Does Racial Diversity Improve Academic Outcomes? A Natural Experiment in Higher Education Classrooms*

Yan Lau†

November 2019

Abstract

I estimate causal effects of classroom racial diversity on academic outcomes by exploiting a natural experiment where first-year college students in a mandatory writing course are assigned to discussion conferences with varying racial compositions. Within-classroom diversity is effectively random conditional on scheduling availability. I find that a higher degree of classroom diversity increases GPA at graduation, improves writing course grades of low SAT scorers, and affects the major choice of white students. My results highlight the potential value of racial diversity in higher education and contribute to the debate over race-based admissions policies.

JEL Codes: I20, I21, I23, I28, J15

*I would like to thank Eleanor Choi, Damon Clark, Songman Kang, Nicholas Lawson, Lars Lefgren, Alex Mas, Yoonsoo Park, Jeff Parker, Devesh Raval, Harvey Rosen, Cecilia Rouse, and various conference and seminar participants for invaluable feedback on this project. Alicia Toshima provided excellent research assistance. The views expressed are those of the author, and do not necessarily reflect those of the Federal Trade Commission or any individual Commissioner.

†Bureau of Economics, Federal Trade Commission [Email: ylau@ftc.gov]

1 Introduction

Educators and policy makers care deeply about the racial diversity of student bodies at tertiary institutions. As the former President of the American Council on Education Molly Corbett Board notes, "leaders of colleges and universities... know firsthand that the educational benefits that flow from a diverse student body are crucial for our graduates," adding that "Education in a vibrant and diverse environment will better prepare... graduates for our increasingly globalized economy." (Broad, 2012) By fostering a diverse environment in which students are exposed to worldviews different from their own, students exchange unfamiliar ideas and challenge prior belief systems, enriching the overall learning experience.

Proponents of affirmative action point to such benefits as justification for college admission preferences favoring minority students. Critics, on the other hand, argue that such race-conscious admission policies are discriminatory and lead to mismatch, whereby weaker students may be better served at less selective institutions (see Sander (2004) and others). Over the years, a host of legal cases concerning affirmative action have been fought over precisely this issue of diversity at universities and colleges (Regents of the University of California v. Bakke, 1978; Hopwood v. State of Texas, 1996; Grutter v. Bollinger, 2003; Gratz v. Bollinger, 2003; Fisher v. University of Texas, 2016).

Given such intense and longstanding interest in diversity in higher education, measuring its impact on educational outcomes is crucial for understanding policy implications. Yet challenges arise when measuring the effects of diversity because of endogeneity. Selection by students into schools, courses, and classrooms based on unobservable characteristics may bias estimates.

This paper contributes to the literature by presenting causal evidence on the positive effects of racial diversity on academic outcomes in a real-world classroom setting. To find out whether students exposed to a more diverse set of classmates achieve better grades, I exploit a quasi-experimental setting where first-year students in a year-long mandatory writing course at a small four-year college are randomly assigned to discussion conference groups. Conferences have varying levels of diversity in terms of the racial composition of conference group members. I argue that this within-classroom diversity is effectively random conditional on students' scheduling availability (vis-a-vis other courses enrolled), given the institutional features determining conference assignment, and the fact that students do not know (ex-ante) and cannot manipulate the racial composition of peers in their enrolled conference. I confirm this exogeneity by analyzing predetermined characteristics of students in conferences of varying diversity. Hence, this identification strategy estimates the causal effect of diversity, avoiding selection bias that potentially arises in other situations where better students may select themselves into more diverse environments.

I find that a higher degree of racial diversity in the classroom causes a statistically significant increase in grade point average (GPA) at graduation. Point estimates suggest that replacing one white student with one minority student in a typical conference increases graduation GPA by an average of 0.02σ. While the average treatment effect on contemporaneous outcomes including first-year GPA and the writing course grade are not statistically significant, these estimates mask heterogeneous effects by SAT score. I find that the beneficial impact of diversity on both writing course grade and GPA at graduation is more positive and statistically significant for students with lower SAT scores. For example, students with an SAT score of 1050 experience an average marginal effect of diversity on the writing course grade that is about four times the average treatment effect point estimate, while students with SAT scores above 1300 experience no statistically significant diversity effect. On the other hand, I do not detect heterogeneous effects by sex or race. I also find that white students in higher diversity classrooms are less likely to take up majors in literature, language and arts, and more likely to take up majors in humanities and social sciences.

The remainder of the paper proceeds as follows. Section 2 relates this paper to the existing literature. Section 3 discusses the institutional background applicable to my identification strategy. Section 4 details the reduced-form empirical analyses conducted and discusses the results obtained. Section 5 considers potential mechanisms that explain the empirical findings and Section 6 concludes.

2 Contribution to the Literature

First and foremost, the main contribution of this paper to the literature is the empirical estimation of causal effects of classroom diversity on academic outcomes. Numerous studies have examined the issue of diversity in higher education settings and its effects on educational outcomes (Bowen and Bok, 1998; Alger et al., 2000; Daniel et al., 2001; Terenzini et al., 2001; Gurin et al., 2002; Hu and Kuh, 2003; Umbach and Kuh, 2006; Denson and Chang, 2009; Arcidiacono and Vigdor, 2010; Hinrichs, 2011; Dills, 2018), as well as on attitudes and perceptions (Rothman et al., 2003; Umbach and Kuh, 2006; Denson and Chang, 2009; Carrell et al., 2019). Beyond tertiary education, several papers have quantified the effects of desegregation on educational outcomes in the United States, both for specific diversity-inducing programs (Angrist and Lang, 2004) and at a more macro level (Guryan, 2004).[1]

However, the extant literature is short on empirical evidence regarding the causal effect of racial diversity on academic outcomes because it is difficult to find exogenous variation in diversity. Many previous studies suffer from selection bias because the observed diversity measures are endogenous to potential choices made by different types of students (e.g. Gurin et al., 2002). Many studies also examine diversity at the institution level (e.g. Daniel et al., 2001; Arcidiacono and Vigdor, 2010; Hinrichs, 2011), which may be less informative of the mechanism behind the diversity effect. While results in some experimental studies that randomize diversity in lab settings do estimate causal effects (e.g. Antonio et al., 2004), these may lack generalizability.

[1]Related literature also considers the impact of affirmative action (race-conscious) versus race-blind college admission policies (Chan and Eyster, 2003; Epple et al., 2008; Fryer et al., 2008; Arcidiacono and Lovenheim, 2016), as well as the interaction between affirmative action and school choice (Hafalir et al., 2013; Alcalde and Subiza, 2014; Ehlers et al., 2014). A further related strand of the literature examines the effects of racial diversity at the workplace on firm outcomes such as productivity (Kahane et al., 2013; Ozgen et al., 2013; Parrotta et al., 2014; and Trax et al., 2015). This paper is also related to the peer effects literature since the diversity effect is in essence a peer effect (through the race of one's peers); for a survey of peer effects theory and empirical evidence, see Epple and Romano (2011) and Sacerdote (2014).

The closest study in the literature to this paper is Dills (2018), which uses a similar classroom assignment setting to identify the causal effects of racial diversity. Dills (2018) finds evidence of a heterogeneous diversity effect on the grade of the course being examined. Her results indicate that white students perform better in classrooms with more minority classmates, while non-white students perform worse, though this effect is limited to below-median SAT scorers. In contrast, I find no heterogeneous effects with respect to race, and more positive diversity effects for students who score lower on the SAT. My study also improves upon Dills (2018) in various dimensions—such as the use of a different diversity measure and examining effects on a range of academic outcomes—discussed below.

Second, in addition to the main contribution of identifying causal effects, the quasi-experimental approach I adopt also offers a greater degree of external validity, since it involves a real-world classroom setting typical of many college courses. While these results are most relevant to similar small four-year colleges, their applicability to larger tertiary institutions (both public and private) should not be discounted. The lecture-plus-conference arrangement of the writing course being examined is similar to courses at many larger institutions which incorporate regular lectures before dividing students into smaller discussion sections. For instance, at the University of Texas at Austin (an oversubscribed public university which has been the subject of an affirmative action lawsuit), 48% of discussion sections forming part of a larger lecture comprise between 10 and 19 students (CollegeData, 2017), approximately the same size as the conferences in the writing course being studied.

Third, this paper further improves upon many previous studies by drawing on the latest methodological concepts and utilizing a novel data source that has numerous advantages. I construct a direct measure of diversity based on the Herfindahl index which I argue more appropriately captures variety in racial composition (as detailed in Section 4.1). While I am not the first to use such a construct (e.g. see Hinrichs, 2011), the majority of the literature uses the proportion of minority students as a proxy for diversity (e.g. Daniel et al., 2001; Denson and Chang, 2009; Arcidiacono and Vigdor, 2010; Dills, 2018). This latter approach only captures exposure to a particular (i.e. non-white) racial grouping as opposed to the overall racial variety within a classroom.[2] I also examine a variety of academic outcomes beyond the grade of the course generating the variation in classroom diversity (cf. Dills, 2018). These include the longer-term dependent variable GPA at graduation and students' choice of major area of study. Lastly, my analysis uses administrative student data in calculating student measures. Many prior studies use survey data with self-reported measures of both diversity and academic outcomes, which may be unreliable and imprecise because of measurement error in both self-reported race and grades (e.g. Gurin et al., 2002; Hu and Kuh, 2003; Umbach and Kuh, 2006; Denson and Chang, 2009). My use of administrative data sidesteps these issues, enabling me to precisely quantify the degree of diversity in each classroom, as well as to examine quantitative grading outcomes.

