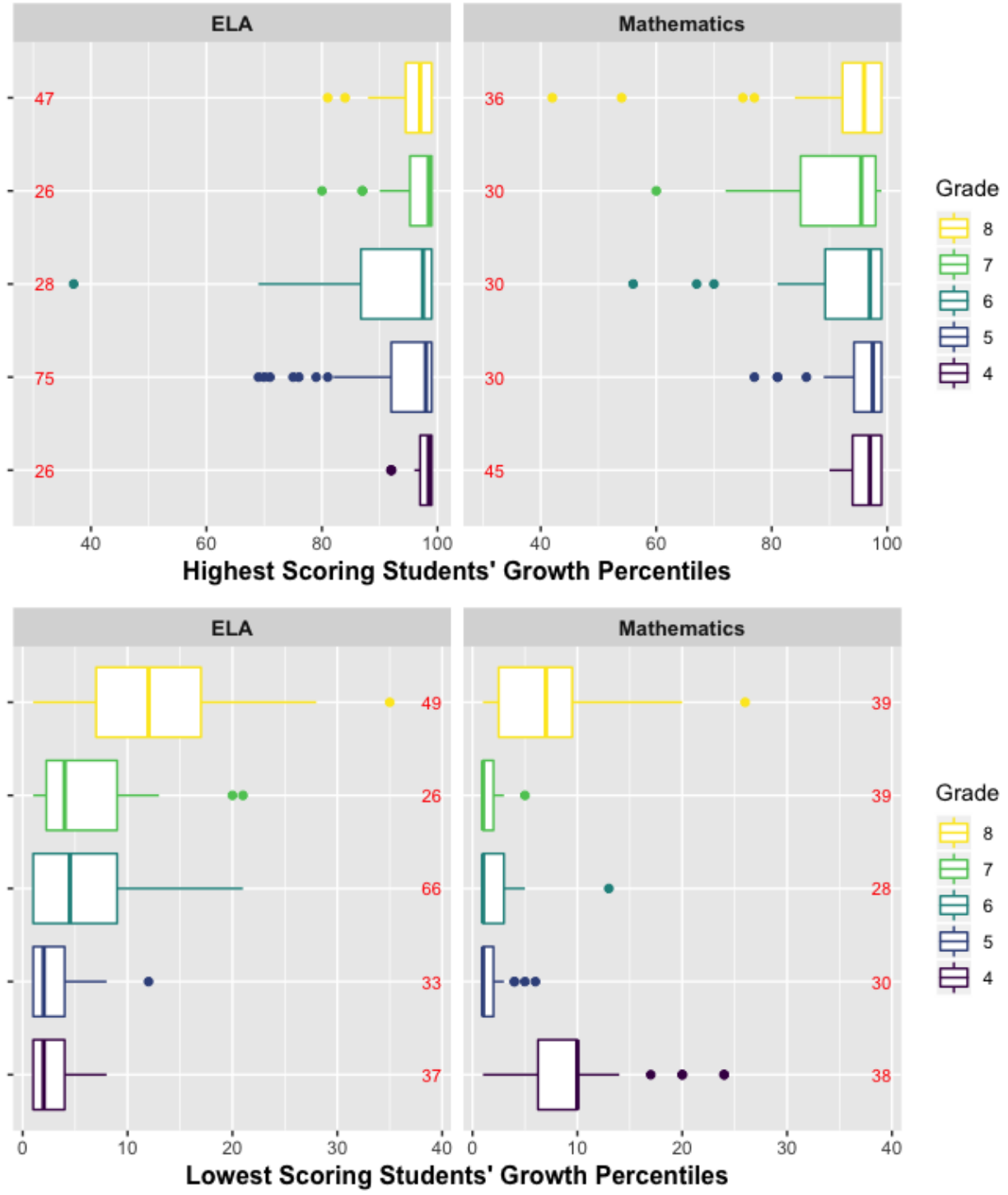
