THE EFFECTS OF DEMOGRAPHIC MISMATCH IN AN ELITE PROFESSIONAL SCHOOL SETTING

Chris Birdsall
School of Public Service
Boise State University
Boise, ID 83725
chrisbirdsall@boisestate.edu

Seth Gershenson
(corresponding author)
American University and IZA
Washington, DC 20016
gershens@american.edu

Raymond Zuniga
Center for Public Administration and Policy
Virginia Tech University
Blacksburg, VA 24061
raymondz@vt.edu

Abstract

Ten years of administrative data from a diverse, private, top-100 law school are used to examine the ways in which female and nonwhite students benefit from exposure to demographically similar faculty in first-year, required law courses. Arguably, causal impacts of exposure to same-sex and same-race instructors on course-specific outcomes such as course grades are identified by leveraging quasi-random classroom assignments and a two-way (student and classroom) fixed effects strategy. Having an other-sex instructor reduces the likelihood of receiving a good grade (A or A–) by 1 percentage point (3 percent) and having an other-race instructor reduces the likelihood of receiving a good grade by 3 percentage points (10 percent). The effects of student–instructor demographic mismatch are particularly salient for nonwhite and female students. These results provide novel evidence of the pervasiveness of demographic-match effects and of the graduate school education production function.

https://doi.org/10.1162/edfp_a_00280

1. INTRODUCTION

A robust literature in the economics of education documents wide-ranging impacts of student–teacher demographic match on both students and teachers. In K–12 classrooms, assignment to an other-race or other-sex teacher has been shown to harm student achievement (Dee 2004, 2007) and increase student absences (Holt and Gershenson 2019).[1] Similarly, racial mismatch lowers teachers' perceptions of student behavior (Dee 2005) and their expectations for students' educational attainment (Gershenson, Holt, and Papageorge 2016). The impact of faculty representation has also been studied in the postsecondary context, particularly among first-year undergraduates (Bettinger and Long 2005; Hoffmann and Oreopoulos 2009; Carrell, Page, and West 2010; Fairlie, Hoffmann, and Oreopoulos 2014). These studies typically find modest effects of having a same-sex or same-race instructor on course grades, the likelihood of dropping a class, and choice of major. Lusher, Campbell, and Carrell (2015) show similar effects of having a same-race teaching assistant (recitation section leader) on course grades and office-hour and course attendance. Even in online environments, instructors, particularly white instructors, are more likely to respond to white male students' comments (Baker et al. 2018).

However, the extant literature has yet to investigate the extent of student–instructor demographic mismatch effects in the postgraduate or professional school setting.[2] The current study contributes to this gap in the literature by showing that the consequences of student–instructor demographic mismatch are just as pronounced in an elite, professional school setting as they are in K–12, community college, and first-year undergraduate classrooms. Doing so is important for at least three reasons.

First, this study enhances our understanding of the production of graduate degrees. Remarkably little is known about the nature of the law school education production function, or that for graduate school more generally.[3] This is troubling, as graduate students comprise a nontrivial segment of the U.S. postsecondary student population: about 15 percent of postsecondary students are graduate students and about 40 percent of outstanding student loan debt was accumulated to finance graduate degrees (Delisle 2014). Graduate degrees themselves facilitate entrance into many high-status and high-paying professions central to the modern economy. The legal profession is one prominent example: Nearly all states require that lawyers hold a Juris Doctor (JD) from an American Bar Association (ABA)-accredited law school, lawyers constitute about 1 percent of the U.S. labor force, and law firm revenues constitute about 1 percent of U.S. gross domestic product (Azmat and Ferrer 2017). The current study provides evidence on some of the educational inputs and environments that affect law school students' achievement, skill development, choice of specialization, and persistence.

1. Mismatch is not universally harmful, however, as Antecol, Eren, and Ozbeklik (2015) find that less-prepared female math teachers reduce female students' achievement but have no such effect on male students.
2. There is a litany of qualitative and anecdotal evidence of such demographic biases in legal education (Banks 1988; Guinier et al. 1994; Darling-Hammong and Holmquist 2015), but to our knowledge there is no credibly identified, quantitative evidence on the impact of law student–instructor demographic match on student outcomes.
3. Exceptions include recent natural experiments involving first-year law students at Stanford, who were randomly assigned to small classes (Ho and Kelman 2014) and at Minnesota, where students were randomly assigned to receive individualized feedback (Schwarcz and Farganis 2017). Neumark and Gardecki (1998) find that increasing female faculty members in economics departments improved time to completion and completion rates for female graduate students.

Second, the current study sheds light on the role that institutions play in perpetuating demographic wage, skill, and partnership gaps in the legal profession. For example, female lawyers earn lower salaries and are less likely to be promoted to partner than their male counterparts, even after conditioning on basic employee and firm characteristics (Wood, Corcoran, and Courant 1993; Dinovitzer, Reichman, and Sterling 2009; Azmat and Ferrer 2017).[4] Azmat and Ferrer (2017) show that performance gaps explain much of the previously unexplained sex gap in lawyers' earnings, though the exact sources of gaps in performance and specialization among practicing lawyers remain unclear. Law school environments and mentoring practices might contribute to this divergence in post–law school productivity, even when male and female students enter law school with similar skills (Bertrand 2011; Ho and Kelman 2014). We test this hypothesis by examining whether the demographic match between law students and instructors affects student outcomes. Doing so will inform law school policy and practice by identifying the malleable factors that influence the success of underrepresented graduate school students and our understanding of the importance that faculty play in the production of graduate education more generally. Indeed, law schools are representative of a broad class of professional graduate schools and programs from which professional service providers are recruited directly into the labor market (e.g., business, engineering; Oyer and Schaefer 2015).

Finally, there are social consequences of demographic gaps in the receipt of law degrees and in the career paths of law school graduates (Holder 2001). For example, the underrepresentation of racial and ethnic minorities in the U.S. judiciary likely contributes to documented demographic disparities in sentencing (Mustard 2001). Indeed, implicit association tests show that white judges often hold implicit (unconscious) biases against nonwhite defendants (Rachlinski and Johnson 2009). In the field, emotional shocks associated with the outcomes of football games have been shown to increase the sentences assigned by judges, particularly for black defendants (Eren and Mocan 2016). And regarding the demographic pay gaps discussed above, a lack of representation among law school faculty and/or how law school faculty interact with and mentor women and students of color can cause sorting into specializations and other behavioral responses that affect prestige, pay, and upward mobility. Ultimately, biases against women and people of color can produce self-fulfilling prophecies in which members of stereotyped groups ultimately conform to what were initially incorrect beliefs (Steele 1997; Loury 2009; Papageorge, Gershenson, and Kang 2016). Institutional factors, such as faculty composition, can therefore perpetuate the underrepresentation of certain demographic groups in the legal profession (Wilkins and Gulati 1996).

Specifically, we use rich administrative data from a top-100 law school in which first-year students are at least quasi-randomly assigned to course sections in conjunction with an array of arguably causal fixed-effects identification strategies to show that having a demographically mismatched first-year law instructor significantly reduces the probability of receiving a "good grade" (A/A–) in the course. Importantly, we find no such effects on the likelihood of dropping a course, which suggests the course-grade

4. This is consistent with "glass ceilings" and pay gaps in top management positions (Bertrand and Hallock 2001), as well as in the labor force more generally (Altonji and Blank 1999).

analyses are not biased by missing grades for courses that students dropped, and is likely due to the relatively rigid first-year requirements for progressing in the program.

Other-race effects tend to be larger in magnitude than other-sex effects, particularly among nonwhite and nonwhite female students, though both are statistically and economically significant. There are cumulative effects of exposure to demographically mismatched first-semester instructors on second-semester course grades in two-course sequences, suggesting that such effects persist, though we find no evidence of contemporaneous spillover effects of exposure to demographically matched faculty on performance in unrelated courses.[5] Classroom environments such as class size and class composition moderate the impact of student–instructor demographic mismatch in ways that hint at the mechanisms through which such effects operate. That we find such effects in an elite professional school setting suggests that the phenomena of implicit bias, stereotype threat, and role-model effects are broad, societal phenomena that permeate beyond relatively vulnerable populations of schoolchildren and community college students, and have implications for all social interactions, even those involving high-achieving individuals. Indeed, a recent field experiment finds that black men are more likely to select preventive services and talk to the doctor about their health problems when the doctor is of the same race (Alsan, Garrick, and Graziani 2018).

The paper proceeds as follows: Section 2 describes the administrative data and institutional details. Section 3 introduces the identification strategy. Section 4 presents the results. Section 5 concludes.

2. DATA AND INSTITUTIONAL DETAILS

This section describes the administrative data analyzed in the current study. We first describe the institutional context and the formation of the analytic sample, then we summarize the analytic sample.

Administrative Data

All analyses use longitudinal administrative data from a private, top-100 law school (LS) located in a major urban center. The LS enrolls approximately 1,000 students per year, on average, and employs approximately 200 full-time and part-time faculty. It is one of the most demographically and geographically diverse top-ranked law schools. The most recent *U.S. News & World Report* (U.S. News) rankings rank the LS in the top 100.[6] Demographically, LS ranks in the top 50 ABA-approved law schools for both racial/ethnic minority and female JD-student enrollment.[7] Thus, although LS is one of the more demographically diverse law schools in the United States, it is not an outlier and is comparable to other highly ranked, national law schools in this regard.

The main analytic sample is restricted to students' first-year required courses for three reasons. First, entering students take the same set of courses during their first two semesters of law school. Most courses are semester-specific, meaning that course A is usually taken in the fall semester and course B is taken in the spring semester.

5. See Appendix table A.1.
6. See http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-law-schools/law-rankings/page+4.
7. Rankings calculated as average percent enrollment from 2009 to 2013 using data obtained from the American Bar Association (www.americanbar.org/groups/legal_education/resources/statistics.html).

Second, the majority of first-year courses are assessed using a blind grading system.[8] This speaks to the mechanisms through which observed mismatch effects operate, as it precludes explicit grading biases of the type documented by Lavy (2008) from being the primary mechanism. Finally, at least in some years, student assignments to specific class sections, made by LS advisors and administrators, were quasi-random.[9] Similarly, the courses taken in each semester of the first year are randomly assigned by school administrators. About three to four sections of each course are offered in a semester in which the course is offered, with the exception of one writing course that has smaller class sizes and thus has about twenty-five sections per semester. We verify, and exploit, this random assignment in the empirical analysis.

