Appendix C to the Fall 2018 PARCC Student 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)

May 31, 2019

1 Introduction

Ceiling and floor effects are a common characteristic of many standardized assessments (i.e. a relatively large proportion of students scoring at/near the scale score extremes). These assessment ceilings/floors can also telegraph onto the Student Growth Percentile (SGP) calculations causing confusing or concerning results.

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. The Center for Assessment has recently added "Ceiling/Test Effects" indicators to the SGP model goodness of fit plots and is providing all clients even more rigorous diagnostic and descriptive analyses through this Appendix to the annual technical report. 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 *only* the highest and lowest achieving students in the current year.[1]

[1] Ranked SIMEX measurement error corrected SGPs are the "official" SGP in Georgia, and are used exclusively in this report.

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[2]. 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[3], 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 End-of-Grade Content Areas

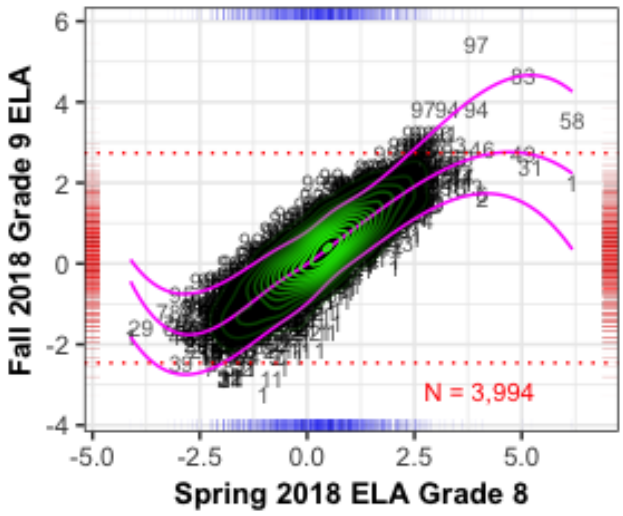
We see very few issues in all content areas and grades in the 2017 and Spring PARCC grade level ELA data. Test ceilings do not appear in either years' data, lessening the concern for growth ceiling effect that could impact a large number of students. However, there are some individual students who appear as outliers in the scatterplots),

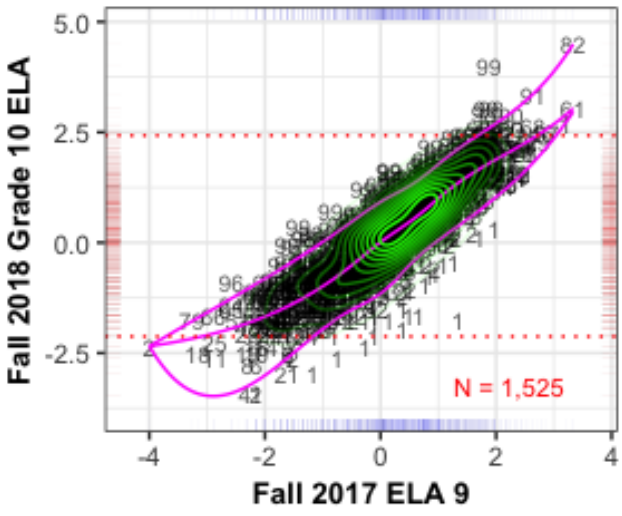
Although ceiling effects are not widespread, there is one individual student that is problematic in the 11^{th} grade ELA analysis that uses spring 2018 Grade 10 ELA as the most recent prior. This student scored near the HOSS in both tests and is a clear outlier in the scatterplot (scoring considerably higher than all other students in consecutive administation periods, and her estimated SGP is 3. This estimate is unexpectedly low for such a high achieving student. A similar student appears in the first 10^{th} grade ELA scatterplot, and this student's SGP estimate is 82, which may be slightly lower than expected, but is still in the high growth range. Manual correction for the first student may be required given the inadequacy of the model to describe growth for outliers such as this.

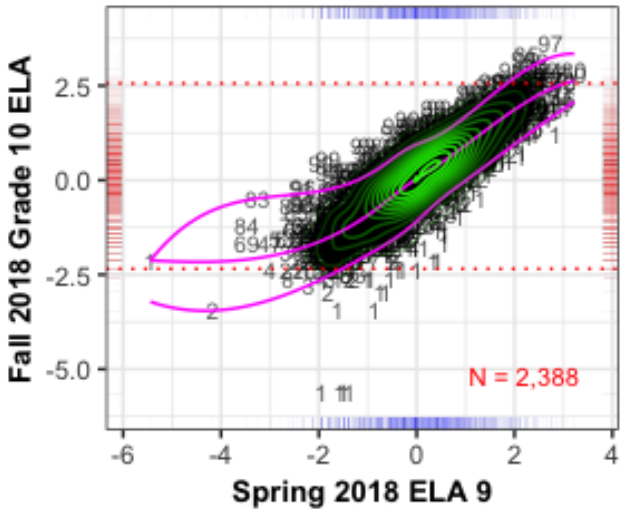
[2]Note that many SGPs are estimated using more than one prior score, and therefore plots may show SGP results from multiple analyses and/or varying SGPs for identical score combinations.

