Online Appendix for: Evolution of Discrimination on Online Platforms

Peibo Zhang
Goizueta Business School, Emory University, peibo.zhang@emory.edu

Xinyu Cao
The Chinese University of Hong Kong Business School, xinyucao@cuhk.edu.hk

Ruomeng Cui
Goizueta Business School, Emory University, ruomeng.cui@emory.edu

Dennis J. Zhang
Olin Business School, Washington University in St. Louis, denniszhang@wustl.edu

---

A. Additional Figures and Tables

A.1. Rating Distribution

Figure A1 shows that among the 23.07% of completed classes that receive ratings, both African-American and White teachers receive a 5-star rating in over 98% of cases, and the rating distributions for African-American and White teachers are similar. These findings suggest that class ratings are less effective as a distinguishing reputation metric. However, students can still infer teachers' reputation based on the cumulative number of classes taught, as displayed on their profile page.

A.2. Cumulative Classes Taught Over Time

Figure A2 plots the average cumulative classes taught from Week 1 to Week 12 for White and African-American teachers separately. The figure clearly illustrates the diverging trajectory: while both groups start at comparable levels in Week 1, the gap in cumulative classes taught widens progressively over the 12-week period, with White teachers accumulating classes at a meaningfully faster rate. This visual evidence complements the regression results in Table 6 of the main manuscript.

Figure A1 Rating Distribution Across White and African-American Teachers' Completed Classes

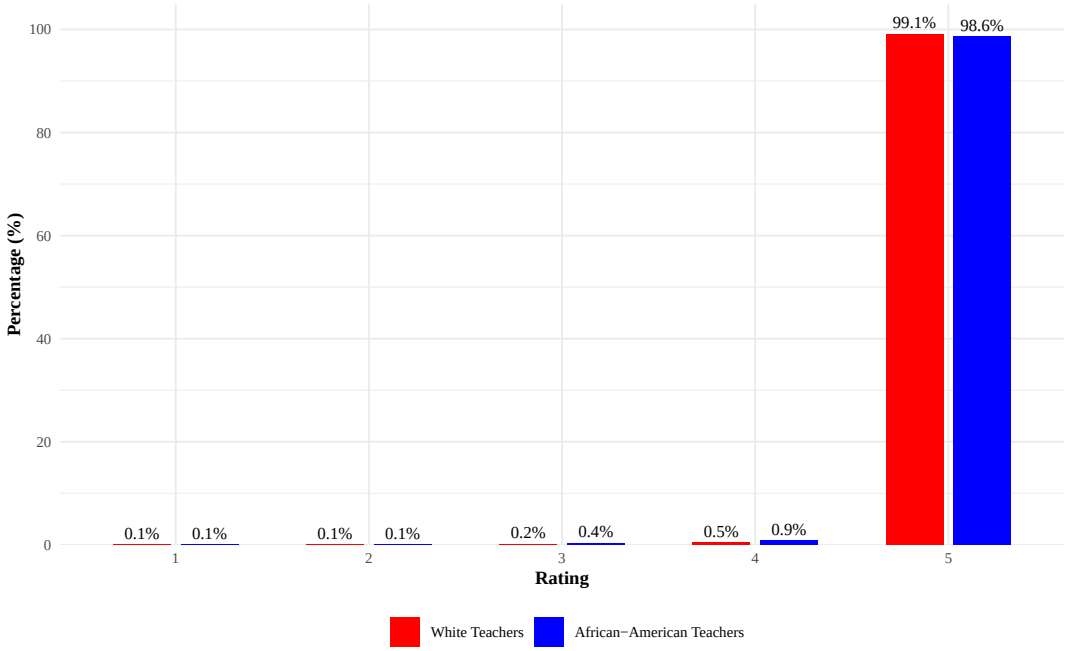
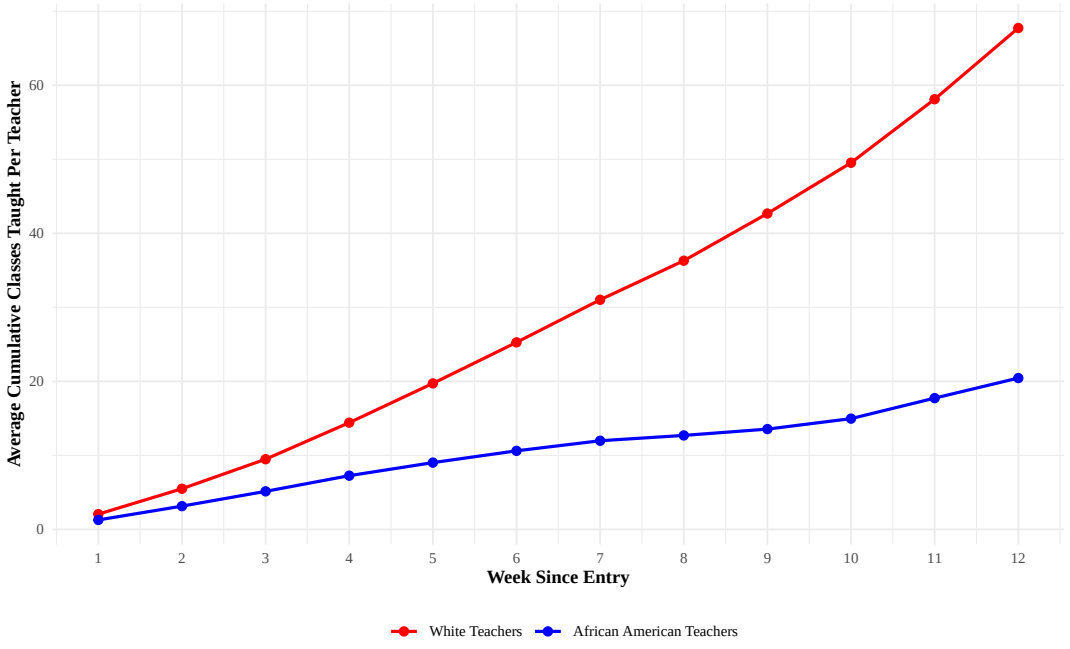