3 Institutional Background

The writing course at the four-year college I examine is a year-long mandatory course that is taken by all first-year and transfer students.[3] The syllabus comprises readings of classical texts from antiquity and the course aims to serve as the foundational writing and critical thinking component of a liberal arts education. All students attend thrice-weekly lectures before breaking up into smaller conferences of approximately 15 students per classroom. Conferences meet either thrice-weekly for 50 minutes or twice-weekly for 80 minutes, during which discussions and debates relating to the current topic and readings are held. These conferences represent a high level of interaction between students in the classroom and the group becomes closely familiar with one another over the course of the year. Instructors leading such conferences have noted that they are a "place for shared conversation [and] collective learning," where students "communicate ideas with others and build from what

[2]Diversity is not necessarily the same as the proportion of minority students in a classroom. For example, with only two race groupings, a classroom with 100% white students is equally diverse as a classroom with 100% black students, both of which are at the opposite end of the diversity spectrum compared to a classroom with a 50-50 split of white and black students.

[3]An exception is that transfer students may opt to take certain other courses in lieu of this course.

others say."[4] If diversity were to play a role in the education of college students, then the diversity of this particular group of students in the highly-interactive conference setting is likely to have a significant impact on academic outcomes.

To estimate an internally-valid causal effect of racial diversity, it is necessary to understand where the variation in diversity comes from. Given that the writing course curriculum is identical across conference classrooms, and its content has been more-or-less constant across years, it is the between-classroom variation in what scholars call structural diversity[5] that identifies our estimated treatment effect. Thus, the relevant question is: How are students assigned to their writing course conferences? During the course sign-up phase immediately before the start of the Fall semester, students select the conference time they prefer, conditional on the times of other enrolled courses in their schedule. Those concerned about selection bias may posit that students select themselves into specific conferences in order to manipulate the degree of racial diversity they experience. In practice, however, the following institutional features make this possibility remote.

1. Students do not observe who else is in a conference, or who the instructor associated with a particular conference time will be. This situation of imperfect information creates a coordination problem for students wanting to choose certain diversity configurations made up of essentially-anonymous peers.

2. Conference enrollment is capped at just above the projected average number of students per conference for that year. Once the cap is reached, students must sign up for another section.

3. Because multiple conference sections may meet concurrently during the same time slot, the system will balance students randomly across conferences held at the same

[4] These comments were quoted from interviews with faculty members.

[5] The distinction between "structural diversity" and "curricular diversity" has been made by many in the existing literature, including Gurin et al. (2002) and Denson and Chang (2009). Structural diversity refers to the numerical makeup of the different racial groups within a student body. On the other hand, curricular diversity refers to students learning content that includes subject matters relating to different peoples and worldviews.

time, should enrollment among them be skewed. Students have no control over this re-balancing mechanism.

4. If conferences in a particular time slot are over-subscribed after sign-up, the system reassigns students to under-subscribed conferences in another time slot, subject to students' scheduling availability. Again, students have no control over this.

5. The order and timing of course sign-up by students is idiosyncratic. First-year students sign up for courses on a single day before the start of classes. However, they are only allowed to access the sign-up system after having met with their advisors and obtained a PIN code, and appointments with advisors (which occur on this same single day) are essentially ordered randomly. This means that unlucky students with later advising appointments have less control over their scheduling. This further complicates the coordination problem, even among friends who know each others' identities and intentions.[6]

6. Students will only have been on campus for a few days prior to signing up. It is unlikely they will base their writing course conference choice on relationships formed so recently.

7. Even if a group of friends somehow manages to get around all the mechanisms stated above and coordinate to enter the same conference, such coordination will most likely be small-scale, affecting at most a couple of students within a conference. In other words, if a grouping of friends is small, the students within the friends group are still subjected to the randomness in diversity generated by conference peers outside the friend group. So unless there is large-scale coordination among a vast majority of students in a particular conference, students cannot precisely manipulate the racial diversity they experience in the writing course conferences.

[6] While it is possible that more motivated students may access the system faster, this heterogeneity is unlikely to matter much. Students are instructed to sign up for courses immediately after the advisor meeting; moreover, the course sign-up system closes at the end of the day, so there is little room for delay.

8. Students are required to attend their assigned conference. A student wishing to switch conferences must go through a tedious petition process to do so, and no conference changes are permitted after the second week of the semester.

Given these reasons, I argue that the racial composition of students in any one particular writing course conference is effectively random. I confirm this empirically in Section 4.5 by analyzing predetermined student-level characteristics.

4 Empirical Analysis & Results

In this section, I document the reduced form empirical analyses conducted to estimate the effect of diversity on educational outcomes, and discuss the results obtained.

4.1 Data

This study uses student-level administrative data of first-year and transfer students who took the writing course between academic years 1995-1996 and 2011-2012. Each year is a cross-section of course students, and I combine these together into a pooled dataset. The analysis is restricted to students for whom a final cumulative GPA at graduation is observed.[7] This is done to maintain a balanced sample consistent across regression specifications. Thus, estimates should be interpreted as the effect of diversity conditional on not dropping out of the sample. (Appendix A considers relaxing this sample restriction, finding that classroom diversity does not affect attrition.)

The dependent variables of interest are standardized versions of the writing course grade,[8]

[7]This is a bit of a misnomer. Observing a cumulative GPA at "graduation" is almost always indicative of having graduated, but not always. In the sample, 20 students have a cumulative GPA at "graduation" reported, but failed their senior year and so did not actually graduate. I will nonetheless include these 20 students in the sample and use the phrase "at graduation" given this small number of exceptions.

[8]The writing course grade is a year-long grade that depends on 7 to 8 course-wide paper assignments over two semesters, as well as a final exam at the end of each semester. Assignments and exams, as well as the grading rubric, are standardized over the entire course, across conferences. Final course grades are assigned by conference instructors; instructor fixed effects are included in specifications to capture any instructor-specific grading effects across years.

cumulative GPA at the end of the first year, and cumulative GPA at graduation. To construct these variables, writing course grades are first numerically scaled using a 4-point grade-point scale comparable to the two GPA variables, where 4.0 points corresponds to grades of A-plus and A, 3.7 points to A-minus, 3.3 points to B-plus, and so forth; this continues until the grades of D and F, which are assigned values of 1.0 and zero points respectively.[9] The three variables (now all in grade-point units) are then normalized to have mean 0 and standard deviation 1 within each cohort year. Thus, all estimated effect sizes are in standard deviation (σ) units. (This normalization also addresses any issues relating to grade inflation.)

The course registration data identifies exactly which students are in which writing course conference with which other students. They also detail the full schedule of other enrolled courses during the first year for any given student. Other academic information of relevance include the total number of units taken in the fall and spring of first year, the total cumulative units at graduation, as well as ex-post major area of study.[10] The data contain student demographic characteristics: most crucially race (white (omitted category), black, Hispanic, Asian / Pacific Islander, and other / multiple race), but also sex, international status, and SAT score (expressed in thousands), from which we can calculate the mean SAT score within each conference.[11]

Table 1 reports summary statistics for most of these variables. Non-standardized mean grades (reported in grade-points) are slightly above the B letter grade. The average student takes 30 units of coursework. 35% of students are non-white, with the bulk of minority students making up the "other / multiple race" category. The student population is 45% male and 5.5% international, with an average incoming SAT score of 1350. In total, the analysis sample includes 4733 students in 401 unique conferences across 17 academic years.

[9] Grades of D-plus and D-minus are not given.

[10] Majors have been grouped into four overarching areas: 1) literature, languages and art; 2) humanities and social sciences; 3) mathematics and natural sciences; and 4) interdisciplinary and double majors.

[11] Raw SAT scores are combined scores out of 1600 comprising verbal and math components. Students with only ACT scores are assigned SAT scores converted based on concordance tables in College Board (2009).

Table 1: Summary Statistics

Variable / Indicator	Mean	Standard Deviation
Writing Course Grade	3.197	(0.517)
Cumulative GPA at End of 1st Year	3.091	(0.475)
Cumulative GPA at Graduation	3.187	(0.390)
Fall Units	3.650	(0.419)
Spring Units	4.033	(0.524)
Cumulative Units at Graduation	30.341	(2.276)
White	0.646	(0.478)
Black	0.020	(0.142)
Hispanic	0.051	(0.220)
Asian / Pacific Islander	0.087	(0.283)
Other / Multiple Race	0.195	(0.396)
Male	0.445	(0.497)
International Student	0.055	(0.229)
Own SAT Score (in thousands)	1.354	(0.113)
Diversity Index	0.621	(0.156)
N Students in Sample	4733	
N Conferences in Sample	401	
Average Conference Size	15.31	(1.253)

Notes: Grades and GPAs are measured in grade-points. Diversity Index refers to the normalized diversity index. Standard deviations reported in parentheses. See also footnote 14 on calculating the diversity index.

The average conference size in the sample is 15.3.[12]

To measure the diversity within a particular writing course conference, I construct a diversity index based on the probability that two randomly selected students from a given conference are of different racial ethnicities.

$$\begin{aligned} Diversity &= \Pr(\text{Two students in same conference have different races}) \\ &= 1 - \Pr(\text{Two students in same conference have same race}) \\ &= 1 - \sum_g (proportion\,(g))^2 \end{aligned}$$

[12]This number is higher than the total number of students divided by the total number of conferences because of excluded and missing data.

where $proportion\,(g)$ is the proportion of students in the conference belonging to each of the 5 race groups g. A higher diversity measure represents conferences that are more diverse in terms of racial composition. The summation term $\sum_g (proportion\,(g))^2$ in the above formulation is commonly known in economics as a Herfindahl index, which measures the concentration of types/groups within different settings.

A common issue with the Herfindahl index is that it ranges from $\frac{1}{G}$ to 1, where $G > 1$ is the number of groups g. A normalization of the index that ranges between 0 and 1 can be computed as

$$\frac{\left[\sum_g (proportion\,(g))^2\right] - \frac{1}{G}}{1 - \frac{1}{G}}$$

Accordingly, a normalized diversity index can be computed as

$$diversity = 1 - \frac{\left[\sum_g (proportion\,(g))^2\right] - \frac{1}{G}}{1 - \frac{1}{G}}$$

which ranges from 0 to 1 and has the similar interpretation whereby a higher value signifies more classroom diversity. All references to the diversity index henceforth pertain to this normalized version. (Appendix B considers the robustness of using alternative measures of diversity in the subsequent analyses.)