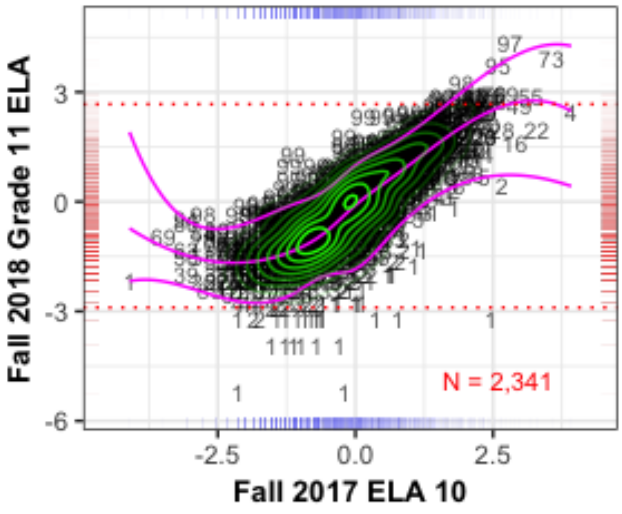
[3]Produced using quantile regression similar to, but not the same as, that used in calculation of the SGPs.

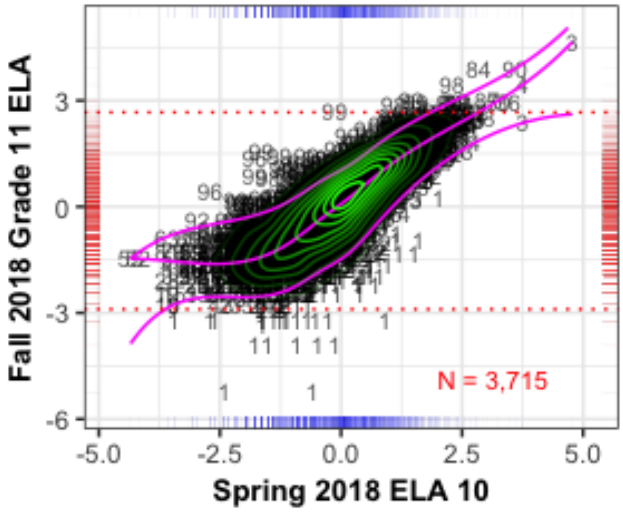
Fig. C.1: Conditional distribution(s) of current and prior scale scores: ELA.











2.2 End-of-Course Test Subjects

The conditional density plots for the EOCT subjects are displayed below. The most recent prior is used in each plot to provide insight on the academic peer group based analyses. Some of these norm groups represent atypical student populations or have particularly low cohort numbers, which can also cause ceiling effects for different reasons.

Overall there is no evidence for concern of ceiling or floor effects for large numbers of students in the PARCC EOCT analyses. There are several cases that appear to be somewhat problematic where students consistently scoring high receive lower than expected SGP estimates. For example, the Algebra I cohort with Geometry as the most recent prior and the Algebra II analysis with Fall 2017 Geometry prior. In both cases, the students' SGPs are below 50 (i.e. below average) and are likely not adequate descriptions of these students' growth given their consistently high achievement. As with the ELA student described above, individual (manual) corrections may be required for these specific cases.

Fig. C.2: Conditional distribution(s) of current and prior scale scores: Algebra I.

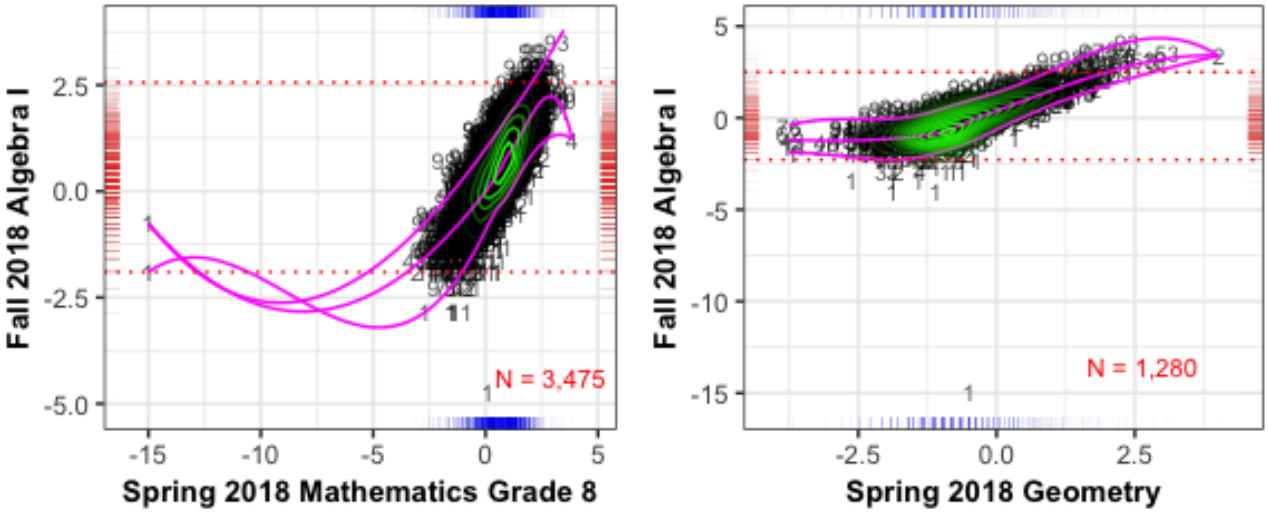


Fig. C.3: Conditional distribution(s) of current and prior scale scores: Geometry.