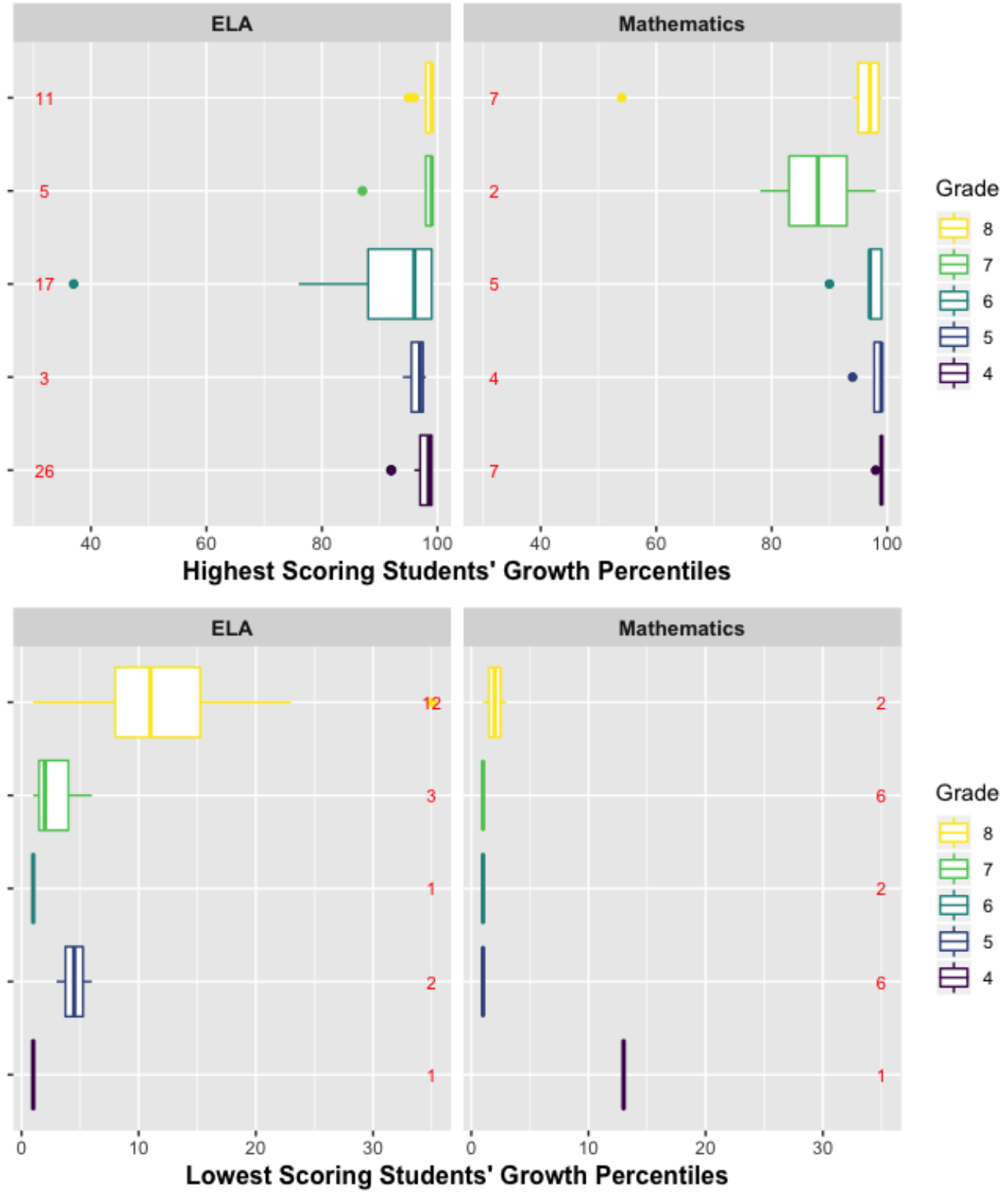