Figure A2 Cumulative Classes Taught Over Time Across White and African-American Teachers



A.3. Pre-Bonferroni Correction Significance Results

In Equations (3) and (4) of the main manuscript, we include a full set of dummy variables for each Week Since Entry term, resulting in more than twenty independent variables in each specification. To address concerns about multiple hypothesis testing, we apply a Bonferroni correction and report

Table A1 Gaps in Daily Bookings and Booking Probability (Pre-Bonferroni Correction)

Dependent Variables:	*Daily Bookings* (1)	*Booked* (2)	*Daily Repeat Bookings* (3)	*Daily New Bookings* (4)
African-American x Week Since Entry 2	−0.160***	−0.024**	−0.050***	−0.111***
	(0.034)	(0.007)	(0.010)	(0.029)
African-American x Week Since Entry 3	−0.266***	−0.039***	−0.117***	−0.149***
	(0.038)	(0.008)	(0.015)	(0.031)
African-American x Week Since Entry 4	−0.380***	−0.059***	−0.195***	−0.185***
	(0.047)	(0.009)	(0.019)	(0.035)
African-American x Week Since Entry 5	−0.571***	−0.086***	−0.323***	−0.248***
	(0.044)	(0.009)	(0.024)	(0.030)
African-American x Week Since Entry 6	−0.673***	−0.103***	−0.417***	−0.256***
	(0.045)	(0.009)	(0.025)	(0.029)
African-American x Week Since Entry 7	−0.683***	−0.104***	−0.500***	−0.183***
	(0.065)	(0.011)	(0.032)	(0.042)
African-American x Week Since Entry 8	−0.736***	−0.114***	−0.554***	−0.182***
	(0.061)	(0.010)	(0.037)	(0.034)
African-American x Week Since Entry 9	−0.880***	−0.135***	−0.644***	−0.236***
	(0.065)	(0.011)	(0.041)	(0.034)
African-American x Week Since Entry 10	−0.932***	−0.150***	−0.716***	−0.217***
	(0.071)	(0.012)	(0.046)	(0.036)
African-American x Week Since Entry 11	−0.997***	−0.157***	−0.781***	−0.216***
	(0.087)	(0.014)	(0.051)	(0.049)
African-American x Week Since Entry 12	−1.088***	−0.175***	−0.829***	−0.259***
	(0.092)	(0.016)	(0.062)	(0.042)
Calendar Week FE	Yes	Yes	Yes	Yes
Teacher FE	Yes	Yes	Yes	Yes
Week Since Entry Terms	Yes	Yes	Yes	Yes
N	252 238	1 556 339	252 238	252 238

