

Appendix C to the 2015 New Jersey PARCC 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)*

January 2016

# 1 Introduction

In the 2014-2015 academic year, New Jersey transitioned from the New Jersey Assessment of Skills and Knowledge (NJASK) to the [Partnership for Assessment of Readiness for College and Careers \(PARCC\)](#). 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). This has occurred despite purported improvements in assessment qualities that should prevent these effects (e.g. adaptive tests). Regardless of the source of assessment ceilings/floors, they can telegraph onto the Student Growth Percentile (SGP) calculations making these estimates questionable.

Although very similar in nature, ceiling effects are somewhat more problematic than floor effects because students that consistently achieve the highest scores receive lower than expected growth percentiles and are therefore negatively impacted. Conversely, the consistently lowest achieving students have higher estimated SGPs than would be expected. Although this could be interpreted as a positive impact for these students by giving them higher SGPs, it can also conceal unacceptably low growth that needs to be identified and addressed.

Essentially these problems are caused by the way in which a “percentile” is 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 tested with a relatively easy test and 20% of the students had a perfect score, these students would be defined as being in the 80<sup>th</sup> percentile because they scored higher than 80% of their peers. This is somewhat misleading however, because their score was *equal to* or greater than 100% of their peers and so could be also described as achieving at the 99.99<sup>th</sup> percentile under an equally valid definition.

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<sup>th</sup> percentile growth. Although there is nothing *technically* incorrect about this estimate because 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 at the highest observable level. 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.

Given these impacts and the difficulty in detecting them using traditional SGP diagnostic tools, the Center for Assessment has 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. Plots of the scale score distributions for the current and prior years’ tests, which may provide an indication of whether a ceiling or floor is present in either (or both) the current or historical data.
2. Box plots showing the range and distribution of SGPs for *only* the highest and lowest achieving students in the current year.

## 2 Prior and Current Year Score Distributions

The marginal and conditional distributions of scale scores can serve as a preliminary indicator of potential ceiling or floor effects in the calculation of student growth percentiles. Some minor problems could present themselves if these characteristics are present in either prior or current year scores, and are particularly likely when present in both. The plots below depict distributions for the current year and the most recent prior year used in the SGP calculations. The marginal (individual or univariate) distributions for each year are shown in the first subsections below, followed by the conditional (joint) distributions. Each content area plots are disaggregated by grade.

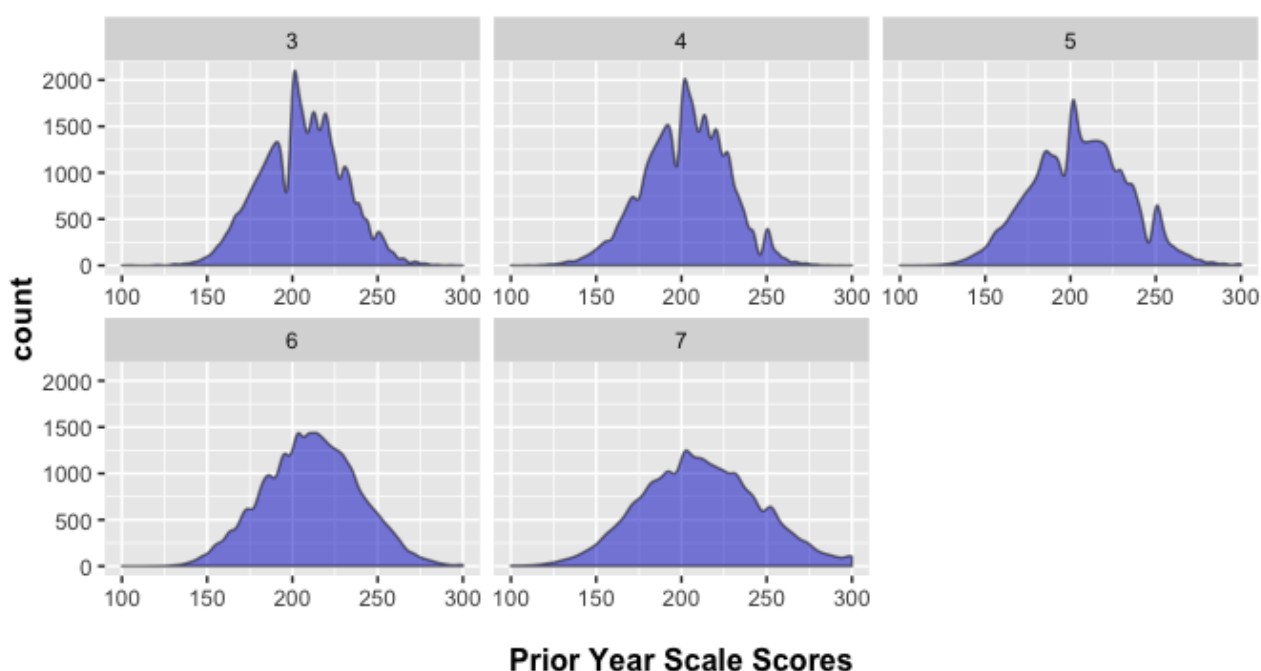
### 2.1 Marginal Distributions

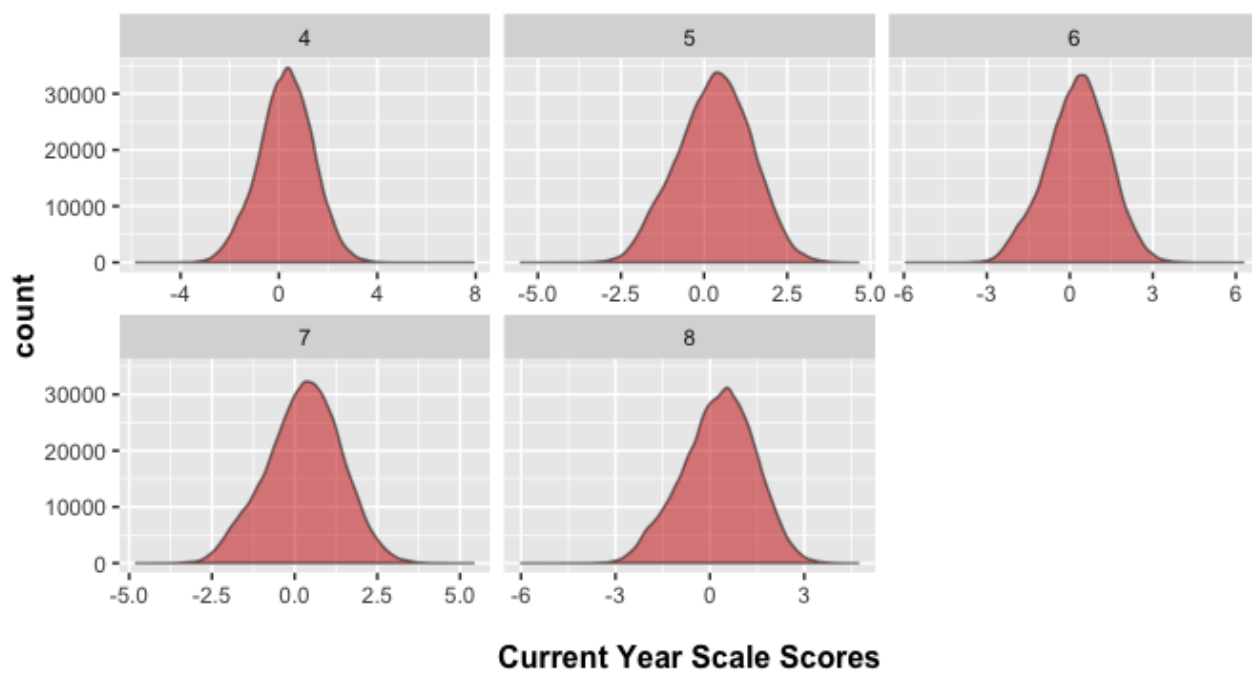
Generally there is evidence of ceiling effects across all grades and in both subjects in the NJASK prior scores (although it is more prominent in the Mathematics tests). These effects appear either as fat, truncated tails or as sharp spikes in the distribution at the extreme right side of the distributions.