The mean diversity within writing course conferences is 0.621.[13] The histogram of this diversity index across individual students in Figure 1 shows that there is broad variation in the value of this measure. This implies that there is variability in the degree of racial diversity experienced by students in different conferences, and suggests that there is sufficient sample variation in the index to identify diversity effects.

To gain a better sense of what changes in the diversity index mean, Table 2 calculates the index for different conference examples. Row (a) calculates the diversity index for a "typical" conference with a racial makeup using the sample-average race proportions from

[13] As mentioned previously, the regression sample is restricted to students for whom a cumulative GPA at graduation is observed. However, when calculating the diversity index and other within-conference statistics, I also include students in the conference who did not graduate.

Figure 1: Histogram of Diversity Index Values

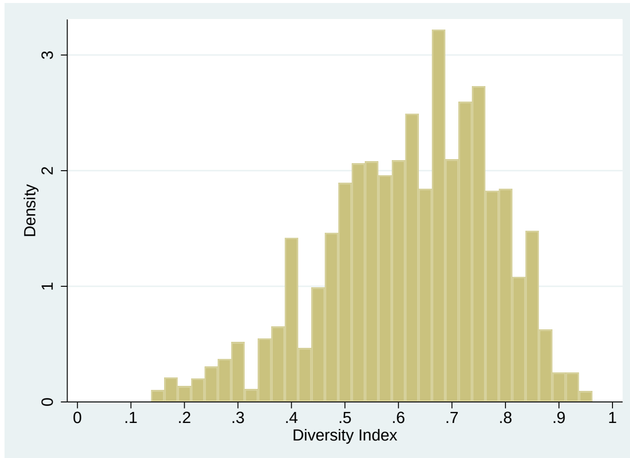


Table 1. The resulting measure of 0.668 for this hypothetical conference is slightly higher than the mean diversity index value.

In rows (b) through (e), I consider what happens to the diversity index when one white student is replaced with one minority student in this "typical" conference. The average conference size of 15.3 students implies that one student represents $\frac{1}{15.3} = 0.065$ of the class. Thus, in row (b), replacing one white student with one black student reduces the proportion of white students in the "typical" conference by 0.065 while increasing the proportion of black students by the same (see bold numbers). This increases the diversity index to 0.759. Similar calculations are carried out for different race replacements in rows (c) through (e). Note that replacing one white student with a student from a more-under-represented minority group (lower initial proportion of students) increases the diversity index by a greater amount. Overall, replacing one white student with one minority student in a "typical" conference increases the diversity index by an average of about 0.08 units. In subsequent discussions, I will use this change in magnitude as a basis for comparison.

Lastly, rows (f) and (g) show the extreme cases of conferences where all races are equally balanced and where there are only white students. Row (h) shows an example of what would happen to the diversity index if two classes with diversity combinations (f) and (g) of equal size were mixed together such that both classes have similar diversity configurations.

Table 2: Examples of Diversity Index Calculations

	Conference Example	White	Black	Hisp.	Asian / PI	Other / mult.	*diversity*
(a)	Sample-average proportions	0.646	0.020	0.051	0.087	0.195	0.668
(b)	Replace 1 white student with 1 black student	**0.581**	**0.085**	0.051	0.087	0.195	0.759
(c)	Replace 1 white student with 1 Hispanic student	**0.581**	0.020	**0.116**	0.087	0.195	0.754
(d)	Replace 1 white student with 1 Asian/PI student	**0.581**	0.020	0.051	**0.152**	0.195	0.748
(e)	Replace 1 white student with 1 other/mult. student	**0.581**	0.020	0.051	0.087	**0.260**	0.731
(f)	Equal proportions of all races	0.200	0.200	0.200	0.200	0.200	1.000
(g)	Only white students	1.000	0	0	0	0	0
(h)	Even mixture of (f) and (g)	0.600	0.100	0.100	0.100	0.100	0.750

Note: When replacing students, I use the average conference size of 15.3 to calculate a 0.065 proportion change. “Hisp.” refers to the Hispanic category; “Asian/PI” refers to the Asian / Pacific Islander category; “other/mult.” refers to other / multiple race category.

4.2 Effects of Diversity

To analyze the effect of diversity on academic outcomes, I estimate variations of the following reduced-form regression specification using ordinary least squares (OLS) for student i in conference c, scheduled at time slot t in year y.

$$\begin{aligned} outcome_{icty} = \beta diversity_{cty} + \sum_{g} \rho_g race\,(g)_{icty} + \alpha_1 SAT_{icty} + \alpha_2 meanSAT_{cty} \\ + \sum_{\hat{y}} \sum_{\hat{t}} \delta_{\hat{t}\hat{y}} free\left(\hat{t}\hat{y}\right)_{icty} + X_{icty}\gamma + \mu_y + \mu_t + Fac_{cty} + \varepsilon_{icty} \quad (1) \end{aligned}$$

where

- $outcome_{icty}$ is one of the (standardized) academic outcomes of interest (writing course grade, cumulative GPA at end of first year, cumulative GPA at graduation) for student i;

- $diversity_{cty}$ is the (normalized) diversity index in conference c at time slot t in year y;[14]

- $race\,(g)_{icty}$ is an indicator variable for student i being in race group g (either black, Hispanic, Asian / Pacific Islander, or other / multiple race; white is the omitted category);

- SAT_{icty} is the SAT score of student i (in thousands);

[14] This exposition is a slight simplification of the actual situation. While uncommon, students are allowed to request a change in conference group from Fall to Spring semester, in the event of a scheduling conflict. Thus, $diversity_{cty}$ here is in fact the average of 1) the diversity index in the Fall semester conference and 2) the diversity index in the Spring semester conference. This accounts for any slight shifts in the index from Fall to Spring should any student switch into or out of conferences. However, switching seldom occurs (8.5% of the time in the data) because course registration and scheduling for both semesters is done at the beginning of the academic year during the course sign-up phase as described in the previous section. This means that almost all students settle on a schedule for the entire academic year at the beginning of the Fall semester. It is only when students rearrange their schedule before the beginning of the Spring semester that switching of conference occurs. To check whether conference diversity in the fall semester has an "effect" on the likelihood of switching, I estimate both linear probability and probit models with an indicator for switching as the dependent variable; in both models, the coefficient on the fall measure of $diversity_{cty}$, the sole explanatory variable, is not statistically significant (with p-values of around 0.35 for both regressions).

- $meanSAT_{cty}$ is the mean SAT score (in thousands) across all students in conference c at time slot t in year y;[15]

- $free\left(\hat{t}\hat{y}\right)_{icty}$ is an indicator variable for whether student i in conference c at time slot t in year y is available (i.e. has no other class scheduled) at time slot $\hat{t}$ specific to the year $\hat{y}$;[16]

- X_{icty} is a vector of covariates included only in some specifications (male indicator, international status, conference size, number of course units, major area of study);

- μ_y are year fixed effects (across all students who took the writing course in year y);

- μ_t are time slot fixed effects (across all students who took the writing course in time slot t)[17] and

- Fac_{cty} are faculty instructor fixed effects.[18]

It is important to distinguish between a writing course conference c (which is the groupings of students within which they experience the structural diversity) and a time slot t (during which multiple conferences can be simultaneously scheduled). For example, a time slot would be Monday-Wednesday-Friday from 10am to 10:50am, and several conferences led by different instructors may be going on concurrently during this particular time slot.

[15] Similar to the situation described in Footnote 14, $meanSAT_{cty}$ here is in fact the average of 1) the mean SAT scores across all students in the Fall semester conference and 2) the mean SAT scores across all students in the Spring semester conference. This accounts for any slight shifts in mean SAT scores from Fall to Spring should any student switch into or out of conferences. Similar to the situation described in Footnote 13, when calculating the mean SAT score for a conference, I also include students in the conference who did not have a cumulative GPA at graduation reported.

[16] Because of the situation described in Footnote 14, the full set of free indicators includes free time slots in both the Fall and Spring semesters for every year $\hat{y}$. The set of time slots $\hat{t}$ comprise only time slots during which a writing course conference is offered, rather than the full universe of time slots available for scheduling at the college.

[17] Fixed effects for time slots in the Fall semester are used. Also note that these time slot fixed effects are 1) regardless of the particular conference c (as there may be multiple conferences going on during the same time slot t) and 2) regardless of the year y (because the time slots in which writing course conferences are scheduled are consistent across years).

[18] These fixed effects use faculty identifiers in the spring semester because writing course grades are assigned in the spring. They span conferences, time slots, and years, because faculty may teach multiple conferences across different time slots within a given year, or they may teach in multiple years.

The estimate of the coefficient β measures the effect of a one (index) unit increase in the racial diversity of a student's writing course conference on the academic outcome of interest, in σ units. As argued previously, this effect estimate has a causal interpretation because assignment of students to writing course conferences, and hence the racial diversity of any particular conference, is effectively random. Importantly, because selection into any particular conference time is conditional on the free time slots during which students' are not enrolled in another course, I account for this by including the full set of $free\left(\hat{t}\hat{y}\right)_{icty}$ indicator variables, which controls for students' scheduling availability. Note that these free indicators are summed across all possible writing time slots $\hat{t}$ over all possible years $\hat{y}$, even though the single observation of student i is enrolled in only one conference c at time t in year y.