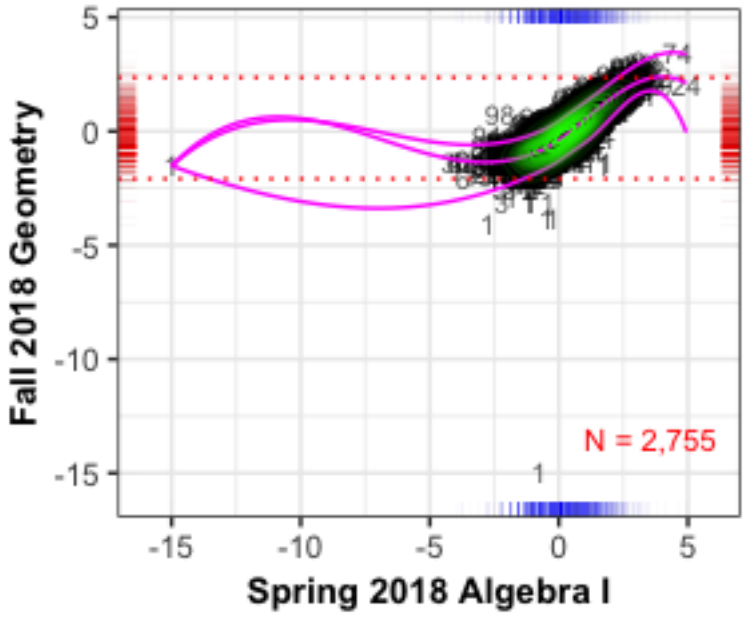
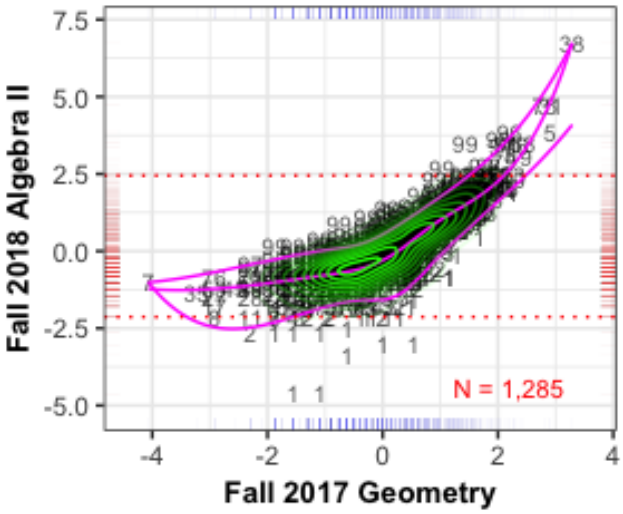
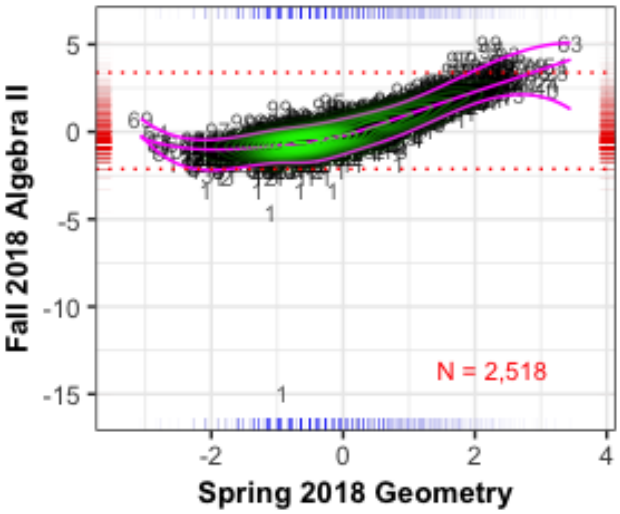


Fig. C.4: Conditional distribution(s) of current and prior scale scores: Algebra II.





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 the distribution of SGPs 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.

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 inter-quartile 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 Grade Level ELA

The scatter plots in the previous section showed that there are no concerns for widespread ceiling or floor effects in either the ELA assessments. Figure C. 5 suggests, however, that there are potential problems in the 10^{th} and 11^{th} grade ELA analyses. Figure C. 6 confirms that a potential ceiling issue does effect a few individual students. Specifically, one student that scored exactly the HOSS in a 11^{th} grade analysis (with a SGP of 3, as discussed above). There are also two notable 10^{th} grade students scoring the HOSS have somewhat lower SGPs, but their growth estimates are still relatively high (82 and 97 - from two seperate analyses) and not cause for concern.