Note: Standard errors, clustered at the teacher level, are in parentheses. Significance: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

the adjusted significance levels in the main manuscript. For completeness, we also present the corresponding results *prior* to Bonferroni adjustment in Appendix Tables A1 and A2.

Table A2 Gap in Daily Openings (Pre-Bonferroni Correction)

Dependent Variables:	*Daily Openings* (1)	*Daily Prime-time Openings* (2)
African-American x Week Since Entry 2	0.415*	0.099
	(0.185)	(0.066)
African-American x Week Since Entry 3	0.595*	0.230**
	(0.235)	(0.082)
African-American x Week Since Entry 4	0.513+	0.177+
	(0.283)	(0.093)
African-American x Week Since Entry 5	0.354	0.115
	(0.305)	(0.098)
African-American x Week Since Entry 6	0.325	0.026
	(0.311)	(0.103)
African-American x Week Since Entry 7	0.071	−0.069
	(0.319)	(0.108)
African-American x Week Since Entry 8	−0.159	−0.097
	(0.327)	(0.113)
African-American x Week Since Entry 9	−0.540	−0.195
	(0.338)	(0.121)
African-American x Week Since Entry 10	−0.550	−0.269*
	(0.375)	(0.130)
African-American x Week Since Entry 11	−0.651	−0.350*
	(0.401)	(0.137)
African-American x Week Since Entry 12	−0.659	−0.400**
	(0.445)	(0.148)
Calendar Week FE	Yes	Yes
Teacher FE	Yes	Yes
Week Since Entry Terms	Yes	Yes
N	252 238	252 238

Note: Standard errors, clustered at the teacher level, are reported in parentheses. Significance at $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$, $^{+}p < 0.1$.

A.4. Additional Analysis for Booking Probabilities

In Section 5.2 of the main manuscript, we showed that the gap in new booking probability widens over time when modeling Week Since Entry as a continuous variable. For completeness, we now examine how this gap evolves across each of the first twelve weeks. Table A3 reports the week-by-week estimates, and the results confirm that the new booking probability gap continues to widen throughout the 12-week period.

Table A3 Gap in New Booking Probability

Dependent Variable:	*New Booked*
	(1)
African-American × Week Since Entry 2	−0.014
	(0.007)
African-American × Week Since Entry 3	−0.020
	(0.007)
African-American × Week Since Entry 4	−0.031**
	(0.008)
African-American × Week Since Entry 5	−0.042***
	(0.007)
African-American × Week Since Entry 6	−0.047***
	(0.007)
African-American × Week Since Entry 7	−0.038***
	(0.008)
African-American × Week Since Entry 8	−0.040***
	(0.007)
African-American × Week Since Entry 9	−0.053***
	(0.008)
African-American × Week Since Entry 10	−0.057***
	(0.008)
African-American × Week Since Entry 11	−0.057***
	(0.010)
African-American × Week Since Entry 12	−0.069***
	(0.010)
Calendar Week FE	Yes
Class Time FE	Yes
Teacher FE	Yes
Week Since Entry Terms	Yes
N	1 431 908

Note: Standard errors, clustered at the teacher level, are reported in parentheses. The significance level in this table is adjusted following the Bonferroni correction. Specifically, Significance at $^{***}p_{adjust} < 0.000043$; $^{**}p_{adjust} < 0.00043$; $^{*}p_{adjust} < 0.002$.