The administrative data include detailed information on course-specific outcomes, such as grades, dropout behavior, and taking an *elective* course in the same concentration in the second year or beyond, as well as student-level outcomes such as persistence, graduation, and engagement with the LS's Law Journals, for every student who entered the JD program between fall 2000 and fall 2011. Additionally, we observe student demographic characteristics, such as sex, age, and race/ethnicity, as well as Law School Admission Test (LSAT) scores, undergraduate grade point average (GPA), and home ZIP Code.[10] We use home ZIP Codes to construct measures of distance from LS and to collect the median income and fraction of adults who have a college degree in each ZIP Code from the 2000 and 2010 U.S. censuses, which proxy for students' socioeconomic status. Administrative data on instructors include rank (e.g., tenure line, tenured, adjunct) and years at LS. Demographic information (i.e., race/ethnicity and sex) and rank of faculty members' JD-granting institutions were determined by reviewing public resumes, curriculum vitae, and Web sites.[11]

Sample and Summary Statistics

Our aim is to estimate the impact of student–instructor demographic match in first-year required courses. The primary unit of analysis is therefore the student-course level. There are ten required courses in the first year, which cover subjects such as procedure, constitutional law, and property law. The main analytic sample includes 36,560 student-course observations from more than 1,000 unique course sections.[12] Panel A of table 1 summarizes the student-course data, separately by students' race and sex. On average, white students have higher first-year course grades than nonwhite students. There is no appreciable sex gap in first-year course grades. Dropping first-year required courses is exceedingly rare, likely because they are required and students are generally forbidden from switching sections. White students and nonwhite students have

8. Unfortunately, the data do not identify which, if any, courses were subject to non-blind grading. Another complication is that students may challenge their grades in some circumstances, at which point the grading is no longer blind, and more advantaged students may feel more confident in challenging grades. Unfortunately, we do not observe which grades were challenged.
9. The assignment protocol changed about midway through the period of study, though both processes were arguably conditionally random. That said, we do not assume or rely on random assignment in the main analysis and instead rely on a quasi-experimental two-way fixed-effects identification strategy. However, a series of balance and Hausman-style tests suggests that assignments were, in fact, as good as random.
10. Unfortunately, LSAT and undergraduate GPA data are missing for a large, nonrandom subset. Accordingly, we rely on these data sparingly and do not report demographic group means for these variables.
11. The rank of instructors' JD programs comes from the usual U.S. News Rankings.
12. We report all sample sizes rounded to the nearest ten.

Table 1. Sample Statistics for First-Year Required Courses

	White		Nonwhite		Male		Female	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: Student-Course Level								
Course grade (0-4)	3.36	0.46	3.14	0.51	3.27	0.50	3.29	0.49
Take another course	0.80		0.81		0.80		0.81	
Dropped course	0.003		0.003		0.003		0.003	
Grade: A	0.40		0.23		0.33		0.34	
Grade: B	0.56		0.67		0.61		0.60	
Grade: C, D, F	0.04		0.10		0.07		0.06	
Other-sex instructor	0.51		0.53		0.42		0.58	
Other-race instructor	0.18		0.95		0.41		0.50	
Same instructor	0.93		0.92		0.93		0.92	
Observations	23,200		13,360		15,250		21,320	
Panel B: Student Level								
Age (first semester)	25.5	2.6	25.4	2.4	25.7	2.7	25.3	2.5
Female student	0.54		0.66		0.00		1.00	
Black student	0.00		0.21		0.05		0.10	
Latinx student	0.00		0.37		0.12		0.14	
Asian student	0.00		0.34		0.10		0.14	
White student	1.00		0.00		0.70		0.59	
Other race student	0.00		0.08		0.03		0.03	
Persist to second year	0.89		0.91		0.88		0.90	
Joined top law review at LS	0.14		0.06		0.12		0.11	
Graduated in 5 years	0.82		0.82		0.81		0.82	
Observations	2,890		1,680		1,910		2,660	
Panel C: Instructor Level								
Nonwhite instructor	0.00		1.00		0.15		0.20	
Black instructor	0.00		0.47		0.08		0.08	
Latinx instructor	0.00		0.23		0.02		0.06	
Asian instructor	0.00		0.30		0.05		0.06	
White instructor	1.00		0.00		0.85		0.80	
Female instructor	0.47		0.57		0.00		1.00	
Years of experience at LS	5.76	9.64	3.00	5.89	7.37	10.90	3.15	6.24
Has JD	0.95		0.97		0.94		0.96	
Rank of JD school	37.6	36.1	42.4	44.1	36.1	39.9	40.6	34.5
Has PhD	0.10		0.03		0.08		0.10	
Has master of laws degree	0.09		0.21		0.13		0.09	
Has bachelor of laws degree	0.04		0.00		0.05		0.01	
Observations	140		30		90		90	

Notes: The Dropped course descriptive statistics are based on slightly larger samples (23,300 for white students, 13,430 for nonwhite students, 15,320 for male students, and 21,400 for female students) because including dropped courses increases the number of student-course level observations for students who drop classes. There are no Other race instructors in the analytic sample. Same instructor in panel A is a binary variable indicating the student had the same instructor in the previous course. JD = Juris Doctor; SD = standard deviation; LSAT = Law School Admission Test; LS = Law School.

Table 2. Sample Statistics for First-Year Required Courses

Course Level Characteristics	Mean	SD	Course Name	Percent
Class size	41.60	34.00	Civil Procedure	5.76
Female student	0.59	0.14	Civil Procedure II	2.30
Age (first semester)	25.90	2.33	Constitutional Law	6.14
Black student	0.08	0.07	Contracts	6.14
Latinx student	0.13	0.10	Torts	5.47
Asian student	0.13	0.10	Legal Writing I	28.60
White student	0.63	0.13	Legal Writing II	31.29
Other student	0.03	0.05	Property	5.85
Female instructor	0.45	0.50	Property II	1.92
Black instructor	0.08	0.26	Criminal Law	6.53
Asian instructor	0.02	0.15	Observations	1,040
Latinx instructor	0.04	0.19		
White instructor	0.87	0.34		
More than one instructor race choice in term	0.74	0.44		
More than one instructor race choice in academic year	0.75	0.43		
More than one instructor sex choice in term	0.94	0.24		
More than one instructor sex choice in academic year	0.95	0.21		
Observations	1,040			

Notes: Classroom level demographics are presented as proportions. SD = standard deviation.

near-equal likelihoods of having an other-sex instructor, whereas female students are more likely than male students to have an other-sex instructor. Nonwhite students are much more likely to have an other-race instructor than are white students, as the majority of instructors are white.

Panel B of table 1 reports descriptive statistics at the student level. The average age of first-year JD students is about 25 years for all demographic groups. Whereas female students form a majority of both white and nonwhite students, the representation of female students is greater among nonwhite students than among white students. Among nonwhite students, 21 percent are black, 37 percent are Latinx, and 34 percent are Asian. Graduation rates are similar across demographic groups, which for students are coded as White, Black, Latinx, Asian, or Other.

Finally, panel C of table 1 reports descriptive statistics at the instructor level, for instructors who taught at least one first-year required course between 2000 and 2012. On average, white instructors have more experience at LS than nonwhite instructors, and male instructors have more experience than female instructors. About 47 percent of white instructors are female, while 57 percent of nonwhite instructors are female. Almost half of nonwhite faculty are black, 23 percent are Latinx, and 30 percent are Asian; unlike for students, there is no “Other race” category for instructors. In the empirical models, same-race is coded as an exact racial group match, as opposed to an indicator for both student and teacher being nonwhite. The average instructor attended a JD program ranked in the top 50 by U.S. News. White and male instructors attended slightly higher-ranked programs, on average, than did nonwhite and female instructors, respectively.

Table 2 reports descriptive statistics at the classroom (i.e., course-section) level. There are 1,040 unique first-year required course offerings in the analytic sample. The average class contained about 42 students, 59 percent of whom were female. The majority (87 percent) of courses were taught by white faculty, 8 percent were taught by black instructors, 4 percent by Latinx instructors, and 2 percent by Asian instructors. Table 2 also reports the frequency of the ten courses that constitute the analytic sample. Some courses appear less often either because they had smaller average class sizes, were merged into a single course, or ceased to be required between 2000 and 2012. Still, outliers here are the legal writing classes, which are overrepresented because of their smaller class size. Subject-specific summary statistics are provided in table A.2, which shows that the average writing class has fifteen students whereas the other classes average fifty to eighty students. Because of the notably smaller class size and the different structure of writing classes, as a sensitivity analysis, we reestimate the baseline model on a sample that excludes the writing classes in table A.3 and confirm the main results are not driven by student outcomes in these unique classes.

3. IDENTIFICATION STRATEGY

This section describes the main identification strategy used to estimate the causal effects of student–instructor demographic match on course-specific outcomes. We first introduce the preferred two-way fixed effects (FE) specification. We then discuss the key identifying assumptions and present a test of the "endogenous sorting" threat to identification. Finally, we describe a three-way FE specification used to identify the effect of mismatch in the first course of two-course sequences on performance in the second course.

Baseline Model

Our primary interest is in how student–instructor demographic match affects outcomes (y) at the student-course level. Specifically, we are interested in δ in the linear regression model:

$$y_{ijcst} = \beta_0 + \beta_1 X_i + \beta_2 W_j + \beta_3 Z_{cst} + \delta Other_{ij} + \epsilon_{ijcst}, \tag{1}$$

where X, W, and Z are vectors of observed student (i), instructor (j), and course-section (cs) characteristics, respectively; t indexes semesters; *Other* is a vector of variables that measure the degree of demographic similarity between student and instructor; and ϵ represents the unobserved determinants of y.[13] We operationalize *Other* in various ways, such as a set of four mutually exclusive race-by-sex indicators (i.e., same race and other sex, same sex and other race, same race and same sex, other race and other sex) and simpler definitions that include binary indicators for other sex and/or other race. However, in all specifications, race matches are coded as specific matches such as black-black, Latinx-Latinx, and so forth, as opposed to "minority-minority."

