2016 ACP IM-ITE Regression Based Individual Analyses

To further investigate the potential impact of administration challenges during the first four days of the 2016 ACP examination on individuals a series of regression models were defined to predict ACP scores based on previous test data. The goal was to ascertain if restarts were associated with under-predicted performance.

Creating predicted ACP Scores

Using 2015 data that was unaffected by test administration incidents, three regression models were created to predict PGY1, PGY2, and PGY3 scores.

Model 1: PGY1 ACP Score = constant + b1*Step2; R^2=0.41
Model 2: PGY2 ACP Score = constant + b1*Step2 + b2*Prior_Yr_ACP_Score; R^2=0.66
Model 3 PGY3 ACP Score = constant + b1*Step2 + b2*Prior_Yr_ACP_Score; R^2=0.68

As expected, the models for PGY 2 and 3 better modelled the outcomes since they contain earlier ACP scores that were relatively recent and are content equivalent relative to USMLE Step2CK scores.

Applying the models to 2016 Data

The models above were then applied to 2016 ACP data from the first 8 days of testing where the first four days reflected administrations that featured multiple restarts and days 5-8 where restart rates were at expected baseline levels.

In order to facilitate comparisons between 2016 ACP Scores and Predicted scores, a preliminary equating based on data from days 5-8 was performed and applied to all ACP scores to create preliminary on scale ACP scores that are on the same metric as the 2016 predicted scores.