We also examined the gap in repeat booking probability in Section 5.2 of the main manuscript. For completeness, we now analyze how this gap evolves across successive repeat classes and report the results in Table A4 using the class-level data. The estimates show that while there is a significant racial gap in rebooking probability after the first and second classes, the gap becomes statistically indistinguishable from zero after three or more classes. Moreover, for each repeat stage, we detect no systematic change in the gap over time. Taken together, these findings provide strong evidence

that there is no widening racial disparity in repeat booking probability between African-American and White teachers.

Table A4 Gap in Repeat Booking Probability

Dependent Variables:	*Rebook (after 1st class)*		*Rebook (after 2nd class)*		*Rebook (after 3+ classes)*	
	(1)	(2)	(3)	(4)	(5)	(6)
African-American	−0.037***		−0.069***		−0.010	
	(0.010)		(0.018)		(0.012)	
African-American × Week Since Entry (Continuous)		0.001		0.002		−0.001
		(0.004)		(0.007)		(0.004)
Included Weeks	12 Weeks	12 Weeks	12 Weeks	12 Weeks	12 Weeks	12 Weeks
Calendar Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Class Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	No	Yes	No	Yes	No
Teacher FE	No	Yes	No	Yes	No	Yes
Week Since Entry Terms	No	Yes	No	Yes	No	Yes
N	92 814	92 814	26 069	26 069	63 145	63 145

Note: Standard errors, clustered at the teacher level, are reported in parentheses. Significance at $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

B. Robustness Analyses

This section presents a comprehensive set of robustness analyses to validate the robustness of our estimates for the initial discrimination and its evolution over time.

B.1. Initial Racial Gap: First 6 Days on the Platform

We begin by analyzing the daily booking gap using data from teachers' first six days on the platform. In Table B1, we use the class-level dataset and define the outcome variable *Booked*, which equals 1 if an opening was booked and 0 otherwise. Columns (1)–(5) report the booking probability gap between African-American and White teachers within the first 2 to 6 days after entry, respectively. We omit Day 1 results due to insufficient observations. As shown in Table B1, a statistically meaningful booking gap emerges as early as Day 3.

Table B1 First 6 Days Gap in Probability of Being Booked by All Students

Dependent Variable:	*Booked*				
	(1)	(2)	(3)	(4)	(5)
African-American	−0.011	−0.022+	−0.035***	−0.034***	−0.034***
	(0.017)	(0.012)	(0.009)	(0.008)	(0.007)
Included Days	First 2 Days	First 3 Days	First 4 Days	First 5 Days	First 6 Days
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Class Time FE	Yes	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	Yes	Yes	Yes
N	4761	14 822	27 938	43 163	60 219

Note: All regressions in this table use the class-level dataset. Teacher characteristics include: base pay per class, teaching experience, education level, gender, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$.

We then replicate the analysis in Table B2, replacing the outcome with *NewBooked*. Here, *NewBooked* = 1 if an opening was booked by a new student and 0 otherwise; repeat bookings are excluded so that only openings available to new students are considered. Columns (1)–(5) again present the gap across the first 2 to 6 days on the platform. Consistent with the previous findings, we observe that the racial gap in new-booking probability also becomes evident by Day 3.