The NJASK ELA scores display fat tails and the Mathematics score distributions are truncated with sharp spikes at the upper end of the distributions, both of which suggest that there may be issues with students consistently achieving maximum scores. This could potentially translate to ceiling effects in growth analyses.

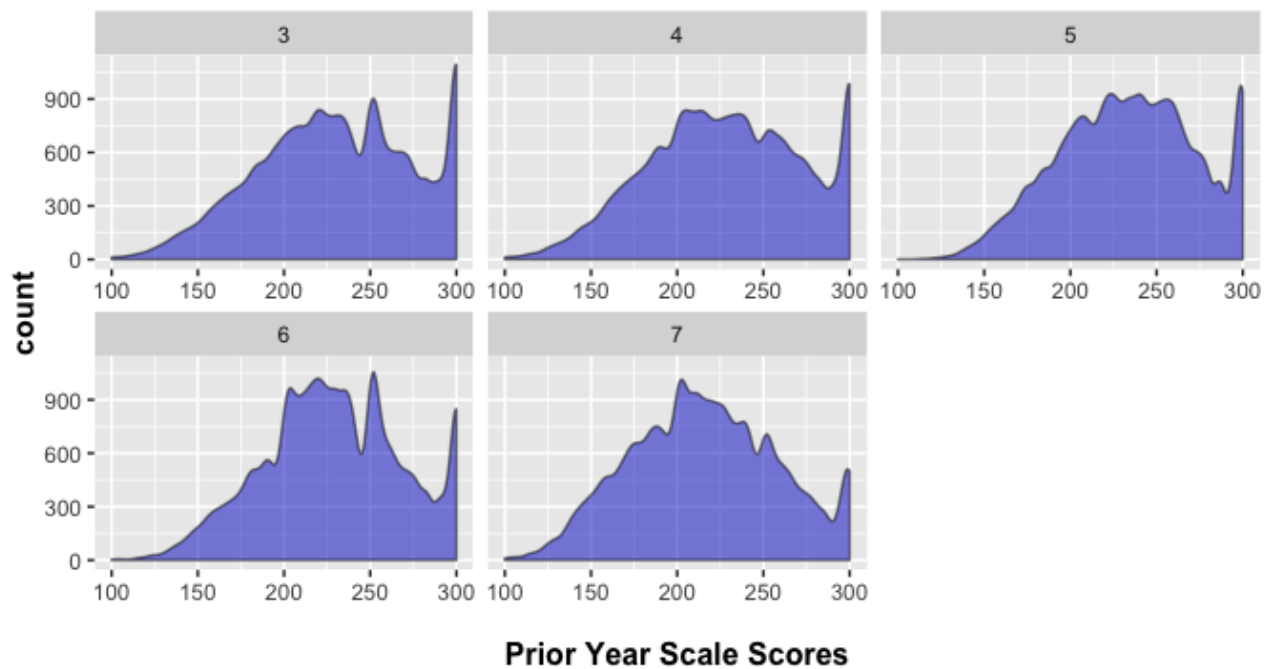
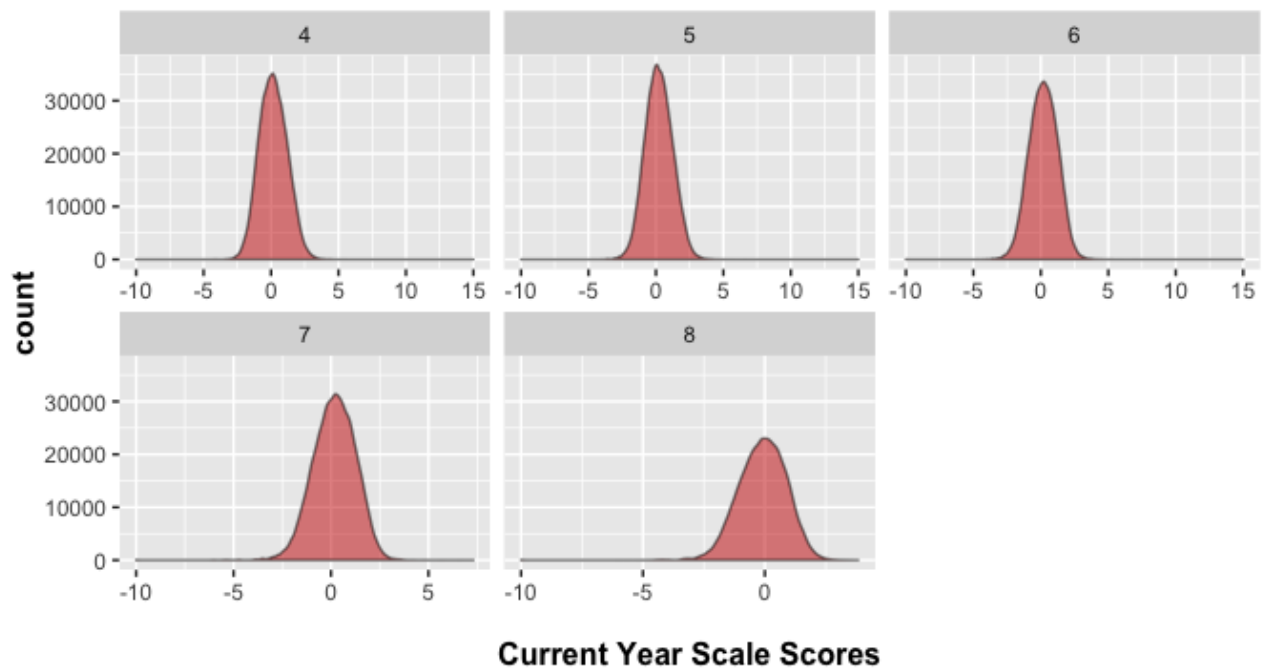
#### 2.1.1 ELA

**Fig. C.1:** Marginal distributions of prior scale scores: EOGT ELA.



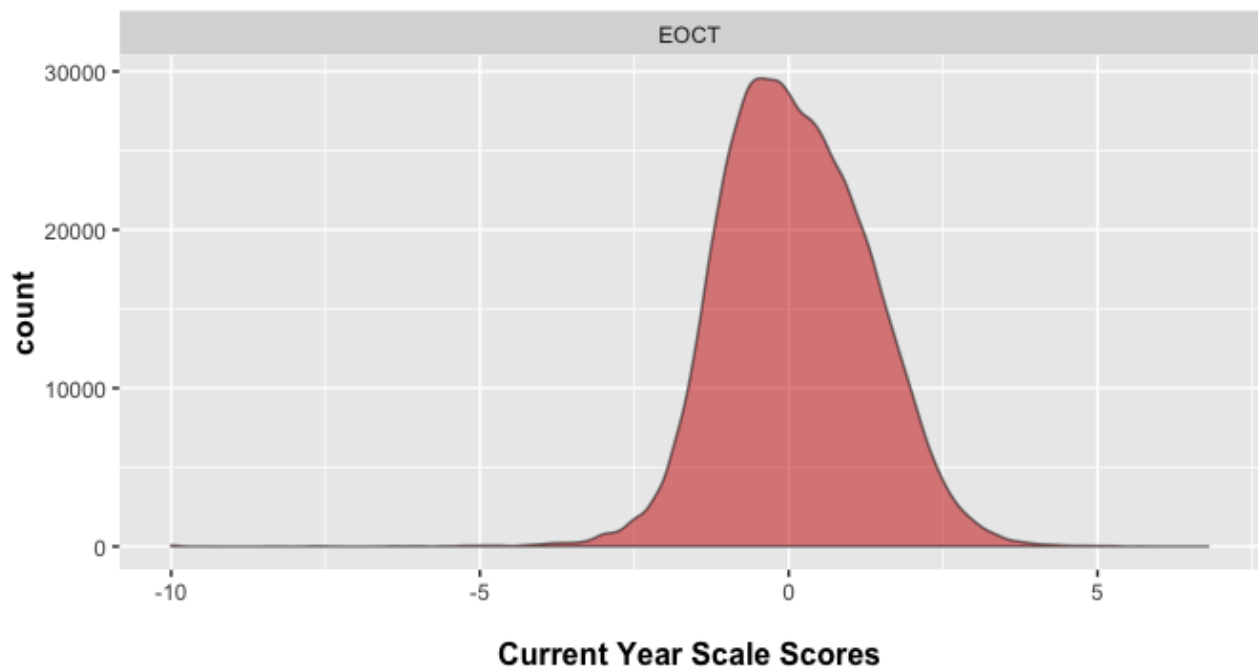
**Fig. C.2:** Marginal distributions of current scale scores: EOGT ELA.