Given that course-section assignments are allegedly conditionally (on X) random, ordinary least squares (OLS) estimates of equation 1 might well be unbiased and

13. We consider models that allow the effect of *Other* to vary by subject, but find no systematic evidence of differential effects by subject, perhaps because we are under-powered to do so. Accordingly, we report estimates of the average effect of student-instructor demographic match that are averages across subjects.

have a causal interpretation. However, if the quasi-random assignment rule is imperfectly followed, these estimates might be biased. For example, unobserved student characteristics might jointly predict outcomes and assignment to an other-race teacher. Similarly, equation 1 fails to control for unobserved instructor attributes, such as grading policies or teaching style. Accordingly, we follow Fairlie, Hoffmann, and Oreopoulos (2014) and augment equation 1 to condition on both student and classroom FE, which yields our preferred specification:

$$y_{ik} = \theta_i + \omega_k + \delta Other_{ik} + \epsilon_{ik}. \tag{2}$$

Several aspects of equation 2 merit attention. First, the vectors X, W, and Z fall out of the model because they are colinear with the FE. Second, we collapse the subscripts $jcst$ into a single k subscript because identification now comes from within-classroom variation in *Other* and classrooms are instructor-, course-, section-, and semester-specific: the classroom FE (ω) subsumes instructor, course, semester, and year FE. Specifically, the classroom FE uniquely identify each course section taught in a given semester and thus control for the course's location (classroom) quality, meeting day(s) and time, class size, and class composition. Thus the classroom FE also ensure that identification comes from students who experienced the same lectures, assignments, and grading practices. Third, equation 2 is only identified for outcomes that vary within students across courses, such as course grades, due to the student FE (θ). Finally, there is a possible sample selection issue for the analyses of course grades, since grades are only observed for students who complete the course, and it is possible that student–instructor demographic mismatch affects the likelihood that students complete the course. This turns out to be a practically unimportant concern, as dropping courses is quite rare (occurs in only 0.6 percent of cases) and we find no evidence that demographic mismatch affects course dropouts.[14] We estimate equation 2 using the estimation routine proposed by Correia (2016) and compute two-way cluster-robust standard errors, which allows for correlation both within instructors across semesters, and within students across courses (Cameron, Gelbach, and Miller 2012).[15]

Sorting Test

Although the two-way FE in equation 2 address many threats to validity, one potential threat remains: differential sorting by student race or sex (Fairlie, Hoffmann and Oreopoulos 2014). For example, the student FE control for scenarios in which high-ability students sort into female-taught courses, but does not adequately control for sex-specific sorting processes in which high-ability female students sort into female-taught courses and high-ability male students sort into male-taught courses. To discern the extent to which differential sorting on unobservables occurs, we follow Fairlie, Hoffmann, and Oreopoulos (2014) in implementing a formal test for differential sorting on observables. The test relies on the intuition of difference-in-differences estimators

14. This is perhaps unsurprising, as we are investigating required first-year courses.
15. Clustering along only one dimension and/or at lower levels yields nearly identical inferences and slightly smaller standard errors for the main course-grade results. Accordingly, we report the more conservative two-way clustered standard errors in the main text. This is motivated by the guidance in Angrist and Pischke (2009), which suggests clustering at the highest level.

and the bounding procedure of Altonji, Elder, and Taber (2005). It is best illustrated via an example. Suppose we want to test for differential sorting by sex. We would first compute the mean of observed student characteristic L (e.g., LSAT score) in classroom k for each sex g: $\bar{L}_k^g$. Then estimate the linear regression

$$\bar{L}_k^g = \gamma_0 + \gamma_1 Female_k + \gamma_2 1\left\{Female = g\right\} + \gamma_3 Female_k \times 1\left\{Female = g\right\}, \qquad (3)$$

where *Female* is a binary indicator equal to 1 if the section-k teacher is female, and zero otherwise; $1\{\cdot\}$ is the indicator function; and γ_3 is the parameter of interest. Specifically, γ_3 represents "the difference-in-differences estimate" of the average difference in observed characteristics between female and male students in female- and male-taught courses. If γ_3 is significantly different from zero, there are differences by student sex in sorting into courses on observables that systematically vary with the sex of the instructor. Alternatively, if the OLS estimate of γ_3 in equation 3 is statistically indistinguishable from zero, there is no evidence of differential sorting on observables, and thus differential sorting on unobservables in a way that would bias the two-way FE estimates of equation 2 is unlikely.

Cross-Semester Effects in Two-Course Sequences

Finally, we consider whether exposure to an other-race or other-sex instructor in the first course of a two-course sequence affects performance in the second course. Naturally, this analysis can only be conducted for the subset of first-year courses that are part of a required two-course sequence.[16] While this question can be addressed using the baseline two-way FE model given in equation 2, it is also possible to further increase the estimates' validity by augmenting equation 2 to condition on a second-semester course FE (φ).[17] Specifically, we estimate three-way FE models of the form

$$y_{is2} = \theta_i + \omega_{s1}^{(i)} + \varphi_{s2}^{(i)} + \delta Other_{is1} + \epsilon_{is}, \qquad (4)$$

where 1 and 2 index semesters and s indexes subjects. Estimates of δ in equation 4 are robust to excluding the second-semester course FE, which is reassuring because it suggests the demographic background of the first-semester instructor does not affect second-semester classroom assignments. Estimates of equation 4 report standard errors clustered along three dimensions: student, semester 1 instructor, and semester 2 instructor.

4. RESULTS

This section presents the empirical results. We first present estimates of the sorting test characterized by equation 3. We then present the baseline two-way FE estimates, followed by tests for heterogeneous impacts of student–instructor demographic mismatch.

16. There are three such sequences: Civil Procedure I & II, Legal Writing I & II, and Property Law I & II.
17. This is similar to the identification strategy used by Figlio, Schapiro, and Soter (2015) to identify the impact of adjunct instructors, though in that case the first-semester course FE were not included because adjunct status varies only at the classroom level.

Table 3. Sorting Test Estimates

Outcome	LSAT	UGPA	Median Income (ZIP)	% Adult w/BA (ZIP)	In/Nearby State	Student Age
			Panel A: Sorting by Race			
Nonwhite instructor	0.078 (0.248)	−0.025 (0.058)	−3,874.264*** (1,268.960)	4.051*** (1.361)	−0.019 (0.015)	0.004 (0.068)
Nonwhite student	−4.848*** (0.107)	−0.149*** (0.033)	−32,14.724*** (577.044)	−2.406*** (0.288)	−0.027*** (0.010)	−0.181*** (0.049)
Nonwhite instructor * Nonwhite student	0.376 (0.263)	−0.067 (0.104)	49.169 (1,210.234)	−1.867** (0.813)	0.025 (0.025)	0.104 (0.104)
Constant	160.790*** (0.091)	3.460*** (0.023)	78,674.542*** (529.096)	37.980*** (0.455)	0.530*** (0.006)	25.560*** (0.033)
Observations	1,820	490	2,010	2,010	2,020	2,020
			Panel B: Sorting by Sex			
Female instructor	0.255 (0.167)	−0.059 (0.060)	−623.512 (1,124.828)	1.168 (0.814)	0.008 (0.013)	−0.093 (0.066)
Female student	−1.048*** (0.115)	0.132*** (0.046)	−1,477.353** (727.315)	0.495 (0.317)	0.040*** (0.012)	−0.369*** (0.057)
Female instructor * Female student	−0.036 (0.181)	0.013 (0.073)	−1,359.981 (1,149.638)	−0.809 (0.513)	−0.010 (0.017)	0.001 (0.081)
Constant	159.560*** (0.115)	3.337*** (0.038)	78,731.551*** (760.778)	36.831*** (0.550)	0.492*** (0.009)	25.738*** (0.046)
Observations	1,860	480	2,090	2,090	2,090	2,090

Notes: Each column represents tests for sorting on a different student background characteristic. In/Nearby State is a binary variable indicating the student's home address is within the same state as the institution or a bordering state. Standard errors in parentheses are clustered by course. LSAT = Law School Admission Test; UGPA = undergraduate grade point average; BA = Bachelor of Arts degree.

$^{**}p < 0.05$; $^{***}p < 0.01$.

Sorting Test Estimates

Table 3 presents estimates of the sorting test characterized by equation 3.[18] Panel A reports estimates for differential sorting by race, comparing the average characteristics of white and nonwhite students. Panel B does the same for differential sorting by sex, comparing the average characteristics of male and female students.

We perform the sorting test for six outcomes: LSAT score, undergraduate GPA, median income in student's home ZIP Code, percent of population with college degree in student's home ZIP Code, a binary indicator equal to one if the student came from the surrounding tristate area, and student age.[19] The LSAT and undergraduate GPA variables likely measure a combination of students' cognitive and noncognitive skills (Heckman and Kautz 2012). The ZIP Code information proxies for the student's socioeconomic background, which is an important predictor of undergraduate college success (Bailey and Dynarski 2011). The "In/Nearby State" indicator provides a crude measure of students' distances from home, which is known to predict undergraduate enrollments (Alm and Winters 2009; Cooke and Boyle 2011).

Only one of the twelve estimates of γ_3 in table 3 is statistically significant, which suggests little differential sorting on observables by sex or race. Given the multiple hypotheses tested, it is possible that the significant result in panel A is spurious: Indeed,

18. The sorting test estimates remain essentially unchanged when course name and year FE are added to the regression.
19. Data on LSAT and undergraduate GPA are missing for many students, so these results should be interpreted with caution.

Table 4. Impact of Demographic Mismatch on First-Year Required Course Outcomes

	Continuous Grade (1)	A Grade (2)	A/A- Grade (3)	C, D, F Grade (4)	Take Another (5)	Dropped Course (6)
Other-sex	-0.016^{**}	-0.008^{**}	-0.013^{**}	0.001	0.016	-0.000
	(0.007)	(0.004)	(0.007)	(0.003)	(0.010)	(0.000)
Other-race	-0.037^{**}	-0.015^{**}	-0.028^{***}	0.009	-0.000	0.001
	(0.016)	(0.007)	(0.011)	(0.007)	(0.008)	(0.001)
Differences in coefficients (*P*)	0.214	0.455	0.186	0.256	0.197	0.867
Observations	36,560	36,560	36,560	36,560	18,620	36,730
Course fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents a different model specification. The outcomes are measured as follows: Continuous Grade measures a student's received grade on a 0–4 scale (F–A); A Grade is a binary indicator for whether a student received an A grade; A or A– Grade is a binary indicator for whether a student received an A or A– grade; C, D, or F Grade is a binary indicator for whether a student received a C, D, or F grade; Take Another is a binary indicator for whether a student takes a subsequent elective course in the same field after his first year; and Dropped Course is a binary indicator for whether a student drops the course before the end of the semester. Column 5 has fewer observations because not all required courses correspond to elective course subjects. Difference in coefficients compares the other-sex effect to the other-race effect. Standard errors in parentheses are clustered by student and instructor.

