Appendix O
Technical Analysis
Using Analysis of Variance and Analysis of Covariance

Introduction

This appendix provides statistical analysis of bar exam data in New York using analysis of variance (ANOVA) and analysis of covariance (ANCOVA) to address the three primary questions from the main report: (1) How do candidate background characteristics compare across bar exam administrations? How do they relate to performance on the bar exam in New York before and after UBE adoption? (2) How do candidates grouped by race/ethnicity and gender perform on the bar exam before and after UBE adoption? (3) How does performance on the New York bar examination compare before and after UBE adoption?

The next section describes the methods, including variables considered, candidate samples and statistical methods used. The section following that provides the results of pre- and post-UBE comparisons and has three major components corresponding to different candidate samples or statistical methods including: (1) all candidates taking the New York bar exam; (2) domestic-educated candidates taking the bar exam for the first time; and (3) how statistically removing differences in candidate background characteristics such as undergraduate grade point average (UGPA), Law School Admission Test (LSAT) scores and law school grade point average (LGPA) affects bar performance, particularly by gender and race/ethnicity. The last two sections of this appendix include a discussion of the results and conclusions.

Methods

Methods are presented in three parts. The first part considers the variables examined, including the outcome variables and predictor variables. The second considers the different samples used. The third part presents the statistical methods employed.

Variables Considered

Outcome Variables

Outcome variables (or dependent variables) are of greatest interest. In the present case, bar exam scores and pass rates are outcome variables of interest along with MBE scores and written scores.

Bar Exam Scores. A description of the New York bar exam before and after UBE adoption was provided in section 2.1 of the main report. As a reminder, bar exam scores prior to UBE adoption in July 2016 were rescaled to the 400-point UBE score scale (see section 2.4) in the analyzed data. The passing score on the 400-point scale is 266.

Pass Rates. Bar pass rate is the percentage of candidates passing the bar exam. For particular groups of candidates (e.g., females), it is the percentage of the group passing the bar exam.

Predictor Variables

Predictor variables (or independent variables) are indicators or characteristics that may be related to or may affect the outcome variables (or dependent variables). For example, a predictor variable of interest in the present case is bar exam administration year, specifically, for the bar administrations before and after the change to UBE in New York in July 2016 to identify the extent to which UBE may have had an impact on bar exam performance overall and by candidates grouped by gender and race/ethnicity. Because candidate performance is related to previous academic performance, and because candidates grouped by gender and race/ethnicity may differ on candidate background characteristics that indicate previous academic performance, like undergraduate grade point average (UGPA), Law School Admission Test (LSAT) scores, and law school grade point average (LGPA), candidate background characteristics are important predictor variables to take into consideration when attempting to determine the potential impact of the UBE on candidate bar exam performance.

UBE. The first UBE was administered in New York in July 2016. Data collected before that administration (July 2015, February 2016) is pre-UBE versus UBE for the data collected on or after July 2016 (July 2016, February 2017 and July 2017). In tables below, UBE data are in bold font. In the analysis below, UBE indicates bar exam administration year.

Bar Exam Administration Month. Bar exam performance in July tends to be better, on average, than performance in February. The July administration has between 70-80% first time takers who are recent law-school graduates whereas the February administration has between 60-70% repeat takers, many of whom did not pass in the previous July administration. In addition, first-time takers in February tend not to perform as well as first time takers in July. Consequently, the analysis below is conducted separately for July and February administrations.

Gender. Gender could be coded as male, female, or omitted. Only candidates not omitting gender were included in the analysis below.

Race/ethnicity. Race/ethnicity had seven categories plus omission and other. Because their numbers were small, results for the Puerto Rican, American Indian/Alaskan Native, and Chicano/Mexican American groups were not separately presented in subsequent analyses. In addition, candidates categorized as other and those omitting race/ethnicity were not included.

Samples[1]

This appendix used slightly modified versions of the two samples of candidates included in the main report above: the *New York State Board of Law Examiners (NYSBLE) sample*, which included all candidates who took the bar exam in New York between February 2015 and July 2017, and the s*chool-based sample*, which included a subset of the NYSBLE sample providing candidate background characteristics (undergraduate grades, Law School Admission Test scores, and law school grades). Section 2.2 of the report above provides additional details regarding these samples. In this appendix, each sample excluded any candidate with missing gender or race/ethnicity because the statistical models included gender and race/ethnicity together and missing information would have complicated the results. As we will see, the slight difference in samples led to some results that differed slightly from the main report. A third sample of data that included national MBE data was included for portions of the analysis below for reference purposes.

Statistical Methods

In statistical analysis, the probability is typically used to determine whether or not a statistical test (e.g., a difference in means, the effect of a predictor on an outcome variable) is statistically meaningful. The probability quantifies the chance that a particular result will occur (e.g., a probability of .1 indicates a 10% chance). The level of probability that is considered statistically meaningful is commonly referred to as alpha, and a value of .05 is often used to determine statistical significance where an observed probability less than or equal to .05 would be determined to be statistically significant (i.e., the probability of a particular result occurring due to chance is at or less than a 5%; this is also referred to as a Type-I error). Often, .05 is a reasonable value to use for alpha, however, when a researcher conducts multiple analyses (or multiple statistical tests) using the same data, the likelihood increases that one or more of the tests will be statistically significant due to chance. For example, if one runs two statistical tests, each at the .05 alpha level, the probability that one or both of the two tests is significant will not be .05, but more like .10 (.05 + .05). This increase from running multiple tests is referred to as alpha (or Type-I error) inflation. There are different ways of keeping alpha inflation in check, but one straightforward way is to use a smaller alpha value for individual statistical tests, such as .01. That way, five tests run at the .01 level will then have an overall probability of .05 that one or more will be statistically significant due to chance. An alpha value of .01 was used in the analysis below.[2] Although they were not considered to be statistically significant, p-values between .01 and .10 were noted in tables provided along with several levels of observed p-

[1] Most of the analysis included in this appendix excluded candidates omitting gender or race/ethnicity, so the data deviates slightly from the main report.

[2] Section 7 of the main report, which conducted statistical analysis using regression models also used an alpha level of .01 to identify statistically meaningful results.

values. A "+" indicated a p-value between .01 and .10, "*" indicated a p-value <.01, "#" indicated a p-value <.001, and "$" indicated a p-value<.0001.

Another issue in statistical analysis is that large sample sizes, like those with New York bar exam data, can complicate the interpretation of statistically significant results. With large samples, say over 1000, even very small differences (e.g., differences in means or proportions) can be identified as statistically significant; a statistically significant result may be small in magnitude and not practically meaningful. To account for this situation, there are two steps that were taken when studying the effects of independent variables (predictor variables) on dependent variables (outcome variables) in the analysis of variance (ANOVA) and analysis of covariance (ANCOVA) statistical analysis techniques used in this appendix.

Part of the first step was to run an overall statistical test of main effects and interactions to determine the statistical significance of predictor variables included in the ANOVA (or ANCOVA) model. A main effect refers to each of the independent variables and whether each had a unique influence on the outcome (dependent variable) being analyzed. For example, if we are interested in the effects of the UBE (bar exam administration year), gender and race/ethnicity on bar exam scores, we would have a statistical test of the independent effects of each of the three variables on bar exam scores, which would be the tests of each of the three main effects (e.g., the main effect test for gender would be whether the mean bar exam score for females differs from that of males). Interactions test whether the impact of a variable depends upon which level it is on another variable and are indicated by an * between the variables being tested for interaction. For example, a significant interaction between gender and UBE for February candidates (designated by UBE*gender in a table of ANOVA results) would occur if males had the same mean score in both February 2016 and 2017, but females had a different mean score in 2017 than they did in 2016. Because of the large number of candidates, statistical tests may be statistically significant, even if the differences between, say males and females is small.

The next portion of the first step was to quantify and review the magnitude of the effects identified as statistically significant. ANOVA provides a breakdown of how much of the differences (called variance) in outcome variable (e.g., bar exam scores) is attributable to each of the effects being tested (e.g., bar exam administration year), so the effect of each can be expressed as a fraction (or percentage) of the total amount of differences (or variance) in the outcome variable, referred to as η^2. If η^2 is less than 1% of the total variance, it indicates that while the differences may be statistically significant, they may be of trivial importance. In this appendix, η^2 represents the effect of a particular predictor after removing (partialling out) the effects of all other predictors in the model.

One issue here is that the statistical test for a variable considers the overall effects for an independent variable. So, for example, when studying UBE (bar exam administration year) in

July, ANOVA not only tests results from the post-UBE years of 2016 and 2017 with the pre-UBE year of 2015, but it also tests for differences between the two UBE years of 2016 and 2017 as well as combinations. A statistically significant finding for the UBE variable does not necessarily reflect statistical significance of the UBE per se. Likewise, η^2 captures the overall magnitude of the effect of UBE (bar exam administration year), not specific year-to-year comparisons. Therefore, for independent variables with more than two levels, a statistically significant test of an overall effect can be followed up by a comparison of the differences in the means for particular groupings of interest (for example, 2016 vs 2015 and 2017 vs 2015), which was the second step conducted in the analysis.

The second step[3] involved reviewing the magnitude of effects between levels of independent variables by expressing them as differences in means (or proportions) in terms of what are called effect sizes (ES). An effect size is defined by the type of statistics being compared across groups. For a difference in means, it is usually defined as the difference between means divided by an appropriate standard deviation (SD). Typically, the default is to use the weighted SD of the two groups involved in the comparison. For purposes of this appendix, when an ES is comparing a subgroup, say males versus females within a given administration, the SD for the total group (also called root mean square error or RMSE) was used in computing the ES. Cohen (1988) describes an ES based upon the absolute value of the difference in means at three levels: small=0.2-0.5, medium=0.5-0.8 and large= >0.8. This approach to describing ES for differences in means was used throughout this appendix. The analogous ES for a difference in proportions (p) is bit more complex because it involves the difference in arcsine transformations of the square root of the proportions (-2 x arcsine square root (p)). The same criteria for small, medium and large differences were used for the proportion ESs, 0.2-0.5, 0.5-0.8, >0.8, respectively. Because the ES values were generally below 0.2, they were not reported in tables to avoid over-crowding, but were referenced in the text or in footnotes if they exceeded 0.2. The η^2 value described earlier is also a type of effect size but different criteria are typically used for interpretation because η^2 takes values between 0 and 1. Cohen's criteria for η^2 would set a small effect between .01 and .06 (1% and 6%), a medium effect between .06 and .14 (6% and 14%) and a large effect greater than 0.14 (>14%).

[3] Part of this second step could have included conducting statistical tests comparing one level of an independent variable to another level (post-hoc comparisons) before obtaining effect sizes. We were primarily interested in the magnitudes of effects, so we didn't conduct follow-up statistical tests.