## 2.1.2 Mathematics

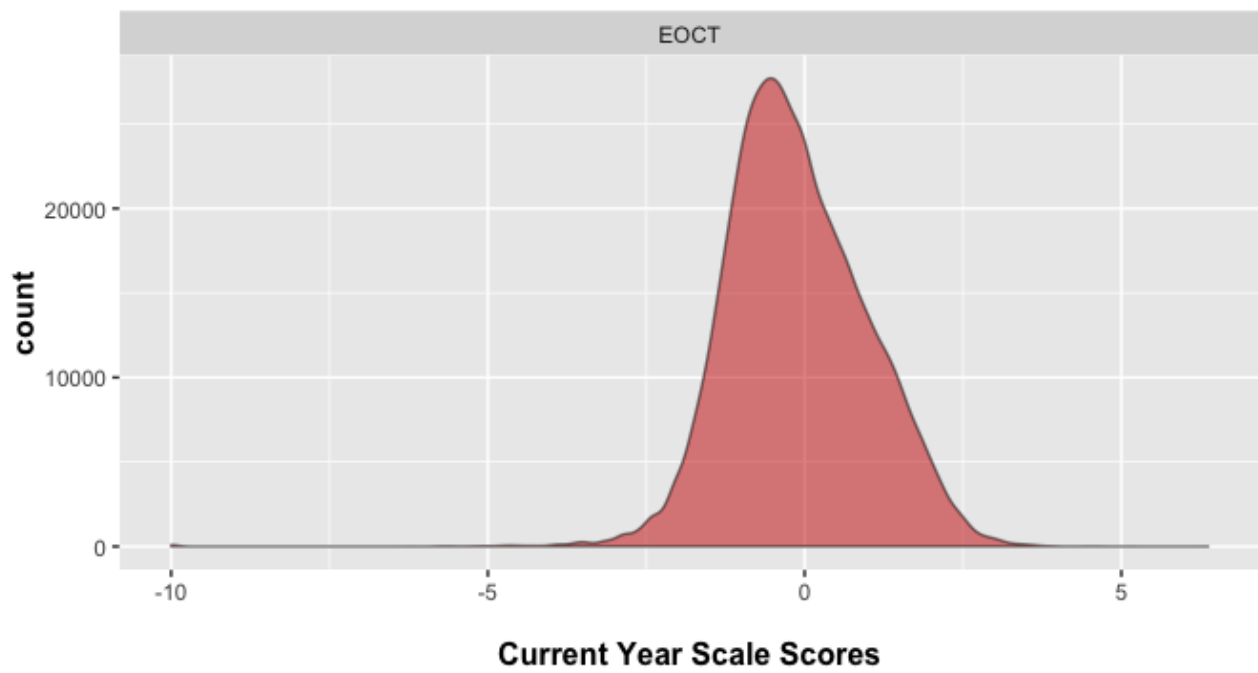
**Fig. C.3:** Marginal distributions of prior scale scores: EOGT Mathematics.**Fig. C.4:** Marginal distributions of current scale scores: EOGT Mathematics.

### 2.1.3 EOCT Mathematics

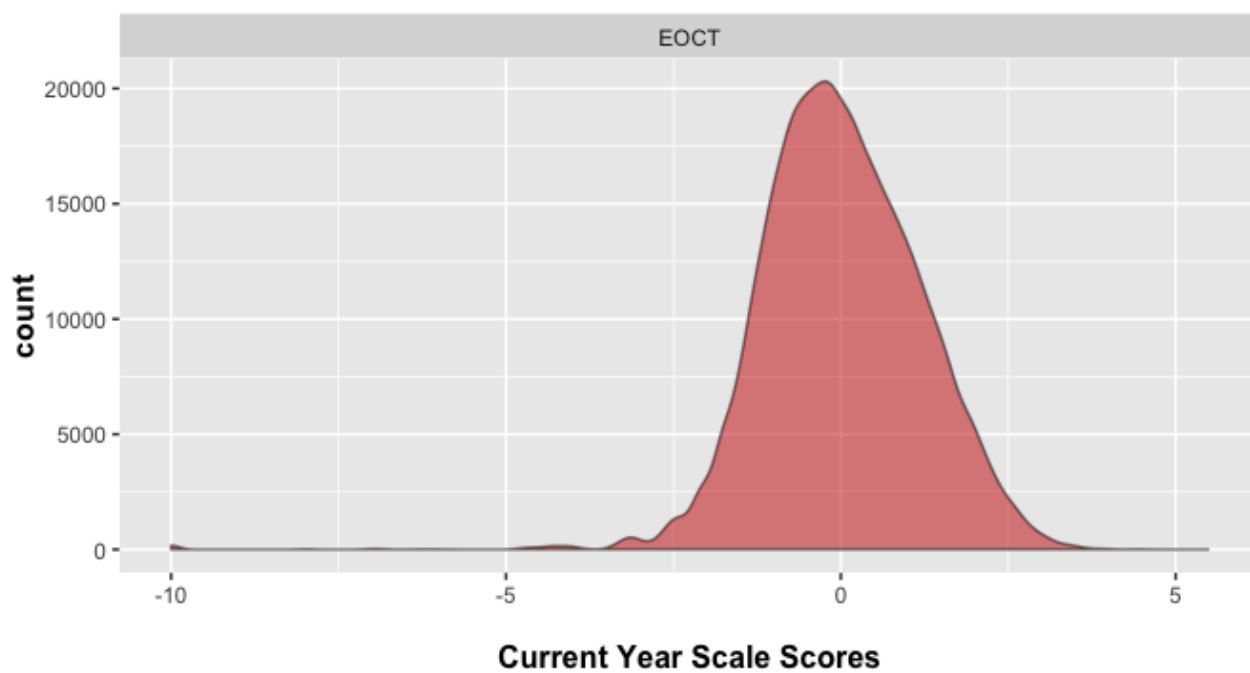
**Fig. C.5:** Marginal distributions of current scale scores: Algebra I.



**Fig. C.5:** Marginal distributions of current scale scores: Geometry.



**Fig. C.5:** Marginal distributions of current scale scores: Algebra II.



## 2.2 Conditional Distributions

The marginal density plots provide a limited amount of information, particularly for the potential for ceiling/floor effects in the calculation of *growth*. In order to provide a more nuanced view of the relationship between the prior and current scale scores, the following plots depict the conditional (joint) distributions for each content area and grade level. These plots start with a basic scatter plot of each student's scores, and on top of this is layered 1) **green contour lines** to provide a sense of the joint density, 2) a **magenta non-linear line** identifying the bivariate relationship between prior and current scores, and 3) **rug plots** that describe the marginal distributions (as above, the prior scores are blue and current scores are red).