The inclusion of SAT_{icty} controls for students' prior abilities. Hence, β can be interpreted as a value-added diversity effect on the assessed outcome—that is, the gain or loss in σ units caused by an increase in racial diversity in the writing course conference. The inclusion of $meanSAT_{cty}$ controls for linear-in-means peer effects emanating from higher-quality classmates in the same conference as student i. This ensures that the diversity effect is purely measuring the effect of structural (racial) diversity, as opposed to having conference peers of a certain race being higher- or lower-quality peers because of systemic inequity and racial segregation in the pre-tertiary education system. Race category indicators $race\left(g\right)_{icty}$ are included as controls because race may impact grades and because $diversity_{cty}$ is a function of student i's own race. Moreover, year fixed effects μ_y account for any academic-year- or cohort-specific differences. Time slot fixed effects μ_t account for any behavioral differences between students who willingly select into particular time slots (e.g. early birds in thrice-weekly Monday-Wednesday-Friday morning conferences, as opposed to late owls in twice-weekly Tuesday-Thursday afternoon conferences), even after conditioning on scheduling availability. Lastly, faculty instructor fixed effects Fac_{cty} capture instructor-specific differences such as instructor demographics and grading harshness.

A subset of regression specifications include additional covariates represented in X_{icty}. Indicators for being male and for being an international student are included because such predetermined characteristics may have an impact on grades. Conference size is included as there is much evidence in the education literature that class sizes matter.[19] Number of units taken are included because busier schedules may have an impact on overall grades. Lastly, some specifications control for the major area of study.[20] This last set of dummy variables can be considered "ex-post" for regressions where the dependent variable is measured in the first year, because students declare their majors only in the third year of study. However, even before declaring, students usually start taking coursework to work towards a particular major even in their first year. Accounting for major area of study is important because some majors may impose harsher grading schemes. Given the potential endogeneity of this last covariate though, it is included in only some specifications.

Regression coefficient estimates in Table 3 show the effect of diversity on writing course grade (columns (1) through (3)), on cumulative GPA at the end of the first year (columns (4) through (6)), and on cumulative GPA at graduation (columns (7) through (9)). Standard errors clustered at the writing course conference level are reported in parentheses. The first column in each set of three shows regression results for the specification of equation (1) without any covariates X_{icty}. The second column in each set shows regression results which add male and international status indicators, as well as conference size and the number of course units over the relevant time frame, as covariates. The third column in each set augments the specifications in the second column with category indicators for major area of study. Within each set of three specifications, estimates of the coefficient on diversity do not change substantially across them.

There does not appear to be any statistically significant average treatment effect of diversity on writing course grade (columns (1) through (3)) or cumulative GPA at the end of the first year (columns (4) through (6)). The point estimates across these different specifications

[19]See Lazear (2001); Krueger (1999); and Angrist and Lavy (1999) among others.

[20]See Footnote 10 for the categories used.

Table 3: Effect of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Writing Course Grade			Cum. GPA at End of 1st Year			Cum. GPA at Graduation		
Diversity	0.103	0.122	0.125	0.094	0.119	0.114	0.288**	0.255**	0.269**
	(0.092)	(0.093)	(0.093)	(0.101)	(0.103)	(0.102)	(0.118)	(0.119)	(0.119)
Own SAT	1.491***	1.454***	1.470***	2.006***	1.975***	1.908***	1.951***	2.165***	2.164***
	(0.119)	(0.118)	(0.118)	(0.113)	(0.112)	(0.111)	(0.140)	(0.138)	(0.139)
Mean Conf. SAT	-0.812*	-0.935*	-0.940*	0.050	-0.068	-0.072	1.506**	1.143*	1.106*
	(0.462)	(0.477)	(0.479)	(0.485)	(0.498)	(0.495)	(0.586)	(0.588)	(0.591)
Male		-0.149***	-0.145***		-0.158***	-0.166***		-0.257***	-0.249***
		(0.027)	(0.027)		(0.026)	(0.026)		(0.032)	(0.032)
International		-0.013	-0.022		0.214***	0.204***		0.252***	0.236***
		(0.056)	(0.056)		(0.060)	(0.059)		(0.071)	(0.071)
Conference Size		-0.023**	-0.022**		-0.010	-0.011		-0.005	-0.004
		(0.011)	(0.011)		(0.012)	(0.012)		(0.015)	(0.015)
1st Year Fall Units		0.068**	0.068**		0.138***	0.132***			
		(0.033)	(0.033)		(0.034)	(0.034)			
1st Year Spring Units		0.305***	0.295***		0.311***	0.324***			
		(0.029)	(0.029)		(0.028)	(0.028)			
Units at Graduation								0.033***	0.030***
								(0.006)	(0.006)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Area Indicators	No	No	Yes	No	No	Yes	No	No	Yes
N	4733	4733	4733	4733	4733	4733	4733	4733	4733
R-square	0.197	0.234	0.237	0.196	0.242	0.250	0.163	0.184	0.189

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: Grades and GPAs are standardized within year in σ units. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the writing course conference level. All regressions include race indicators, free indicators, year, time slot, and instructor fixed effects.

are all positive and around 0.1σ, but not statistically significant. This implies that diversity in writing course conferences does not have contemporaneous effects on outcomes during the first year while the students are in the writing course. However, the next set of estimates suggests diversity might have longer-term effects.

The effect of diversity on cumulative GPA at graduation is positive and statistically significant (columns (7) through (9)). The estimates suggest that a one-unit increase in the diversity index of the writing course conference (i.e. from a single-race classroom to one with equal proportions) increases graduation GPA by approximately 0.26σ (the preferred and most conservative estimate in column (8)). Alternatively, replacing one white student with one minority student in a "typical" conference increases graduation GPA by 0.02σ. In this student-replacement scenario, the change represents a 0.008 grade-point increase in cumulative GPA at graduation.

A brief examination of the coefficients on other explanatory variables reveals nothing unanticipated. Own SAT score (in thousands) is positively and strongly correlated with grades across all specifications and dependent variables examined. Mean SAT score of the conference (in thousands) negatively affects outcomes from the first year (and with statistical significance only for the writing grade); on the other hand, mean SAT score does have a positive impact on cumulative GPA at graduation. This is not as surprising as it seems. Grading for the writing course is likely to be curved within conference instructor, so ceteris paribus, ending up in a conference with better peers will make one's own performance look relatively worse. On the other hand, better peers would improve one's own human capital, affecting longer-term outcomes that depend on these earlier human capital accumulations—hence the positive and statistically significant estimates for cumulative GPA at graduation. Overall, male students do worse academically by all three measures. International students do better on average in two of the three measures, the exception being the writing course grade, in which they do no worse; this may be due to language barrier issues. As expected, larger conferences are associated with poorer writing grades, but have no impact on the other two non-writing GPA outcomes. Moreover, taking a greater number of units is positively associated with higher academic performance across all three dependent variables. This is likely because higher-ability students opt to take a greater number of units, but still manage to perform better in spite of the tougher workload.

4.3 Heterogeneity

I consider whether there is heterogeneity in the effects of diversity by augmenting the above regressions based on equation (1) with interaction terms. These interaction terms multiply the diversity index with the heterogeneous dimension being considered. In particular, I investigate whether an individual student's own sex, ability, or race has an impact on the magnitude of the diversity effect. These results are presented in Table 4. In these regressions, I include as controls X_{icty} a male indicator, international status, conference size, and number of course units, but not category dummies for major area of study; this is my preferred specification from columns (2), (5), and (8) in Table 3.

In Table 4, columns (1) through (3) show estimates for specifications with the writing course grade as the dependent variable. Columns (4) through (6) show estimates for specifications with cumulative GPA at the end of the first year as the dependent variable. Finally, columns (7) through (9) show estimates for specifications with cumulative GPA at graduation as the dependent variable. Within each set of three columns, the first column investigates heterogeneity by sex, using a male indicator interaction term. The second column within the set of three investigates heterogeneity by ability, using an SAT score interaction term. Lastly, the third column investigates heterogeneity by race, using a non-white minority indicator interaction term.

Being male does not reduce diversity's effect on grades by a statistically significant amount (columns (1), (4), and (7)). The coefficient estimates on the interaction terms between the diversity index and the male indicator are all negative, but not statistically distinguishable from zero.

Table 4: Heterogeneous Effects of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Writing Course Grade			Cum. GPA at End of 1st Year			Cum. GPA at Graduation		
Diversity	0.169	1.732*	0.130	0.178	1.433	0.137	0.289**	1.420	0.244*
	(0.117)	(0.974)	(0.105)	(0.126)	(0.934)	(0.116)	(0.140)	(1.136)	(0.131)
Diversity×Male	-0.111			-0.137			-0.080		
	(0.174)			(0.175)			(0.201)		
Diversity×Own SAT		-1.185			-0.967			-0.858	
		(0.721)			(0.692)			(0.845)	
Diversity×Minority			-0.031			-0.074			0.041
			(0.180)			(0.184)			(0.214)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Area Indicators	No	No	No	No	No	No	No	No	No
N	4733	4733	4733	4733	4733	4733	4733	4733	4733
R-square	0.234	0.234	0.234	0.242	0.242	0.242	0.184	0.184	0.184

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: Grades and GPAs are standardized within year in σ units. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the writing course conference level. All regressions include race indicators, free indicators, year, time slot, and instructor fixed effects, as well as other controls (own SAT score, mean SAT score in conference, male indicator, international status, conference size, number of course units).