To demonstrate the consequences of the initial gap in booking probability for teachers' overall booking performance, we next examine differences in the daily number of total bookings, new bookings, and repeat bookings. Table B3 reports these results using teacher–day–level data. Consistent with the probability-based findings, the gap in daily bookings emerges as early as Day 3 and is driven primarily by differences in new-student bookings. This gap then persists throughout the first six days on the platform. Taken together, these results indicate that the initial booking disadvantage faced by African-American teachers is unlikely to be the result of random fluctuations in early activity; rather, it reflects a stable and systematic difference between the two groups.

Table B2 First 6 Days Gap in Probability of Being Booked by New Students

Dependent Variable:	*NewBooked*				
	(1)	(2)	(3)	(4)	(5)
African-American	−0.010	−0.023*	−0.034***	−0.033***	−0.033***
	(0.017)	(0.011)	(0.008)	(0.007)	(0.006)
Included Days	First 2 Days	First 3 Days	First 4 Days	First 5 Days	First 6 Days
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Class Time FE	Yes	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	Yes	Yes	Yes
N	4751	14 782	27 790	42 867	59 718

Note: All regressions in this table use the class-level dataset. Teacher characteristics include: base pay per class, teaching experience, education level, gender, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Table B3 First 6 Days Gap in Daily Bookings

Panel A. Daily Total Bookings						
Dependent Variable:	*Daily Bookings*					
	(1)	(2)	(3)	(4)	(5)	(6)
African-American	0.002	-0.016	-0.039^{+}	-0.068**	-0.079**	-0.099***
	(0.002)	(0.014)	(0.021)	(0.026)	(0.026)	(0.025)
Panel B. Daily New Bookings						
Dependent Variable:	*Daily New Bookings*					
	(1)	(2)	(3)	(4)	(5)	(6)
African-American	0.002	-0.016	-0.040*	-0.066**	-0.075**	-0.092***
	(0.002)	(0.013)	(0.020)	(0.024)	(0.024)	(0.023)
Panel C. Daily Repeat Bookings						
Dependent Variable:	*Daily Repeat Bookings*					
	(1)	(2)	(3)	(4)	(5)	(6)
African-American	0.000	-0.000	0.001	-0.003	-0.004	-0.007^{+}
	(0.000)	(0.001)	(0.002)	(0.003)	(0.004)	(0.004)
Included Days	First Day	First 2 Days	First 3 Days	First 4 Days	First 5 Days	First 6 Days
Calendar Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
N	3514	7028	10542	14056	17570	21084

Note: All regressions use the teacher-day-level dataset. Teacher characteristics include base pay per class, teaching experience, education level, gender, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and reported in parentheses. Statistical significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$.

B.2. Initial Racial Gap: Sensitivity Analysis for Unobserved Variables

Even after controlling for a comprehensive set of teacher characteristics, one might argue that our estimates could still suffer from omitted-variable bias. To address this concern, we follow a well-established body of economic literature and conduct a sensitivity analysis to assess the robustness of our results to potential unobserved factors (Lin et al. 1998, Imbens 2003, Altonji et al. 2005, Cinelli and Hazlett 2019, Oster 2019). In particular, we adopt the widely used frameworks of Altonji et al. (2005) and Oster (2019).

This method evaluates how vulnerable the estimated race coefficient is to omitted variables by comparing a "short" regression (including only the race indicator) with a "long" regression (including race and the full set of observed teacher characteristics). Following Altonji et al. (2005) and Oster (2019), the key assumption is that selection on unobservables is proportional to selection on observables. Let $\hat{\beta}_0$ and $\hat{R}_0$ denote the coefficient and R^2 from the short regression, and let $\hat{\beta}$ and $\hat{R}$ denote the coefficient and R^2 from the long regression.

Under proportional selection, the bias-adjusted coefficient is given by

$$\beta^* = \hat{\beta} - \delta \left(\hat{\beta}_0 - \hat{\beta}\right) \frac{R_{\max} - \hat{R}}{\hat{R} - \hat{R}_0},$$

which represents what the race effect would be if unobserved variables were δ times as important as the observed controls.