Results

NYSBLE Sample Overall

Table O.1 shows the February and July results of the ANOVA testing the effects of UBE (bar exam administration year), gender and race/ethnicity on the bar exam, MBE, and written score in the NYSBLE sample. Separate analysis was performed on February and July scores since they are known to have different candidate characteristics.

The main finding was that UBE in its various combinations with gender and race/ethnicity accounted for less than 1% of the variance in bar exam scores, MBE scores, or written scores. Of all predictors, race/ethnicity accounted for the most variance, with 5.7% (high end of a small effect size) in February and 11.7% (high end of a medium effect size) in July. In July, the interaction of UBE with race/ethnicity was also statistically significant but it accounted for only 0.12% of the variance in bar exam scores.

Table O.2 shows the means and SDs of bar exam scores plus pass rates for each of the bar exam administrations. The values in bold font were obtained after the UBE was adopted. The most noticeable pattern was that scores on both the MBE and written components as well as bar exam scores and pass rates increased across years. Average bar exam score increased by 8.17 points between July 2015 and July 2017, a difference producing an effect size of approximately 0.24, a value on the low end of being a small difference. Pass rates increased from 41.4% in February 2016 to 44.0% in February 2017. The July pass rate rose from 61.8% in 2015 to 64.2% in 2016, the first UBE administration, and 69.2% in 2017.

Table O.3 shows New York versus the national performance on the MBE between February 2016 and July 2017. In February, the New York mean MBE scores rose 1.2 points between 2016 and 2017 while the national mean MBE score declined by 0.89 points. Since the New York mean was 5.59 points below the national mean in 2016, the difference was reduced to 3.51 points in 2017, the first UBE February administration. In July, compared with 2015, the 2016 MBE mean rose 2.05 points in New York while nationally it rose 0.42 points. In 2017, the MBE mean rose another 2.20 points in New York while nationally the mean MBE score rose 1.35 points. Since New York mean MBE scores were lower in July across these years, the difference that began at 2.81 points in July 2015 was reduced to 0.33 points in July 2017. These results show that New York candidates overall have performed as well or better since the UBE was implemented.

Table O.1
Analysis of Variance (ANOVA) of Bar Exam Scores by UBE year (UBE or U), Gender (G) and Race/Ethnicity (E) New York State Board of Law Examiners Sample

February		Bar Exam				MBE				Written			
Effect	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE	1	5885.57	7.59	*	.0010	2758.27	11.79	#	.0015	636.00	2.65		.0003
G	1	18.95	0.02		-	8138.44	34.80	$	.0045	7247.07	30.22	$	.0039
E	3	341631.67	146.94	$	.0570	75996.81	108.33	$	.0421	96289.53	133.83	$	.0521
U*G	3	122.99	0.16		-	45.65	0.20		-	198.21	0.83		.0001
U*E	3	774.22	0.33		.0001	1271.17	1.81		.0007	79.02	0.11		-
G*E	3	5170.21	2.22		.0009	1530.79	2.18	+	.0008	1309.18	1.82		.0007
U*G*E	3	1392.96	0.60		.0002	462.71	0.66		.0003	331.84	0.46		.0002
Total	**7287**	**5997785.331**			**.0604**	**1803564.938**			**.0571**	**1847021.681**			**.0558**

July		Bar Exam				MBE				Written			
Effect	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE	2	169030.18	81.04	$	.0051	54498.14	92.34	$	.0059	33867.06	56.15	$	.0035
G	1	10536.14	10.10	*	.0003	42610.35	144.39	$	.0046	8978.28	29.77	$	.0009
E	3	3907063.14	1248.80	$	.1172	848760.56	958.71	$	.0912	1133752.21	1253.05	$	.1187
U*G	2	2482.71	1.19		.0001	648.08	1.10		.0001	947.11	1.57		.0001
U*E	6	39601.05	6.33	$	.0012	14724.72	8.32	$	.0016	9712.93	5.37	$	.0010
G*E	3	36816.15	11.77	$	.0011	9317.14	10.52	$	.0010	9324.53	10.31	$	.0010
U*G*E	6	7212.23	1.15		.0002	2739.81	1.55		.0003	1903.80	1.05		.0002
Total	**27625**	**33322530.14**			**.1361**	**9310572.91**			**.1251**	**9554303.86**			**.1287**

+ p<.10, * p<.01, # p<.001, $ p<.0001

SS: Sums of Squares is the type III version in SAS which is obtained for the effect beyond that including all other effects.

F: The family of statistics used to obtain the statistical significance of the particular effect. For example, the UBE in July had an F value of 81.04. The $ that follows it indicates that the likelihood of the observed difference in mean bar exam scores between July 2015 and 2017 being as large as was obtained by chance was less than .0001, or 1 in 10,000.

η^2: The partial η^2 obtained by dividing the SS value for the effect by the total SS in the last line. For example, the partial η^2 for bar exam scores for the UBE main effect in July was obtained by dividing 169030.18 by 33322530.14=.0051. The interpretation is that after accounting for gender, ethnicity, and their interactions, the UBE main effect accounts for .0051 or 0.51% (one half of one percent) of the variance in bar exam scores.

Table O.2
Pass Rates and Means and Standard Deviations of Bar Exam, MBE, and Written Scores
New York State Board of Law Examiners Sample*

Admin		February			July				
Year		2016	**2017**	DF17**	2015	**2016**	**2017**	DJ16	DJ17**
N		3,627	**3,661**		9,531	**9,249**	**8,846**		
Bar Exam	Mean	258.61	**260.81**	2.20	274.23	**277.79**	**282.40**	3.72	8.17
	SD	28.70	**28.64**		33.40	**34.95**	**35.40**		
MBE	Mean	129.36	**130.55**	1.19	137.11	**139.16**	**141.36**	2.05	4.25
	SD	15.81	**15.63**		17.92	**18.35**	**18.58**		
Written	Mean	129.24	**130.21**	0.97	137.09	**138.58**	**140.98**	1.63	3.89
	SD	15.87	**15.96**		17.97	**18.76**	**18.88**		
Pass Rate		41.4%	**44.0%**	2.6%	61.8%	**64.2%**	**69.2%**	2.4%	7.4%

*Values in bold are UBE. DF17 is the difference between the February 2017 UBE and February pre-UBE 2016 score. DJ16 is the difference between the July 2016 UBE score and the July 2015 pre-UBE score. DJ17 is the difference between the July 2017 UBE score and the July 2015 pre-UBE score. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**Effect sizes for the observed differences in mean bar exam component scores never exceeded 0.10 in February and 0.24 in July. A value of 0.24 would be considered on the low end of a small effect size via Cohen's criteria.

Table O.3
Means and Standard Deviations for the MBE
New York State Board of Law Examiners (NYSBLE) Sample and National Sample*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
NYSBLE	Mean	129.36	**130.55**	1.19	137.11	**139.16**	**141.36**	2.05	4.25
	SD	15.81	**15.63**		17.95	**18.35**	**18.58**		
	N	3,627	**3,661**		9,531	**9,249**	**8,846**		
National	Mean	134.95	**134.06**	-0.89	139.92	**140.34**	**141.69**	0.42	1.77
	SD	15.01	**14.70**		16.09	**16.70**	**16.78**		
	N	23,325	**22,269**		48,384	**46,518**	**46,627**		
Mean Difference		5.59	**3.51**		2.81	**1.18**	**0.33**		

*Values in bold are UBE. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.

NYSBLE Sample by Gender

Gender accounted for less than one percent of the total variance in bar exam scores (Table O.1). While the overall performance and pass rates increased across years, gender showed the unusual pattern of being not statistically significant for bar exam scores, but being statistically significant for the MBE and written scores individually. Table O.4 shows means and pass rates for females versus males as a function of bar exam administration (interaction of gender and UBE). The most notable findings were that (a) males scored higher on the MBE, on average, compared to females and (b) average written scores in February were higher for females than males in February and were higher for males than females in July.[4] In addition, the gender difference decreased by over a point with the UBE in February 2017 and by over a point and a half on bar exam scores. The results in July were mixed with over a point increase in the gender difference, with males scoring higher than females, for bar exam scores in 2016 but a net decrease in the difference by over half a point in 2017. The larger difference between females and males in 2016 was primarily due to a three quarter of a point increase in the difference in average written score between males and females. Patterns of pass rates were generally consistent with patterns of bar exam scores, with males passing from 1-4% higher in February and 4-8% higher in July. There was no indication that the UBE exacerbated average bar exam score differences or pass rate differences between females and males. In fact, the results in July 2017 showed the smallest differences across the three-year period.

[4] In the main report, females tended to do better than males on the written component, so this result is somewhat anomalous given the results in the main report and other research that has been conducted studying gender differences on written exams. The difference between the samples used in this appendix and those used in the main report was that in this appendix, any candidates without a valid gender or race/ethnicity were excluded from the samples used in the analysis here. This was done because complete information on gender and race/ethnicity were needed for inclusion in the ANOVA and ANCOVA analysis. Since the data included in the main report is more complete, we would point the reader to the complete NYSBLE results as better representing the performance of all candidates in New York (e.g., section 4.2).

Table O.4
Means of Bar Exam, MBE, Written, and Pass Rates
New York State Board of Law Examiners Sample by Gender*

Admin	February			July				
Year	2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
Bar Exam								
Female	257.43	**260.39**	2.96	271.50	**274.50**	**280.13**	3.00	8.63
Male	260.04	**261.38**	1.34	277.18	**281.42**	**285.04**	4.24	7.86
Range**	2.61	**0.99**		5.68	**6.92**	**4.91**		
MBE								
Female	127.54	**129.31**	1.77	134.34	**136.40**	**139.13**	2.06	4.79
Male	131.55	**132.21**	0.66	140.11	**142.21**	**143.96**	2.10	3.85
Range**	4.01	**2.90**		5.77	**5.81**	**4.83**		
Written								
Female	129.69	**131.03**	1.34	136.92	**138.05**	**140.95**	1.13	4.03
Male	128.69	**129.10**	0.41	137.28	**139.17**	**141.03**	1.89	3.75
Range**	1.00	**1.93**		0.36	**1.12**	**0.08**		
Pass Rates								
Female	39.5%	**43.5%**	4.0%	59.0%	**60.6%**	**67.1%**	1.6%	8.1%
Male	43.6%	**44.6%**	1.0%	64.9%	**68.3%**	**71.6%**	3.4%	6.7%
Range**	4.1%	**1.1%**		5.9%	**7.7%**	**4.5%**		

*Values in bold are UBE. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**Range is the difference between the largest and smallest mean. Since there are only two values for gender, it is the difference between males and females.