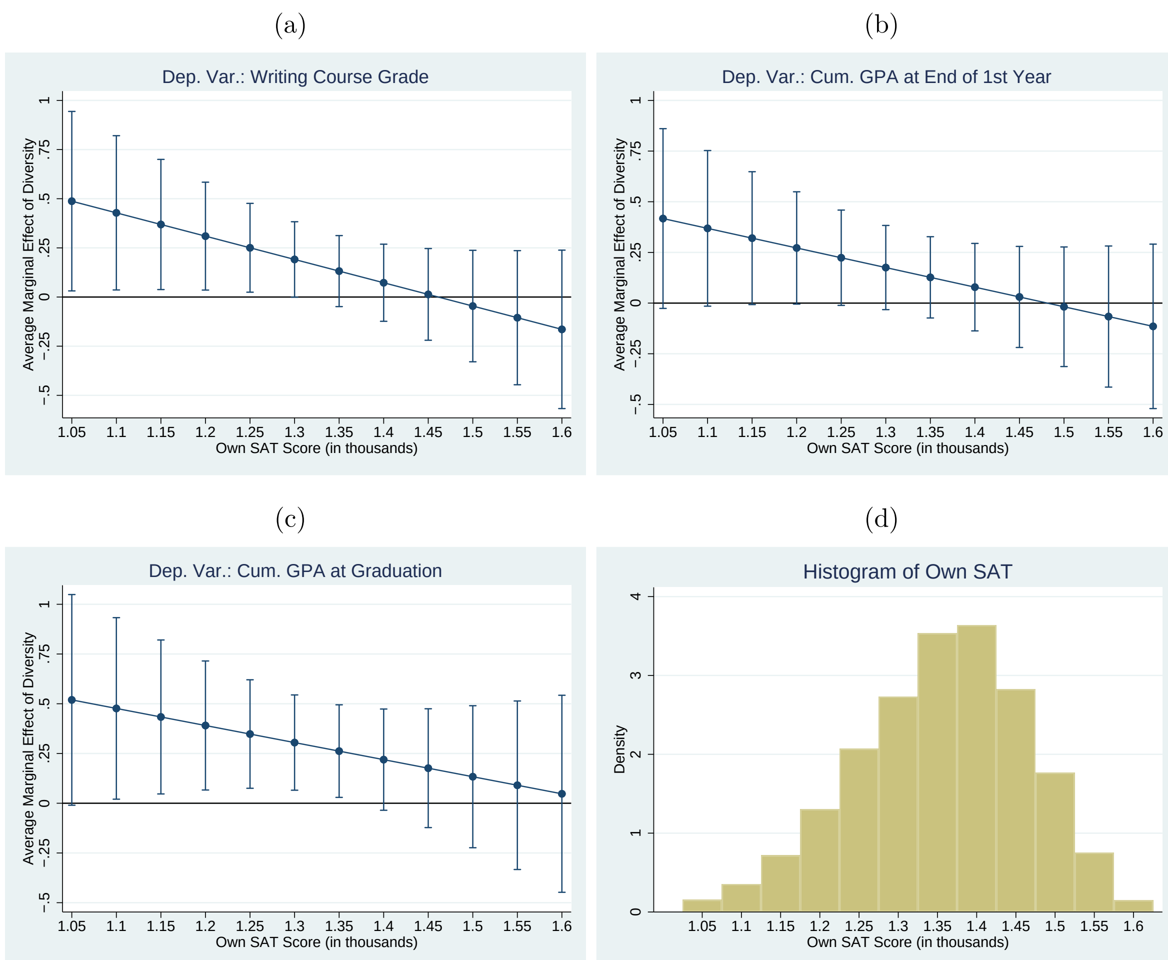
Lower ability students (where ability is proxied for by own SAT score) benefit more academically from the positive effects of diversity. This is not immediately obvious from the estimates in columns (2), (5) and (8) of Table 4. To see the heterogeneous effects, I use these regression results to calculate the average marginal effects (AMEs) of diversity on the three dependent variables at specific own SAT scores. I plot these AMEs, along with 95% confidence intervals, in panels (a) through (c) of Figure 2. Additionally, panel (d) is a histogram showing the distribution of own SAT scores.

Panel (a) of Figure 2 plots AMEs of diversity on the writing course grade. The downward sloping graph implies that diversity effects are stronger for students with lower SAT scores. For a student with a score of 1050, the AME of diversity on the writing course grade is 0.5σ, which is four times the point estimate in the uninteracted regression specification (column (2) of Table 3). On the other hand, students with SAT scores above 1300 experience no statistically significant diversity effect on the writing course grade.

Similar patterns are observed in panels (b) and (c) for the cumulative GPA outcomes at the end of first year and at graduation, respectively. For panel (b), point estimates of the diversity effect on cumulative GPA at the end of first year also decline as SAT scores increase, but none are statistically significant at the 5% level—even those at lower SAT score levels. Since a component of first year GPA is the writing course grade, this muted effect suggests that diversity in the classroom does not have contemporaneous spillover effects on other course grades in the first year.

On the other hand, diversity has statistically significant heterogeneous effects by ability beyond the first year, as displayed in the downward-sloping plot in panel (c). Here, we see again that the diversity effect on cumulative GPA at graduation declines as SAT scores increase, and all but one of the AMEs below the SAT score of 1400 are significant at the 5% level. For a student with a score of 1100, the AME of diversity on the writing course grade is just under 0.5σ, which is almost twice the estimated average treatment effect in the uninteracted regression specification (column (8) of Table 3).

Figure 2: Average Marginal Effects of Diversity by SAT Score



Notes: Grades and GPAs are standardized within year in σ units. SAT scores are expressed in thousands. For panels (a) through (c), bars represent 95% confidence intervals. Average marginal effects are calculated using regression estimates from columns (2), (5), and (8) of Table 4. To keep the x-axis scaling consistent across panels, the histogram in panel (d) excludes 55 observations (out of a regression sample of 4,733) with SAT scores below 1025.

Non-white minorities do not experience larger diversity effects compared to their white counterparts (columns (3), (6), and (9) in Table 4). When interacting the diversity index with an indicator for being a minority (either black, Hispanic, Asian / Pacific Islander, or other / multiple race), none of the coefficient estimates on this interaction term are statistically significant.

4.4 Major Choice

Interacting with racially-diverse classmates may influence the choice of major. To investigate this possibility, I run a multinomial logit regression with major area of study as the categorical dependent variable. To capture possibly heterogeneous effects of diversity across different student race groups, I include the diversity index as well as its interactions with each race group (white being the omitted category) as covariates in the multinomial logit specification. Other covariates include own SAT score, mean SAT score in conference, male indicator, international status, conference size, number of first year course units, race indicators, free indicators, as well as year, time slot, and instructor fixed effects.

Columns (1) and (2) of Table 5 report estimates of the marginal effects of the $diversity_{cty}$ variable on the probability of being in each of the four major areas of study for white students and minority students respectively. These average marginal effects are evaluated for white students and (non-white) minority students separately, based on estimates from a single multinomial logit regression.

The two statistically significant estimates within columns (1) and (2) imply that when white students are placed in higher diversity classrooms, they are less likely to take up majors in literature, language and arts, and more likely to take up majors in humanities and social sciences. This pattern of white students flowing from one major area to another may be the result of their being exposed to worldviews and social networks different from their own when placed in more racially diverse classrooms.

Table 5: Effect of Diversity on Major Area of Study

	(1)	(2)
Dependent Variable:	Marginal Effects of Diversity	
Pr (Major Area = . . .)	White	Minority
Lit., Lang. & Arts	-0.189***	-0.080
	(0.056)	(0.093)
Humanities & Social Sciences	0.149**	0.069
	(0.063)	(0.098)
Math & Natural Sciences	0.055	-0.010
	(0.055)	(0.077)
Inter-disciplinary / Double Major	-0.016	0.022
	(0.031)	(0.051)
N	3058	1675

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: Each column calculates average marginal effects based on estimates from the same multinomial logit regression, where each estimate reported is the marginal effect of a change in the diversity index on the probability of being in the major area of study for a given row. Standard errors in parentheses are calculated using the delta method. All specifications include as covariates the diversity index, its interaction with each race category, own SAT score, mean SAT score in conference, male indicator, international status, conference size, number of first year course units, race indicators, free indicators, as well as year, time slot, and instructor fixed effects.

4.5 Randomization Check

To confirm that the variation in racial diversity is indeed effectively random and exogenous for the identification strategy, I use a modified form of equation (1) in order to investigate whether the diversity index is correlated with certain predetermined student-level covariates. I use OLS to estimate the linear probability regression

$$x_{icty} = \beta diversity_{cty} + \sum_{g} \rho_g race\,(g)_{icty} + \sum_{\hat{y}} \sum_{\hat{t}} \delta_{\hat{t}\hat{y}} free\left(\hat{t}\hat{y}\right)_{icty} + \mu_y + \mu_t + \varepsilon_{icty} \qquad (2)$$

where

- x_{icty} is a predetermined covariate; and
- all other variables are as before.

If $diversity_{cty}$ is exogenous, then it should have no "impact" on any predetermined covariate x_{icty} conditional on the other factors mentioned previously. That is, the estimate of the coefficient β in equation (2) should be zero.

Table 6 presents the results from these randomization checks for six dependent variables. Four of these are from the vector X_{icty}: an indicator for being male, an indicator for being an international student, own SAT score and conference size.[21] Two additional dependent variables used to check for randomization are high school GPA and an admissions rating.[22] None of the estimates of the coefficient on $diversity_{cty}$ are statistically significant, consistent with racial diversity in writing course conferences being exogenous.

[21]Race category indicators cannot be used as a dependent variable in this check because a student's own race contributes to the calculation of the diversity index measure, thereby generating a mechanical relationship.

[22]These two variables were not included in X_{icty} in the main regression specifications because of collinearity with own SAT score.

Table 6: Randomization Checks

Dep. Var.:	(1) Male	(2) Int'l	(3) Own SAT	(4) Conf. Size	(5) High Sch. GPA	(6) Admit. Rating
Diversity	-0.049	0.017	-0.020	-0.045	0.002	0.053
	(0.053)	(0.031)	(0.014)	(0.401)	(0.065)	(0.057)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	4733	4733	4733	4733	3778	4706
R-square	0.143	0.200	0.236	0.586	0.206	0.402

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the writing course conference level. All regressions include race indicators, free indicators, year, time slot, and instructor fixed effects.