$^{**}p < 0.05$; $^{***}p < 0.01$.

it loses its statistical significance after adjusting for multiple comparisons (Schochet 2009). Moreover, this result suggests sorting in the "wrong" direction, in the sense that nonwhite students assigned to nonwhite faculty are from *lower* socioeconomic backgrounds, which would bias *against* finding a positive impact of demographic match on student outcomes. In sum, the general lack of sorting on observables observed in table 3 suggests differential sorting on unobservables is unlikely to bias two-way FE estimates of equation 2. The lack of endogenous sorting is unsurprising given LS's claims that students were at least quasi-randomly assigned to classrooms. We further test this claim below by examining the sensitivity of the baseline estimates to controlling for student FEs.

Main Results

Table 4 reports two-way FE estimates of equation 2 using a simple definition of *Other*: binary indicators for whether or not the student had an other-sex and other-race instructor. The first four columns of table 4 use different definitions of the course grade as the outcome. Column 1 uses a continuous measure of the course grade, which is measured on a 0 to 4 scale. Having an other-sex and other-race teacher significantly reduced the student's course grade by 0.02 and 0.04, respectively, though these estimates are not significantly different from one another. These effects represent small (≈1 percent) changes from the average course grade of 3.36. Although small in magnitude, recall that these are course-specific effects that might add up to nontrivial differences in cumulative GPA that preclude underrepresented students from prestigious internships after the first year or alter class rankings in ways that affect initial job placements and starting salaries.

Additionally, these small effects could be due to the effect of student–instructor demographic mismatch operating on particular margins of the course-grade distribution. Accordingly, in columns 2 and 3 we estimate linear probability models in which the outcomes are binary indicators for "good" grades, defining a good grade as an A or an A or A−, respectively. Consistent with the results in column 1, columns 2 and 3 show significant, negative effects of demographic mismatch on the probability that students receive a good grade regardless of how good grade is coded. That the effect on having an A is smaller than that on the more inclusive definition of good grade suggests that demographic match effects operate on both the A/A− and A−/B+ margins. Column 4 shows that there is no effect of student–instructor demographic mismatch on the likelihood of receiving a "bad grade" (less than B−). These results show that demographic mismatch affects grades, primarily by affecting the likelihood of receiving top grades (A or A−). Racial mismatch effects tend to be larger than sex mismatch effects, but these differences are not statistically significant. These effects are arguably economically significant, as the other-race effect of 0.03 constitutes 9 percent of the sample average "good-grade" rate and might add up to have a nontrivial effect on cumulative GPA. The remaining columns of table 4 show that there are neither effects of mismatch on the likelihood that the student takes an elective course in the subject in the second year or beyond nor on the likelihood the student drops the course.[20] The latter null result is important, as it suggests that the sample selection inherent in the course-grade analyses is negligible.

Because an important contribution of the current paper is the identification of causal effects of same-race and same-sex instructors on course outcomes, we now leverage the alleged quasi-random assignment of students to course sections to cross-validate the baseline two-way FE estimates. The intuition of the Hausman test (Hausman 1978) suggests that if student assignments to course sections were conditionally random, then the estimates should be robust to the inclusion of student FE, as the claim is that students are randomly assigned to course sections (classrooms). Similar intuition motivates the common practice of verifying that experimental estimates of causal effects are robust to conditioning on predetermined characteristics in treatment-effect regressions (Angrist and Pischke 2009).

In table 5, we show that the baseline estimates are quite robust to the inclusion of student and/or classroom FE. This lends additional support to a causal interpretation of the baseline estimates and to the claim that students were randomly assigned to first-year courses.[21] Specifically, column 2 shows that the "naive OLS" mismatch effects in column 1 are robust to controlling for observed student characteristics such as LSAT score, undergraduate GPA, socioeconomic status, and distance to the law school. This suggests students were, in fact, randomly assigned to course sections.[22] Columns 3 and 4 compare student random effects and student FE estimators, in the spirit of the original Hausman test, and again find the point estimates are robust to controlling for unobserved student heterogeneity. Finally, columns 5 and 6 show the results are robust to conditioning on classroom FE, which means that the mismatch effects are not driven by differential teacher or classroom characteristics, such as teaching or grading

20. The sample size for subsequent course taking is smaller because there are not subsequent courses in all required first-year courses.

21. Because table 5 shows the pooled OLS estimates can be given a causal interpretation, we can also estimate pooled logit models to verify that the baseline linear model provides reasonable approximations of the partial effects of interest. Accordingly, table A.4 reports logit average partial effects (APE) that are comparable to the linear estimates reported in table 4. The logit APE are quite similar to the linear coefficient estimates, suggesting that the main results are robust to the functional form choice.

22. We include missing-data dummies to allow use of the full sample.

Table 5. Impact of Demographic Mismatch on First-Year Required Course Outcomes

	A/A– Grade					
	(1)	(2)	(3)	(4)	(5)	(6)
Other-sex	−0.012**	−0.011**	−0.012***	−0.013*	−0.012*	−0.013**
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
Other-race	−0.024**	−0.027***	−0.031***	−0.033***	−0.023*	−0.028***
	(0.009)	(0.009)	(0.009)	(0.012)	(0.013)	(0.011)
Female instructor	0.041***	0.029***	0.031***	0.031*		
	(0.005)	(0.005)	(0.005)	(0.018)		
Nonwhite instructor	0.009	0.021***	0.017**	0.015		
	(0.008)	(0.008)	(0.008)	(0.016)		
Nonwhite student	−0.152***	−0.088***	−0.085***		−0.088***	
	(0.011)	(0.011)	(0.011)		(0.013)	
Female student	0.038***	0.044***	0.042***		0.042***	
	(0.008)	(0.008)	(0.008)		(0.009)	
Observations	36,560	36,560	36,560	36,560	36,560	36,560
Cohort class dummies	Yes	Yes	Yes	No	No	No
Course subject type dummies	No	Yes	Yes	Yes	No	No
Student characteristics	No	Yes	Yes	No	Yes	No
Course FE	No	No	No	No	Yes	Yes
Student RE	No	No	Yes	No	No	No
Student FE	No	No	No	Yes	No	Yes

Notes: Each column represents a different model specification. The outcome A/A– Grade is a binary indicator for whether a student received an A or A– grade. Course Subject Types are Civil Procedure, Constitutional, Contracts, Criminal, Legal Writing, Property, and Torts. Student Characteristics are age, Law School Admission Test, undergraduate grade point average, median income, and percent of adults with Bachelor of Arts degree in home ZIP Code, in/nearby state, and missing data indicators for each. Standard errors in parentheses are clustered by student in columns 1–4 and by student and instructor in columns 5 and 6. FE = fixed effects; RE = random effects.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

practices, or the physical location or condition of the classroom. Column 6 replicates the baseline two-way FE estimates of equation 2.

Heterogeneity

Having established arguably causal impacts of student–instructor demographic mismatch on course grades, we now test for possible heterogeneity in such effects. First, we investigate possible heterogeneity by student background and by the precise type of demographic mismatch, because understanding the determinants of success for students from historically underrepresented groups is of paramount policy interest.[23] Second, we investigate whether these demographic mismatch effects are moderated by the demographic composition or the size of specific classrooms, as classroom environments might moderate the impact of mismatch (Inzlicht and Ben-Zeev 2000; Ho and Kelman 2014).[24]

23. We find no evidence of heterogeneity along other observable student dimensions, such as students' ability (LSAT score), age, home region, and ZIP Code socioeconomic status. Nor do we find evidence of heterogeneity by observable instructor characteristics, such as experience, rank of JD program, or faculty rank (i.e., adjunct, teaching-track, tenure-line, tenured). These null results are not reported in tabular form in the interest of brevity.
24. A relevant question here is whether class characteristics vary by subject. Table A.2 reports mean course characteristics by subject. The primary outlier is leal writing, which has significantly smaller classes than the other

Table 6. Impact of Demographic Mismatch on First-Year Required Course Outcomes

	All Students (1)	Female Students (2)	Nonwhite Students (3)	Nonwhite Female Students (4)
		Panel A		
Other-sex	−0.013* (0.007)	−0.035* (0.019)	−0.017* (0.009)	−0.041** (0.019)
Other-race	−0.033*** (0.012)	−0.031** (0.013)	−0.046* (0.024)	−0.052* (0.028)
Female faculty	0.031* (0.018)		0.033* (0.017)	
Nonwhite faculty	0.015 (0.016)	0.028 (0.018)	0.011 (0.016)	0.020 (0.019)
		Panel B		
Same race, Mismatch sex (1)	−0.013 (0.008)	−0.038 (0.023)	−0.093*** (0.032)	−0.094* (0.049)
Mismatch race, Same sex (2)	−0.033** (0.013)	−0.034** (0.017)	−0.083** (0.032)	−0.078* (0.043)
Mismatch race, Mismatch sex (3)	−0.046*** (0.015)	−0.066*** (0.024)	−0.096*** (0.033)	−0.116** (0.045)
Female faculty	0.031* (0.018)		0.033* (0.017)	
Nonwhite faculty	0.015 (0.016)	0.028 (0.018)	0.012 (0.016)	0.021 (0.019)
Difference in Coefficients (*P*)				
1 = 2	0.111	0.876	0.666	0.599
1 = 3	0.018**	0.048**	0.931	0.473
2 = 3	0.155	0.083*	0.176	0.052*
Observations	36,560	21,320	13,360	8,790
Course fixed effects	No	No	No	No
Student fixed effects	Yes	Yes	Yes	Yes

Notes: Each column in each panel represents a different model specification. The outcome A or A− Grade is a binary indicator for whether a student received an A or A− grade. The omitted category in panel B is Same Race, Same Sex. Estimates are not shown for course subject dummies. In some samples, estimates are not shown for certain instructor effects because they are perfectly colinear with other-sex or other-race parameters. Standard errors in parentheses are clustered by student and instructor.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Panel A of table 6 estimates the baseline student-FE specification, sans classroom FE, to enable identification of mismatch effects for specific demographic subgroups of the sample. We feel comfortable making this trade-off because table 5 shows that the full-sample estimates are robust to omitting the classroom FE. Column 1 of table 6 repeats the estimates shown in column 4 of table 5 to facilitate comparisons. Columns 2 and 3 estimate this specification separately for female and nonwhite students, respectively. We might expect these groups to be particularly affected by faculty representation, given the general overrepresentation of white men in the legal profession. These models yield two key findings. First, as expected, the other-sex effect is driven by female students' grades and the other-race effect is driven by nonwhite students' grades. Specifically, for female students, the likelihood of receiving an A/A– increases by 3.5 percentage points (10 percent) when taught by a female instructor, compared with an

subjects. However, we find no evidence of systematic differences between legal writing and other subjects in tests for subject heterogeneity.

overall sex-match effect of 1.3 percentage points (3 percent) in the full sample. Similarly, the race-match advantage for nonwhite students is 4.6 percentage points (20 percent), compared with 3.3 percentage points (9 percent) in the full sample.[25] Second, the other-sex effect is similar for both white and nonwhite students, and the other-race effect is similar for both male and female students. This lack of heterogeneity is also interesting. Finally, column 4 shows that the harmful effects of demographic mismatch are most pronounced for nonwhite female students, although these differences are not significantly different from the overall effects of sex representation for women or of racial representation for nonwhite students.