NYSBLE Sample by Race/Ethnicity

Race/ethnicity accounted for the single largest percentage of variance among UBE, gender, and race/ethnicity predictor variables, with 5.7% accounted for in February and 11.7% accounted for in July. In the February administration there was no indication that race/ethnicity effects differed by year, meaning that how a particular racial/ethnic group performed compared to the other groups did not differ depending upon the particular year, say pre-UBE versus UBE. If the UBE did change how any particular racial/ethnic group performed relative to the other groups, say Black/African American versus Caucasian/White, it would be expected that the interaction between UBE and race/ethnicity would be statistically significant (the U*E effect in Table O.1) and account for a substantial amount of variance in bar exam scores. Whereas February showed no such interaction, July results had a statistically significant interaction between UBE and race/ethnicity, although it accounted for only one tenth of one percent of the variance in bar exam scores. Table O.5 shows the average bar exam performance and pass rates for different racial/ethnic groups as a function of year of administration. Figure O.1 shows the same information for the bar exam score in graph form. Across all administrations, the Caucasian/White group obtained the highest mean bar exam score while the Black/African

American group obtained the lowest. In February, the difference in the means between these two groups declined by about a point after UBE adoption. In July, there was a 2.36-point increase in the difference between the two groups in 2016 that declined slightly to only a 1.38-point increase in 2017. The increase in the range after UBE adoption was primarily due to the written score since the range on the MBE decreased by a fraction of a point in both 2016 and 2017 compared to pre-UBE 2015. Pass rates were generally consistent with patterns in the bar exam scores, with the exception of the Black/African American group between July 2015 and July 2016, where average bar exam scores increased but pass rates decreased (see Section 4.2, specifically Figure 4.2.29 for an explanation of this contradictory result, which has to do with the distribution of scores for the Black/African American group).

Table O.5
Pass Rates and Means of Bar Exam, MBE, and Written Scores
New York State Board of Law Examiners Sample by Race/Ethnicity*

Admin	February			July				
Year	2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17**
Bar Exam Score								
Asian/Pacific Islander	252.57	**255.42**	2.85	259.49	**265.21**	**272.91**	5.72	13.42
Black/African American	250.01	**252.12**	2.11	257.59	**258.40**	**263.54**	0.81	5.95
Caucasian/White	266.60	**267.75**	1.15	285.39	**288.56**	**292.82**	3.17	7.43
Hispanic/Latino	255.10	**258.06**	2.96	263.22	**267.79**	**270.67**	4.57	7.45
Range (max-min)	16.59	**15.63**		27.80	**30.16**	**29.28**		
MBE								
Asian/Pacific Islander	126.17	**128.11**	1.94	129.89	**133.58**	**137.26**	3.69	7.37
Black/African American	124.84	**126.55**	1.71	128.07	**129.97**	**132.05**	1.90	3.98
Caucasian/White	133.58	**133.66**	0.08	142.81	**144.08**	**146.18**	1.27	3.37
Hispanic/Latino	127.43	**129.53**	2.10	130.90	**134.40**	**135.66**	3.50	4.76
Range(max-min)	8.74	**7.11**		14.74	**14.11**	**14.13**		
Written								
Asian/Pacific Islander	126.51	**127.25**	0.74	129.37	**131.57**	**135.59**	2.20	6.22
Black/African American	124.99	**125.51**	0.52	129.36	**128.39**	**131.44**	-0.97	2.08
Caucasian/White	133.06	**134.03**	0.97	142.70	**144.43**	**146.60**	1.73	3.90
Hispanic/Latino	127.20	**128.46**	1.26	132.04	**133.36**	**134.96**	1.32	2.92
Range(max-min)	8.07	**8.52**		13.34	**16.04**	**15.16**		
Pass Rate								
Asian/Pacific Islander	33.9%	**37.5%**	3.6%	44.6%	**51.0%**	**60.4%**	6.4%	15.8%
Black/African American	27.3%	**30.4%**	3.1%	41.0%	**39.5%**	**48.7%**	-1.5%	7.7%
Caucasian/White	52.5%	**53.4%**	0.9%	75.2%	**76.3%**	**79.6%**	1.1%	4.4%
Hispanic/Latino	36.5%	**41.0%**	4.5%	48.1%	**53.8%**	**57.0%**	5.7%	8.9%
Range(max-min)	25.2%	**23.0%**		34.2%	**36.8%**	**30.9%**		

*Values in bold are UBE. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**The maximum effect size between any difference in change of scores between 2015 and 2017 was at most 0.30, a value in the low to middle range of the small region.

The interaction between UBE and race/ethnicity noted above is reflected in Figure O.1 by lines that are not parallel to one another. Lines that cross are particularly indicative of an interaction. Interaction in July occurred at two points. Between 2015 and 2016, the Black/African American change in means lagged behind the other three groups, rising only 0.81 points compared to 3 to 6 points for the other three groups. In 2017, however, the 5.14 rise for the Black/African American group was quite large, although the 7.7-point rise by the Asian/Pacific Islander group out-paced that of the other three groups. Between July 2016 and 2017, the Asian/Pacific Islander group average bar exam scores increased at a faster rate than the other groups, in particular the Hispanic/Latino group, resulting in the lines tracing their change in scores crossing.

Figure O.1
Mean Bar Exam Scores
New York State Board of Law Examiners by Race/Ethnicity

A, a = Asian/Pacific Islander,
B, b = Black/African American,
H, h = Hispanic/Latino,
W, w = Caucasian/White.

Summary for NYSBLE Sample

Overall, New York bar exam scores increased across the three years in both February and July, outpacing the increase at the national level over both periods. Since the pre-UBE New York MBE scores were 5.59 points lower than the national mean in February and 2.81 points lower in July, their faster rise in the following bar administrations brought them closer to the national mean in 2017.

Differences in average bar exam scores between males and females, decreased across the studied years. In February, the difference declined by over 1.5 points between 2016 and 2017 and in July it decreased by slightly over 0.75 points between 2015 and 2017. Males scored higher than females, on average, on the MBE in February and July. In addition, males scored higher than females, on average, on the written component in July in the sample studied in this appendix, although the complete NYSBLE sample in the main report showed a different pattern for the written component, with females tending to score higher, on average, than males in July. Females scored higher, on average, than males on the written component in February.

Scores for racial/ethnic groups increased, on average, from pre-UBE to the end of the study period in 2017, but their rate of increase and relative values were substantially different in some cases. The Caucasian/White group had the highest average bar exam scores by over 9 points in February and 20 points in July. The Asian/Pacific Islander group's average score increased more rapidly than other groups by 2017. The Black/African American group had the lowest mean scores at each bar exam administration and showed a modest gain in July 2016 compared to other groups, but showed larger gains in 2017. The Asian/Pacific Islander group showed the largest gains in July, especially between 2016 and 2017.

It is important to note that all of these differences in total account for less than 15% of the variance in bar exam scores and the effect size for any of the particular differences between UBE and pre-UBE levels never exceeded 0.30, a value that would be considered a small effect. Further, the candidates were a mix of first-takers and repeaters as well as domestic-educated and foreign-educated individuals. In addition, the analysis to this point has not accounted for potential shifts in candidate background characteristics across bar exams before and after UBE adoption, which may explain the increasing bar exam performance observed across years. The influence of candidate background characteristics on bar exam performance is studied in subsequent sections. The next section focuses domestic-educated candidates taking the New York bar exam for the first time.

Domestic-educated First-time Takers in the NYSBLE Sample (US-1)

Domestic-educated first-time takers in the NYSBLE sample (US-1 for purposes of this appendix), which is the subset of New York candidates who graduated from U.S.-based law schools and who took the New York bar exam for the first time, are the largest group of candidates in July and were studied here.[5]

ANOVA results in Table O.6, showed that the combined percentage of variance accounted for by the UBE, race/ethnicity and gender totaled 8.2% of the variance in February bar exam scores, an increase of approximately 2% over that obtained when all New York candidates were included (see Table O.1). The 2% increase in variance accounted for in February was counterbalanced in July where the percentage of variance accounted for decreased by almost 4%, from 13.6% to 9.7%. The focus on domestic-educated first-time takers reduced the February sample by almost 80% from 7,288 to 1,496. The effect on the July sample was less pronounced, decreasing it by about 40% from 27,626 to 16,670. Of the three variables, race/ethnicity had the strongest relationship with bar exam scores (and MBE and written scores) accounting for 7.1% of the variance in February and July. UBE was statistically significant for the US-1 sample for bar exam scores (and MBE and written scores) in July and for written scores in February. UBE accounted for, at most, 0.7% of the variance in bar exam scores. Gender was statistically significant in July for MBE and written scores but accounted for less than one half of one percent of the variance in either. UBE had a statistically significant interaction with race/ethnicity for the written score in July, accounting for 0.15% of the variance in bar exam scores. Despite statistically significant results in some instances, the variance accounted for by UBE and gender were quite small.

[5] And had valid gender and race/ethnicity information.

Table O.6
Analysis of Variance (ANOVA) of Bar Exam Component Scores
by UBE year (UBE or U), Gender (G) and Race/Ethnicity (E)
New York State Board of Law Examiners Sample
Domestic-Educated First-Time Takers (US-1)

February		Bar Exam Score				MBE				Written			
Effect	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE	1	1834.17	2.27		.0014	11.39	0.05		-	1742.48	6.70	*	.0042
G	1	638.26	0.79		.0005	90.19	0.39		.0002	1003.66	3.86	+	.0024
E	3	92610.83	38.18	$	.0711	26528.33	38.29	$	.0708	21461.66	27.52	$	.0520
U*G	1	23.69	0.03		-	5.02	0.02		-	17.29	0.07		-
U*E	3	1046.62	0.43		.0008	330.27	0.48		.0009	315.78	0.40		.0008
G*E	3	804.64	0.33		.0006	963.52	1.39		.0026	309.98	0.40		.0008
U*G*E	3	2304.31	0.95		.0018	730.62	1.05		.0019	663.36	0.85		.0016
Total	**1495**	**1303450.135**			**.0819**	**374689.8325**			**.0878**	**413030.6861**			**.0687**

July		Bar Exam Score				MBE				Written			
Effect	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE	2	101723.16	68.62	$	.0074	23146.29	53.67	$	.0059	25822.71	56.63	$	.0062
G	1	2405.74	3.25	+	.0002	17227.43	79.90	$	.0044	5389.75	23.64	$	.0013
E	3	972808.04	437.50	$	.0712	224959.04	347.77	$	.0569	266538.65	389.66	$	.0643
U*G	2	4149.47	2.80	+	.0003	1126.85	2.61	+	.0003	1687.79	3.70	+	.0004
U*E	6	11891.63	2.67	+	.0009	1226.19	0.95		.0003	6407.76	4.68	$	.0015
G*E	3	13148.97	5.91	#	.0010	4285.93	6.63	$	.0011	2260.67	3.30	+	.0005
U*G*E	6	3710.78	0.83		.0003	2249.83	1.74		.0006	763.93	0.56		.0002
Total	**16669**	**13655016.29**			**.0965**	**3956560.088**			**.0928**	**4146052.937**			**.0846**

+ p<.10, * p<.01, # p<.001, $ p<.0001

Table O.7 shows the mean scores and pass rates across years for both February and July. As was the case with the New York sample studied above, mean scores tended to increase across years in February and July. The exception for the US-1 sample was for the MBE in February, where the mean score declined by 0.17 in 2017 whereas all other scores rose. Unlike the entire New York sample where MBE scores were below the national mean in February and July, Table O.8 shows that the means for the US-1 group were 5.4 to 6.1 points higher in February and 4.5 to 7 points higher than the national mean in July and became progressively more so across all three years. Even in February 2017 when the MBE mean decreased 0.17 points for the US-1 group, the decline was less than the 0.89-point drop nationally. The pass rates were consistent with bar exam score increases across years.