Residuals were computed for each individual. Residual = Actual-Predicted

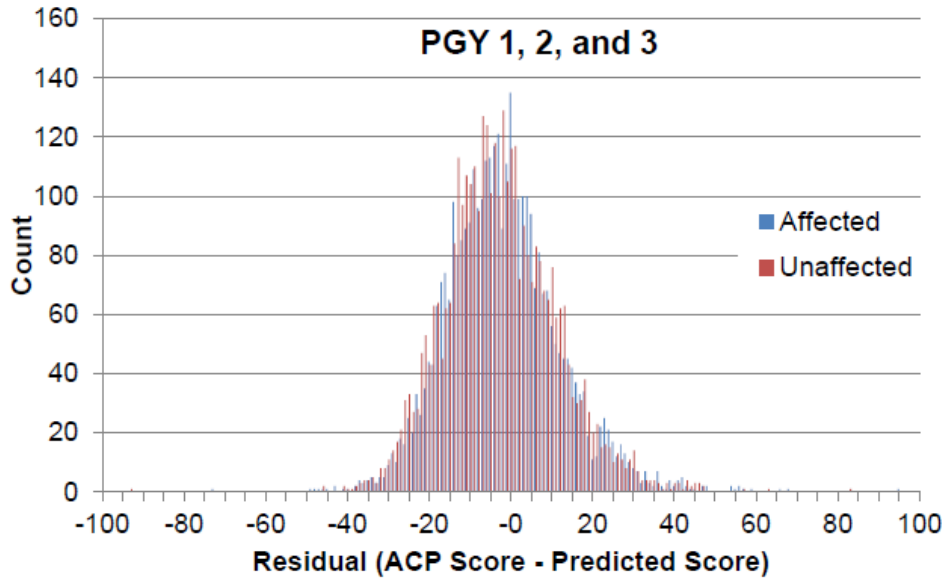
Summary of Residuals

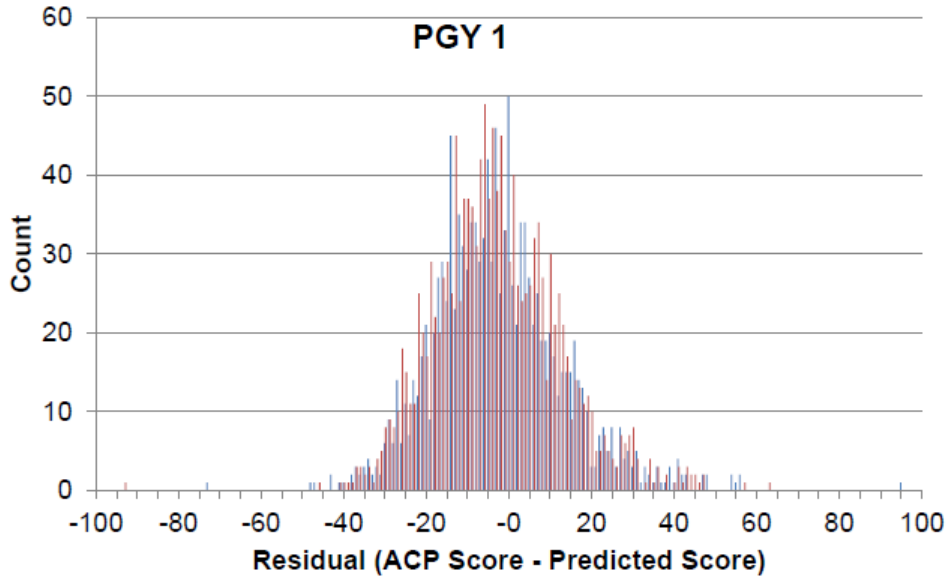
A summary of the actual, predicted, and residual scores (actual – predicted). No major difference between those who were or were not impacted, in fact, those that were affected had slightly lower residuals by 1 point.

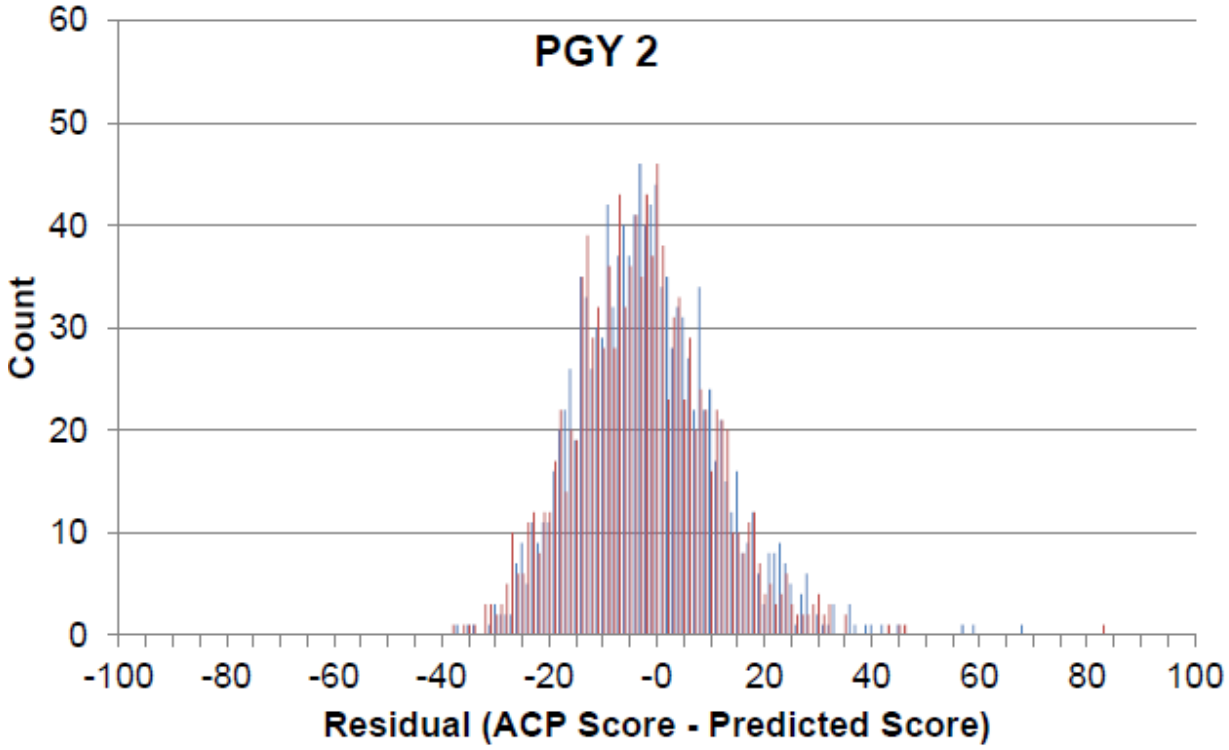
PGY	Affected	Actual			Predicted			Residual		
		Mean	N	SD	Mean	N	SD	Mean	N	SD
1	N	185	1380	20	188	1380	12	-3	1380	15
	Y	186	1248	20	189	1248	11	-2	1248	16
	Total	186	2628	20	189	2628	11	-3	2628	16
2	N	205	1158	22	207	1158	16	-3	1158	13
	Y	205	1207	21	207	1207	16	-2	1207	13
	Total	205	2365	21	207	2365	16	-2	2365	13
3	N	215	1157	23	217	1157	18	-2	1157	13
	Y	215	1113	22	217	1113	17	-2	1113	13
	Total	215	2270	22	217	2270	17	-2	2270	13
Total	N	201	3695	25	203	3695	19	-3	3695	14
	Y	202	3568	24	204	3568	19	-2	3568	14
	Total	201	7263	25	203	7263	19	-2	7263	14