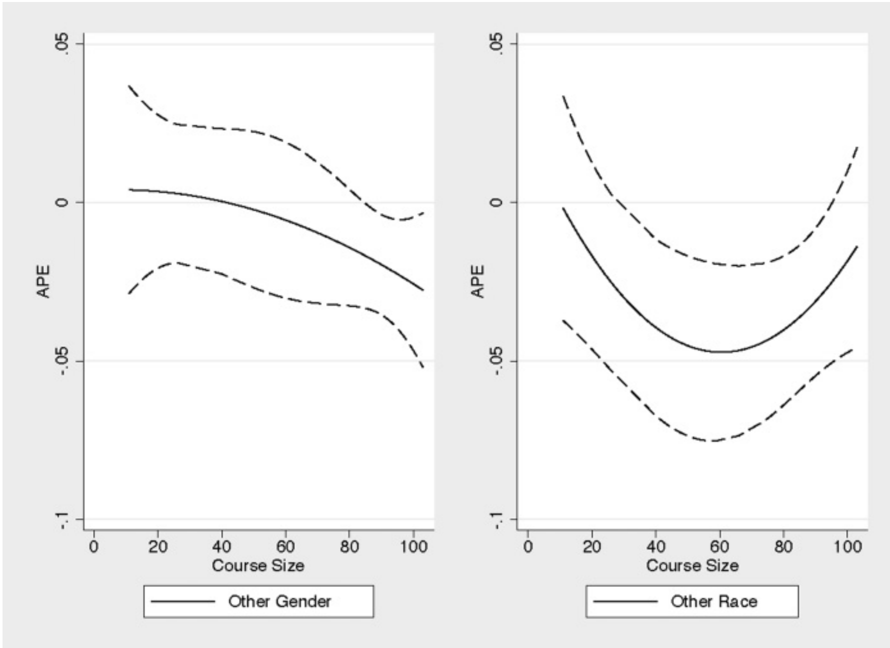
Panel B of table 6 generalizes the models estimated in panel A by allowing for multiplicative effects of having both an other-race *and* other-sex instructor. Here, *Other* is specified as a set of four mutually exclusive categorical indicators, with same-sex *and* same-race serving as the omitted reference category. Column 1 shows that overall, relative to students whose instructors are of the same race and sex, any type of demographic mismatch leads to a lower likelihood of receiving a good grade. However, having a different-race *and* different-sex instructor is significantly worse than instances in which demographic mismatch occurs along only one dimension. Column 2 shows that this is true for the female subsample as well, which is consistent with the results presented in panel A, and shows the effect of having a different-race *and* different-sex instructor is more pronounced for female students than for male students. However, column 3 shows that nonwhite students are similarly harmed by any type of student–instructor demographic mismatch. Finally, and again consistent with the results presented in panel A, column 4 of panel B shows that nonwhite female students benefit the most from intersectional demographic representation (i.e., having both a same-race and same-sex instructor).

Next, we test for heterogeneity in the impact of student–instructor demographic mismatch by classroom characteristics, such as class size and class composition. Whether larger classrooms magnify or dampen the mismatch effects documented previously is theoretically ambiguous, as smaller classrooms could either shine a spotlight on implicit biases or facilitate relationships that supersede stereotypes. We also allow the effect of mismatch to vary with the demographic composition of classrooms, as the impact of an other-race or other-sex instructor might be more pronounced in less diverse settings in which female or nonwhite students feel isolated. Given the exploratory nature of this analysis, we model the heterogeneity using quadratics in class size and percent female (nonwhite). The quadratics are at least marginally jointly significant in both cases.[26]

Appendix table A.5 reports the coefficient estimates for these models, though for ease of interpretation we plot the marginal effects as functions of class size and percent female (nonwhite). Figure 1 plots the marginal effects (and corresponding 95 percent confidence intervals) on the probability of receiving an A/A– of having an other-race or other-sex instructor as a function of class size for the range of class sizes observed in the analytic sample. Interestingly, there is essentially no effect of mismatch in the

25. The nonwhite effect itself is almost entirely driven by black students' responses to black instructors, which is consistent with Fairlie, Hoffmann, and Oreopoulos (2014), although we focus on the aggregate nonwhite effect because the race-specific analysis is underpowered due to the small share of Asian and Latinx instructors.

26. Cubic and nonparametric specifications yield qualitatively similar results.



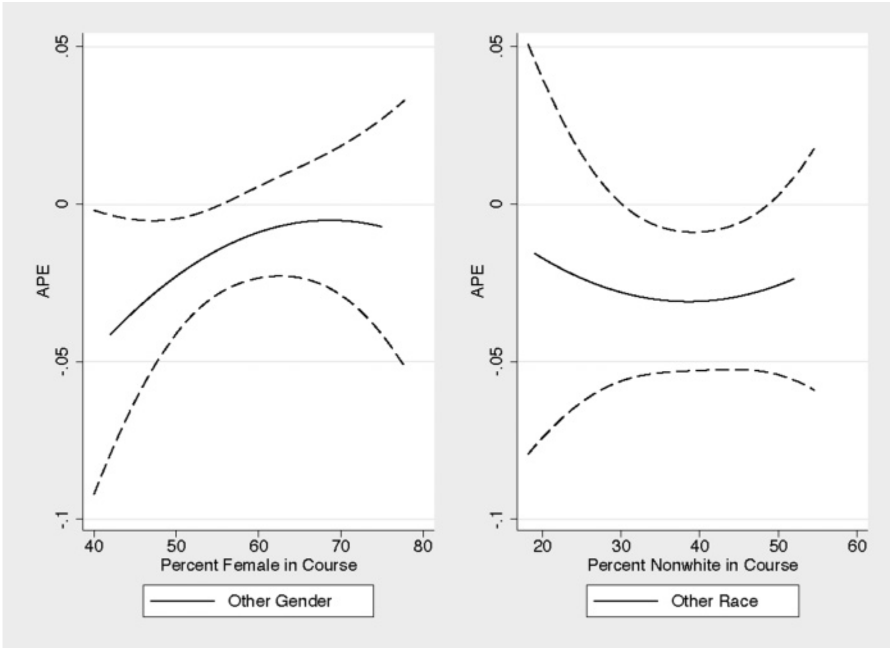
Notes: Good Grade is defined as an A or A–. Each graph represents a different model specification.

Figure 1. Average Partial Effects (APE) of Student–Instructor Mismatch on the Probability of Receiving a Good Grade as a Function of Class Size

smallest classes. The other-sex effect monotonically increases in magnitude with class size, though at a relatively slow pace, and only becomes statistically significant in relatively large classes. The other-race effect, meanwhile, exhibits a U-shaped pattern. The deleterious effect of having an other-race instructor is largest in classrooms of about sixty students. One possible interpretation of this pattern is that the personal connections and relative anonymity in very small and very large classes, respectively, mitigate the harm associated with having an other-race instructor.

Similarly, figure 2 plots the marginal effects on the probability of receiving an A/A– of having an other-race (other-sex) instructor as a function of the fraction of the classroom that is nonwhite (female). The other-race effect is fairly constant at about −0.03 or −0.04, regardless of the proportion of nonwhite students in the class. However, the other-sex effect is less linear. Intuitively, it is most pronounced when female students make up less than half the class. The other-sex effect approaches zero when 60 to 70 percent of the class is female. This is suggestive of stereotype threat,[27] whereby females disengage with law school when they perceive themselves as outsiders, and consistent with experimental evidence that shows the sex ratio of a classroom affects a female student's test performance but not a male student's (Inzlicht and Ben-Zeev 2000).

27. *Stereotype threat* occurs when the presence of a white or male instructor triggers historically underrepresented students' recognition of their outgroup status, which in turn causes emotional responses that hinder their academic performance and ultimately lessens their engagement with school (Steele 1997).



Notes: Good Grade is defined as an A or A–. Each graph represents a different model specification.

Figure 2. Average Partial Effects (APE) of Student–Instructor Mismatch on the Probability of Receiving a Good Grade as a Function of Class Composition

Cross Semester Effects

Finally, table 7 reports estimates of equation 4, which show the impact on performance in the second course of having an other-sex or other-race instructor in the first course of a required two-course sequence. These are all required first-year courses that take place in the fall and spring semesters of the first year. The model is estimated for three outcomes: course grade, a binary indicator for "good grade" (i.e., A or A−), and a binary indicator for "bad grade" (i.e., < B−). These models can only be estimated for the subset of courses that are part of a two-course sequence.

Panel A of table 7 shows results that are broadly similar to the baseline two-way FE estimates reported in table 4: There are negative effects of student–instructor mismatch in the first course on grades, and on the probability of receiving a good grade in the second course. Once again, the other-race effect is about twice as large as the other-sex effect, though here only the other-sex effect is statistically significant at traditional confidence levels; this is due to the larger standard errors associated with the smaller sample of courses that constitute two-course sequences and the fact that there are more female faculty than nonwhite faculty in two-course sequences. That these estimates are qualitatively similar to the baseline estimates reported in table 4 lends further credence to a causal interpretation of the relationship between student–instructor demographic mismatch and course grades. Moreover, this similarity sheds some light on the mechanisms at work, as the cross-semester effects documented in table 7 suggest increased subject-specific learning that persists into the subsequent semester.