Fig. C.5: Grade Level SGP distributions for highest and lowest 25+ scale scores by content area and grade level.

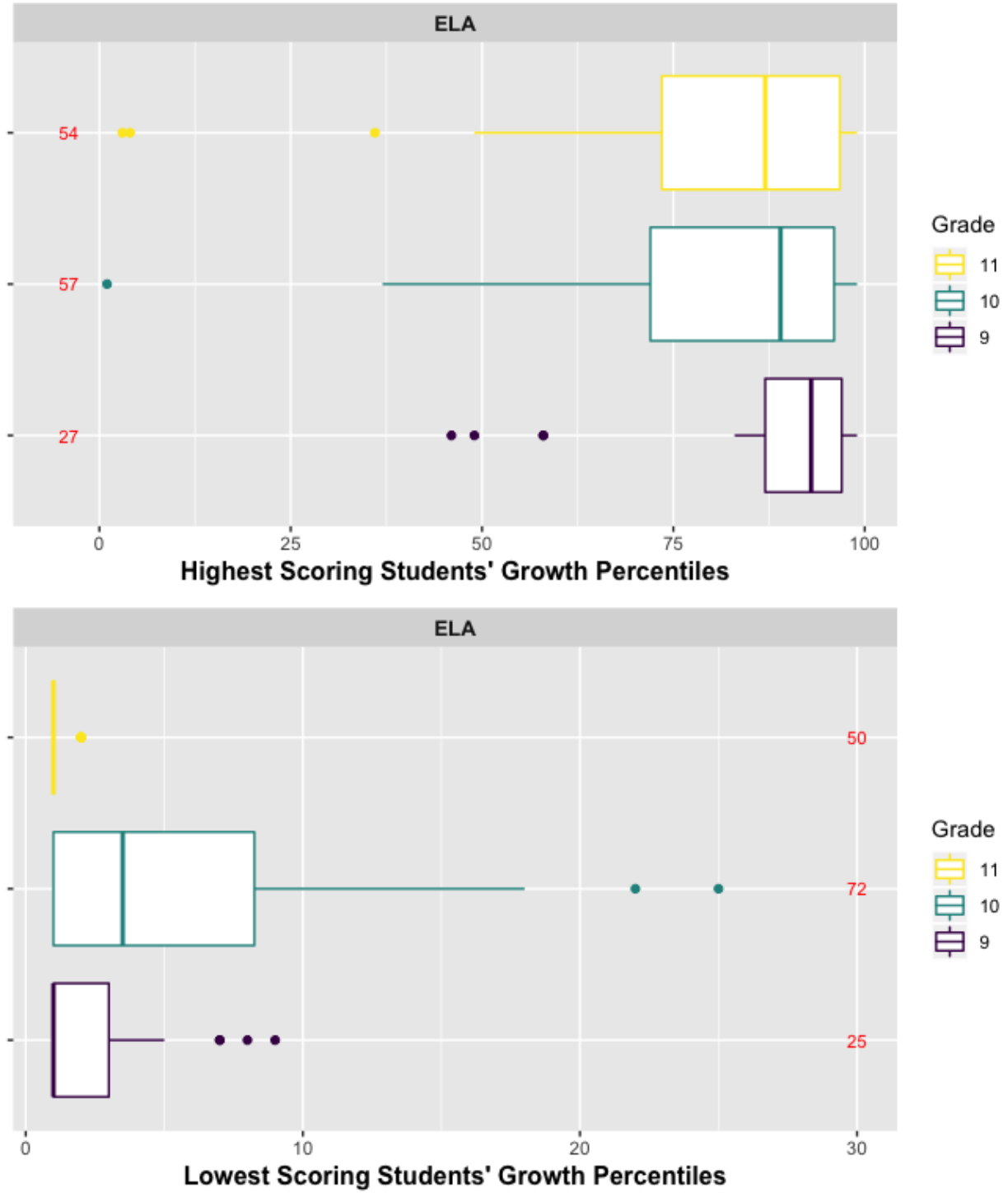
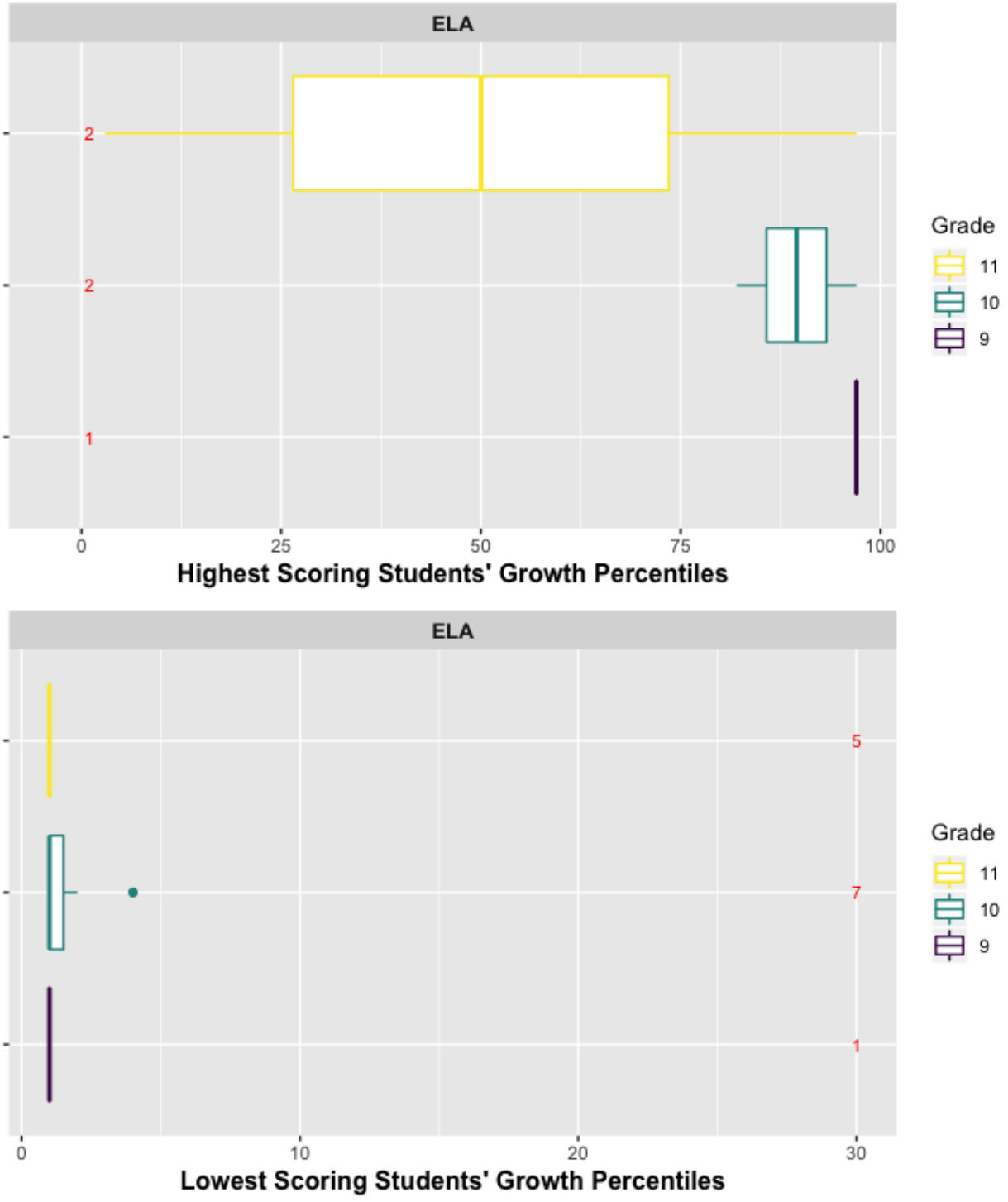