The value of δ that drives the adjusted coefficient to zero, i.e. the amount of selection needed to fully eliminate the estimated booking gap between African-American and White teachers, is

$$\delta^* = \frac{\hat{\beta}}{(\hat{\beta}_0 - \hat{\beta})} \cdot \frac{\hat{R} - \hat{R}_0}{R_{\max} - \hat{R}}.$$

Two quantities therefore summarize robustness: (i) $|\delta^*|$, which measures how strong selection on unobservables would have to be (relative to observed selection) to reduce the race coefficient to zero, and (ii) $\beta^* \equiv \beta^*(\delta = 1)$, the implied effect if unobserved variables were as important as the observed controls.

Following Oster (2019), we set $R_{\max} = 1.3\hat{R}$. Empirically, this benchmark strikes a balance: it allows for substantial unobserved heterogeneity while still imposing a plausible upper bound consistent with typical regression settings where the addition of further controls yields diminishing returns.

Following this approach, we find that unobserved factors in our context would need to be more than 6.19 times as strongly related to daily bookings as the full set of observed controls in order to fully eliminate the estimated racial gap. This value substantially exceeds 1, the benchmark suggested by Oster (2019) for indicating robustness, and lies on the higher end of values reported

in other studies using the same framework (Garcia-López et al. 2020, Bazzi et al. 2020, Elfers and Koenraadt 2022, Bao et al. 2024). We also find that if unobserved variables were as influential as the observed characteristics ($\delta = 1$), African-American teachers would still receive approximately 0.091 fewer bookings per day than comparable White teachers.

Taken together, these results indicate that the estimated racial disparity in daily bookings is highly robust to potential omitted-variable bias, and that the Week 1 booking gap is most likely a causal effect of race.

B.3. Evolution of Initial Gap Over Time

We conduct two additional robustness checks on the evolution of the initial gap to address concerns related to the impact of the COVID-19 pandemic and the choice of functional form.

Table B4 shows that, using data with class dates up to December 5, the pattern of an increasing booking gap is fully consistent with our main findings. African-American teachers continue to receive statistically significantly fewer bookings than White teachers beginning in Week 1, and this gap widens over time.

Table B4 Pre-Pandemic Gap in Daily Bookings between African-American and White Teachers

Dependent Variable:	*Daily Bookings*			
	(1)	(2)	(3)	(4)
African-American	−0.120**	−0.304***		
	(0.046)	(0.044)		
African-American × Week Since Entry2			−0.235***	
			(0.064)	
African-American × Week Since Entry3			−0.280***	
			(0.084)	
African-American × Week Since Entry4			−0.437*	
			(0.211)	
African-American × Week Since Entry (continuous)				−0.149***
				(0.041)
Included Weeks	First Week	4 Weeks	4 Weeks	4 Weeks
Calendar Week FE	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	No	No
Teacher FE	No	No	Yes	Yes
N	7552	16 009	16 009	16 009

Note: All regressions in this table use the teacher-day-level dataset. Controls include teacher characteristics: base pay per class, teaching experience, education level, gender, round 1 interview score, round 2 interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

In addition, in the main analysis of daily bookings in Section 4.2 of the main manuscript, we use raw booking counts rather than a log transformation. We make this choice for two reasons: (1) it allows the results to be interpreted directly in their natural units, and (2) it simplifies the subsequent calculation of percentage contributions from daily new bookings and daily repeat bookings in the mechanism analyses. For completeness and to demonstrate the robustness of our findings, we replicate the main analysis using $\log(1 + \text{DailyBookings}_{it})$ (to accommodate zero values). As shown in Table B5, the results remain fully consistent with our main findings, confirming that our conclusions are not sensitive to this specification choice.