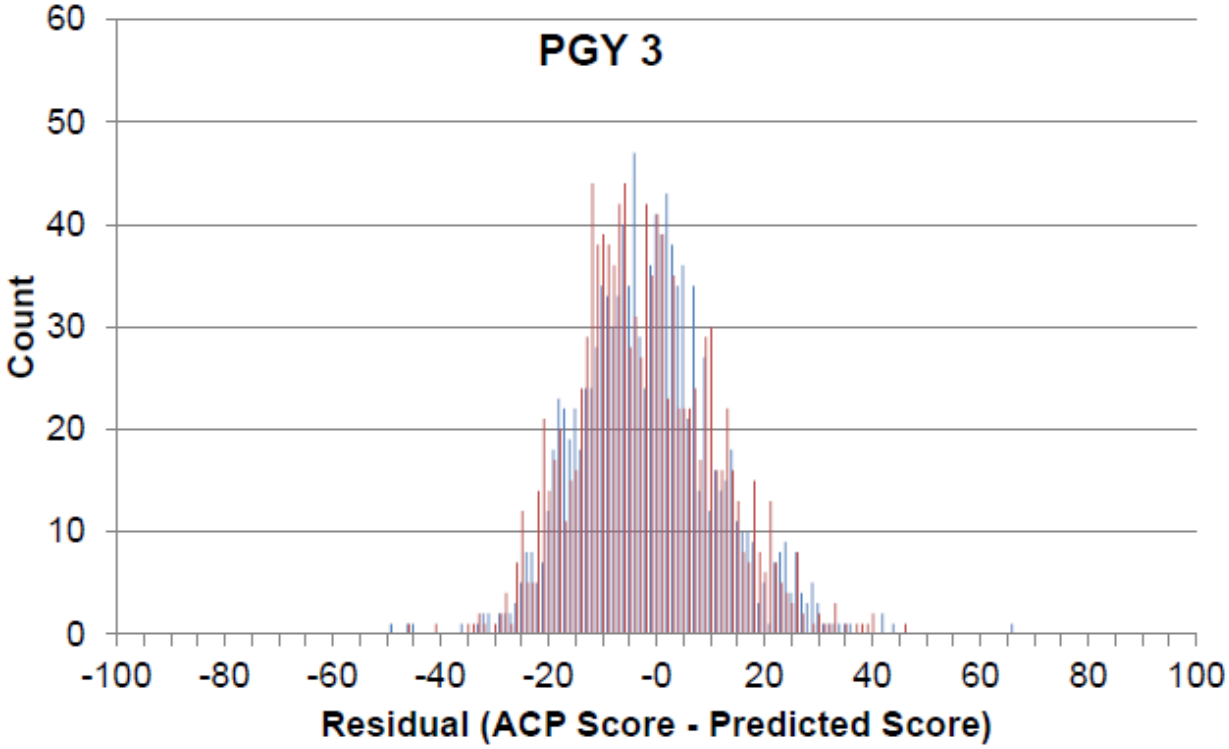
Plot of Residual for Affected and Unaffected Examinees