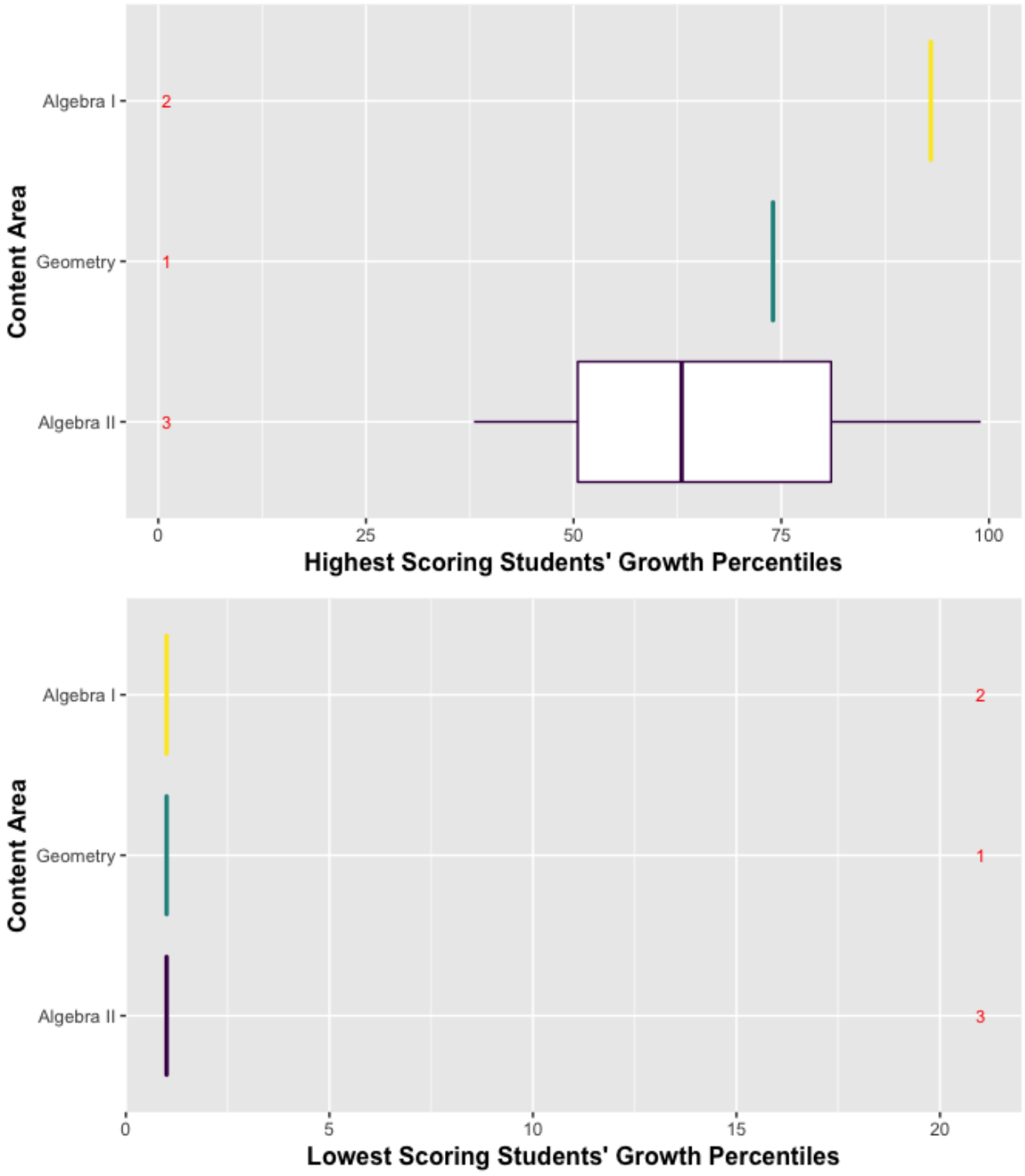
Fig. C.6: Grade Level SGP distributions for the HOSS and LOSS scores by content area and grade level.