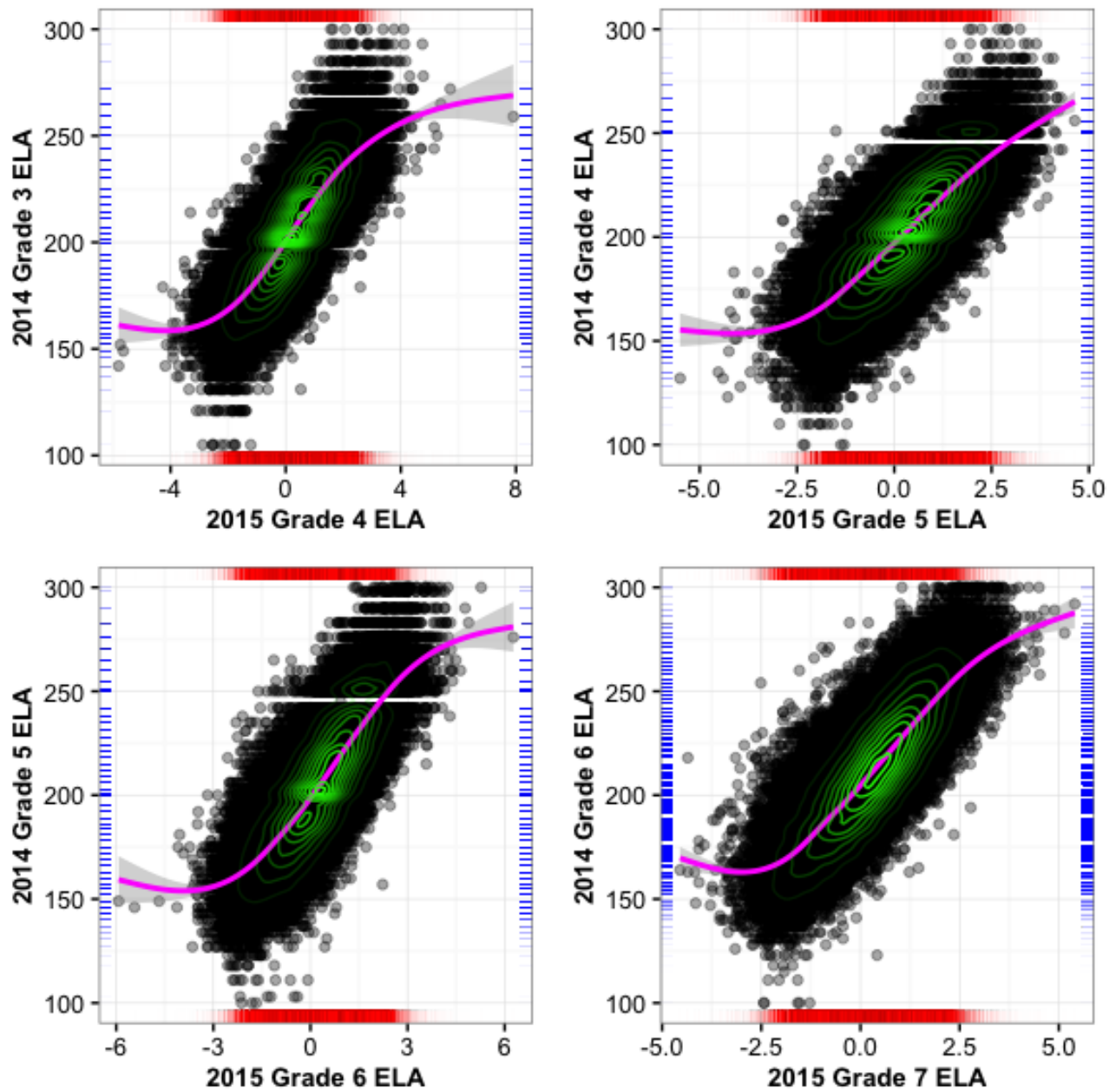
When ceiling or floor effects are present, they present themselves as dark shaded points in the extreme top-right or bottom-left corners of the plots. This suggests that staying at the extremes from year to year is not uncommon, which may lead to odd growth estimates for these chronically high/low achieving students. For the 2014 and 2015 New Jersey data, we again see very few issues in all content area and grade combinations. Where minor ceiling/floor effects appear in the prior year (NJASK) data, the 2015 score distributions for these students are well distributed. That is, in plots in which a minor shaded point in the upper right corners are present the other students who scored at the HOSS on the 2014 NJASK tests had a wide range of PARCC scores, which lessens the concern for a ceiling effect<sup>1</sup>.

---

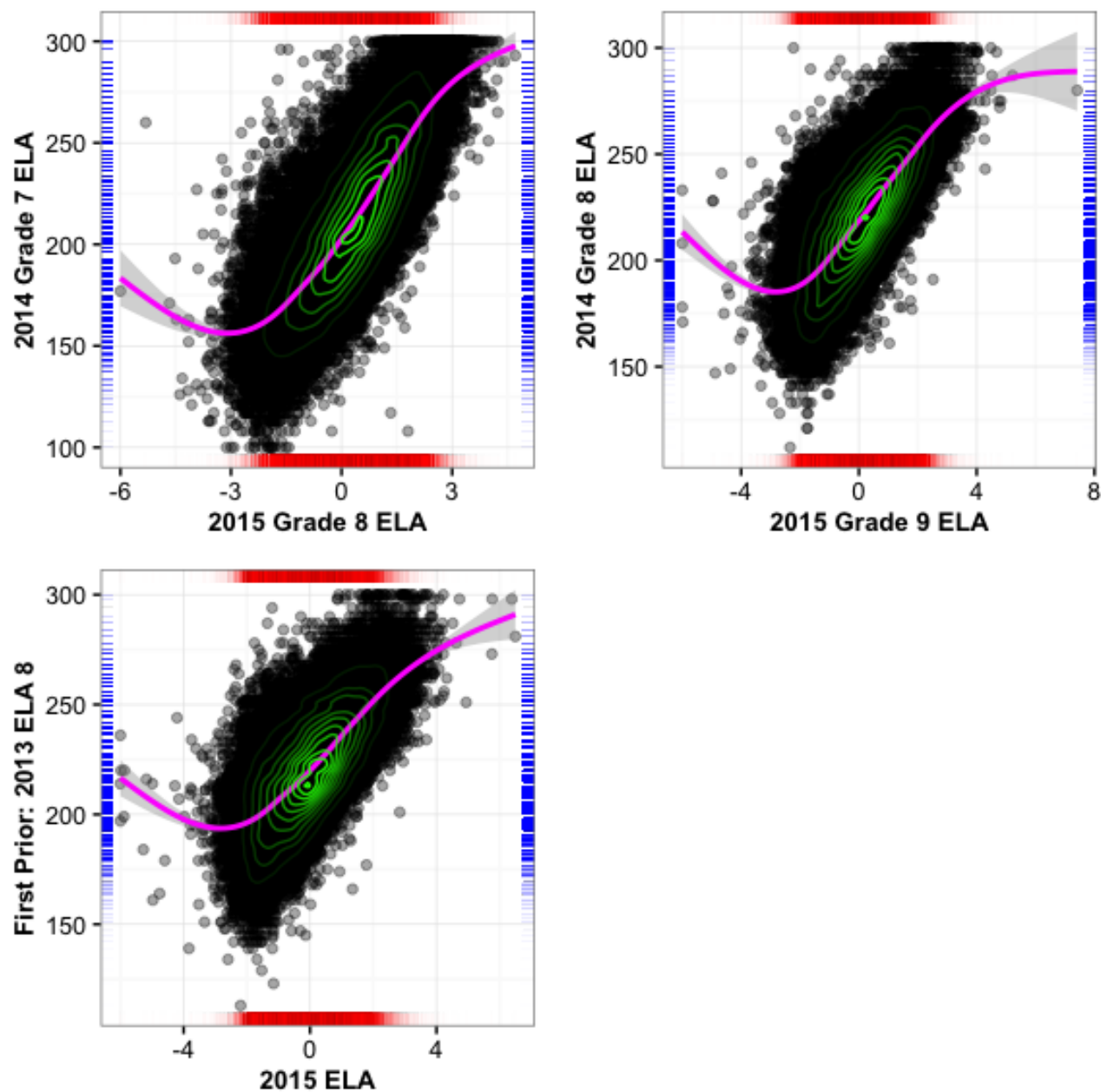
<sup>1</sup>Again, only when these academic peer groups are clustered in the HOSS/LOSS year after year do we encounter problems.