Histograms of the Affected and Unaffected examinees overall any PGY were completed. They demonstrate that the quality of the prediction models is strikingly similar for both the affected and unaffected individuals. It is quite suggestive that scores are equally well predicted for both affected and unaffected groups and consequently suggest no meaningful impact of administration challenges affecting 2016 scores.



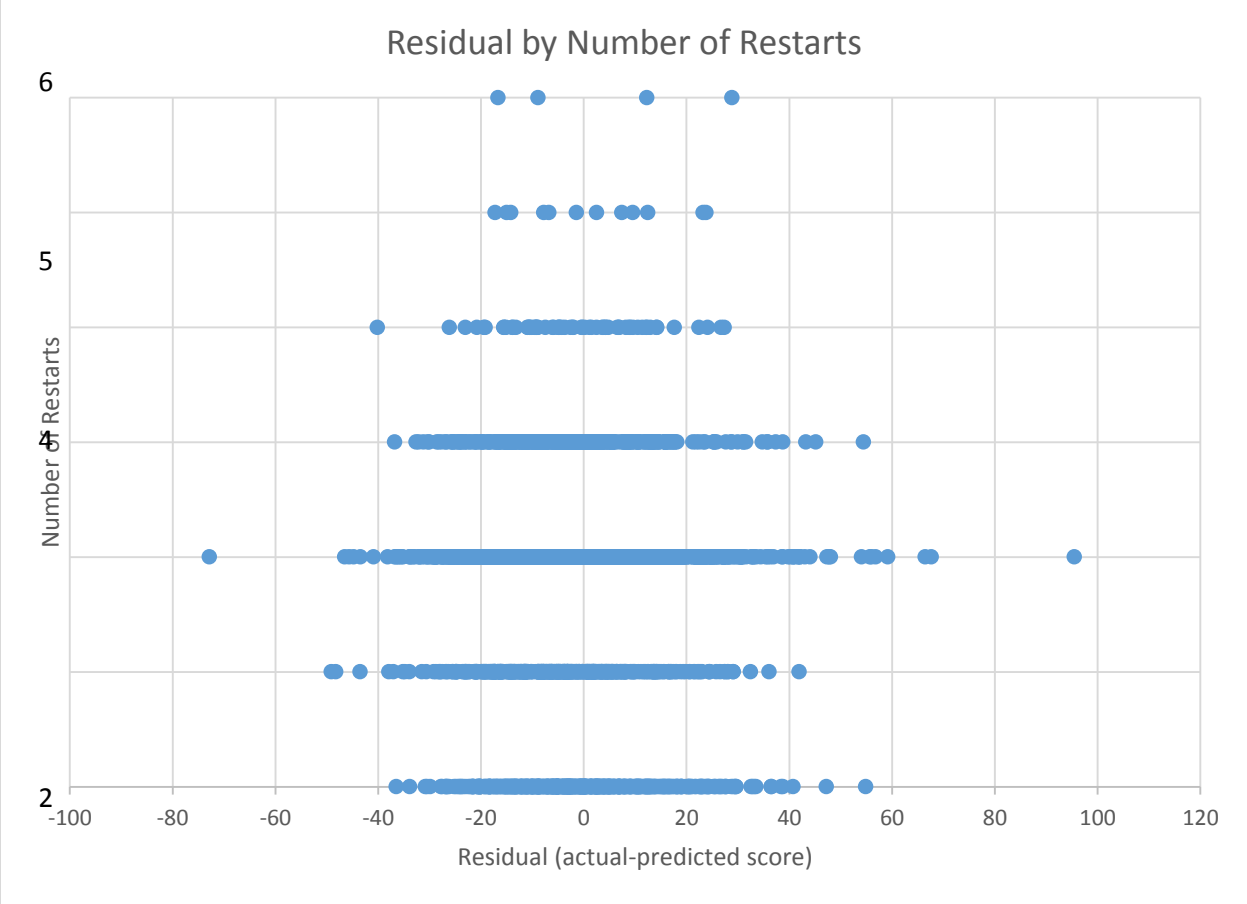



PGY 2
Count
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
-0
20
40
60
80
100
Residual (ACP Score - Predicted Score)


PGY 3
Count
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
-0
20
40
60
80
100
Residual (ACP Score - Predicted Score)