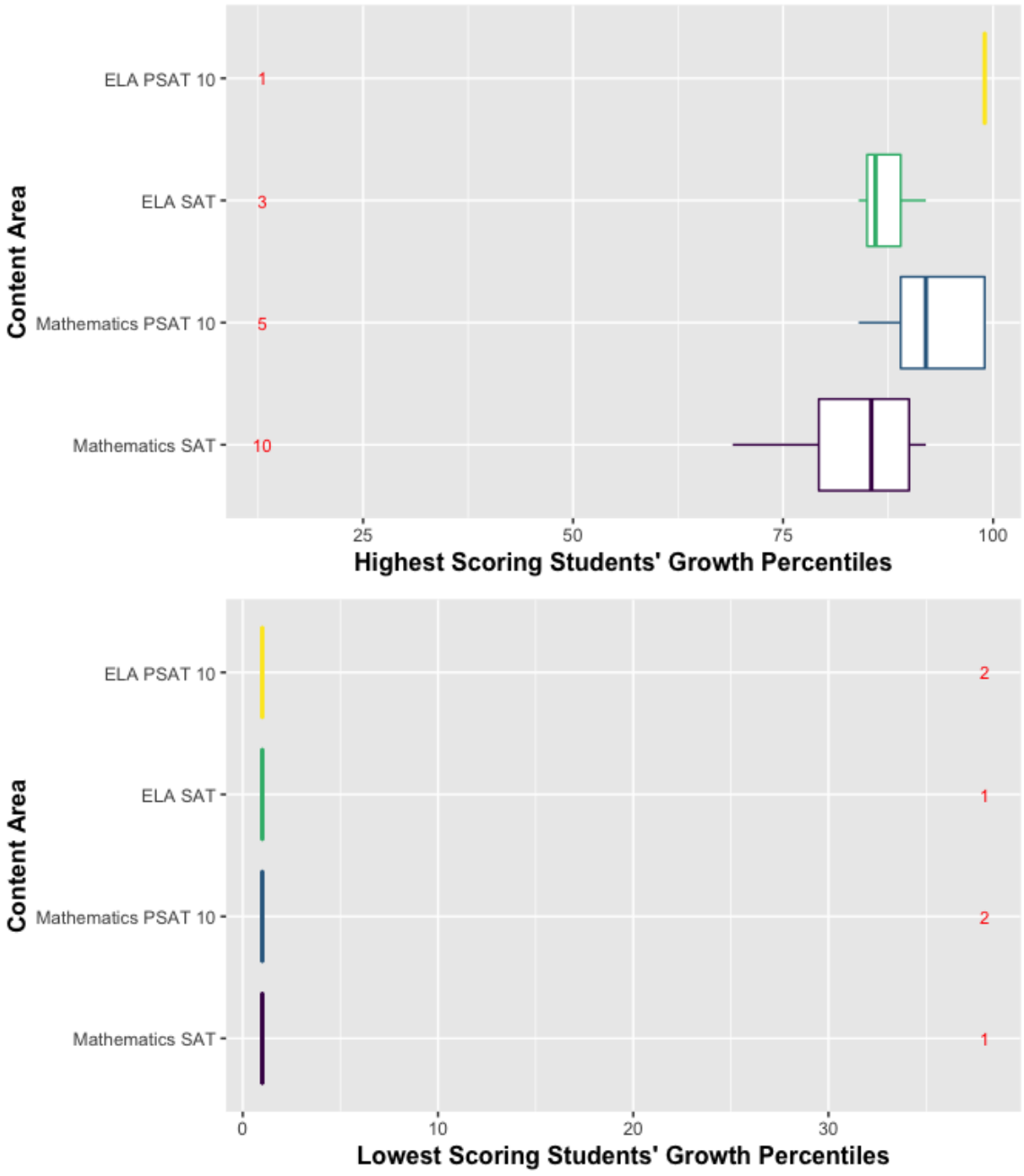
Fig. C.8: Grade Level SGP distributions for the HOSS and LOSS scores by content area and grade.