Table O.7
Pass Rates and Means and Standard Deviations of Bar Exam, MBE, and Written Scores
New York State Board of Law Examiners Sample
Domestic-Educated First-Time Takers (US-1)*

Admin		February			July				
Year		2016	**2017**	DF17**	2015	**2016**	**2017**	DJ16	DJ17**
N		698	**798**		5,901	**5,627**	**5,142**		
Bar Exam	Mean	277.53	**280.77**	3.24	288.52	**293.13**	**298.01**	4.61	9.49
	SD	28.66	**28.72**		27.00	**28.50**	**29.45**		
MBE	Mean	140.33	**140.16**	-0.17	144.46	**146.41**	**148.65**	1.95	4.19
	SD	15.91	**15.58**		14.68	**15.20**	**15.57**		
Written	Mean	137.64	**140.55**	2.91	144.29	**146.71**	**149.30**	2.42	5.01
	SD	17.27	**15.91**		15.10	**15.75**	**16.09**		
Pass Rate		67.6%	**71.2%**	3.6%	80.1%	**83.0%**	**86.5%**	2.9%	6.4%

*Values in bold are UBE. DF17 is the difference between the February 2017 UBE and February pre-UBE 2016 score. DJ16 is the difference between the July 2016 UBE score and the July 2015 pre-UBE score. DJ17 is the difference between the July 2017 UBE score and the July 2015 pre-UBE score. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**Effect sizes for the observed differences in mean bar exam component scores never exceeded 0.19 in February and 0.34 in July. A value of 0.34 would be considered near the mid-point of a small effect size via Cohen's criteria.

Table O.8
Means and Standard Deviations for the MBE
New York State Board of Law Examiners Sample
Domestic-Educated First Time Takers (US-1) and National Sample*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
US-1	Mean	140.33	**140.16**	-0.17	144.46	**146.41**	**148.65**	1.95	4.19
	SD	15.91	**15.58**		14.68	**15.20**	**15.57**		
	N	698	**798**		5,901	**5,627**	**5,142**		
National	Mean	134.95	**134.06**	-0.89	139.92	**140.34**	**141.69**	0.42	1.77
	SD	15.01	**14.70**		16.09	**16.70**	**16.78**		
	N	23,325	**22,269**		48,384	**46,518**	**46,627**		
Mean Difference		5.38	**6.10**		4.54	**6.07**	**6.96**		

*Values in bold are UBE. DF17 is the difference between the February 2017 UBE and February pre-UBE 2016 score. DJ16 is the difference between the July 2016 UBE score and the July 2015 pre-UBE score. DJ17 is the difference between the July 2017 UBE score and the July 2015 pre-UBE score. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.

Domestic-educated First-time Takers in the NYSBLE Sample (US-1) by Gender

For bar exam scores, gender was not statistically significant in either February or July and accounted for less than one percent of the total variance in bar exam scores for domestic-educated first-time takers. While the overall performance and pass rates were rising across years, gender again showed the unusual pattern of not being statistically significant for bar exam scores, but being statistically significant for the MBE and written scores, but only in July (this is because higher performance of males on the MBE tended to be cancelled out by higher performance of females on the written component for domestic-educated first-time takers).

Table O.9 shows means and pass rates for females versus males as a function of exam administration (which reflects the interaction of gender and UBE) for domestic-educated first-time takers. In 2017, the difference between females and males in bar exam scores decreased by 0.02 points in February from pre-UBE levels (decreasing from 2.23 to 2.21) and by 0.5 points in July (decreasing from 3.22 in 2015 to 2.72 in 2017). For the MBE, the difference decreased by 0.77 points in February and 1.09 points in July. For written scores, females tended to score higher than males, except in 2016, when males scored higher. Otherwise, in absolute differences, the written score gender difference decreased during the UBE years from the pre-UBE year. For pass rates, the gender difference generally mirrored bar exam scores except for February 2017 where bar exam scores increased by a similar amount for males and females, but the male pass rate increased by 5.6% and the females pass rate increased by 1.8%. The resulting difference in February 2017 was over twice what it was in 2016.

Table O.9
Pass Rates and Means of Bar Exam, MBE, and Written Scores
New York State Board of Law Examiners Sample
Domestic-Educated First-Time Takers (US-1) by Gender*

Admin	February			July					
Year	2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17	**
Bar Exam									
Females	276.41	**279.70**	3.29	286.86	**290.29**	**296.68**	3.43	9.82	**
Males	278.64	**281.91**	3.27	290.08	**296.06**	**299.40**	5.98	9.32	
Range	2.23	**2.21**		3.22	**5.77**	**2.72**			
MBE									
Females	138.34	**138.62**	0.28	141.98	**143.86**	**146.83**	1.88	4.85	
Males	142.28	**141.79**	-0.49	146.81	**148.96**	**150.57**	2.15	3.76	
Range	3.94	**3.17**		4.83	**5.10**	**3.74**			
Written									
Females	138.22	**141.00**	2.78	144.85	**146.38**	**149.80**	1.53	4.95	
Males	137.08	**140.06**	2.98	143.76	**147.05**	**148.79**	3.29	5.03	
Range	1.14	**0.94**		1.09	**0.67**	**1.01**			
Pass Rate									
Females	66.2%	**68.0%**	1.8%	78.8%	**80.4%**	**85.5%**	1.6%	6.7%	
Males	69.0%	**74.6%**	5.6%	81.2%	**85.7%**	**87.5%**	4.5%	6.3%	
Range	2.8%	**6.6%**		2.4%	**5.3%**	**2.0%**			

*Values in bold are UBE, DF17 is the difference between the February 2017 UBE and February pre-UBE 2016 score. DJ16 is the difference between the July 2016 UBE score and the July 2015 pre-UBE score. DJ17 is the difference between the July 2017 UBE score and the July 2015 pre-UBE score. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**No effect size exceeded 0.35, a value that would be in the middle of the small effect size range.

Domestic-educated First-time Takers in the NYSBLE Sample (US-1) by Race/Ethnicity

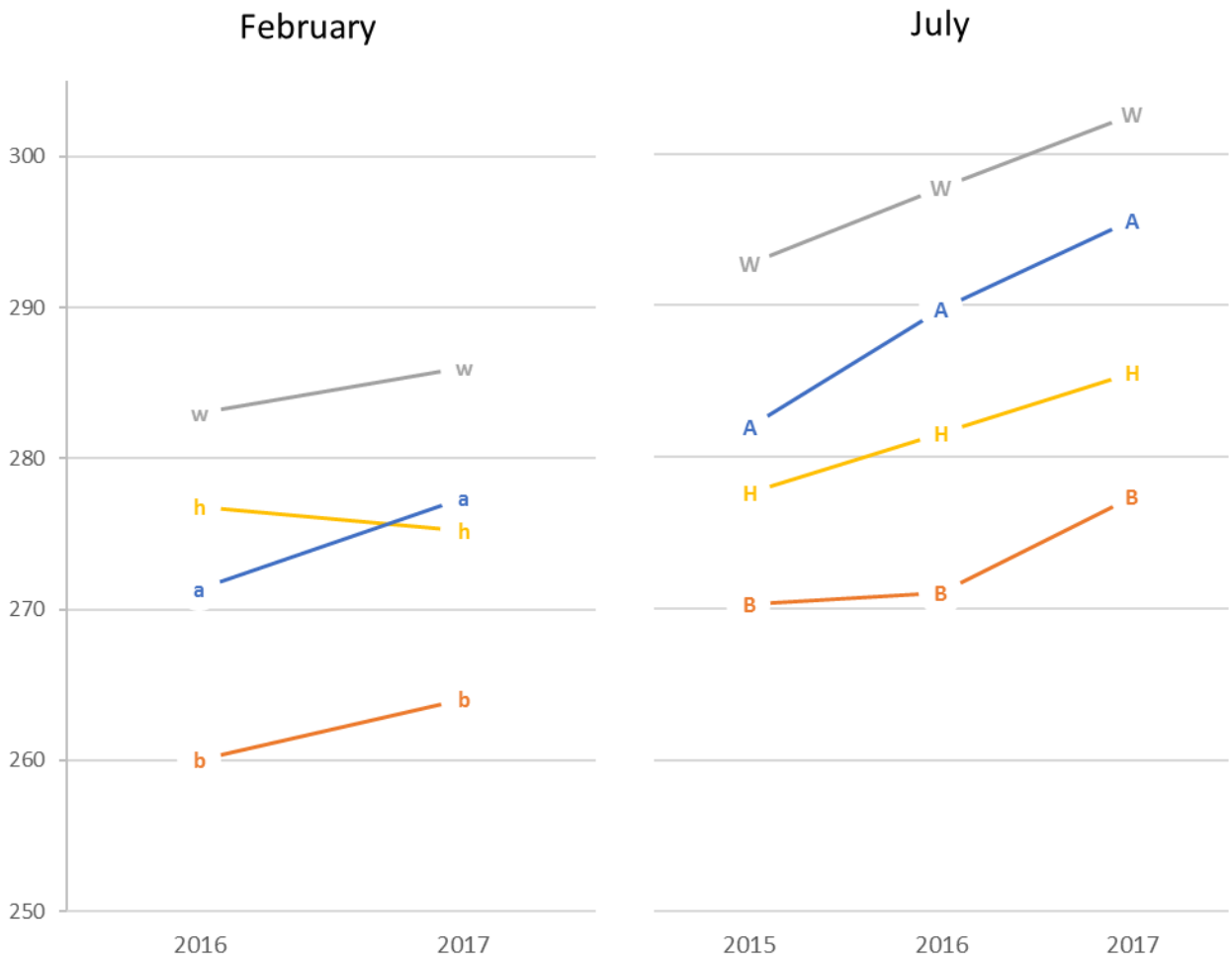
Race/ethnicity accounted for the single largest percentage of variance in bar exam scores among UBE, gender, and race/ethnicity predictor variables, with 7.1% accounted for in both February and July (see Table O.6). In the February administration there was no indication that race/ethnicity effects differed by year (i.e., there was not a statistically significant U*E effect in Table O.6). July results, however, had a statistically significant interaction by year for written scores, although it accounted for less than one fifth of one percent of the variance in written scores. Table O.10 shows the mean scores on the bar exam components for different racial/ethnic groups as a function of bar exam administration year. Figure O.2 graphically shows the same results for the bar exam score.