Residual Plots by Restarts

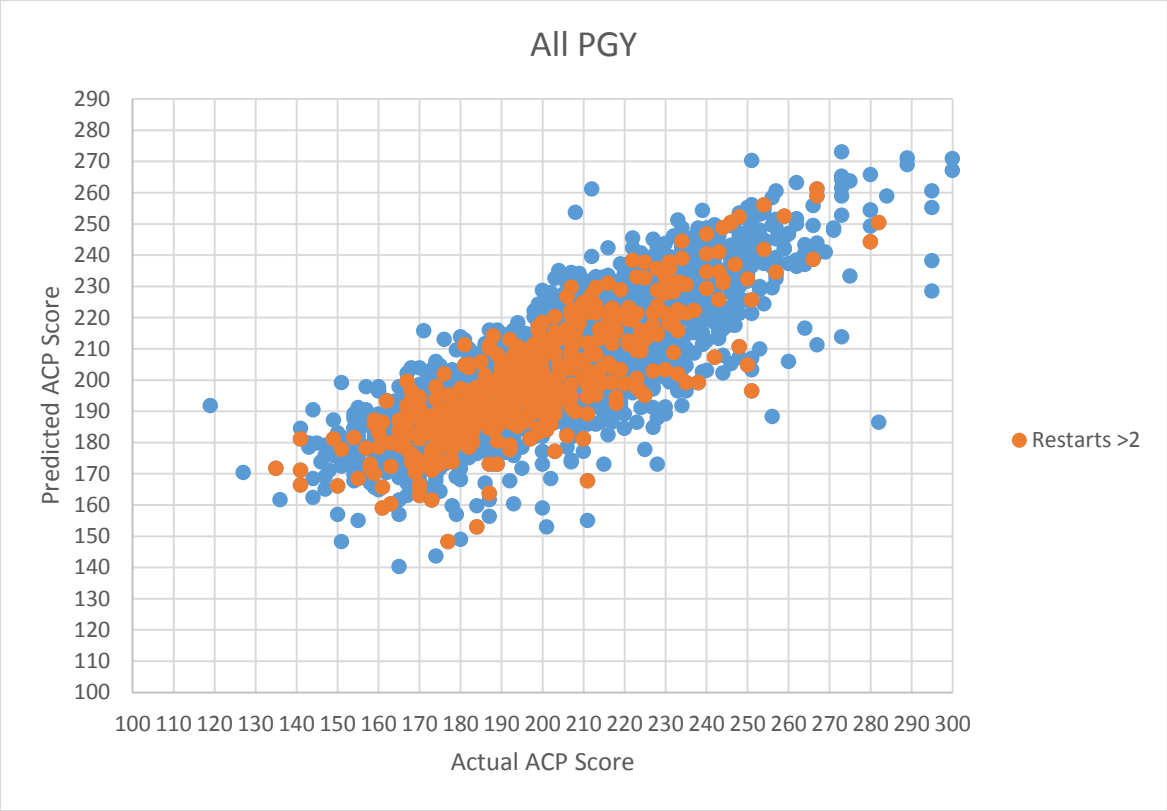
Next, a visualization of the residuals against a measure of administration challenge was created. The plot below shows the residual for affected examinees in the first four days by the number of restarts needed. As can be seen, in each case approximately equal numbers of examinees have positive and negative residuals. Further, the candidates with the most restarts have relatively smaller residuals. This is suggestive that administrative challenges were not associated with negative impact on scores. (A disproportionate number of relatively large negative residual would be anticipated if administration challenges were associated with restarts.)