Table 7. Cross-Semester Effects of Demographic Mismatch in Two-Course Sequences

	Continuous Grade (1)	A/A– Grade (2)	C, D, F Grade (3)
Panel A: Cross-Semester Effects			
Other-sex (OS)	−0.032** (0.012)	−0.034** (0.016)	0.004 (0.008)
Other-race (OR)	−0.077 (0.050)	−0.072 (0.048)	0.016 (0.021)
Course 1 FE	Yes	Yes	Yes
Course 2 FE	Yes	Yes	Yes
Student FE	Yes	Yes	Yes
Panel B: Cross-Semester Effects – Same Instructor			
Other-sex	−0.074 (0.063)	−0.120** (0.060)	0.005 (0.031)
Other-race	−0.206*** (0.063)	−0.046 (0.070)	0.073* (0.043)
OS * Same instructor	0.047 (0.065)	0.095* (0.057)	−0.001 (0.030)
OR * Same instructor	0.145*** (0.054)	−0.029 (0.060)	−0.064 (0.039)
Course 1 Fixed Effects	Yes	Yes	Yes
Course 2 Fixed Effects	Yes	Yes	Yes
Student Fixed Effects	Yes	Yes	Yes

Notes: $N = 4{,}340$. Each column represents a different model specification. The outcomes are measured as follows: Continuous Grade measures a student's received grade on a 0–4 scale (F–A). A/A– Grade is a binary indicator for whether a student received an A or A– grade. C, D, F Grade is a binary indicator for whether a student received a C, D, or F grade. Same Instructor indicates the student had the same instructor for both Courses 1 and 2 in two-course sequences. Estimates are not shown for Same Instructor in even numbered columns because it is perfectly correlated with the Courses 1 and 2 fixed effects. Standard errors in parentheses are clustered three ways: by student, first instructor, and second instructor.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Panel B of table 7 augments equation 4 to allow the cross-semester demographic match effects to vary by student–instructor familiarity. Specifically, we interact the demographic mismatch indicators with indicators for whether the student had the same instructor in both semesters.[28] This idea is motivated by recent research by Hill and Jones (2018) in the primary school setting, who show that students, particularly nonwhite students, benefit from having the same classroom teacher in consecutive years. Intuitively, having the same instructor in consecutive semesters would foster a stronger relationship and better understanding of expectations and learning styles, which in turn might mitigate the harmful effects of demographic mismatch. Indeed, this is precisely what the interaction terms in panel B of table 7 show: The other-sex mismatch effects on good grades are significantly smaller, and indistinguishable from zero, for students who had the same instructor in both courses of the two-course sequence.

28. The familiarity indicator itself is subsumed by the "Course-2 FE."

5. CONCLUSION

We use rich student–instructor matched administrative data from a large, private, top-100 law school to provide novel evidence on the causal relationship between student–instructor demographic match and student outcomes in the law school context. Two-way student and course fixed-effects models provide arguably causal estimates of the impact of such mismatch on short-run (course-specific) outcomes, such as course grades. Sorting and balance tests provide no evidence of endogenous sorting on observables into classrooms, which is consistent with at least quasi-random assignment of students to course sections in the law school, buttressing a causal interpretation of these results.

The baseline estimates suggest that having an other-race or other-sex instructor in a first-year required course significantly reduces the likelihood of earning a good grade (i.e., A or A−) in the course. Specifically, having an other-sex instructor reduces the likelihood of receiving a good grade by 1 percentage point (3 percent) and having an other-race instructor reduces the likelihood of receiving a good grade by 3 percentage points (10 percent). The most comparable estimate in the extant literature comes from Fairlie, Hoffmann, and Oreopoulos (2014), who find that having a same-race community college instructor increases the probability of having a good grade (≥ B) in first-year undergraduate courses by about 3 percentage points (5 percent).[29] That we document similarly sized effects in first-year courses at a top-100 law school suggests that even high-achieving college graduates' graduate and professional school outcomes are influenced by the demographic representation of their instructors.

This result has the potential to contribute to pay gaps, as Oyer and Schaefer (2019) provide descriptive evidence of a wage-class rank gradient in law schools outside the top ten.[30] However, we find no effects of student–instructor demographic mismatch on dropping courses or taking subsequent courses in the same field, nor do we find effects at other points of the grade distribution. Consistent with previous research in the K–12 context, these effects are stronger for underrepresented groups such as female and nonwhite students. The effects are most pronounced for nonwhite female students.

What behaviors drive these results? Unfortunately, the mechanisms at work cannot be precisely identified with these administrative data.[31] However, we can make some informed speculation and perhaps rule out some possible channels. For example, the importance of blindly graded written exams in determining course grades suggests that instructors' grading biases, conscious or not, are not driving these results (Lavy 2008; Hanna and Linden 2012). Of course, this does not rule out the possibility that implicit biases affect how instructors interact with students, which could in turn affect student engagement, and ultimately academic performance. That racial mismatch effects are observed in medium-sized but not in small or large classes suggests that stereotype threat is not the sole explanation, as such effects should not vanish in classrooms of a

29. Grade inflation in graduate school accounts for the different definitions of "good grade," as a C is often considered failing in graduate and professional schools.
30. Our own analyses of the publicly available *After the JD* survey data confirm the positive association between law school GPA and earnings both overall, and for specific demographic groups, for lawyers who attended non-top ten law schools. See Appendix B for details.
31. Dee (2004), Ferguson (2003), Gershenson, Holt, and Papageorge (2016), Papageorge, Gershenson, and Kang (2016), and Dee and Gershenson (2017) provide rich discussions of the channels through which demographic representation might affect student outcomes.

certain size. Similarly, because sex mismatch effects primarily exist in classrooms with fewer female students suggests, whatever the channel, they are more salient in less representative classroom environments. Moving forward, it is important that future research, in all academic environments, seeks to better understand the specific channels through which student–instructor demographic match effects operate and to use this information to better design instructor-facing interventions and instructor training.

Finally, these results suggest that diversity in the legal profession, and the status of women and people of color in the legal profession, would be improved by increasing the diversity of law school faculty. However, whether and how these results would generalize to other law schools, particularly those with less diverse student and faculty populations, remains an open question worthy of future exploration. There are also questions regarding the general equilibrium responses to the hiring of a more diverse faculty, particularly in the law context, which might exacerbate demographic gaps in law offices and the judiciary. There are also potential supply-side limitations of such faculty in the short run. For these reasons, another potential policy response is to provide law school (and university) faculty with theoretically informed implicit bias training, which has proven to be effective in some early pilots (Carnes et al. 2015). Similarly, Darling-Hammong and Holmquist (2015) provide suggestions to law school faculty on how to better serve historically underrepresented students, many of which echo the theoretically-informed, "WISE" interventions and strategies advocated by social psychologists (Walton 2014; Okonofua, Paunesku, and Walton 2016).

ACKNOWLEDGMENTS

Scott Carrell, Stephen B. Holt, Michal Kurlaender, Nicholas Papageorge, and participants at the 2016 APPAM Fall Conference, 2016 Access Group Legal Education Research Symposium, 2017 Royal Economic Society Symposium for Junior Researchers, and 2017 Society of Labor Economists Annual Meeting provided many helpful comments. The authors are thankful for financial support from the Research Grant Program of the AccessLex Institute/Association for Institutional Research (AIR). Opinions reflect those of the authors and not necessarily those of the granting agency. We also thank Stephanie Cellini and three anonymous referees for their helpful comments and suggestions. Kimberly Trocha provided excellent research assistance. An earlier draft of this paper was circulated as IZA Discussion Paper No. 10459, "Stereotype Threat, Role Models, and Demographic Mismatch in an Elite Professional School Setting."

REFERENCES

Alm, James, and John V. Winters. 2009. Distance and intrastate college student migration. *Economics of Education Review* 28(6): 728–738.

Alsan, Marcella, Owen Garrick, and Grant C. Graziani. 2018. Does diversity matter for health? Experimental evidence from Oakland. NBER Working Paper No. 24787.

Altonji, Joseph G., and Rebecca M. Blank. 1999. Race and gender in the labor market. *In Handbook of labor economics*, Vol. 3, edited by Orley Ashenfelter and David Card, pp. 3143–3259. New York: Elsevier.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1): 151–184.

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik. 2015. The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics* 33(1): 63–89.

Azmat, Ghazala, and Rosa Ferrer. 2017. Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy* 125(5): 1306–1355.

Bailey, Martha J., and Susan M. Dynarski. 2011. Gains and gaps: Changing inequality in U.S. college entry and completion. NBER Working Paper No. 17633.

Baker, Rachel, Thomas Dee, Brent Evans, and June John. 2018. Bias in online classes: Evidence from a field experiment. Stanford, CA: CEPA Working Paper No. 18-03.

Banks, Taunya Lovell. 1988. Gender bias in the classroom. *Journal of Legal Education* 38(1/2): 137–146.

Bertrand, Marianne. 2011. New perspectives on gender. *In Handbook of labor economics*, Vol. 4, edited by David Card and Orley Ashenfelter, pp. 1543–1590. New York: Elsevier.

Bertrand, Marianne, and Kevin F. Hallock. 2001. The gender gap in top corporate jobs. *ILR Review* 55(1): 3–21.

Bettinger, Eric P., and Bridget Terry Long. 2005. Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review* 95(2): 152–157.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2012. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29(2): 238–249.

Carnes, Molly, Patricia G. Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E. Ford, Patrick Forscher, Carol Isaac, Anna Kaatz, Wairimu Magua, Mari Palta, and Jennifer Sheridan. 2015. Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine* 90(2): 221–230.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics* 125(3): 1101–1144.

Cooke, Thomas J., and Paul Boyle. 2011. The migration of high school graduates to college. *Educational Evaluation and Policy Analysis* 33(2): 202–213.

Correia, Sergio. 2016. *A feasible estimator for linear models with multi-way fixed effects*. Available from http://scorreia.com/research/hdfe.pdf. Accessed 22 May 2020.

Darling-Hammong, Sean, and Kristen Holmquist. 2015. Creating wise classrooms to empower diverse law students: Lessons in pedagogy from transformative law professors. *Berkeley La Raza Law Journal* 25(1): 1–67.

Dee, Thomas S. 2004. Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics* 86(1): 195–210.

Dee, Thomas S. 2005. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review* 95(2): 158–165.

Dee, Thomas S. 2007. Teachers and the gender gaps in student achievement. *Journal of Human Resources* 42(3): 528–554.

Dee, Thomas S., and Seth Gershenson. 2017. *Unconscious bias in the classroom: Evidence and opportunities*. Available http://services.google.com/fh/files/misc/unconscious-bias-in-the-classroom-report.pdf. Accessed 9 December 2019.

Delisle, Jason. 2014. The graduate student debt review: The state of graduate student borrowing. Washington, DC: New America Education Policy Program.

Dinovitzer, Ronit, Nancy Reichman, and Joyce Sterling. 2009. The differential valuation of women's work: A new look at the gender gap in lawyers' incomes. *Social Forces* 88(2): 819–864.

Eren, Ozkan, and Naci Mocan. 2016. Emotional judges and unlucky juveniles. NBER Working Paper No. 2261.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review* 104(8): 2567–2591.

Ferguson, Ronald F. 2003. Teachers' perceptions and expectations and the black–white test score gap. *Urban Education* 38(4): 460–507.