Table O.10
Pass Rates and Means of Bar Exam, MBE, and Written Scores
New York State Board of Law Examiners Sample
Domestic-Educated First-Time Takers (US-1) by Race/Ethnicity*

Admin		February			July					
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17	
Bar Exam Score										
Caucasian/White	Mean	283.04	**285.99**	2.95	292.77	**297.78**	**302.62**	5.01	9.85	
Asian/Pacific Islander	Mean	271.35	**277.40**	6.05	281.95	**289.74**	**295.62**	7.79	13.67	**
Black/African American	Mean	260.05	**264.06**	4.01	270.26	**271.02**	**277.41**	0.76	7.15	
Hispanic/Latino	Mean	276.81	**275.21**	-1.60	277.66	**281.58**	**285.58**	3.92	7.92	
	Range	22.99	**21.93**		22.51	**26.76**	**25.21**			
MBE										
Caucasian/White	Mean	143.45	**142.98**	-0.47	146.75	**148.66**	**151.00**	1.91	4.25	
Asian/Pacific Islander	Mean	136.49	**137.44**	0.95	141.06	**144.29**	**146.85**	3.23	5.79	
Black/African American	Mean	130.42	**131.57**	1.15	134.42	**136.23**	**139.21**	1.81	4.79	
Hispanic/Latino	Mean	140.65	**137.81**	-2.84	138.68	**140.97**	**142.34**	2.29	3.66	
	Range	13.03	**11.41**		12.33	**12.43**	**11.79**			
Written										
Caucasian/White	Mean	140.18	**142.92**	2.74	146.34	**149.07**	**151.57**	2.73	5.23	
Asian/Pacific Islander	Mean	135.28	**139.93**	4.65	140.95	**145.39**	**148.71**	4.44	7.76	**
Black/African American	Mean	129.48	**132.46**	2.98	135.80	**134.74**	**138.14**	-1.06	2.34	
Hispanic/Latino	Mean	136.45	**137.31**	0.86	139.07	**140.57**	**143.18**	1.50	4.11	
	Range	10.70	**10.46**		10.54	**14.33**	**13.43**			
Pass Rate										
Caucasian/White		75.8%	**77.9%**	2.1	85.1%	**87.5%**	**90.1%**	2.4	5.0	
Asian/Pacific Islander		58.6%	**63.9%**	5.3	73.0%	**81.5%**	**85.0%**	8.5	12.0	
Black/African American		43.8%	**51.3%**	7.5	58.6%	**57.8%**	**68.5%**	-0.8	9.9	
Hispanic/Latino		61.4%	**66.1%**	4.7	65.6%	**73.0%**	**77.6%**	7.4	12.0	
	Range	32.0	**26.6**		26.5	**29.7**	**21.6**			

*Values in bold are UBE. DF17 is the difference between the February 2017 UBE and February pre-UBE 2016 score. DJ16 is the difference between the July 2016 UBE score and the July 2015 pre-UBE score. DJ17 is the difference between the July 2017 UBE score and the July 2015 pre-UBE score. The NYSBLE sample here only includes candidates with valid gender and race/ethnicity, which is different from the NYSBLE sample in the main report.
**Effect size reached 0.50 level, a value in the medium effect size category by Cohen's criteria.

Figure O.2
Mean Bar Exam Scores
New York State Board of Law Examiners
Domestic-Educated First-Time Takers (US-1) by Race/Ethnicity



A, a = Asian/Pacific Islander,
B, b = Black/African American,
H, h = Hispanic/Latino,
W, w = Caucasian/White.

Across all administrations, the Caucasian/White group had the highest mean bar exam score and the Black/African American group had the lowest mean bar exam scores. In February, the difference in the means between these two groups declined by about a point after UBE adoption, the same as for the overall New York group. In July, the difference in mean bar exam scores between the Caucasian/White group and Black/African American group increased in 2016 before decreasing in 2017. The statistically significant interaction of race/ethnicity with UBE for the written score in July appears to be mostly due to the divergent performance of the different groups in 2016. The Black/African American group's written scores decreased by more than a point (declining from 135.80 to 134.74) while the Asian/Pacific Islander group's written scores increased by over four points (from 140.95 to 145.39). Between 2016 and 2017, scores for all groups increased, with increases between 2.50 for the Caucasian/White group and 3.40 for the Black/African American group. The Black/American group, thereby, made up a little for the smaller increase in bar exam performance observed in 2016 compared to other groups. The percentage passing increased for all groups after UBE adoption in year 2017. The 0.8% decrease in pass rate for the Black/African American group in July 2016 was offset by a 10.7% increase in 2017, an increase that was over double that of any other group (the Hispanic/Latin American group had the next largest percent gain between 2016 and 2017 of 4.6%)

Summary for Domestic-Educated First-Time Taker Candidates in the NYSBLE Sample (US-1)

Overall, the main difference in results between the US-1 and the overall NYSBLE samples was that the scores of US-1 candidates were well above the national average for all components of the bar exam instead of being well below as was the case for the overall NYSBLE sample scores. Similar to the NYSBLE sample, scores for the US-1 sample increased across the three years in July, outpacing the increase at the national level. February results were somewhat different as they declined by 0.2 points instead of rising a bit as they did in New York overall. But, that decline was smaller than the 0.9-point decline nationally. The gender difference, where males have higher bar exam scores than females, decreased over this period as well, particularly in July 2017. The gains observed were not shared equally among all race/ethnic groups, and the same pattern was observed as for the overall NYSBLE sample. Although the rate of change differed from 5% for Caucasian/Whites to 12% for both the Asian/Pacific Islanders and the Hispanic/Latin Americans, all groups experienced a rise in the percentage passing in both February and July administrations from pre-UBE to 2017.

It needs to be kept in mind that all of these differences in total account for less than 10% of the variance in bar exam scores and the effect size for any of the particular differences between UBE and pre-UBE levels never exceeded 0.55, a value on the low end of the range of what is considered a medium effect size. The other important factor that has not yet been considered here is the extent to which the observed differences among groups defined by bar

exam administration year (UBE), gender, and race/ethnicity were influenced by candidate background characteristics, specifically UGPA, LSAT scores, and LGPA. The patterns of average performance in these background characteristics tended to differ across years and across groups, and the influence of these factors was studied in the next section which used the school-based sample, a subset of domestic-educated New York candidates for whom candidate background characteristics were available from law schools.

School-based Sample (LS or Sb)[6]

The analysis of the school-based sample data was intended to study the effect of UBE controlling for the background characteristics of the candidates. This required using a different type of analysis than used above. The approach used is referred to as Analysis of Covariance (ANCOVA). The main comparison of interest in this study was examining bar exam scores before and after UBE adoption. ANCOVA involved statistically controlling for three predictor scores (UGPA, LGPA and LSAT scores), and can be thought of as using the following process. To control for LSAT scores on the bar exam scores, for example, the best fitting line between candidates would be obtained between the LSAT scores and bar exam scores for each year. A line has two things that define it: the slope (change in bar exam scores for each increment in LSAT score) and the intercept (the bar exam score that would be expected for an LSAT score of 0). The slope and intercept for the line would be used to obtain a predicted score for each candidate solely based upon their LSAT score. The difference between a candidate's actual score and the predicted score would then be added to the overall mean to obtain their adjusted score after statistically removing the effect of LSAT scores.

Separate analyses were conducted for the February administrations and the July administrations. ANCOVA provides essentially the same type of analysis as was provided by ANOVA, but it also includes the variance from the background characteristics so one can determine to what extent controlling for them had an effect. For these analyses, the ANCOVA controlled for candidate background characteristics: UGPA, LGPA, and LSAT scores. As with the ANOVA, because there are a relatively large number of tests run, an alpha level of .01 was adopted. Similar to the ANOVA tables presented above, ANCOVA tables also include η^2 values. The ANCOVA value of η^2 accounts for the background characteristics (UGPA, LGPA[7], and LSAT scores) compared to a similar ANOVA model, such that the difference in the η^2 values between the ANCOVA and ANOVA models indicate how much the background characteristics contribute to accounting for the variance of the outcome variable.

[6] This data-set is referred to as either Law School and abbreviated LS or School-based and abbreviated as Sb. Both terms are used interchangeably.

[7] 4-point LGPA was used throughout this appendix. See Section 2.4 of the main report for an explanation of 4-point LGPA.

Table O.11 shows the results of the ANCOVA for bar exam scores in February and July for the school-based sample (which included all available domestic-educated candidates). As noted earlier, February results are based upon relatively few candidates and appeared to be unstable for several comparisons being tested. Therefore, they will not be discussed further after this first consideration. The ANOVA η^2 values show that the combined contribution of the UBE, gender and race/ethnicity accounted for 3.9% of the variance in bar exam scores in February and 10.3% in July. These are about 2% to 3% less than what was found for the entire NYSBLE sample in both February and July. The ANCOVA η^2 values show the combined contribution of the background characteristics of UGPA, LGPA, and LSAT scores in addition to UBE, gender and race/ethnicity and accounted for 21.4% of the February variance and 61.4% of the July variance. Thus, the difference between the ANCOVA and ANOVA η^2 values indicated that the background characteristics accounted for 17.5% of the variance in bar exam scores in February compared to only 3.9% of the variance attributable to the UBE, gender and race/ethnicity combined. In July background characteristics accounted for 47.8% of the variance while UBE, gender and race/ethnicity accounted for 10.3%. This indicates that UGPA, LSAT scores, and LGPA together were at least four times more influential in explaining performance on the bar exam than the total combined effects of the UBE, gender, and race/ethnicity.