4.6 Discussion

How economically significant are these diversity effects? Consider the scenario of replacing one white student with one minority student in the "typical" conference, thereby increasing GPA at graduation by 0.008 grade-points (0.02σ). Jones and Jackson (1990) report that a 1 grade-point increase in GPA is associated with a 9% increase in earnings.[23] Assuming linearity, in this scenario, replacing one white student with one minority student would lead to an increase in annual earnings of 0.072%. While this estimate may seem small, it is by no means immaterial. First, one must remember that the effect impacts all 15.3 students in one conference. The national average annual earnings after attending college is $33,400 (United States Education Department, 2016); hence, the annual earnings increase is roughly $24.05 per student, or $368 per conference. Second, these calculated numbers are annual figures, so the total lifetime increase in earnings will be much higher. Moreover, these figures may be underestimating the impact in particular for minority students, who are known to have

[23] Loury and Garman (1993) report similar estimates separately for white and black males, and find that white male students with higher GPAs earn 6% more per grade-point, while black male students earn 27% more per grade-point.

higher returns on GPA (Loury and Garman, 1993).

Furthermore, these results highlight the possibility of costless yet allocative efficiency-enhancing reconfigurations of classroom diversity between different conferences that improve aggregate outcomes. As a thought experiment, consider the example calculations in rows (f), (g), and (h) in Table 2. Suppose there are currently two equal-sized conferences: one in which there are equal proportions of all races (row (f) with $diversity = 1$), and the other in which there are only white students (row (g) with $diversity = 0$). Now, suppose the students in these two conferences are reassigned such that there is an even mixture of diversity in both (row (h) with $diversity = 0.75$). That is, the new mixed classes have similar diversity configurations. Each student from the former class (f) loses 0.25 units of diversity, but each student from former class (g) gains 0.75 units of diversity. While this is by no means a Pareto improvement, one could conceivably create a system where winners compensate losers, though it is unclear who should obtain the initial "property rights" to being in a diverse classroom environment. What is surprising though is that just by rearranging students between two conferences, there is an efficiency-enhancing net average diversity gain of 0.50 units per student. For two average-sized conferences totaling 30.6 students, this costless intervention represents a \$4,600 net increase in aggregate annual earnings in our thought experiment.[24]

In addition to the need to spread minority students out between conferences in order to maximize aggregate diversity among all students and conferences, the heterogeneity results offer additional insights. First, that there are no differential effects between white students and minority students allays the anecdotal notion that white students are somehow hurt academically by being in more racially-diverse classrooms, all else being equal. Second, that lower ability students benefit more from racially-diverse classrooms means that they should receive priority when being assigned to high-diversity conferences. Undoubtedly,

[24]That is, 30.6 students $\times$ 0.50 net diversity gain per student $\times$ 0.10 grade-point effect (0.255σ) per diversity unit $\times$ 9% increase in earnings per grade point $\times$ \$33,400 annual earnings. As before, this calculation uses the coefficient estimates for cumulative GPA at graduation and assumes linearity.

implementing an efficient conference-assignment policy based on optimizing classroom racial diversity while conditioning on student ability will be complicated by student scheduling constraints and other factors.

The policy relevance of these results goes beyond prescriptions for how to assign a fixed set of students between conferences. Assuming there is not a saturation of minority students, these positive estimates of the diversity effect offer modest justification for race-based admissions policies favoring minority students, ceteris paribus. Implementing such policies would increase both the diversity of the admitted cohort as a whole, and the diversity within individual classrooms, as long as the admitted students are not segregated into classrooms by race post-matriculation. On the other hand, efficiency considerations aside, such a policy raises equity concerns given that there will be winners (admitted students who now benefit from a more diverse learning environment) and losers (students who are no longer admitted).

5 Mechanisms

Numerous mechanisms could explain the positive effect of diversity on academic outcomes. Below, I consider some possibilities and whether their explanations are consistent with the empirical findings.

First, diversity can result in **human capital complementarities**. Suppose there are different types of human capital used as inputs in education production. (These can be thought of as specific skills such as reading comprehension or writing skills, but also as specific sets of knowledge or experiences that students have.) Students of different races may possess different relative amounts of these different human capital types, and these different types may complement one another in the education production function. If human capital types are transmitted from one student to another through peer effects within the classroom, then a more racially diverse classroom will gain improved complementarities in education production as the students share a more diverse pool of human capital through peer effects. This is consistent with prior research that suggests more racially diverse environments stimulate more complex and novel thinking (Antonio et al., 2004).

Second, a diverse classroom can combat **stereotype threat** (Spencer et al., 2016; Dills, 2018). Stereotype threat arises when students worry that their actions reaffirm certain negative stereotypes about their own racial group. The resulting distress leads to worse academic performance, above and beyond the typical pressures to perform faced by students who do not face stereotypes. When a classroom is more diverse, the racial group being stereotyped may be large enough to dispel the threat of stereotyping. In effect, by having a greater number of dissimilar peers within the stereotyped group, it is more difficulty for stereotypes to stick, thereby lowering the worry created by such stereotypes. With this worry gone, members of the stereotyped group (most often the minority group) can now focus efforts on improving academic outcomes.

Third, diversity can further expand the development of **social networks** beyond writing conference peers. Scholars have classified diversity as arising either formally or informally in educational settings (Denson and Chang, 2009). The first route (also known in the literature as classroom diversity) stems from experiencing diversity within formalized settings established by the institution, such as in lectures, conferences, or seminars. The second route (also known in the literature as informal interactional diversity) describes situations outside formalized settings where diverse experiences can be had, such as while living together in dormitories or attending social events. Having diverse peers in formal settings may increase the diversity of peers in informal settings, insofar as the writing conference peer groups are correlated with social circles developed beyond the classroom. This in turn amplifies any diversity effects due to other mechanisms, which ultimately lead to improvements in academic outcomes. If these social networks persist over the students' college lives, then diversity effects will be observed over the long term and on course grades other than those of the writing course.

The above empirical findings are most consistent with the first and third mechanisms.

That low SAT scorers see positive diversity effects on their writing course grade aligns with the human capital complementarities story. High SAT scoring students may already have high levels of human capital across all human capital types, so it is the low SAT scorers who benefit the most from classroom diversity. The positive effects of diversity on the longer term and more broadly-encompassing outcome of GPA at graduation suggests that the diversity effects from the writing course are spreading beyond the writing conference peer group to other social settings, consistent with the third mechanism. Expanded social networks can also explain why white students who experience more diversity make different choices with respect to certain major areas of study, insofar as student social networks have influence over one's academic interests.

On the other hand, stereotype threat does not appear to be the mechanism driving the empirical results. If stereotype threat were in fact being mitigated by classroom racial diversity, one would expect the stereotyped groups—in this case, the non-white minorities—to enjoy larger treatment effects of diversity (relative to non-stereotyped white students). However, I do not detect heterogeneous effects by race in any of the academic outcomes examined.

6 Conclusion

The findings in this paper suggest that a greater degree of racial diversity in the classroom causes a statistically significant increase in the cumulative grade point average (GPA) at graduation. I also find that students with lower incoming SAT scores experience statistically significant benefits from racial diversity in the classroom in terms of writing course grade and GPA at graduation. On the other hand, I do not detect heterogeneous average treatment effects between male and female students, or between white and minority students. Furthermore, I find that white students in higher diversity classrooms are less likely to take up majors in literature, language and arts, and more likely to take up majors in humanities and social sciences.

This paper makes several contributions. First, I exploit a quasi-experimental identification strategy which generates internally-valid causal effects, given the exogenous variation in racial diversity. Second, the writing course conference offers a real-world classroom setting in which racial diversity plays a vital human capital role through in-class discussions and debates. This setting is similar to numerous other classroom contexts in higher education and lends credence to the external validity of the estimated effects. This contribution is especially pertinent in comparison to previous experimental studies that identify causal effects, but which were conducted in more controlled settings. Lastly, my use of relevant methodological concepts and administrative data in constructing appropriate measures enables me to precisely quantify both diversity and academic outcomes.

Future research avenues include examining the effect of classroom diversity on outcomes besides academic performance, such as social or (post-graduation) labor market outcomes. Moreover, the findings and mechanisms developed here in this classroom setting could be applicable to workplace settings as well, where diversity effects on worker productivity or earnings could stem from workplace racial diversity. The positive effect of diversity on academic outcomes found in this paper contributes but one important piece to the larger picture concerning the value of racial diversity in higher education and potentially in society more broadly.

References

Alcalde, Jose and Begona Subiza (2014) "Affirmative action and school choice," *International Journal of Economic Theory*, Vol. 10, No. 3, pp. 295–312.

Alger, Jonathan R., Jorge Chapa, Roxane Harvey Gudeman, Patricia Marin, Geoffrey Maruyama, Jeffrey F. Milem, Jose F. Moreno, and Deborah J. Wilds (2000) "Does Diversity Make a Difference? Three Research Studies on Diversity in College Classrooms," report, American Council on Education and American Association of University Professors, Washington, DC.

Angrist, Joshua D. and Kevin Lang (2004) "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *The American Economic Review*, Vol. 94, No. 5, pp. 1613–1634.

Angrist, Joshua D. and Victor Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 533–575.

Antonio, Anthony Lising, Mitchell J. Chang, Kenji Hakuta, David A. Kenny, Shana Levin, and Jeffrey F. Milem (2004) "Effects of Racial Diversity on Complex Thinking in College Students," *Psychological Science*, Vol. 15, No. 8, pp. 507–510.

Arcidiacono, Peter and Michael Lovenheim (2016) "Affirmative Action and the Quality-Fit Trade-Off," *Journal of Economic Literature*, Vol. 54, No. 1, pp. 3–51, March.