Figlio, David N., Morton O. Schapiro, and Kevin B. Soter. 2015. Are tenure track professors better teachers? *Review of Economics and Statistics* 97(4): 715–724.

Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2016. Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review* 52:209–224.

Guinier, Lani, Michelle Fine, Jane Balin, Ann Bartow, and Deborah Lee Stachel. 1994. Becoming gentlemen: Women's experiences at one ivy league law school. *University of Pennsylvania Law Review* 143(1): 1–110.

Hanna, Rema N., and Leigh L. Linden. 2012. Discrimination in grading. *American Economic Journal: Economic Policy* 4(4): 146–168.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46(6): 1251–1271.

Heckman, James J., and Tim Kautz. 2012. Hard evidence on soft skills. *Labour Economics* 19(4): 451–464.

Hill, Andrew J., and Daniel B. Jones. 2018. A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review* 64: 1–12.

Ho, Daniel E., and Mark G. Kelman. 2014. Does class size affect the gender gap? A natural experiment in law. *Journal of Legal Studies* 43(2): 291–321.

Hoffmann, Florian, and Philip Oreopoulos. 2009. A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources* 44(2): 479–494.

Holder, Eric H. 2001. The importance of diversity in the legal profession. *Cardozo Law Review* 23(6): 2241–2251.

Holt, Stephen B., and Seth Gershenson. 2019. The impact of demographic representation on absences and suspensions. *Policy Studies Journal* 47(4): 1069–1099.

Inzlicht, Michael, and Talia Ben-Zeev. 2000. A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science* 11(5): 365–371.

Lavy, Victor. 2008. Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics* 92(10): 2083–2105.

Loury, Glenn C. 2009. *The anatomy of racial inequality*. Cambridge, MA: Harvard University Press.

Lusher, Lester, Doug Campbell, and Scott Carrell. 2015. TAs like me: Racial interactions between graduate teaching assistants and undergraduates. NBER Working Paper No. 21568.

Mustard, David B. 2001. Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *Journal of Law and Economics* 44(1): 285–314.

Neumark, David, and Rosella Gardecki. 1998. Women helping women? Role model and mentoring effects on female Ph.D. students in economics. *Journal of Human Resources* 33(1): 220–246.

Okonofua, Jason A., David Paunesku, and Gregory M. Walton. 2016. Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences U.S.A.* 113(19): 5221–5226.

Oyer, Paul, and Scott Schaefer. 2015. Firm/employee matching: An industry study of U.S. lawyers. *ILR Review* 69(2): 378–404.

Oyer, Paul, and Scott Schaefer. 2019. The returns to elite degrees: The case of American lawyers. *ILR Review* 72(2): 446–479.

Papageorge, Nicholas W., Seth Gershenson, and Kyungmin Kang. 2016. Teacher expectations matter. IZA Discussion Paper No. 10165.

Rachlinski, Jeffrey J., Sheri Lynn Johnson. 2009. Does unconscious racial bias affect trial judges? *Notre Dame Law Review* 84(3): 1195–1246.

Schochet, Peter Z. 2009. An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review* 33(6): 539–567.

Schwarcz, Daniel, and Dion Farganis. 2017. The impact of individualized feedback on law student performance. *Journal of Legal Education* 67(1): 139–175.

Steele, Claude M. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52(6): 613–629.

Walton, Gregory M. 2014. The new science of wise psychological interventions. *Current Directions in Psychological Science* 23(1): 73–82.

Wilkins, David B., and G. Mitu Gulati. 1996. Why are there so few black lawyers in corporate law firms? An institutional analysis. *California Law Review* 84(3): 493–625.

Wood, Robert G., Mary E. Corcoran, and Paul N. Courant. 1993. Pay differences among the highly paid: The male–female earnings gap in lawyers' salaries. *Journal of Labor Economics* 11(3): 417–441.

APPENDIX A

Table A.1. Baseline Model Allowing for Contemporaneous Spillovers

	Continuous Grade		A/A– Grade	
	(1)	(2)	(3)	(4)
Other-sex (OS)	-0.017^{**}	0.017	-0.012^{*}	−0.004
	(0.007)	(0.029)	(0.007)	(0.028)
Another OS	−0.003	0.033	0.009	0.021
	(0.013)	(0.025)	(0.011)	(0.019)
Other-race (OR)	-0.037^{**}	-0.053^{***}	-0.028^{***}	−0.028
	(0.016)	(0.019)	(0.011)	(0.019)
Another OR	−0.003	−0.010	−0.001	−0.001
	(0.011)	(0.014)	(0.010)	(0.011)
OS * Another OS		-0.048^{*}		−0.016
		(0.028)		(0.028)
OR * Another OR		0.023		−0.000
		(0.022)		(0.022)
Observations	36,560	36,560	36,560	36,560
Course fixed effects	Yes	Yes	Yes	Yes
Student fixed effects	Yes	Yes	Yes	Yes

Notes: Each column represents a different model specification. The outcomes are measured as follows: Continuous Grade measures a student's received grade on a 0–4 scale (F–A). A/A– Grade is a binary indicator for whether a student received an A or A– grade. Standard errors in parentheses are clustered by student and instructor.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.2. Mean Course Characteristics by Subject

Subject	*N*	Course Size	Female Students	White Students	Female Faculty	White Faculty
Civil Procedure	80	83.1	0.60	0.65	0.21	0.82
Constitutional	60	83.1	0.58	0.63	0.14	0.69
Contracts	60	75.0	0.59	0.62	0.31	0.84
Criminal Law	70	76.8	0.59	0.65	0.57	0.62
Legal Writing	620	15.5	0.59	0.63	0.50	0.92
Property	80	85.0	0.59	0.64	0.60	0.79
Torts	60	79.2	0.58	0.63	0.42	1.00

Notes: The unit of analysis is course sections. Each statistic reported is the average across course sections: average size, percent female students in class, percent white students in class, percent of course sections taught by female faculty, and percent of course sections taught by white faculty.

Table A.3. Impact of Demographic Mismatch on Non-Writing First-Year Required Course Outcomes

	Continuous Grade	A Grade	A/A– Grade	C, D, F Grade	Take Another	Dropped Course
	(1)	(2)	(3)	(4)	(5)	(5)
Other-sex	−0.011	−0.005	−0.012	−0.001	0.016	0.000
	(0.008)	(0.004)	(0.008)	(0.003)	(0.010)	(0.000)
Other-race	-0.042^{**}	-0.017^{**}	-0.034^{***}	0.011	0.000	0.001
	(0.017)	(0.007)	(0.011)	(0.008)	(0.008)	(0.001)
Differences in coefficients (*P*)	0.071^{*}	0.151	0.061^{*}	0.119	0.198	0.271
Observations	28,290	28,290	28,290	28,290	18,620	28,400

Table A.3. Continued.

	Continuous Grade	A Grade	A/A– Grade	C, D, F Grade	Take Another	Dropped Course
	(1)	(2)	(3)	(4)	(5)	(5)
Course fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Student fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents a different model specification. The outcomes are measured as follows: Continuous Grade measures a student's received grade on a 0–4 scale (F–A); A Grade is a binary indicator for whether a student received an A grade; A or A– Grade is a binary indicator for whether a student received an A or A– grade; C, D, F Grade is a binary indicator for whether a student received a C, D, or F grade; Take Another is a binary indicator for whether a student takes a subsequent elective course in the same field after his first year; and Dropped Course is a binary indicator for whether a student drops the course before the end of the semester. Column 5 has fewer observations because not all required courses correspond to elective course subjects. Difference in coefficients compares Other-sex effect against Other-race effect. Standard errors in parentheses are clustered by student and instructor.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.4. Impact of Demographic Mismatch on First-Year Required Course Outcomes (Logit Average Partial Effects)

	A/A– Grade	C, D, F Grade	Take Another	Dropped Course
	(1)	(2)	(3)	(4)
Other-sex	−0.011**	0.002	0.012**	0.001
	(0.005)	(0.002)	(0.006)	(0.000)
Other-race	−0.029***	0.007	−0.008	0.001
	(0.010)	(0.004)	(0.010)	(0.001)
Female instructor	0.029***	−0.001	0.002	−0.000
	(0.005)	(0.003)	(0.007)	(0.001)
Nonwhite instructor	0.025***	−0.002	−0.009	0.001
	(0.009)	(0.003)	(0.009)	(0.001)
Nonwhite student	−0.088***	0.030***	0.038***	−0.001
	(0.012)	(0.005)	(0.013)	(0.001)
Female student	0.044***	−0.019***	0.000	−0.000
	(0.008)	(0.004)	(0.009)	(0.001)
Observations	36,560	36,560	18,540	36,730

Notes: Each column represents a different model specification. The outcomes are measured as follows: A/A– Grade is a binary indicator for whether a student received an A or A– grade. C, D, F Grade is a binary indicator for whether a student received a C, D, or F grade. Take Another is a binary indicator for whether a student takes a subsequent course in the same field. Dropped Course is a binary indicator for whether a student drops the course before the end of the semester. Column 3 has fewer observations because not all required courses correspond to elective course subjects. Standard errors in parentheses are clustered by student and instructor.

$^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.5. Heterogeneity by Course Size and Percent Female (Nonwhite) in Course

	Course Size		Percent Female (Nonwhite) in Course	
	(1)	(2)	(3)	(4)
Other-sex (OS)	0.003828		−0.245622	
	(0.024974)		(0.176855)	
OS * Course size	0.000052			
	(0.001052)			
OS * Course size (Sq)	−0.000003			
	(0.000009)			

Table A.5. Continued.

	Course Size		Percent Female (Nonwhite) in Course	
	(1)	(2)	(3)	(4)
OS * Percent female			0.007005 (0.005837)	
OS * Percent female (Sq)			−0.000051 (0.000048)	
Other-race (OR)		0.020580 (0.024309)		0.028642 (0.094617)
OR * Course size		−0.002240** (0.000928)		
OR * Course size (Sq)		0.000018** (0.000008)		
OR * Percent nonwhite				−0.003086 (0.004580)
OR * Percent nonwhite (Sq)				0.00004 (0.000056)
p value for joint significance tests	0.076*	0.005***	0.088*	0.052*
Observations	36,560	36,560	36,560	36,560
Course fixed effects	Yes	Yes	Yes	Yes
Student fixed effects	Yes	Yes	Yes	Yes

Notes: The covariates are defined as follows: Other-sex is a binary indicator for whether the student's sex is different from the instructor. Other-race is a binary indicator for whether the student's race is different from the instructor. Course size measures the total number of students in the classroom. Percent Female measures the percentage of female students in the classroom. Percent Nonwhite measures the percentage of nonwhite students in the classroom. Sq indicates quadratic term. Standard errors in parentheses.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

APPENDIX B

This appendix uses publicly available data from *After the JD* (AJD) to document the descriptive relationship between law school grades and early-career salaries for individuals who earned JDs from non–top 10 law schools.[32] The motivation for this appendix is to show the impacts of student–instructor mismatch on course grades documented in the current study likely translate into demographic pay gaps among early-career law professionals.