## 2.2.1 ELA

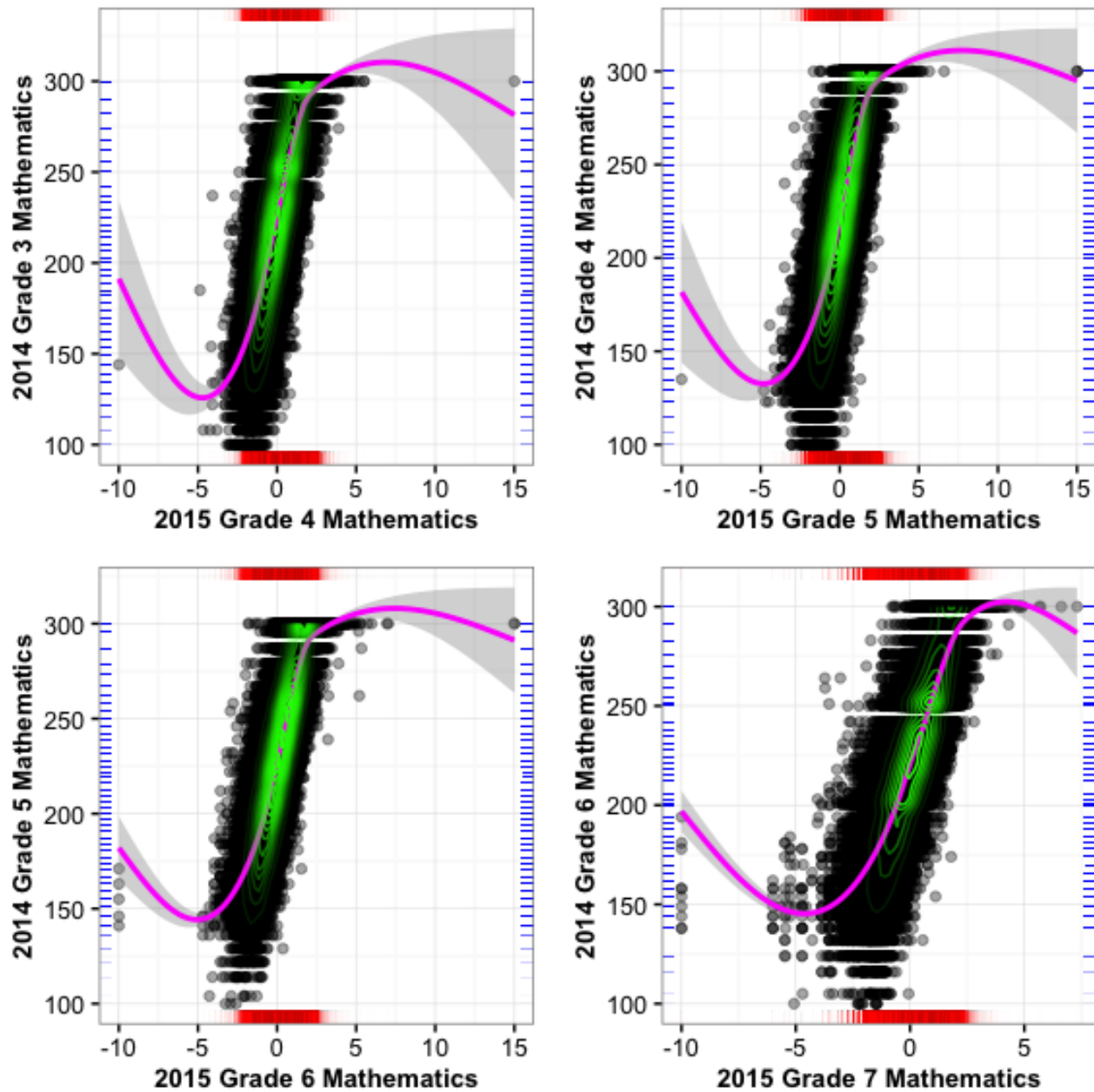
**Fig. C.6:** Conditional distribution(s) of current and prior scale scores: EOGT ELA.