Table O.11
Analysis of Covariance (ANCOVA) Results
School-Based Sample (LS)

February		Bar Exam				MBE				Written			
Source	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE(U)	1	5712.22	18.37	$	.0129	946.06	8.59	#	.0063	1703.48	13.28	#	.0099
Gender(G)	1	11.97	0.04		-	1354.23	12.30	#	.0090	1329.34	10.36	*	.0077
Ethnicity(E)	3	1747.86	1.87		.0039	1509.03	4.57	*	.0101	497.30	1.29		.0029
U*G	1	14.78	0.05		-	5.31	0.05		-	1.04	0.01		-
U*E	3	2398.25	2.57	+	.0054	613.03	1.86		.0041	882.46	2.29	+	.0051
G*E	3	1845.04	1.98		.0041	640.53	1.94		.0043	895.15	2.33	+	.0052
U*G*E	3	1558.64	1.67		.0035	755.54	2.29		.0050	134.05	0.35		.0008
Total Sums of squares		443452.59				150096.66				172388.2583			
η² ANCOVA					.2138				.1776				.1659
η² ANOVA					.0390				.0386				.0512
η² NY					.0604				.0571				.0558
η² US-1					.0819				.0878				.0687

July		Bar Exam				MBE				Written			
Source	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE(U)	2	48767.73	69.36	$	.0056	9501.20	45.05	$	.0038	13961.20	50.09	$	.0054
Gender(G)	1	234.50	0.67		-	5711.21	54.16	$	.0023	6911.22	49.60	$	.0027
Ethnicity(E)	3	20307.98	19.25	$	.0023	9453.58	29.88	$	.0038	4159.75	9.95	$	.0016
U*G	2	1766.20	2.51	+	.0002	125.24	0.59		.0001	837.66	3.01	+	.0003
U*E	6	5134.05	2.43	+	.0006	698.67	1.10		.0003	2171.35	2.60	+	.0008
G*E	3	1076.95	1.02		.0001	65.53	0.21		-	622.27	1.49		.0002
U*G*E	6	656.70	0.31		-	264.19	0.42		.0001	400.26	0.48		.0002
Total Sums of squares		8666103.081				2491940.72				2567325.095			
η² ANCOVA					.6143				.5977				.483974
η² ANOVA					.1027				.0966				.0938
η² NY					.1361				.1251				.1287
η² US-1					.0965				.0928				.0846

+ p<.10, * p<.01, # p<.001, $ p<.0001

The July results in Table O.11 for the bar exam scores were statistically significant for the UBE and for race/ethnicity main effects. Gender was statistically significant for both the MBE and the written scores which cancelled each other out so there was no statistically significant effect for the bar exam scores. No interactions were statistically significant.

The biggest difference between the school-based sample results and the previous analyses of the NYSBLE sample was the drop in percentage of variance accounted for by race/ethnicity after accounting for UGPAs, LSAT scores, and LGPAs using the school-based sample. Whereas it accounted for from 5-12% of the variance in the analyses of various samples (NYSBLE overall, domestic-educated first-time takers, and school-based sample) before adjusting for background characteristics (i.e., the ANOVA results), the maximum it accounted for in the school-based sample after accounting for candidate background characteristics was 1% in February (MBE component score) and 0.4% in July (i.e., the ANCOVA results). This indicated

that much of the effect of race/ethnicity was accounted for by UGPAs, LSAT scores, and LGPAs.

Table O.12 shows the adjusted (Adj Mean) and unadjusted mean (Mean) scores for each bar exam component across the different years (UBE main effect). Adjusted mean scores are estimates of means after statistically accounting for UGPAs, LSAT scores, and LGPAs. Unadjusted means scores are the calculated means for a particular variable (i.e., a typically calculated average). Adjusted and unadjusted mean scores increased across years in both February and July. The biggest difference between adjusted and unadjusted average scores occurred in July 2017 where the unadjusted means were higher than the adjusted means, which indicated that the candidates' actual scores were, on average, higher than what would be expected from their performance on background characteristics (2.10 points higher for the bar exam, 0.78 points higher for the MBE, and 1.29 points higher for the written component).

Table O.12
MBE Means and Adjusted Means Controlling for UGPA, LSAT, LGPA
School-Based Sample*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
Bar Exam	Mean	256.90	**261.32**	4.42	285.80	**291.33**	**297.69**	5.53	11.89
	Adj Mean	257.07	**262.69**	5.62	285.91	**291.13**	**295.59**	5.22	9.68
	Diff	0.17	**1.37**		0.11	**-0.20**	**-2.10**		
MBE	Mean	128.62	**129.90**	1.28	143.02	**145.38**	**148.39**	2.36	5.37
	Adj Mean	128.65	**130.94**	2.29	143.31	**145.73**	**147.61**	2.42	4.30
	Diff	0.03	**1.04**		0.29	**0.35**	**-0.78**		
Written	Mean	128.40	**131.36**	2.96	143.05	**145.91**	**149.25**	2.86	6.20
	Adj Mean	128.63	**131.70**	3.07	142.84	**145.39**	**147.96**	2.55	5.12
	Diff	0.23	**0.34**		-0.21	**-0.52**	**-1.29**		

*Values in bold are UBE. The school-based sample here only includes candidates with valid gender and race/ethnicity, which is different from the school-based sample in the main report.

Table O.13 shows the adjusted and unadjusted MBE means for candidates in the school-based sample compared to the national MBE means. The most obvious difference was that the February results were below the national mean by (a) 4.2 to 6.3 points for the unadjusted means and (b) 3.1 to 6.3 for the adjusted means. The pattern was reversed in July with school-based sample means above the national mean by (a) 3.1 to 6.7 points for the unadjusted means and (b) 3.4 to 5.9 points for the adjusted means. The reversal of mean scores from well below the national mean in February to well above the national mean in July was not observed in the overall NYSBLE sample or domestic-educated first-time takers (US-1) group. This suggests that

in addition to having relatively small sample sizes, the February results for the school-based sample did not well represent the entire NYSBLE sample.

Table O.13
MBE Means and Adjusted Means Controlling for UGPA, LSAT, LGPA
School-based Sample (LS) and National Sample*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
LS	Mean	128.62	**129.90**	1.28	143.02	**145.38**	**148.39**	2.36	5.37
	Adj Mean	128.65	**130.94**	2.06	143.31	**145.73**	**147.61**	2.42	4.30
National	Mean	134.95	**134.06**	-0.89	139.92	**140.34**	**141.69**	0.42	1.77
Mean Diff		-6.33	**-4.16**		3.10	**5.04**	**6.70**		
Adj Mean Diff		-6.30	**-3.12**		3.39	**5.39**	**5.92**		

*Values in bold are UBE.

School-Based Sample Gender Effects

Table O.14 shows adjusted and unadjusted mean bar exam scores for females and males across bar exam administrations. The February results were erratic, showing a small difference in the unadjusted means in 2016 in the direction of males having higher scores but a difference of over 3 points in the opposite direction in 2017. The adjusted means reversed the difference in both years and moved the 3-point difference in 2017 to be almost identical for the females and males. In July, the difference between males and females for the unadjusted means were all in the direction of males having higher scores and differed by from 1.3 to 5.1 points. The difference between males and females in the adjusted means were smaller and ranged from 1.22 to 1.61 points with females having higher scores then males in 2015 and 2017. On the MBE and written component, the effect of the adjustment on the MBE scores was similar to the total in that the male and female adjusted means moved closer together for the July results. The written score showed a different pattern in both February and July. Scores favored females in February 2016 and 2017 by 2.7 and 3.6 points (2.8 and 2.7 points for adjusted means). In July, the unadjusted means for females ranged from 0.01 points lower than the males in 2016 to 2 points higher in 2017. After adjustment, the female means were all above their male counterparts by from 1.6 to 3.5 points.

Based upon what would be expected from their UGPAs, LSAT scores, and LGPAs, in July male scores were under-predicted[8] by from 0.72 to 1.38 points on the MBE, 1.27 to 1.96 points on the written component, and 2.04 to 3.36 points on the bar exam. Female performance

[8] Under-predicted here indicates that the adjusted mean was lower than the observed mean. Conversely, over-predicted indicates that the adjusted mean was higher than the observed mean.

was over-predicted in July 2015 and July 2016 (1.34 or 1.5 points on the MBE, 0.28 or 0.85 points on the written component, and 1.73 or 2.38 points on the bar exam) and under-predicted in July 2017 (0.14 points on the MBE, 0.65 points on the written component, and 0.82 points on the bar exam).

Table O.14
MBE Means and Adjusted Means Controlling for UGPA, LSAT, LGPA
School-Based Sample (LS) by Gender*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
Bar Exam									
Female	Mean	256.75	**262.78**	6.03	284.33	**288.76**	**297.02**	4.43	12.69
	Adj Mean	257.34	**262.68**	5.34	286.71	**290.49**	**296.20**	3.78	9.49
	Diff	0.59	**-0.10**		2.38	**1.73**	**-0.82**		
Male	Mean	257.06	**259.58**	2.52	287.28	**293.81**	**298.34**	6.53	11.06
	Adj Mean	256.80	**262.71**	5.91	285.10	**291.77**	**294.98**	6.67	9.88
	Diff	0.26	**3.13**		-2.18	**-2.04**	**-3.36**		
	M-F Diff.	0.31	**-3.20**		2.95	**5.05**	**1.32**		
	Adj M-F Diff.	-0.54	**0.03**		-1.61	**1.28**	**-1.22**		
MBE									
Female	Mean	126.87	**129.72**	2.85	140.60	**142.81**	**146.71**	2.21	6.11
	Adj Mean	127.18	**129.64**	2.46	141.94	**144.31**	**146.57**	2.37	4.63
	Diff	0.31	**-0.08**		1.34	**1.50**	**-0.14**		
Male	Mean	130.41	**130.11**	-0.30	145.45	**147.86**	**150.03**	2.41	4.58
	Adj Mean	130.12	**132.23**	2.11	144.68	**147.14**	**148.65**	2.46	3.97
	Diff	-0.29	**2.12**		-0.77	**-0.72**	**-1.38**		
	M-F Diff.	3.54	**0.39**		4.85	**5.05**	**3.32**		
	Adj M-F Diff	2.94	**2.59**		2.74	**2.83**	**2.08**		
Written									
Female	Mean	129.74	**133.01**	3.27	143.76	**145.90**	**150.26**	2.14	6.50
	Adj Mean	130.04	**133.03**	2.99	144.61	**146.18**	**149.61**	1.57	5.00
	Diff	0.30	**0.02**		0.85	**0.28**	**-0.65**		
Male	Mean	127.02	**129.39**	2.37	142.34	**145.91**	**148.27**	3.57	5.93
	Adj Mean	127.22	**130.36**	3.14	141.07	**144.61**	**146.31**	3.54	5.24
	Diff	0.20	**0.97**		-1.27	**-1.30**	**-1.96**		
	M-F Diff	-2.72	**-3.64**		-1.42	**0.01**	**-1.99**		
	Adj M-F Diff	-2.82	**-2.67**		-3.54	**-1.57**	**-3.30**		

*Values in bold are UBE.