The public-use AJD data report annual earnings in eight bins: <\$40,000, \$40,000-\$49,999, \$50,000-\$59,999, \$60,000-\$74,999, \$75,000-\$99,999, \$100,000-\$124,999, \$125,000-\$149,999, and >\$150,000. Accordingly, we estimate descriptive ordered-logit models in which this categorical annual-earnings variable is the dependent variable. Appendix table B.1 reports the ordered-logit coefficients for the full sample. The parsimonious specifications in columns 1 and 2 document the unconditional female pay gap and wage-GPA gradient, respectively. The omitted reference category for the GPA variable is <3.0. Column 4 shows that these patterns are robust to controlling for law school quality.

Because the ordered-logit coefficients are not directly interpretable, table B.2 reports the average partial effects (APE) of these covariates on the probability of being in each earnings band for the fully-specified, full-sample estimates reported in column 4

32. The AJD is a representative survey of new law-school graduates, conducted by the American Bar Foundation, in 2002, 2007, and 2010. See www.americanbarfoundation.org/research/project/118 for further information.

Table B.1. Descriptive Ordered-Logit Earnings Regressions: Coefficient Estimates

	(1)	(2)	(3)	(4)
Female	-0.45^{***}		-0.49^{***}	-0.49^{***}
	(0.07)		(0.07)	(0.07)
Black	−0.02		0.11	0.02
	(0.11)		(0.12)	(0.11)
Latinx	−0.10		0.04	−0.01
	(0.11)		(0.12)	(0.12)
Asian	0.62^{***}		0.63^{***}	0.42^{***}
	(0.13)		(0.14)	(0.13)
Other race	0.06		0.10	0.06
	(0.19)		(0.19)	(0.20)
> 3.75 grade point average (GPA)		1.82^{***}	1.90^{***}	1.68^{***}
		(0.16)	(0.16)	(0.18)
3.5–3.74 GPA		1.59^{***}	1.65^{***}	1.42^{***}
		(0.13)	(0.13)	(0.14)
3.25–3.49 GPA		1.04^{***}	1.09^{***}	0.87^{***}
		(0.13)	(0.13)	(0.13)
3.0–3.24 GPA		0.56^{***}	0.59^{***}	0.48^{***}
		(0.12)	(0.12)	(0.12)
Missing GPA		1.34^{***}	1.35^{***}	1.16^{***}
		(0.10)	(0.10)	(0.10)
Top 10 law school				2.21^{***}
				(0.15)
11–20 law school				1.45^{***}
				(0.13)
21–100 law school				0.38^{***}
				(0.07)
Observations	3,785	3,892	3,785	3,755

Notes: Each column represents a different model specification. Cut points not shown. Standard errors in parentheses.

$^{***}p < 0.01$.

Table B.2. Descriptive Ordered-Logit Earnings Regressions: Average Partial Effects

	0–39K	40–49K	50–59K	60–74K	75–99K	100–124K	125–149K	>150K)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.04^{***}	0.04^{***}	0.02^{***}	0.00^{**}	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.04^{***}
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Black	−0.00	−0.00	−0.00	−0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Latinx	0.00	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Asian	-0.03^{***}	-0.04^{***}	-0.02^{***}	-0.00^{*}	0.02^{***}	0.02^{***}	0.02^{***}	0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Other race	−0.00	−0.00	−0.00	−0.00	0.00	0.00	0.00	0.00
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)
> 3.75 GPA	-0.13^{***}	-0.15^{***}	-0.07^{***}	-0.01^{**}	0.06^{***}	0.07^{***}	0.08^{***}	0.13^{***}
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
3.5–3.74 GPA	-0.11^{***}	-0.13^{***}	-0.06^{***}	-0.01^{**}	0.05^{***}	0.06^{***}	0.07^{***}	0.11^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
3.25–3.49 GPA	-0.07^{***}	-0.08^{***}	-0.03^{***}	-0.00^{**}	0.03^{***}	0.04^{***}	0.04^{***}	0.07^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
3.0–3.24 GPA	-0.04^{***}	-0.04^{***}	-0.02^{***}	-0.00^{*}	0.02^{***}	0.02^{***}	0.02^{***}	0.04^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Missing GPA	-0.09^{***}	-0.10^{***}	-0.05^{***}	-0.01^{**}	0.04^{***}	0.05^{***}	0.06^{***}	0.09^{***}
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)

Table B.2. Continued.

	0–39K (1)	40–49K (2)	50–59K (3)	60–74K (4)	75–99K (5)	100–124K (6)	125–149K (7)	>150K) (8)
Top 10 law school	−0.17***	−0.20***	−0.09***	−0.01**	0.08***	0.09***	0.11***	0.18***
	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
11-20 law school	−0.11***	−0.13***	−0.06***	−0.01**	0.06***	0.06***	0.07***	0.12***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
21-100 law school	−0.03***	−0.03***	−0.02***	−0.00**	0.01***	0.02***	0.02***	0.03***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Observations	3,755	3,755	3,755	3,755	3,755	3,755	3,755	3,755

Notes: Standard errors in parentheses.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.3. Descriptive Ordered-Logit Earnings Regressions by Demographic Background: Coefficient Estimates

	Male (1)	Female (2)	White (3)	Black (4)	Latinx (5)	Asian (6)
Female			−0.49***	−0.97***	−0.65**	−0.28
			(0.08)	(0.23)	(0.27)	(0.25)
Black	0.15	−0.09				
	(0.16)	(0.16)				
Latinx	0.04	−0.09				
	(0.17)	(0.18)				
Asian	0.27	0.53***				
	(0.20)	(0.18)				
Other Race	0.13	−0.03				
	(0.28)	(0.29)				
> 3.75 GPA	1.62***	1.70***	1.69***	17.99***	0.61	1.44**
	(0.29)	(0.22)	(0.19)	(1.07)	(0.84)	(0.59)
3.5–3.74 GPA	1.54***	1.31***	1.48***	3.34***	2.86***	0.62
	(0.19)	(0.20)	(0.16)	(0.68)	(0.81)	(0.56)
3.25–3.49 GPA	0.92***	0.83***	0.87***	2.41***	0.40	0.48
	(0.18)	(0.19)	(0.15)	(0.64)	(0.53)	(0.35)
3.0–3.24 GPA	0.43**	0.53***	0.51***	0.83**	0.42	−0.10
	(0.17)	(0.17)	(0.14)	(0.39)	(0.43)	(0.41)
Missing GPA	1.15***	1.18***	1.27***	1.00***	0.71**	
	(0.14)	(0.15)	(0.12)	(0.29)	(0.33)	
Top 10 law school	2.26***	2.15***	2.15***	2.49***	3.02***	2.08***
	(0.20)	(0.24)	(0.18)	(0.41)	(0.50)	(0.45)
11–20 law school	1.47***	1.43***	1.51***	1.53***	0.85	1.03***
	(0.17)	(0.20)	(0.16)	(0.38)	(0.54)	(0.34)
21–100 law school	0.34***	0.41***	0.42***	0.36	0.33	0.13
	(0.10)	(0.11)	(0.08)	(0.28)	(0.30)	(0.30)
Observations	1,995	1,760	2,703	330	312	341

Notes: Each column represents a different model specification. Cut points not shown. Standard errors in parentheses.

$^{**}p < 0.05$; $^{***}p < 0.01$.

of table B.1. Here we see that female lawyers are 2 to 4 percentage points more likely than male lawyers to be in the lowest-earning categories and 2 to 4 percentage points less likely than male lawyers to be in the highest-earning categories. The APE for the categorical GPA indicators show that each 0.25 increase in GPA is associated with a 2 to 4 percentage point increase in the probability of being in one of the high-earnings

Table B.4. Descriptive Ordered-Logit Earnings Regressions by Law School Rank: Coefficient Estimates

	Top 10	11-20	21-100	Outside 100
	(1)	(2)	(3)	(4)
Female	-0.55^{**}	-0.44^{**}	-0.44^{***}	-0.60^{***}
	(0.24)	(0.20)	(0.10)	(0.12)
Black	0.07	0.11	-0.07	0.06
	(0.33)	(0.28)	(0.17)	(0.23)
Latinx	0.42	-0.60	-0.00	0.10
	(0.37)	(0.37)	(0.17)	(0.27)
Asian	0.57	0.23	0.33^{*}	0.72^{***}
	(0.37)	(0.30)	(0.20)	(0.27)
Other race	1.13^{*}	-0.36	-0.28	0.67
	(0.64)	(0.46)	(0.26)	(0.41)
$>$ 3.75 GPA	-0.12	2.72^{***}	1.98^{***}	1.54^{***}
	(1.02)	(0.43)	(0.26)	(0.31)
3.5–3.74 GPA	0.80	2.15^{***}	1.62^{***}	1.33^{***}
	(0.85)	(0.38)	(0.21)	(0.26)
3.25–3.49 GPA	0.97	1.50^{***}	1.07^{***}	0.50^{**}
	(0.82)	(0.42)	(0.20)	(0.21)
3.0–3.24 GPA	0.42	1.99^{***}	0.42^{**}	0.41^{**}
	(0.81)	(0.41)	(0.19)	(0.18)
Missing GPA	0.52	2.40^{***}	1.33^{***}	0.90^{***}
	(0.78)	(0.33)	(0.16)	(0.16)
Observations	370	467	1,737	1,181

Notes: Each column represents a different model specification. Cut points not shown. Standard errors in parentheses.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

brackets, and a symmetric decrease in the probability of being in a low-earning bracket. Importantly, this suggests that even a relatively small change in GPA attributable to student–instructor demographic mismatch in first-year law courses might substantively affect early-career earnings.

Table B.3 estimates the fully specified ordered-logit model separately by sex and race. The key results here are: (1) the sex pay gap exists for white, black, and Latinx lawyers and (2) the wage-GPA gradient exists in the male, female, white, and black subsamples. Table B.4 similarly shows that the wage–GPA gradient exists for graduates of all law schools outside the U.S. News Top 10. This is consistent with results reported in Oyer and Schaefer (2019). The U.S. News rank of the law school studied in the current paper falls in the 21–100 range (column 3), for which grades are quite important.