**Fig. C.7:** Conditional distribution(s) of current and prior scale scores: EOGT ELA *Continued*.

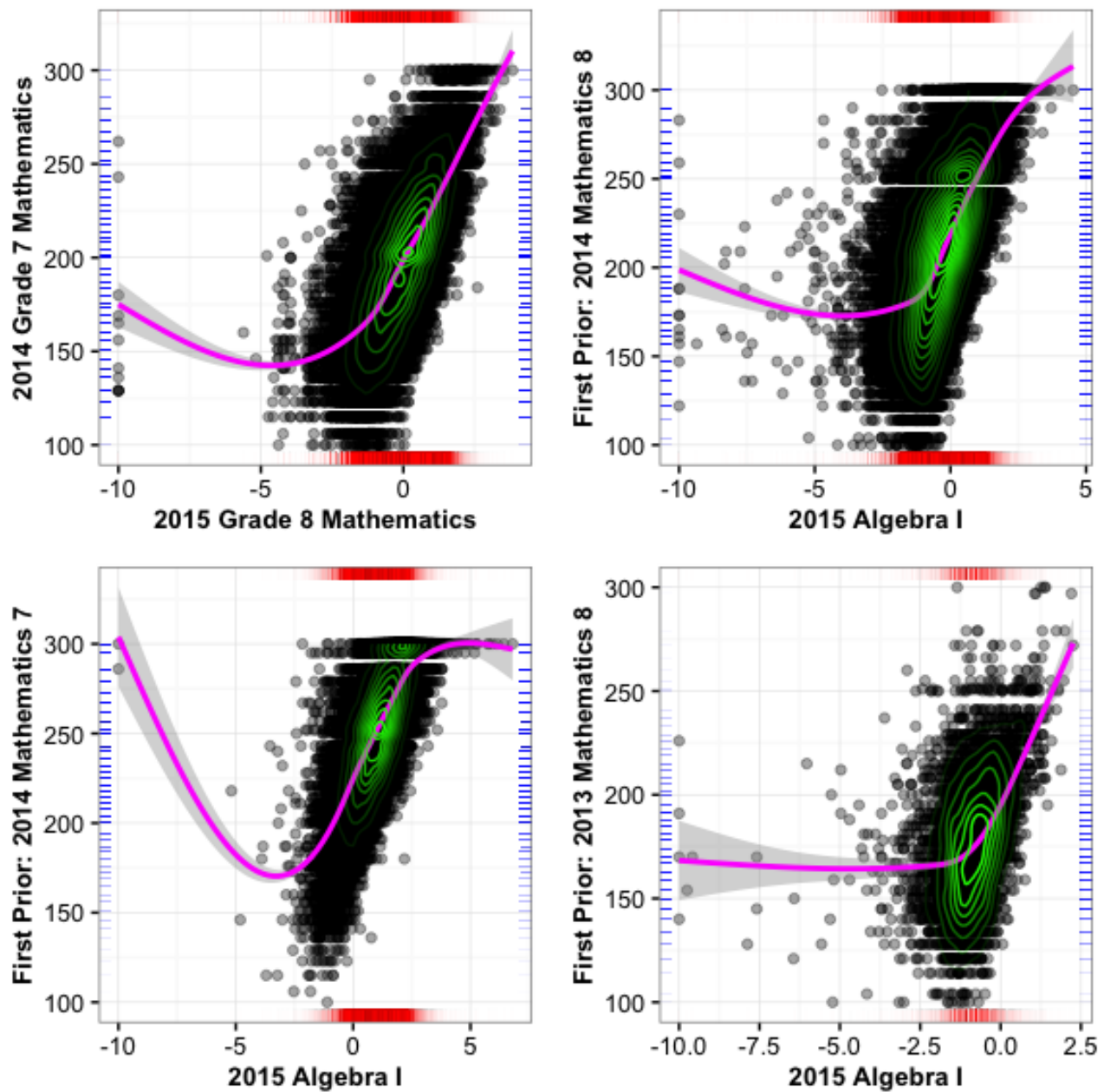


### 2.2.2 Grade Level and EOCT Mathematics

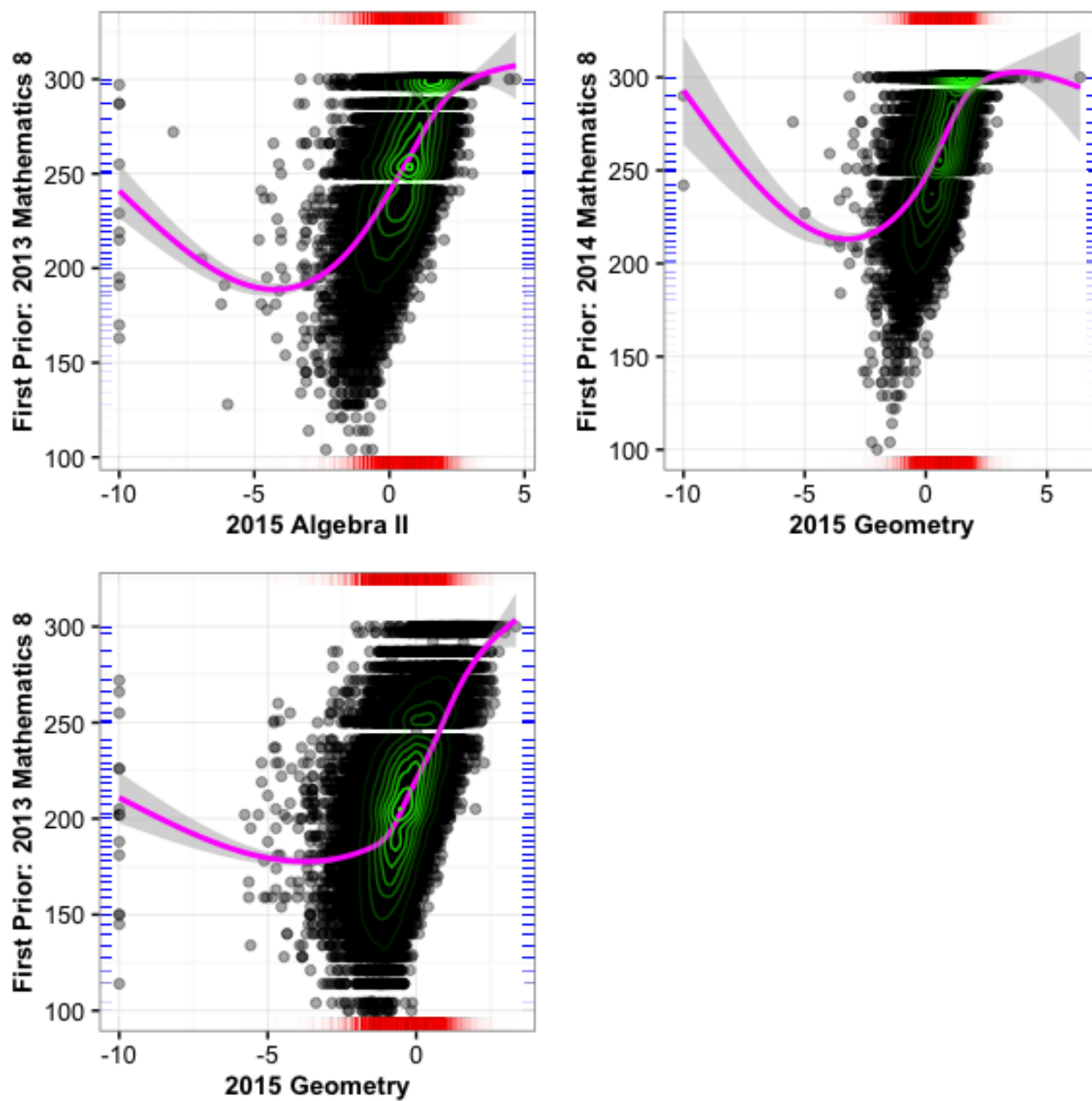
**Fig. C.8:** Conditional distribution(s) of current and prior scale scores: Grades 4-7 Mathematics.



**Fig. C.9:** Conditional distribution(s) of current and prior scale scores: Grade 8 Mathematics and EOCT Algebra I.



**Fig. C.10:** Conditional distribution(s) of current and prior scale scores: EOCT Algebra II and Geometry.



### 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 SGPs for students scoring at the highest/lowest obtainable scale score (HOSS/LOSS - i.e. the actual ceiling/floor). However, in order to assure that an adequate number of students are included in these plots, the first set of plots in each subsection uses, at a *minimum*, the highest/lowest 50 scores. Note that this roughly corresponds to the number of students used in the SGP model goodness of fit plots, and this is why these plots are provided as a starting point for this part of the investigation. 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 50 (the exact number is shown at the margins in red text).

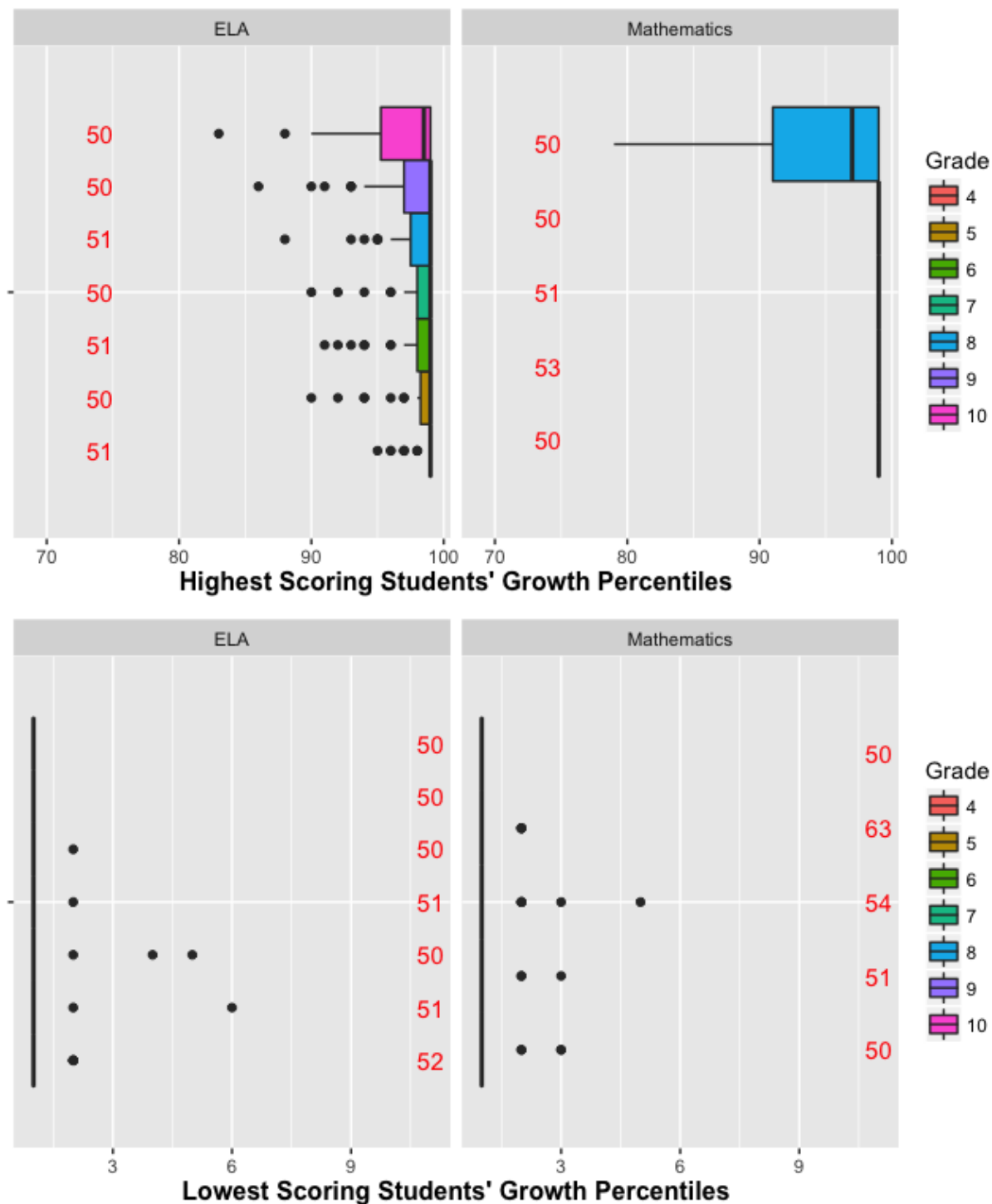
The second set of box plots in the following subsections isolate *only* those students scoring the HOSS/LOSS. These plots may then incorporate a widely varying number of students depending on the prevalence of a ceiling/floor in the current year.

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<sup>th</sup> and 75<sup>th</sup> percentiles). The upper whisker extends from the hinge to the highest value that is within  $1.5 \times \text{IQR}$  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  $1.5 \times \text{IQR}$  of the hinge. Data beyond the end of the whiskers are outliers and plotted as 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  $\text{SGP} = 99/1$ .