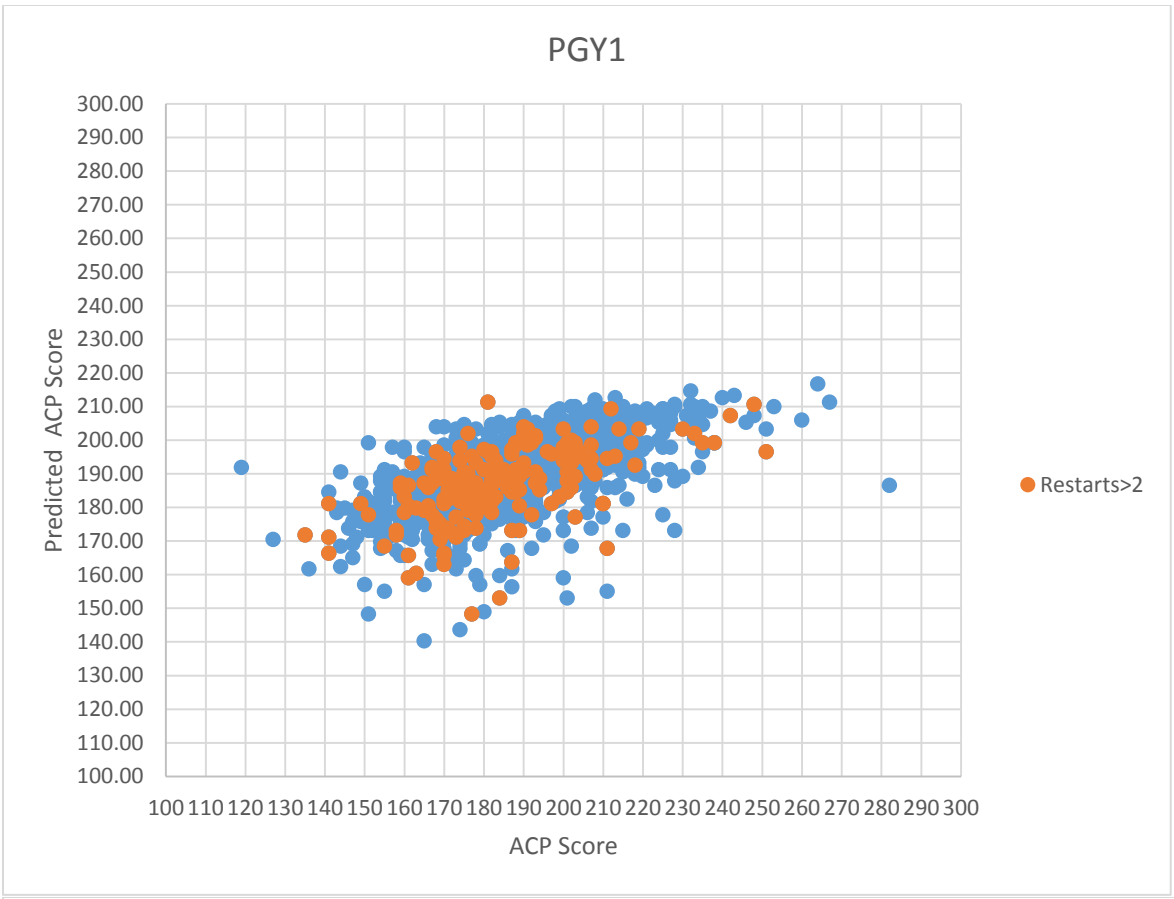


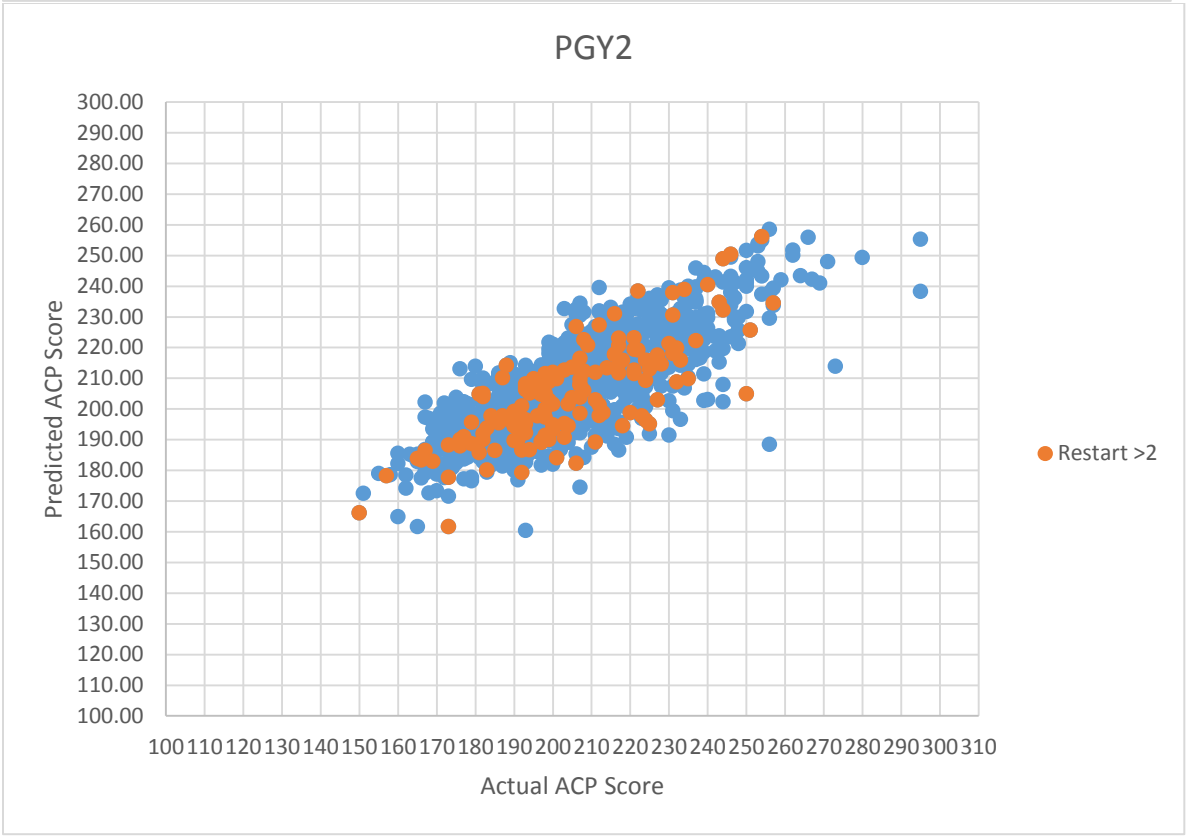
Scatterplot of Actual and Predicted Scores Overall and by PGY with Restarts Notated