Arcidiacono, Peter and Jacob L Vigdor (2010) "Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College," *Economic Inquiry*, Vol. 48, No. 3, pp. 537–557.

Bowen, William G and Derek Bok (1998) *The shape of the river: Long-term consequences of considering race in college and university admissions*: Princeton University Press.

Broad, Molly C. (2012) "Sunday Dialogue: Using Race in Admissions," in *New York Times Sunday Review*. Letters to the Editor, October 6, 2012. Retrived from `https://www.nytimes.com/2012/10/07/opinion/sunday/sunday-dialogue-using-race-in-admissions.html`.

Carrell, Scott E., Mark Hoekstra, and James E. West (2019) "The Impact of College Diversity on Behavior towards Minorities," *American Economic Journal: Economic Policy*, Vol. 11, No. 4, pp. 159–182, November.

Chan, Jimmy and Erik Eyster (2003) "Does Banning Affirmative Action Lower College Student Quality?" *The American Economic Review*, Vol. 93, No. 3, pp. 858–872.

College Board (2009) "ACT and SAT Concordance Tables," Research Note RN-40, College Board Office of Research and Development.

CollegeData (2017) "College Profile: University of Texas at Austin." Retrived from `http://www.collegedata.com/cs/data/college/college_pg04_tmpl.jhtml?schoolId=788` on June 19, 2017.

Daniel, Kermit, Dan A Black, and Jeffrey Smith (2001) "Racial Differences in the Effects of College Quality and Student Body Diversity on Wages," in Gary Orfield ed. *Diversity Challenged: Evidence on the Impact of Affirmative Action*: Harvard Education Publishing Group, pp. 221–231.

Denson, Nida and Mitchell J. Chang (2009) "Racial Diversity Matters: The Impact of Diversity-Related Student Engagement and Institutional Context," *American Educational Research Journal*, Vol. 46, No. 2, pp. 322–353.

Dills, Angela K. (2018) "Classroom Diversity and Academic Outcomes," *Economic Inquiry*, Vol. 56, No. 1, pp. 304–316.

Ehlers, Lars, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim (2014) "School choice with controlled choice constraints: Hard bounds versus soft bounds," *Journal of Economic Theory*, Vol. 153, pp. 648–683.

Epple, Dennis and Richard E. Romano (2011) "Peer Effects in Education: A Survey of the Theory and Evidence," in *Handbook of Social Economics*, Vol. 1B: Carnegie Mellon University, pp. 1053–1163.

Epple, Dennis, Richard Romano, and Holger Sieg (2008) "Diversity and Affirmative Action in Higher Education," *Journal of Public Economic Theory*, Vol. 10, No. 4, pp. 475–501.

Fisher v. University of Texas (2016) *579 U.S. —*: Supreme Court.

Fryer, Roland G., Glenn C. Loury, and Tolga Yuret (2008) "An Economic Analysis of Color-Blind Affirmative Action," *Journal of Law, Economics, and Organization*, Vol. 24, No. 2, pp. 319–355.

Gratz v. Bollinger (2003) *539 U.S. 244, 123 S. Ct. 2411, 156 L. Ed. 2d 257*: Supreme Court.

Grutter v. Bollinger (2003) *539 U.S. 306, 123 S. Ct. 2325, 156 L. Ed. 2d 304*: Supreme Court.

Gurin, Patricia, Eric Dey, Sylvia Hurtado, and Gerald Gurin (2002) "Diversity and Higher Education: Theory and Impact on Educational Outcomes," *Harvard Educational Review*, Vol. 72, No. 3, pp. 330–367.

Guryan, Jonathan (2004) "Desegregation and Black Dropout Rates," *The American Economic Review*, Vol. 94, No. 4, pp. 919–943.

Hafalir, Isa E., M. Bumin Yenmez, and Muhammed A. Yildirim (2013) "Effective affirmative action in school choice," *Theoretical Economics*, Vol. 8, No. 2, pp. 325–363.

Hinrichs, Peter (2011) "The Effects of Attending a Diverse College," *Economics of Education Review*, Vol. 30, No. 2, pp. 332–341.

Hopwood v. State of Texas (1996) *78 F.3d 932 5th Cir.*: Court of Appeals, 5th Circuit.

Hu, Shouping and George D Kuh (2003) “Diversity experiences and college student learning and personal development,” *Journal of College Student Development*, Vol. 44, No. 3, pp. 320–334.

Jones, Ethel B. and John D. Jackson (1990) “College Grades and Labor Market Rewards,” *The Journal of Human Resources*, Vol. 25, No. 2, pp. 253–266.

Kahane, Leo, Neil Longley, and Robert Simmons (2013) “The Effects of Coworker Heterogeneity on Firm-Level Output: Assessing the Impacts of Cultural and Language Diversity in the National Hockey League,” *Review of Economics and Statistics*, Vol. 95, No. 1, pp. 302–314.

Krueger, Alan B. (1999) “Experimental Estimates of Education Production Functions,” *The Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 497–532.

Lazear, Edward P. (2001) “Educational Production,” *The Quarterly Journal of Economics*, Vol. 116, No. 3, pp. 777–803.

Loury, Linda Datcher and David Garman (1993) “Affirmative Action in Higher Education,” *The American Economic Review*, Vol. 83, No. 2, pp. 99–103.

Ozgen, Ceren, Peter Nijkamp, and Jacques Poot (2013) “The impact of cultural diversity on firm innovation: evidence from Dutch micro-data,” *IZA Journal of Migration*, Vol. 2, No. 1, p. 18.

Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova (2014) “Labor diversity and firm productivity,” *European Economic Review*, Vol. 66, pp. 144–179.

Regents of the University of California v. Bakke (1978) *438 U.S. 265, 98 S. Ct. 2733, 57 L. Ed. 2d 750*: Supreme Court.

Rothman, Stanley, Seymour Martin Lipset, and Neil Nevitte (2003) "Does Enrollment Diversity Improve University Education?" *International Journal of Public Opinion Research*, Vol. 15, No. 1, pp. 8–26.

Sacerdote, Bruce (2014) "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics*, Vol. 6, No. 1, pp. 253–272.

Sander, Richard H. (2004) "A Systemic Analysis of Affirmative Action in American Law Schools," *Stanford Law Review*, Vol. 57, No. 2, pp. 367–484, November.

Spencer, Steven J., Christine Logel, and Paul G. Davies (2016) "Stereotype Threat," *Annual Review of Psychology*, Vol. 67, No. 1, pp. 415–437. PMID: 26361054.

Terenzini, Patrick T., Alberto F. Cabrera, Carol L. Colbeck, Stefani A. Bjorklund, and John M. Parente (2001) "Racial and Ethnic Diversity in the Classroom: Does It Promote Student Learning?" *The Journal of Higher Education*, Vol. 72, No. 5, pp. 509–531.

Trax, Michaela, Stephan Brunow, and Jens Suedekum (2015) "Cultural diversity and plant-level productivity," *Regional Science and Urban Economics*, Vol. 53, pp. 85–96.

Umbach, Paul D. and George D. Kuh (2006) "Student Experiences with Diversity at Liberal Arts Colleges: Another Claim for Distinctiveness," *The Journal of Higher Education*, Vol. 77, No. 1, pp. 169–192.

United States Education Department (2016) "College Scorecard." Data updated September 13, 2016. Retrieved from `https://collegescorecard.ed.gov/`.

Supplemental Appendices

A Expanded Sample

The analyses in the main paper are restricted to students for whom a final cumulative GPA at graduation is observed. This is done to maintain a balanced sample consistent across regression specifications. Doing so may introduce sample attrition bias, if diversity affects the likelihood of observing a cumulative GPA at graduation. To address this potential issue, I relax this sample restriction in this appendix, and present two sets of findings.

A.1 Degree Completion

First, I investigate whether diversity affects degree completion (i.e. sample non-attrition). I define degree completion as having a cumulative GPA at graduation observed in the data and a passing senior thesis grade. As mentioned in Footnote 7, observing a cumulative GPA at graduation is almost always indicative of degree completion, but not always. In the sample, 20 students have a cumulative GPA at graduation reported, but failed their senior year and so did not actually graduate. (Including / excluding these 20 students does not change the findings in this section.) This exception notwithstanding, the most common reasons for degree non-completion include dropping out of college or transferring to another institution.

Table 7 shows summary statistics calculated for students who completed the degree and those who did not complete the degree. (The only difference compared to statistics in Table 1 is that the ones here exclude the 20 exception students.) The statistics indicate that on average, students who do not complete the degree receive worse grades and take fewer units in their first year. Demographically however, they do not seem to be different from completers, except that non-completers are more likely to be male. This suggests that observations are not dropping out of the analysis sample predominately because of specific predetermined student characteristics.

Table 7: Summary Statistics by Degree Completion

Variable / Indicator	(1) Completed Degree	(2) Non-completers
Writing Course Grade	3.198	2.813
	(0.517)	(0.819)
Cumulative GPA at End of 1st Year	3.092	2.708
	(0.475)	(0.721)
Cumulative GPA at Graduation	3.191	
	(0.386)	
Fall Units	3.650	3.535
	(0.419)	(0.474)
Spring Units	4.034	3.787
	(0.523)	(0.568)
Cumulative Units at Graduation	30.345	
	(2.275)	
White	0.646	0.651
	(0.478)	(0.477)
Black	0.021	0.023
	(0.142)	(0.149)
Hispanic	0.051	0.056
	(0.219)	(0.23)
Asian / Pacific Islander	0.087	0.075
	(0.282)	(0.264)
Other / Multiple Race	0.195	0.195
	(0.396)	(0.396)
Male	0.444	0.499
	(0.497)	(0.500)
International Student	0.056	0.048
	(0.229)	(0.213)
Own SAT Score (in thousands)	1.354	1.338
	(0.113)	(0.118)
Diversity Index	0.621	0.518
	(0.156)	(0.133)
N Students in Sample	4713	1195

Notes: Grades and GPAs are measured in grade-points. Diversity Index refers to the normalized diversity index. Standard deviations reported in parentheses.