3.2 EOCT Subjects

The end-of-course subject results are shown here *only for students scoring exactly the HOSS and LOSS* respectively. There are several subjects for which potential ceiling effects are evident. All EOCT subjects are disaggregated further by the most recent prior test included in each analyses in order to adequately address any concerns.

Fig. C.7: EOCT SGP distributions for the HOSS and LOSS scores by content area.



The EOCT subject box plots can be disaggregated further by the most recent prior to reflect their constituent norm groups more closely. The following box plots disaggregate each EOCT subject by norm groups.

Fig. C.8: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Algebra I.

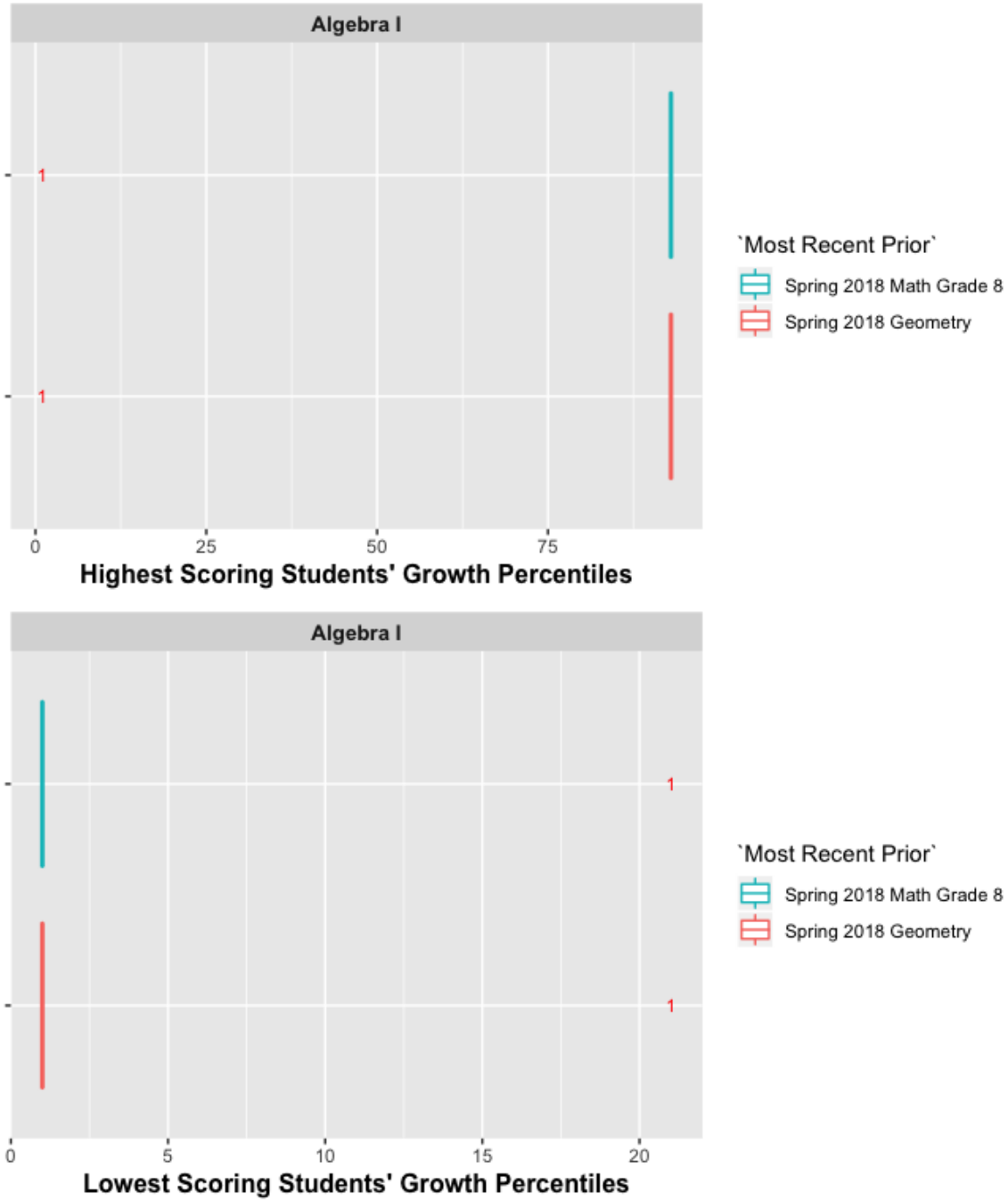
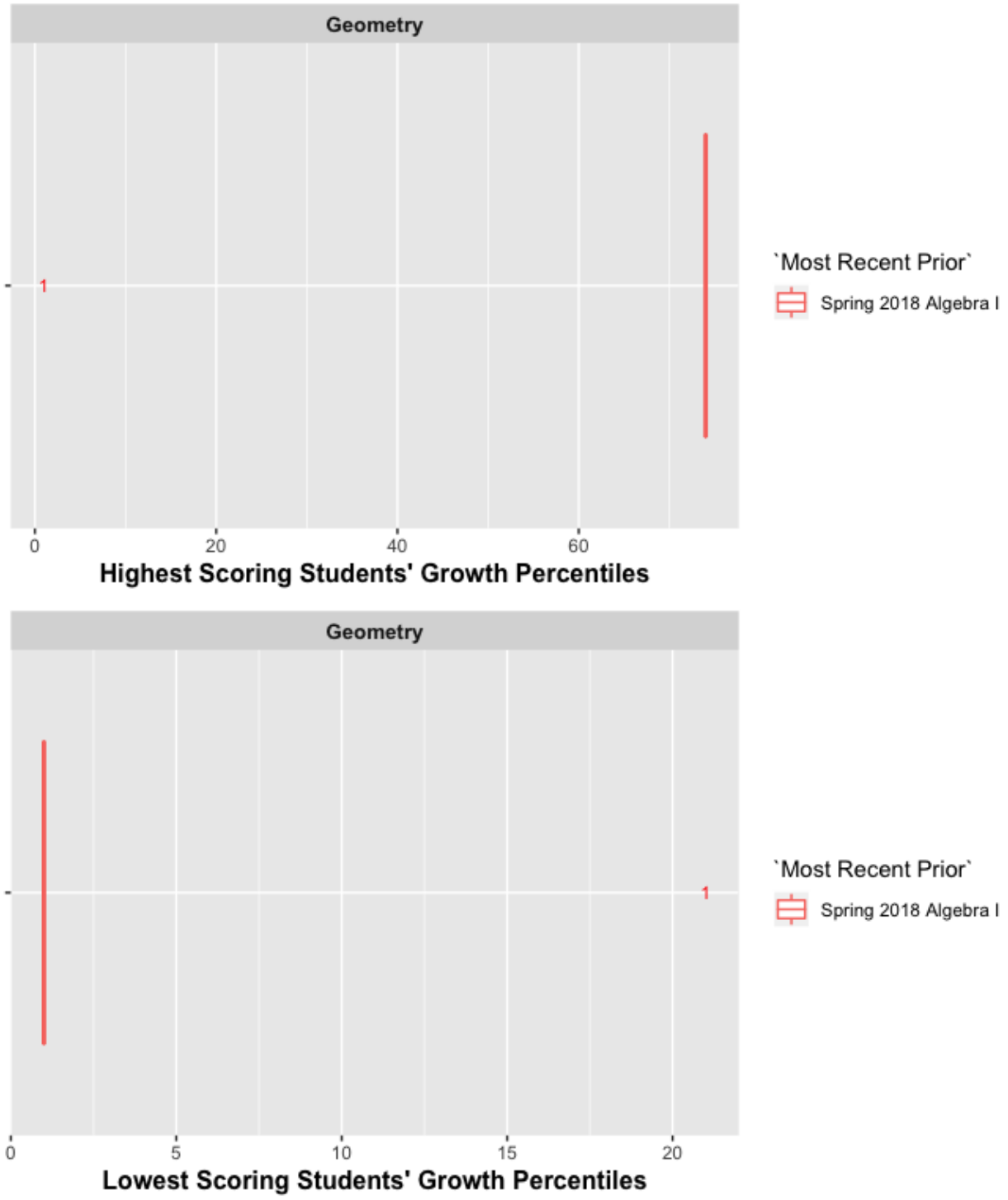
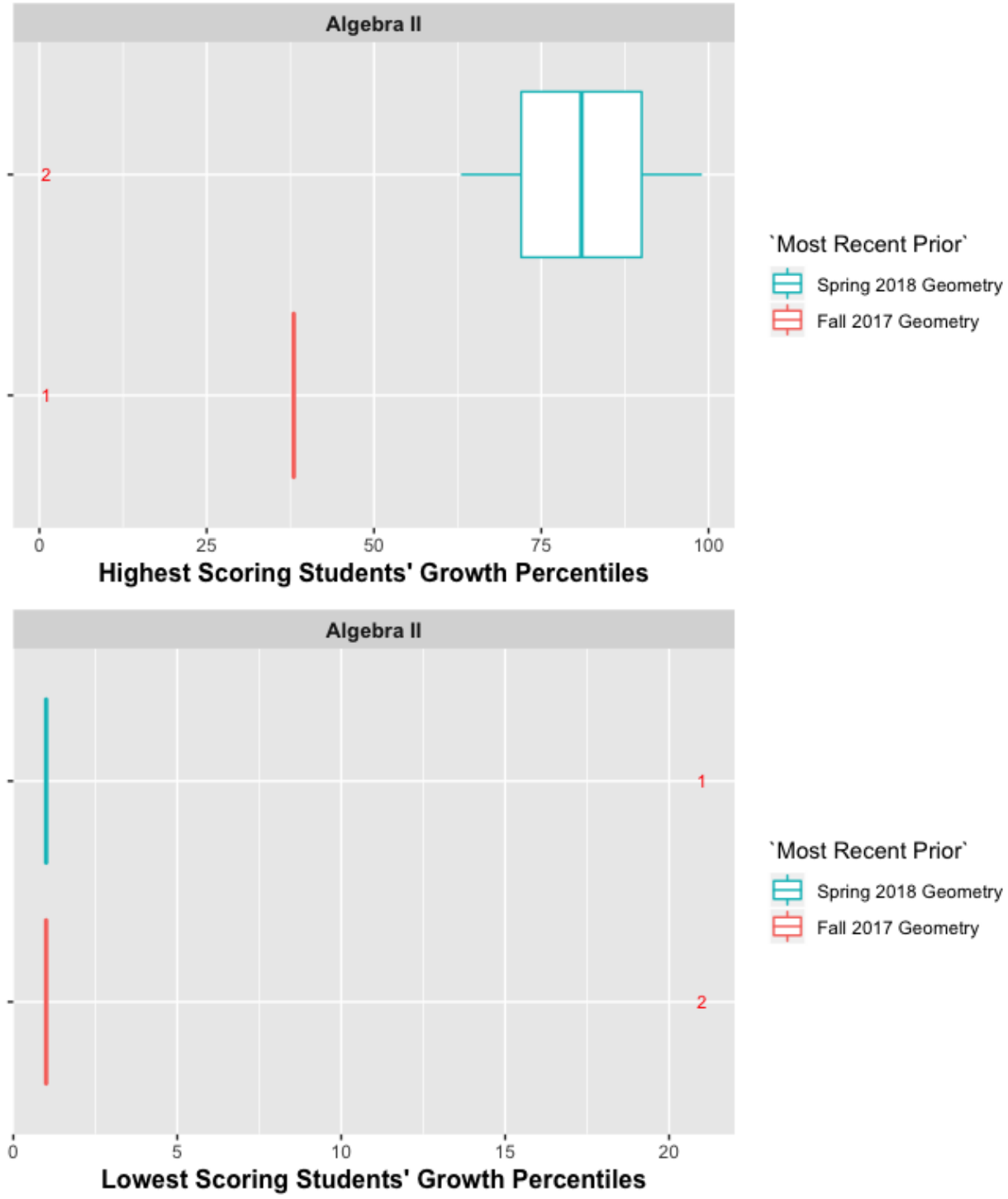


Fig. C.9: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Geometry.



The plots for both Algebra II progressions are notable. This progression can also be seen in Figure C.4 above. In the top bar plot we see that two students scored the HOSS, and their estimated SGPs were 63 and 99. The first student's prior score was also relatively high resulting in an unexpectedly low (although still relatively high) SGP estimate. In the second "bar", the student scored near the HOSS consecutively has an excessivly low SGP estimate of 36.

Fig. C.10: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Algebra II.



4 Discussion

Overall there is little evidence of widespread floor or ceiling effects in the Fall 2018 PARCC SGP analyses. Some unexpectedly low SGP estimates suggest that a few minor problems may exist that could require changes. When ceiling or floor effects are encountered, there are several ways in which they can be "corrected" manually or analytically. These include (but 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). However, as is shown in this report, some IRT estimates may be too extreme and create ceiling issues for some outliers.
3. Leave the results without a correction.