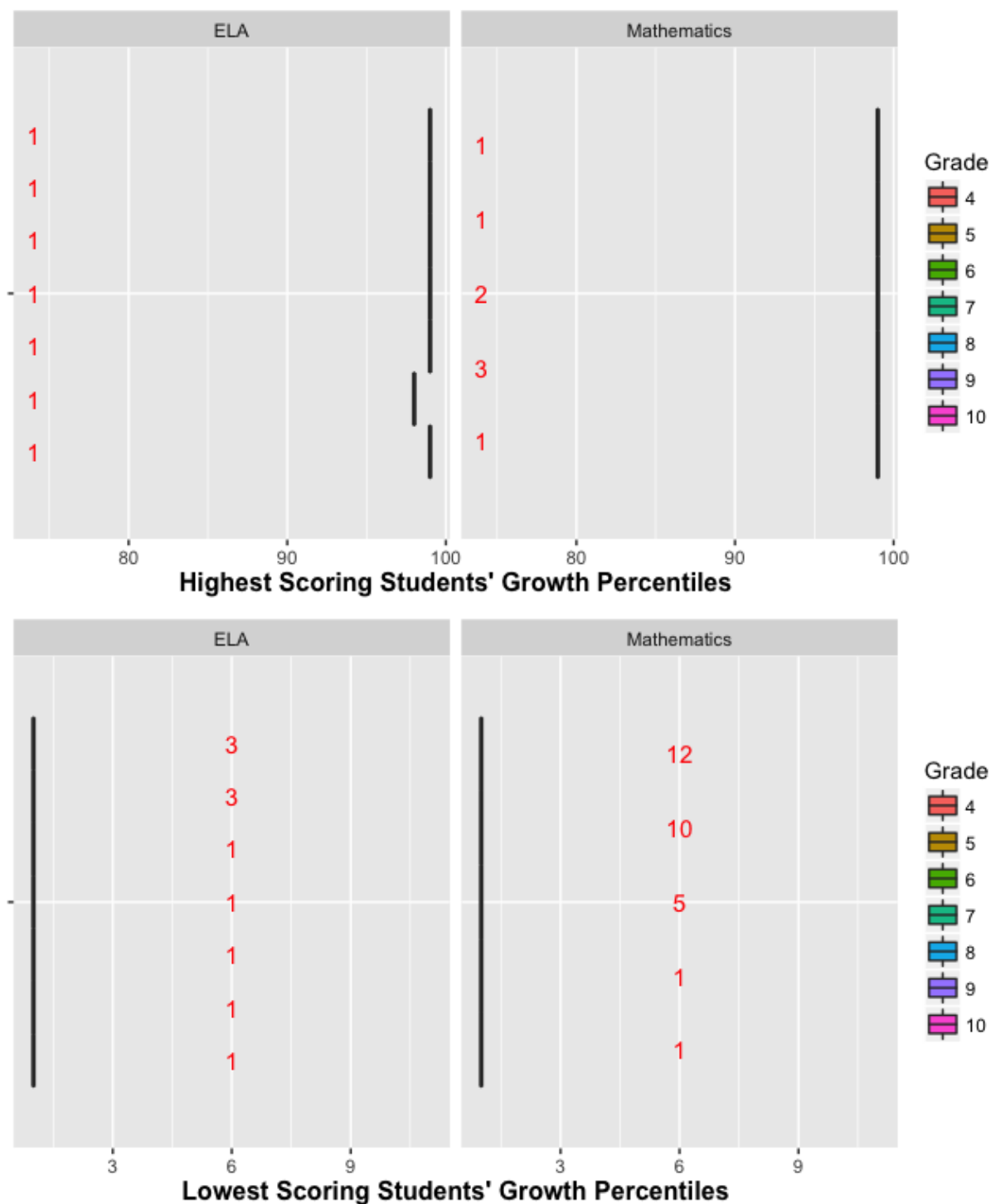
In the 2015 New Jersey EOGT SGP analyses, we see evidence of a very minor ceiling effect in 6th grade mathematics and some floor effects in Algebra I. However, these issues are present when considering the top 50 students' scores, but both disappear when considering only the HOSS/LOSS scores.

### 3.1 EOGT Content Areas

**Fig. C.10:** SGP distributions for highest and lowest 50+ scale scores by content area and grade level.



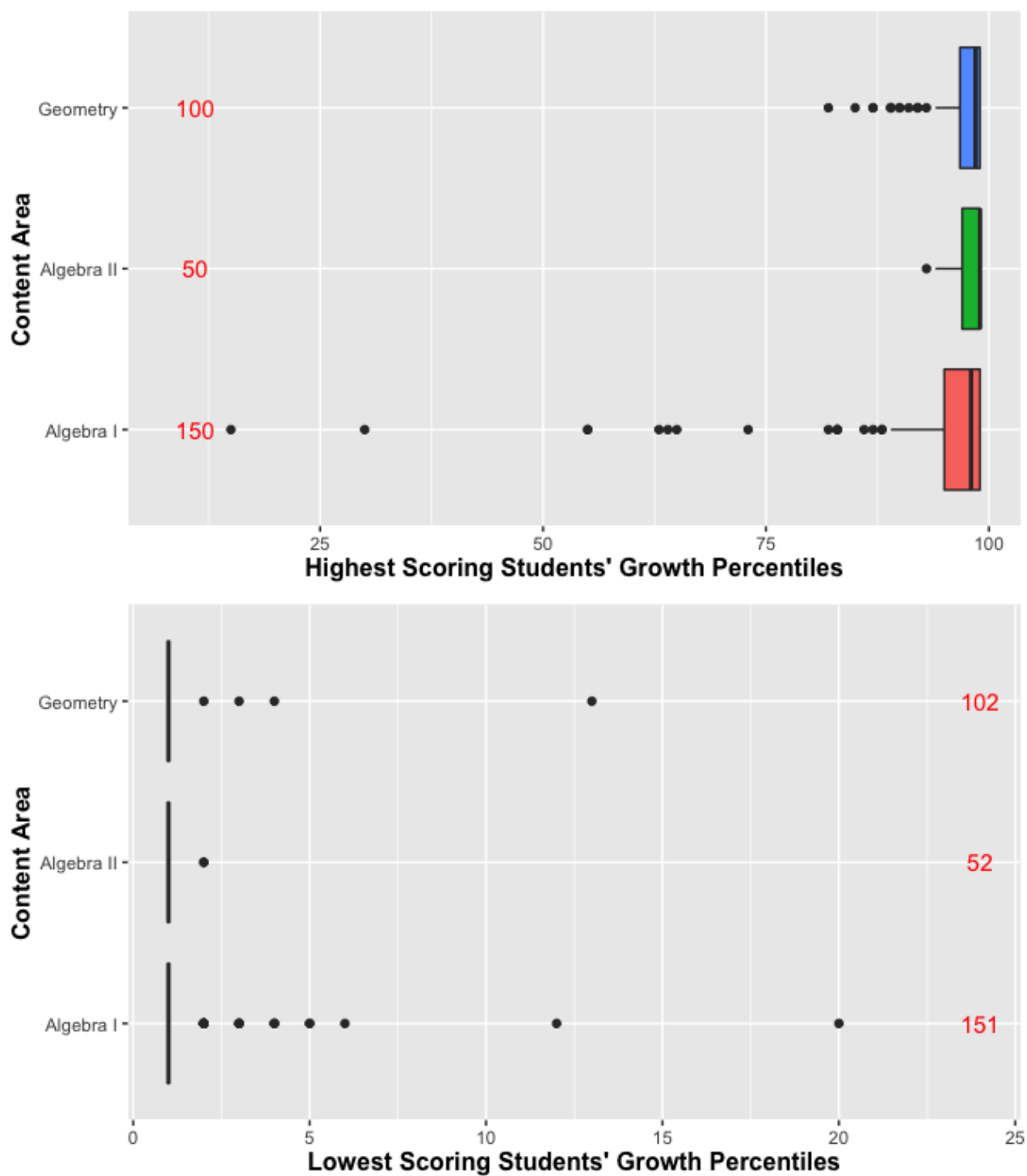
**Fig. C.11:** SGP distributions for the HOSS and LOSS scores by content area and grade level.