School-based Sample Race/Ethnicity Effects

Table O.15 shows adjusted, and unadjusted average bar exam scores, MBE scores, and written scores from the February and July bar exam administrations. Figure O.3 shows the same data for average bar exam scores but in graphic form. The unadjusted means are on the left represented by lower case letters while the adjusted means are on the right in capital letters. The first letter of the racial/ethnic group is used to designate values for that group (w, W= Caucasian/White, b, B=Black/African American, h, H=Hispanic/Latino, a, A=Asian/Pacific Islander).

The most noticeable feature in the figure is that while the unadjusted means in the left panel show substantial separation between the different racial/ethnic groups, the adjusted means (after statistically accounting for UGPA, LSAT, and LGPA) in the right panel are much more similar. The difference between the mean of the highest scoring racial/ethnic group and lowest scoring group (range) was 23.56 in 2015 and which was reduced to 9.55 after adjusting for the candidate background characteristics. The reduction in range was even greater for the subsequent UBE administrations decreasing from 28.80 to 3.88 in 2016 and from 24.44 to 4.14 in 2017. Also, of note, the adjusted means for the Black/African American group were highest in 2015 and 2017 and were similar to other groups in 2016. These results indicate that after adjusting for differences in UGPA, LSAT, and LGPA, racial/ethnic group differences in bar exam mean scores largely dissipate. This is particularly true after UBE adoption in years 2016 and 2017. However, care should be exercised in making generalizations from the school-based sample to the entire NYSBLE sample due to potential differences between the two samples. That said, the school-based sample did provide evidence that background candidate characteristics matter and accounting for them reduces (but does not eliminate) differences among racial/ethnic groups.

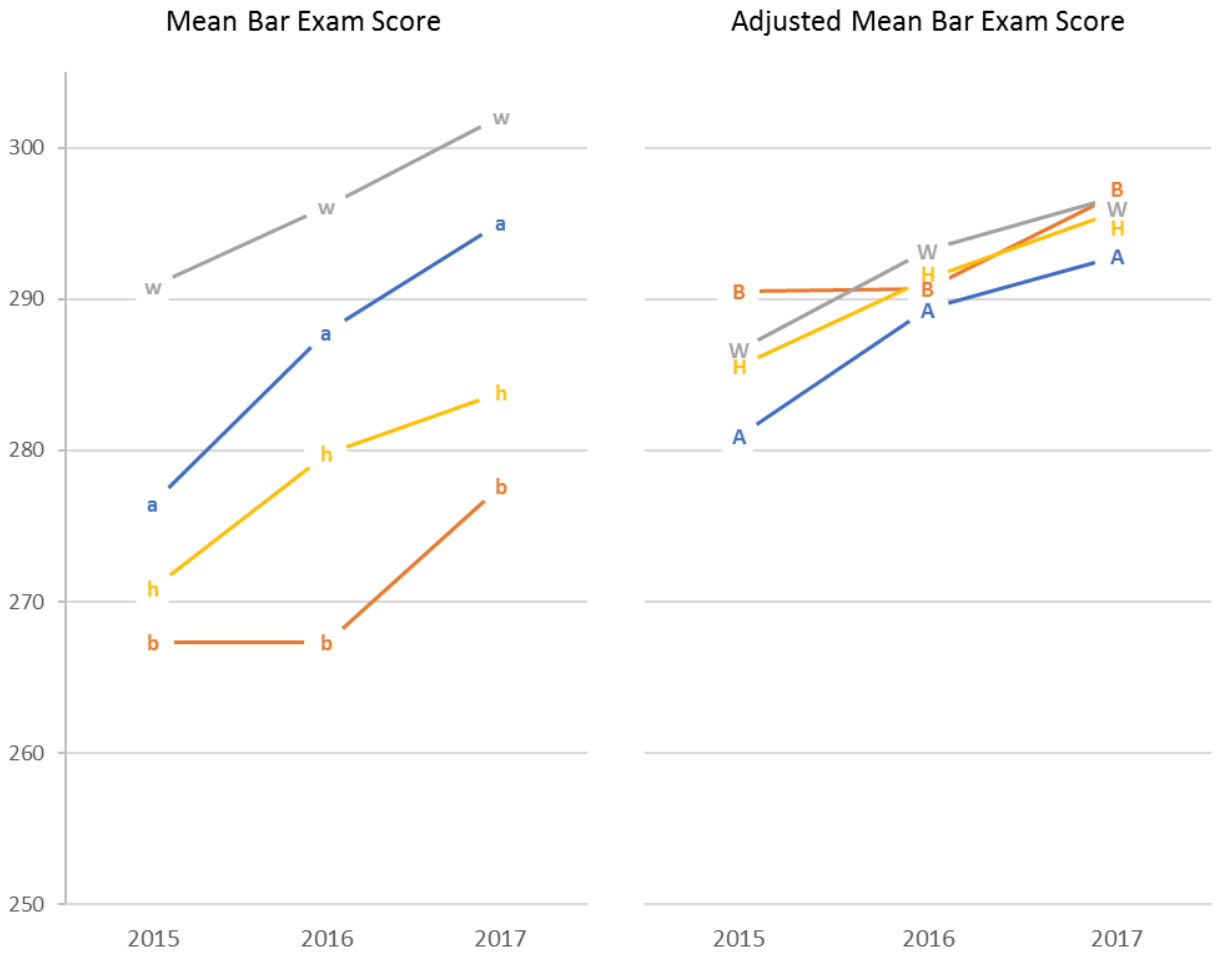
Table O.16 provides ANCOVA results for first-time takers in the school-based sample for July. As noted earlier, February results were omitted because they were based upon too few candidates. η^2 values from the previous analyses are provided at the bottom of the table to provide a context for interpreting these results (All LS). Aside from reducing the percentage of variance in bar exam scores accounted for, the pattern of results for first-time takers did not change compared to the entire school-based sample described above.

Table O.15
MBE Means and Adjusted Means Controlling for UGPA, LSAT, LGPA
School-Based Sample (LS) by Race/Ethnicity*

Admin		February			July				
Year		2016	**2017**	DF17	2015	**2016**	**2017**	DJ16	DJ17
Bar Exam Score									
Caucasian/White	Mean	258.69	**261.45**	2.76	290.86	**296.10**	**302.08**	5.24	11.22
	Adj Mean	257.14	**259.44**	2.30	286.65	**293.20**	**296.82**	6.55	10.17
	Diff	-1.55	**-2.01**		-4.21	**-2.90**	**-5.26**		
Asian/Pacific Islander	Mean	253.86	**264.61**	10.75	276.48	**287.83**	**295.09**	11.35	18.61
	Adj Mean	253.32	**263.61**	10.29	280.95	**289.32**	**292.82**	8.37	11.87
	Diff	-0.54	**-1.00**		4.47	**1.49**	**2.27**		
Black/African American	Mean	254.19	**257.23**	3.04	267.30	**267.30**	**277.64**	0.00	10.34
	Adj Mean	258.92	**262.71**	3.79	290.50	**290.65**	**296.96**	0.15	6.46
	Diff	4.73	**5.48**		23.20	**23.35**	**19.32**		
Hispanic/Latino	Mean	257.20	**263.86**	6.66	270.89	**279.78**	**283.83**	8.89	12.94
	Adj Mean	258.91	**265.01**	6.10	285.53	**291.35**	**295.77**	5.82	10.24
	Diff	1.71	**1.15**		14.64	**11.57**	**11.94**		
Range	Means	4.83	**7.38**		23.56	**28.80**	**24.44**		
	Adj Mean	5.60	**5.57**		9.55	**3.88**	**4.14**		
MBE									
Caucasian/White	Mean	129.92	**130.03**	0.11	145.69	**147.76**	**150.66**	2.07	4.97
	Adj Mean	129.03	**129.05**	0.02	143.38	**146.17**	**147.82**	2.79	4.44
	Diff	-0.89	**-0.98**		-2.31	**-1.59**	**-2.84**		
Asian/Pacific Islander	Mean	126.62	**130.02**	3.40	138.27	**143.02**	**146.42**	4.75	8.15
	Adj Mean	126.05	**129.35**	3.30	140.36	**143.70**	**145.19**	3.34	4.83
	Diff	-0.57	**-0.67**		2.09	**0.68**	**-1.23**		
Black/African American	Mean	126.34	**128.22**	1.88	133.06	**133.89**	**139.09**	0.83	6.03
	Adj Mean	129.00	**131.17**	2.17	145.99	**146.54**	**149.61**	0.55	3.62
	Diff	2.66	**2.95**		12.93	**12.65**	**10.52**		
Hispanic/Latino	Mean	129.04	**132.58**	3.54	135.19	**140.13**	**141.19**	4.94	6.00
	Adj Mean	130.51	**134.18**	3.67	143.51	**146.50**	**147.81**	2.99	4.30
	Diff	1.47	**1.60**		8.32	**6.37**	**6.62**		
Range	Means	3.58	**4.36**		12.63	**13.84**	**11.57**		
	Adj Mean	4.46	**5.13**		5.63	**2.84**	**4.42**		
Written Score									
Caucasian/White	Mean	128.92	**131.36**	2.44	145.52	**148.28**	**151.37**	2.76	5.85
	Adj Mean	128.23	**130.32**	2.09	143.59	**146.97**	**148.94**	3.38	5.35
	Diff	-0.69	**-1.04**		-1.93	**-1.31**	**-2.43**		
Asian/Pacific Islander	Mean	127.47	**134.54**	7.07	138.31	**144.77**	**148.63**	6.46	10.32
	Adj Mean	127.47	**134.17**	6.70	140.67	**145.55**	**147.53**	4.88	6.86
	Diff	0.00	**-0.37**		2.36	**0.78**	**1.10**		
Black/African American	Mean	127.59	**128.95**	1.36	134.13	**133.37**	**138.49**	-0.76	4.36
	Adj Mean	129.79	**131.55**	3.79	144.64	**144.18**	**147.40**	-0.46	2.76
	Diff	2.20	**2.60**		10.51	**10.81**	**8.91**		
Hispanic/Latino	Mean	128.60	**131.20**	2.60	135.92	**139.60**	**142.59**	3.68	6.73
	Adj Mean	129.01	**130.75**	1.74	142.45	**144.87**	**147.97**	2.42	5.52
	Diff	0.41	**-0.45**		6.53	**5.27**	**5.38**		
Range	Means	1.45	**5.59**		11.39	**14.91**	**12.88**		
	Adj Mean	2.32	**3.85**		3.97	**2.79**	**1.54**		

*Values in bold are UBE.