Table B5 Gap in Daily Bookings between African-American and White Teachers (Log Transformed)

Dependent Variable:	*log(1+Daily Bookings)*			
	(1)	(2)	(3)	(4)
African-American	−0.052***	−0.232***		
	(0.010)	(0.013)		
African-American × Week Since Entry (continuous)				−0.029***
				(0.002)
African-American × Week Since Entry2			−0.069***	
			(0.013)	
African-American × Week Since Entry3			−0.103***	
			(0.014)	
African-American × Week Since Entry4			−0.135***	
			(0.017)	
African-American × Week Since Entry5			−0.190***	
			(0.017)	
African-American × Week Since Entry6			−0.228***	
			(0.017)	
African-American × Week Since Entry7			−0.234***	
			(0.020)	
African-American × Week Since Entry8			−0.254***	
			(0.020)	
African-American × Week Since Entry9			−0.287***	
			(0.022)	
African-American × Week Since Entry10			−0.299***	
			(0.023)	
African-American × Week Since Entry11			−0.317***	
			(0.026)	
African-American × Week Since Entry12			−0.337***	
			(0.027)	
Included Weeks	First Week	12 Weeks	12 Weeks	12 Weeks
Calendar Week FE	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	No	No
Teacher FE	No	No	Yes	Yes
Week Since Entry Terms	No	No	Yes	Yes
N	24 598	252 238	252 238	252 238

Note: All regressions in this table use the teacher-day-level dataset. Controls include teacher characteristics: base pay per class, teaching experience, education level, gender, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

C. Evolution of Gender Gap Over Time

While our analysis focuses on racial discrimination as the source of the initial booking gap, the amplification mechanisms we document do not rely on the initial disparity being driven by race per se. In this section, we demonstrate that the same amplification dynamics apply to other types of disparities, such as gender differences. Specifically, we replicate the analyses underlying Tables 3, 5, and 6 of the main manuscript using gender as the group attribute, and report the corresponding results in Tables C1, C2, and C3, respectively. The model specifications remain identical to those in the original tables. In each specification, *Female* is a dummy variable equal to 1 for female teachers and 0 for male teachers. We also use Week Since Entry as a continuous variable to illustrate how the gender gap evolves over time.

From Column (1) of Table C1, we observe that female teachers have 0.078 fewer bookings in the first week compared to their male counterparts, with the effect size roughly 3/4 of that observed in the race analysis. In Column (2), when the observation window is extended to 12 weeks, the average gap widens to 0.343 bookings. Column (3) of Table C1 further shows that the booking gap between female and male teachers widens over time, mirroring the pattern observed in the race analysis.

Table C1 Gap in Daily Bookings between Female and Male Teachers

Dependent Variable:	*Daily Bookings*		
	(1)	(2)	(3)
Female	−0.078*	−0.343***	
	(0.039)	(0.086)	
Female × Week Since Entry (Continuous)			−0.050**
			(0.016)
Relative Effect Size	17.69%	26.24%	
Included Weeks	First Week	12 Weeks	12 Weeks
Calendar Week FE	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	No
Teacher FE	No	No	Yes
Week Since Entry Terms	No	No	Yes
N	24 598	252 238	252 238

Note: All regressions in this table use the teacher-day-level dataset. Teacher characteristics include: base pay per class, teaching experience, education level, race, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Second, Table C2 shows that although female teachers open fewer courses in the first week than male teachers, this opening gap does not significantly widen over time, whether measured across all time slots or only during prime-time slots. Taken together with the previous result, this evidence suggests that the growing booking gap for female teachers is not driven by supply-side behavior but is instead likely attributable to increasing demand-side discrimination, as in the race case.