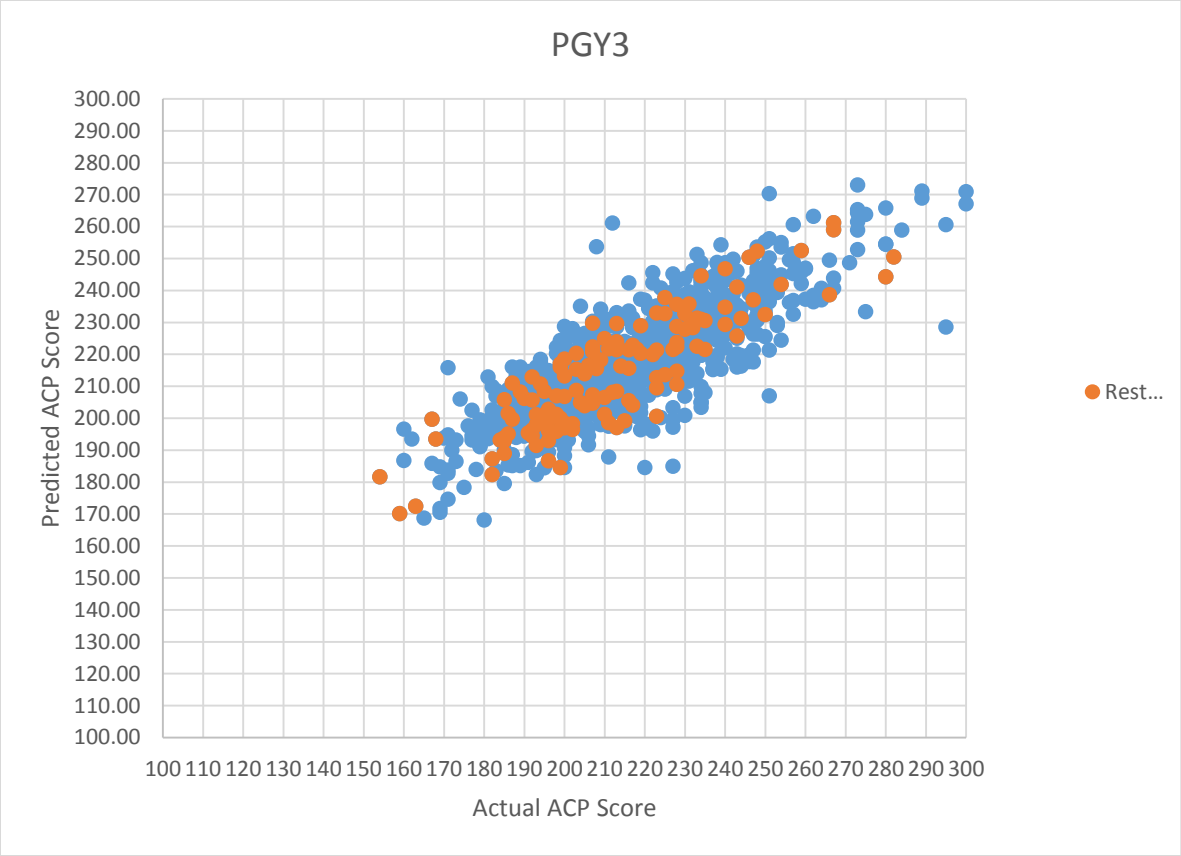
Finally a series of scatterplots of all affected examinees and by each PGY were computed with a separate highlight for examinees with more than 2 restarts. Several observations are noteworthy, as the model fit indices suggest, the prediction equations are better. This can be seen by a narrower oval shape containing data points in PGY3 and PGY2 than PGY 1. In each case, the examinees with the most restarts are represented throughout the distribution and are

not the most substantial outliers or disproportionately represented in under predictions. This suggests again that the administration challenges did not appreciably affect examinee scores.




PGY1
Predicted ACP Score
300.00
290.00
280.00
270.00
260.00
250.00
240.00
230.00
220.00
210.00
200.00
190.00
180.00
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
ACP Score
Restarts>2


PGY2
Predicted ACP Score
300.00
290.00
280.00
270.00
260.00
250.00
240.00
230.00
220.00
210.00
200.00
190.00
180.00
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310
Actual ACP Score
Restart >2


PGY3
Predicted ACP Score
300.00
290.00
280.00
270.00
260.00
250.00
240.00
230.00
220.00
210.00
200.00
190.00
180.00
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
Actual ACP Score
Rest...