3.2 PSAT/SAT Subjects

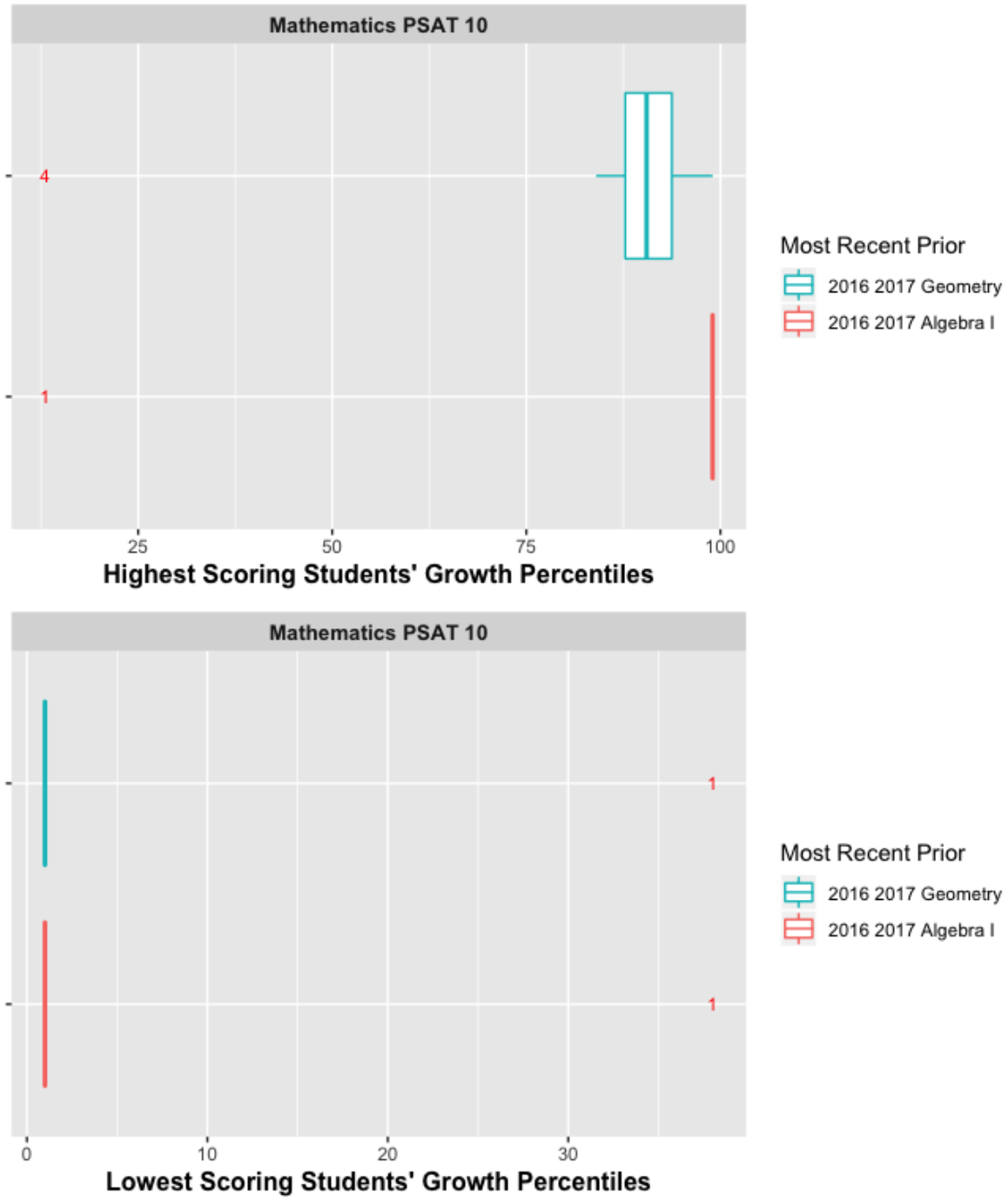
The PSAT/SAT subject results are shown here *only for students scoring exactly the HOSS and LOSS* respectively. Small ceiling effects are evident in the Mathematics SAT analysis.

Fig. C.9: PSAT/SAT SGP distributions for the HOSS and LOSS scores by content area.



The following box plots disaggregates the Mathematics PSAT 10 analyses by the most recent prior to reflect their constituent norm groups more closely.

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



4 Discussion

Overall there is little evidence of floor or ceiling effects in the 2017-2018 Rhode Island RICAS/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.