Table C2 Gap in Daily Openings between Female and Male Teachers

Dependent Variables:	*Daily Openings*			*Daily Prime-Time Openings*		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	−0.488*	−0.470		−0.075	−0.062	
	(0.218)	(0.294)		(0.081)	(0.090)	
Female × Week Since Entry (Continuous)			0.021			0.003
			(0.041)			(0.014)
Included Weeks	First Week	12 Weeks	12 Weeks	First Week	12 Weeks	12 Weeks
Calendar Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	No	Yes	Yes	No
Teacher FE	No	No	Yes	No	No	Yes
Week Since Entry Terms	No	No	Yes	No	No	Yes
N	24 598	252 238	252 238	24 598	252 238	252 238

Note: All regressions in this table use the teacher-day-level dataset. Controls include teacher characteristics: base pay per class, teaching experience, education level, race, interview score, permitted class level, recruitment channel, and nationality. Standard errors are clustered at the teacher level and are reported in parentheses. Statistical significance levels are indicated as follows: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Third, to directly test whether discrimination intensifies over time, we examine booking probabilities of new students by gender in Table C3. Column (1) shows that female teachers have a slightly lower and marginally significant booking rate compared to their male counterparts, with an effect size roughly 1/5 of that observed in the race analysis. Column (3) further shows that the booking probability gap grows over time; although the trend is statistically insignificant, its magnitude is about 2/5 of that in the race case. Again, given the smaller effect sizes and standard errors comparable to those in the race analysis, we interpret this insignificance as a consequence of limited statistical power.

In summary, we find that female teachers exhibit lower bookings and lower booking probabilities from new students in the first week, with the booking numbers significantly widening and booking-probability gaps directionally widening over time. These patterns closely mirror those in the race analysis. The main difference lies in the smaller effect sizes, approximately one-fifth as large, which render some estimates statistically insignificant. We interpret these insignificant results as arising from limited statistical power and view the overall evidence as consistent with our conceptual framework.

Table C3 Gap in Probability of Being Booked by New Students

Dependent Variables:	*NewBooked*		
	(1)	(2)	(3)
Female	−0.005	−0.014*	
	(0.007)	(0.006)	
Female × Week Since Entry			−0.002
			(0.001)
Week Since Entry			0.008***
			(0.002)
Included Weeks	First Week	12 Weeks	12 Weeks
Calendar Week FE	Yes	Yes	Yes
Class Time FE	Yes	Yes	Yes
Teacher Characteristics	Yes	Yes	No
Teacher FE	No	No	Yes
N	77 458	1 431 908	1 431 908

Note: All regressions in this table use the class-level dataset. We exclude all repeat bookings, leaving only openings available for new students at booking time. For each opening, *NewBooked*= 1 if it was booked (necessarily by a new student), and *NewBooked*= 0 if not booked. Standard errors, clustered at the teacher level, are reported in parentheses. Significance: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

References

Altonji J, Elder T, Taber C (2005) Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy* 113(1):151–184, ISSN 00223808, 1537534X, URL `http://www.jstor.org/stable/10.1086/426036`.

Bao L, Huang D, Lin C (2024) Can artificial intelligence improve gender equality? evidence from a natural experiment. *Management Science* .

Bazzi S, Fiszbein M, Gebresilasse M (2020) Frontier culture: The roots and persistence of "rugged individualism" in the united states. *Econometrica* 88(6):2329–2368.

Cinelli C, Hazlett C (2019) Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82(1):39–67, ISSN 1369-7412, URL `http://dx.doi.org/10.1111/rssb.12348`.

Elfers F, Koenraadt J (2022) What you don't know won't hurt you: Market monitoring and bank supervisors' preference for private information. *Journal of Banking & Finance* 143:106572.

Garcia-López MÀ, Jofre-Monseny J, Martínez-Mazza R, Segú M (2020) Do short-term rental platforms affect housing markets? evidence from airbnb in barcelona. *Journal of Urban Economics* 119:103278.

Imbens GW (2003) Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review* 93(2):126–132, URL `http://dx.doi.org/10.1257/000282803321946921`.

Lin DY, Psaty BM, Kronmal RA (1998) Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 948–963.

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2):187–204.