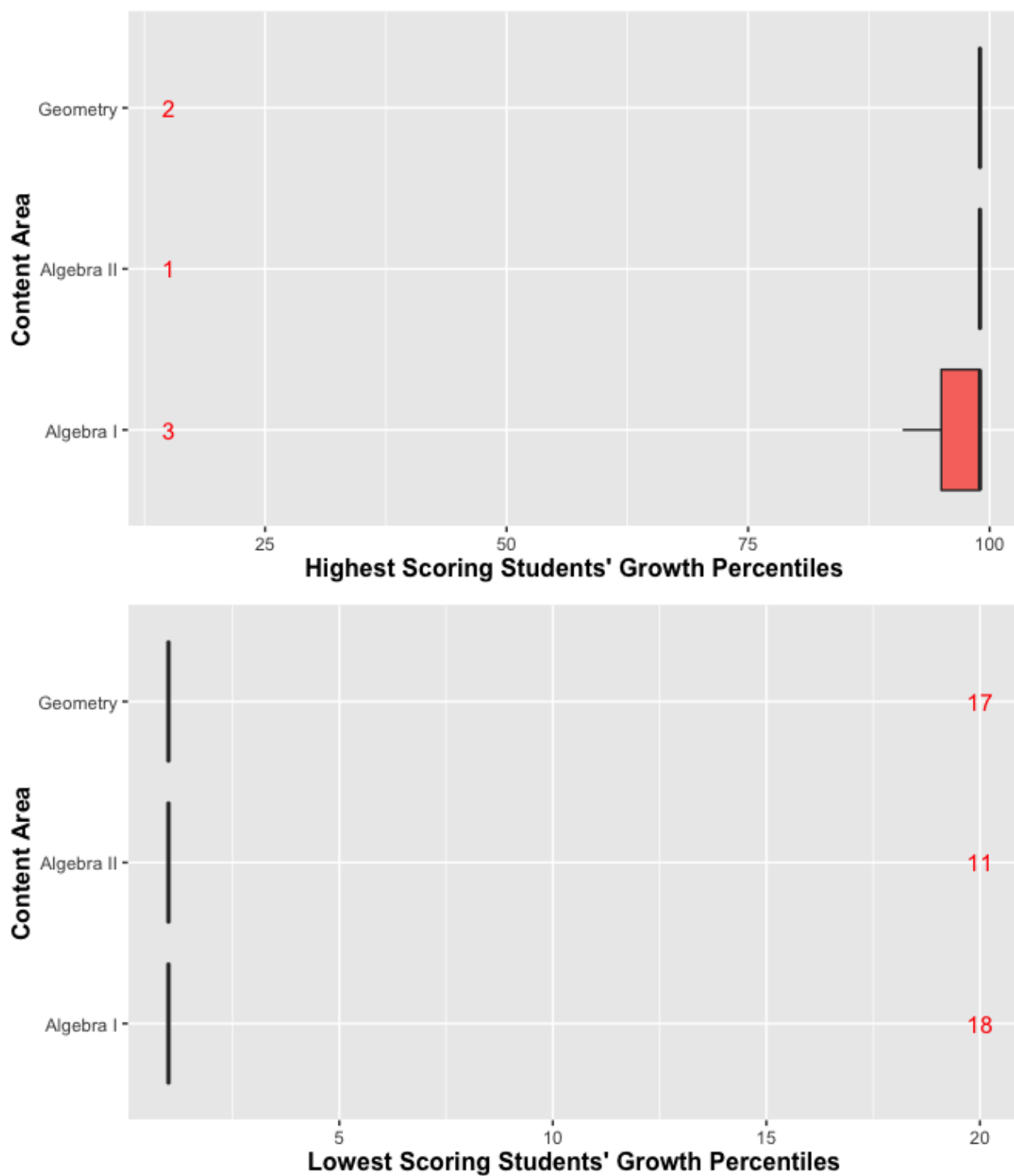


### 3.2 EOCT Content Areas

**Fig. C.16:** EOCT SGP distributions for highest and lowest 50+ scale scores by content area.

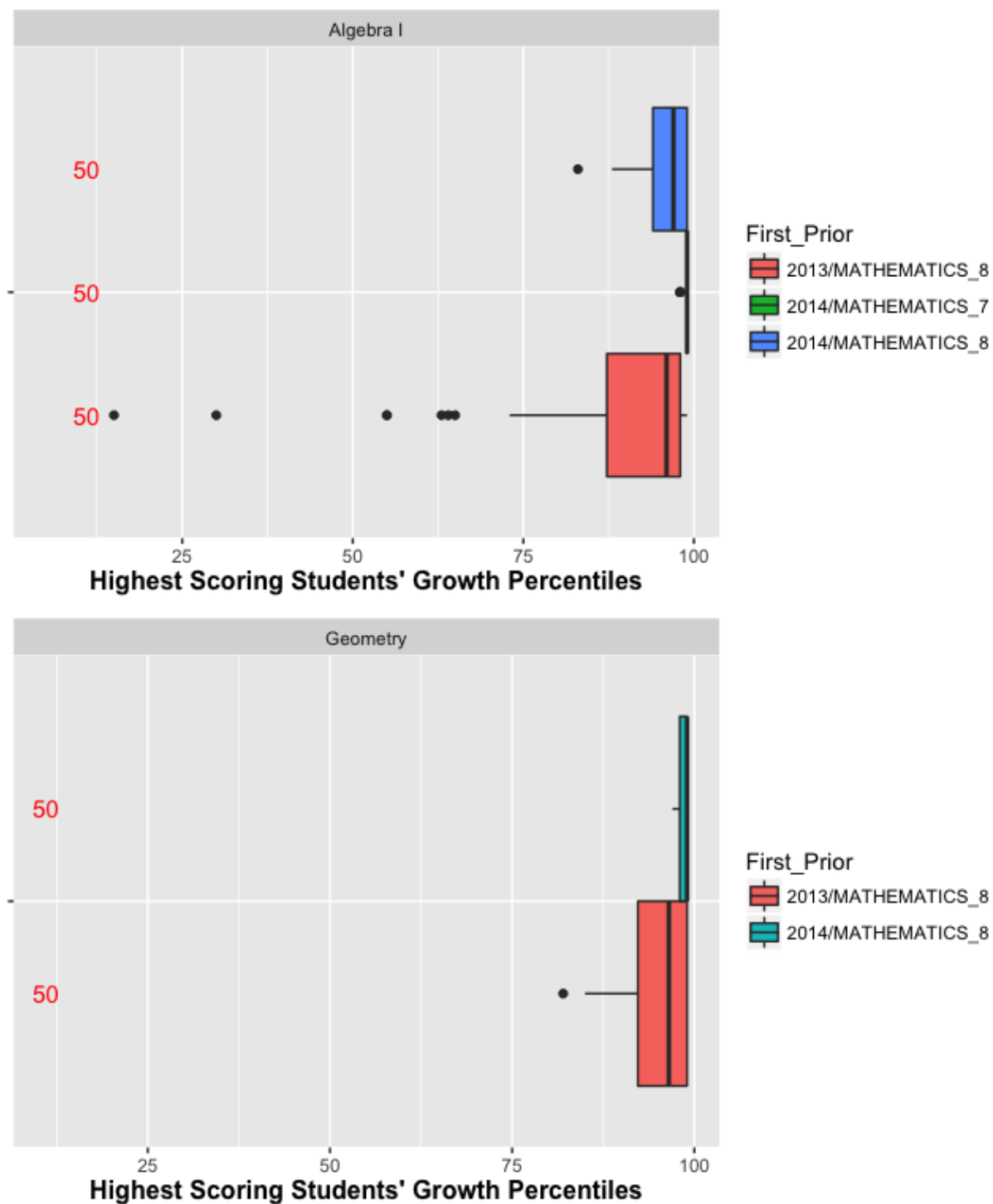


**Fig. C.17:** EOCT SGP distributions for the HOSS and LOSS scores by content area.



The Algebra I and Geometry box plots can be disaggregated further by the first available prior (i.e. their constituent norm groups)<sup>2</sup>. The following two plots show those for completeness of the report.

**Fig. C.16:** EOCT SGP distributions for highest and lowest 50+ scale scores by content area.



<sup>2</sup>Note that Algebra II only has a single norm group and so is not presented here

## 4 Discussion

Overall there is little evidence of floor or ceiling effects in the 2015 New Jersey PARCC 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 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  $\theta$  estimates).
3. Leave the results without a correction.