Figure O.3
Mean and Adjusted Mean Bar Exam Scores
School-Based Sample (LS) by Race/Ethnicity



A, a = Asian/Pacific Islander,
B, b = Black/African American,
H, h = Hispanic/Latino,
W, w = Caucasian/White.

Table O.16
July Analysis of Covariance (ANCOVA) Results
School-Based Sample First-Time Takers (LS-1)

		Bar Exam Score				MBE				Written			
Source	df	SS	F		η²	SS	F		η²	SS	F		η²
UBE(U)	2	48620.13	72.32	$	.0070	9500.08	47.40	$	.0047	14067.14	51.71	$	.0067
Gender(G)	1	16.95	0.05		-	5840.44	58.29	$	.0029	4097.69	30.12	$	.0019
Ethnicity(E)	3	17226.36	17.08	$	.0025	8213.42	27.32	$	.0041	3683.23	9.03	$	.0017
U*G	2	1110.79	1.65		.0002	305.29	1.52		.0002	590.54	2.17		.0003
U*E	6	7409.35	3.67	*	.0011	917.81	1.53		.0005	3281.62	4.02	#	.0016
G*E	3	566.19	0.56		.0001	9.79	0.03		-	405.42	0.99		.0002
U*G*E	6	1481.76	0.73		.0002	721.82	1.20		.0004	451.92	0.55		.0002
Total	8627	6973774.69				2025776.203				2111520.774			
η² ANCOVA					.5854				.5746				.4459
η² ANOVA					.1015				.0961				.0907
All LS													
η² ANCOVA					.6143				.5977				.483974
η² ANOVA					.1027				.0966				.0938
η² NY					.1361				.1251				.1287
η² US-1					.0965				.0928				.0846

+p<.10, *p<.01, #p<.001, $p<.0001.

Summary for School-based Sample

It is unfortunate that the February school-based sample had relatively few candidates and did not appear to well represent the entire group of candidates in New York. While the July data also may not have perfectly represented entire group in New York, the July data was adequate for adjusting bar exam performance for the candidate background characteristics. The adjustment using analysis of covariance (ANCOVA) statistical techniques showed that much of the difference between racial/ethnic groups could be explained by differences in UGPAs, LSAT scores, and LGPAs. Comparing the η^2 values for the adjusted and unadjusted racial/ethnic data indicated that approximately 97% of the variance in bar exam scores by different racial/ethnic groups could be attributed to candidate background characteristics (% attributable to prior academic indicators = (1-(adjusted η^2 /unadjusted η^2)) x 100 = (1- (0.0025/0.0823)) x 100 = (1-0.0304)*100 = 0.9696x100 = 96.96%). A certain proportion of gender differences, where males tended to score higher, on average, on the MBE and females tended to score higher, on average, on the written component, was also attributable to candidate background characteristics. A somewhat surprising finding was that part of the difference that resulted in males scoring higher than females, on average, was due to males actually scoring higher on the MBE and written component than would be predicted given their performance on background characteristics.

As far as the UBE is concerned, mean scores increased on bar exams administered after UBE adoption in both February and July in the school-based sample, just as they did for the entire NYSBLE sample and domestic-educated first-time takers in the NYSBLE sample. When the means were adjusted for prior academic performance, they still increased across years, but not as rapidly. Comparing actual versus predicted performance accounting for prior academic indicators, performance on the written component exceeded what would be predicted by prior academic performance for all three July administrations by a progressively larger amount, particularly in 2017 where the mean was 1.29 points higher than predicted. On the MBE, July means were below predicted levels by a small amount in 2015 and 2016 (0.29 to 0.35 points), but in 2017 exceeded the predicted mean by 0.78 points. The 2017 results were consistent with some analyses of the national results indicating that the rise in scores in July 2017 were not what would be predicted from prior indicators. The results from the school-based sample analysis need to be interpreted with some caution because they may not well represent the domestic-educated candidates taking the New York bar exam. It also needs to be kept in mind that all of the differences related to UBE, gender, and race/ethnicity in total accounted for less than 4% (in February) and 11% (in July) of the variance in bar exam scores, MBE scores, or written scores. More than four times the amount of variance was accounted for by UGPAs, LSATs, and LGPAs.

Summary and Discussion

Summary

The combined effect of (a) the UBE across years, (b) gender, and (c) race/ethnicity never accounted for more than 14% of the variance in bar exam scores. For first-time takers, the percentage of variance accounted for dropped to less than 10%. The variance attributable to the UBE across study years was less than 3%. Race/ethnicity accounted for up to 12% of the variance in bar exam scores, but after adjusting for UGPAs, LSAT scores, and LGPAs, the percentage of variance accounted for by race/ethnicity dropped to at most 2%.

Given this as context, New York bar exam scores rose across all three study years from July 2015 to July 2017 in both February and July administrations. The data showed that (a) males tended to perform better on the MBE and females tended to perform better on the written component than males and (b) there were differences in performance across racial/ethnic groups. However, despite differences observed across groups, scores increased for each group across years in February and July, before and after UBE adoption in New York. The Black/African American group showed a somewhat smaller gain in July 2016 compared to other groups but showed a larger gain in July 2017 that resulted in overall gains at July 2017 consistent with those experienced by the other racial/ethnic groups.

Over the three years observed, average bar exam scores and average MBE scores rose for New York candidates, with average MBE scores rising at a somewhat faster pace than average scores nationally. Average MBE scores of New York candidates were typically lower than the national mean because of the relatively large proportion of foreign-educated candidates taking the New York bar exam. Overall, average MBE scores for New York candidates were 5.6 points below the average MBE scores nationally before UBE adoption in February 2016 and 2.8 points lower in July 2015. The rise in MBE scores across years moved the New York candidate performance closer to national results.

When scores were adjusted for candidate background characteristics as represented UGPAs, LSAT scores, and LGPAs, differences among racial/ethnic groups largely disappeared as did some of the differences between females and males. The remaining difference between females and males after adjusting for candidate background characteristics was primarily due to males scoring higher than predicted on the MBE and written components.

Discussion

Given that UBE, gender, and race/ethnicity accounted for at most 14% of the variance in bar exam scores, care must be exercised in interpreting the differences that were observed. This is particularly true since the performance on background characteristics that candidates bring with them to the bar exam was found to account for 50% or so of the variance in bar exam scores.

Within this context, scores for New York candidates on the bar exam, MBE, and written component, increased after implementation of the UBE, more so on the MBE than observed at the national level. The results throughout this appendix indicate that the UBE did not appear to have much, if any, impact on the performance of New York candidates. However, there are other factors that need to be considered before drawing conclusions from the analysis above.

The UBE provides candidates with a portable score which allows them to take the bar exam in any UBE jurisdiction. This benefit could change the characteristics of candidates who decided to sit for the bar exam in a particular jurisdiction. In July 2017, 9,932 candidates sat for the New York bar exam, a 3.54% decline from the 10,297 that sat the previous July. Nationally, the number of candidates who took the bar increased by a slight amount (0.23%). When looking at the numbers taking the exam by different racial/ethnic categories, the decline for Caucasian/White candidates between 2016 and 2017 was 8.79%, down a total of 12.62% from pre-UBE July 2015. Given that the New York decline is counter that observed nationally, the UBE may have had an impact on the candidates choosing to take the bar exam in New York.

Potential limitations in the representativeness of the domestic-educated candidates included in the school-based sample used to control for prior academic performance should also be considered when drawing conclusions from the results of this appendix. The Pre-UBE February and July data were only 29% and 33%, respectively, of all of the domestic-educated candidates who sat for the exam. The UBE data, in contrast, is represented by from at least 38% in February 2017 while the July UBE data is based upon 74% and 67% in 2016 and 2017, respectively. The result was that the February data was insufficient to support statistical analyses, particularly for first-time takers, and those that were done on February data tended to be unstable. Further, the school-based sample data contained a disproportionately large number of first-time takers and Caucasian/White candidates. Given that the school-based sample data was used to control for differences in prior academic performance, the disproportionate representation of the different racial/ethnic groups could make the adjustment less than ideal for those groups not proportionately represented. A related matter is that small sample sizes for three of the seven racial/ethnic group categories (American Indian/Alaskan Native, Chicano/Mexican American, and Puerto Rican) led to data from these groups being excluded from the analyses.

Another factor to note is that bar exam scores before UBE adoption were based upon a different mix of components including weighting the MBE score 40% (instead of 50%), a New York developed multiple-choice exam, and essays developed by New York. Scores on the New York bar exam pre-UBE were made comparable by a simple multiplier that reduced the maximum score from 1,000 to the UBE maximum score of 400, which was reasonable for aligning pre-UBE scores with UBE scores for purposes of this study but should not be interpreted as suggesting that scores from either exam are necessarily strictly interchangeable.

While these are all limitations to what can be concluded from the results of this appendix, there are some factors that mitigate these limitations. Adjusting for candidate background characteristics removed a large amount of the difference in performance by race/ethnicity and gender. Also, gender differences decreased somewhat over the course of the UBE in July, which suggests that differences may have abated somewhat. The increase in MBE scores over the study period beyond the increase at the national level indicated that the increase in bar exam scores was likely due, at least in part, to changes in the characteristics of candidates taking the New York bar exam. Ultimately, as was found in the main report, the results here supported the conclusion that UBE adoption had, at most, small effects on candidates taking the New York bar exam.

Conclusions

Returning to the three research questions addressed in the main report:

How do candidate background characteristics compare across bar exam administrations? How do they relate to performance on the bar exam in New York before and after UBE adoption?

Candidate background characteristics as reflected in UGPAs, LSAT scores, and LGPAs accounted for approximately 50% of the variance in bar exam scores. Differences in performance by gender and race/ethnicity were greatly reduced after accounting for UGPAs, LSAT scores, and LGPAs.

How do candidates grouped by race/ethnicity and gender perform on the bar exam before and after UBE adoption?

Race/ethnicity, gender and bar exam administration year together accounted for at most a total of 14% of the variance in bar exam scores. For gender and the four race/ethnic groups that were in sufficient numbers to support analysis, bar exam scores and pass rates increased after UBE adoption. Scores for the Black/African American group rose at a slower rate in July 2016 than other groups but then rose more quickly than other groups in July 2017. The difference between females and males for July exams progressively decreased somewhat each year after UBE adoption. The small increases in bar exam scores and the decrease in differences between females and males were consistent with changes in MBE scores over the same period, indicating that shifts in performance were likely due to changes in the composition of the candidate population and/or changes in exam preparation.

How does performance on the New York bar examination compare before and after UBE adoption?

Candidate bar exam scores and pass rates increased after adopting the UBE. The increases in bar exam scores and pass rates were likely due to changes in the composition of the candidate population sitting for the exam in New York or to changes in exam preparation, rather than to changes in the exam after UBE adoption.