Columns (1) through (3) of Table 8 show linear probability regression results for the specification of equation (1), with the degree completion indicator as the dependent variable. These three columns of specifications are structured identically to the three-column sets in Table 3, where further covariates are progressively added.[25] The first two coefficient estimates on the diversity variable are statistically insignificant, while the third is significant only at the 10% level. This suggests that diversity in the classroom has little impact on degree completion, and that attrition bias in the main sample is not a serious concern.

The remaining coefficient estimates in columns (1) through (3) indicate other interesting patterns. Male students are more likely to drop out. On the other hand, students who take more units in their first year are more likely to remain in the sample and complete the degree. As mentioned before, better students on average take a greater number of course units, and are more likely to graduate.

I do not consider correlations to between these other covariates and degree completion to be a pressing issue with respect to sample attrition bias for two reasons. First, these variables are included as controls in most of the analyses in the main paper; this inclusion accounts for any bias generated from differential attrition. Second, classroom diversity is uncorrelated with these variables and hence still exogenous, so the diversity effect estimates remain internally valid.

A.2 First-year Outcomes

Next, I re-include into the analysis sample students for whom a cumulative GPA at graduation are not reported. Using this expanded sample, I rerun OLS regressions for the specification of equation (1) for two dependent variables: the writing course grade and cumulative GPA at the end of the first year. Columns (4) through (6) of Table 8 show results from specifications where the writing course grade is the dependent variable, while columns (7) through (9) show results from specifications where cumulative GPA at the end of the first year is the

[25]For the major areas categorical variable in the third column, the category of "undeclared" is added.

Table 8: Diversity Effect of Expanded Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Degree Completion			Writing Course Grade			Cum. GPA at End of 1st Year		
Diversity	-0.079	-0.066	-0.090*	-0.015	0.026	0.016	0.018	0.056	0.039
	(0.054)	(0.052)	(0.047)	(0.096)	(0.095)	(0.096)	(0.102)	(0.101)	(0.100)
Own SAT	0.085	0.052	0.095*	1.391***	1.33***	1.378***	1.855***	1.809***	1.798***
	(0.056)	(0.057)	(0.052)	(0.118)	(0.114)	(0.114)	(0.115)	(0.111)	(0.111)
Mean Conf. SAT	-0.222	-0.228	-0.271	-0.904*	-1.007**	-1.063**	-0.097	-0.268	-0.299
	(0.218)	(0.218)	(0.199)	(0.488)	(0.479)	(0.477)	(0.505)	(0.483)	(0.481)
Male		-0.032***	-0.021*		-0.220***	-0.206***		-0.239***	-0.236***
		(0.012)	(0.011)		(0.027)	(0.026)		(0.027)	(0.026)
International		0.014	0.022		-0.034	-0.033		0.209***	0.203***
		(0.025)	(0.024)		(0.066)	(0.066)		(0.067)	(0.067)
Conference Size		-0.002	-0.002		-0.023**	-0.024**		-0.010	-0.010
		(0.006)	(0.005)		(0.011)	(0.011)		(0.011)	(0.011)
1st Year Fall Units		0.075***	0.060***		0.167***	0.158***		0.247***	0.234***
		(0.015)	(0.013)		(0.037)	(0.036)		(0.036)	(0.035)
1st Year Spring Units		0.094***	0.070***		0.420***	0.393***		0.436***	0.429***
		(0.012)	(0.011)		(0.032)	(0.031)		(0.030)	(0.030)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Area Indicators	No	No	Yes	No	No	Yes	No	No	Yes
N	5914	5914	5914	5908	5908	5908	5913	5913	5913
R-square	0.107	0.131	0.273	0.152	0.214	0.226	0.150	0.229	0.242

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: Grades and GPAs are standardized within year in σ units. Degree completion indicator is dependent variable of a linear probability model estimated using OLS. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the writing course conference level. All regressions include race indicators, free indicators, year, time slot, and instructor fixed effects.

dependent variable. As before, these sets of three columns are structured identically to the three-column sets in Table 3.

Using the expanded sample, the point estimates of the effect of diversity on both outcomes are still statistically insignificant. Compared to estimates from the main specifications, the magnitude of the effect estimates are much smaller. This implies that the effect of diversity is being dampened by a non-positive diversity effect from the now-included non-completers. In regressions not reported here, I run the same specifications but including only non-completers in the sample; these lead to negative but statistically insignificant point estimates. Thus, while diversity does not affect the likelihood of dropping out for this non-completing group, we also cannot reject the null hypothesis that diversity has no effect on the writing course grade and cumulative GPA at the end of the first year. The remaining coefficient estimates for these two outcomes follow a similar pattern as those estimates obtained before in Table 3.

The findings in this appendix suggest that there are heterogeneous effects for these two groups when examining the two particular first-year outcomes in question. They also reiterate the need to qualify the estimates in the main analysis as the effects of diversity conditional on not dropping out of the sample; that is, diversity effects conditional on graduating.[26]

B Alternative Measures of Diversity

This appendix considers two alternate measures of diversity: a peer diversity index and Shannon entropy. Overall, the results presented in the main paper are robust to the use of these alternative measures of diversity.

Table 9 contains three panels of regression results. The columns of specifications are structured identically to the columns in Table 3; however, coefficients on covariates besides the diversity measure are omitted for compactness of exposition. Panel (A) reproduces the regression results in Table 3 using the original diversity index for comparison purposes.

[26]These are effectively equivalent; see Footnote 7.

Table 9: Effect of Alternate Measures of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Writing Course Grade			Cum. GPA at End of 1st Year			Cum. GPA at Graduation		
Panel (A)									
Diversity	0.103	0.122	0.125	0.094	0.119	0.114	0.288**	0.255**	0.269**
	(0.092)	(0.093)	(0.093)	(0.101)	(0.103)	(0.102)	(0.118)	(0.119)	(0.119)
R-square	0.197	0.234	0.237	0.196	0.242	0.250	0.163	0.184	0.189
Panel (B)									
Peer Diversity	0.082	0.096	0.099	0.080	0.100	0.094	0.275**	0.242**	0.253**
	(0.085)	(0.085)	(0.086)	(0.093)	(0.095)	(0.094)	(0.110)	(0.112)	(0.111)
R-square	0.197	0.234	0.237	0.196	0.242	0.250	0.163	0.184	0.189
Panel (C)									
Shannon Entropy	0.043	0.059	0.063	0.043	0.061	0.051	0.175**	0.158**	0.165**
	(0.059)	(0.06)	(0.06)	(0.066)	(0.067)	(0.067)	(0.078)	(0.079)	(0.078)
R-square	0.197	0.234	0.237	0.196	0.242	0.250	0.163	0.183	0.189
Panel (D)									
Prop. Non-White	0.121	0.123	0.123	0.127	0.141	0.145	0.402***	0.356**	0.371**
	(0.118)	(0.118)	(0.118)	(0.124)	(0.127)	(0.126)	(0.147)	(0.148)	(0.148)
R-square	0.197	0.234	0.237	0.196	0.242	0.25	0.163	0.184	0.189
N (all specifications)	4733	4733	4733	4733	4733	4733	4733	4733	4733
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Major Area Indicators	No	No	Yes	No	No	Yes	No	No	Yes

Significance Levels: *** = 1%; ** = 5%, * = 10%

Notes: Grades and GPAs are standardized within year in σ units. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the writing course conference level. All regressions include race indicators, free indicators, year, time slot, and instructor fixed effects. Other controls are own SAT score, mean SAT score in conference, male indicator, international status, number of course units.

Panel (B) of Table 9 replaces the diversity index with a peer diversity index. The calculation of this peer diversity index is identical to the original diversity index *except* that the student's own race is excluded in the race proportions used. That is, if there were 15 students in the conference inclusive of the student of the current observation, then the peer diversity index uses race proportions of the remaining 14 students (the peers of the "own" student) to calculate the new index. This is analogous to the use of peer average test scores (excluding own test score) as a measure of peer quality in the peer effects literature. Comparing Panels (A) and (B), the coefficient estimates found using the peer diversity index are indistinguishable from the estimates found using the original diversity index.

I present the regression results using this peer diversity index measure for completeness, to parallel with the peer effects literature. However, the reason I prefer the original diversity index over the peer diversity index is because a student's own race clearly contributes to the diversity of the conference as a whole. For instance, consider the simplified case where you are one of a total of 3 students in a conference. Suppose you are black and the other two students are white. The peer diversity index disregards your own race and measures the diversity of the conference as 0, even though you yourself are a black student. On the other hand, the original diversity index takes your own race into account and measures the diversity of the conference as 0.556 (for the 5 race category case). This latter measure seems more sensible given that being a conference of 1 black and 2 white students, there is clearly some degree of diversity in the conference being experienced by yourself (the black student) even if your peers are all white. Furthermore, the production of education output may depend on one's own human capital (and thus, one's own race) as well as the human capital / race of one's peers. In this sense, the production complementarities gained from racial diversity depend on a measure of the diversity of the entire group, and not just that of the peer group. Regardless, the estimates of the diversity effect for both diversity indices are nearly identical. Moreover, since one's own race is included as a covariate in all regression specifications, using either measure does not make much difference econometrically—the only difference is in the interpretation of the coefficient estimates.

Panel (C) of Table 9 replaces the diversity index with the Shannon entropy measure of diversity. Shannon entropy is a popular measure of diversity among ecologists, and is calculated as

$$Shannon = -\sum_{g} \left[proportion\,(g) \times \ln\left(proportion\,(g)\right)\right]$$

where $proportion\,(g)$ is the proportion of students in the conference belonging to race group g. The higher the Shannon entropy, the greater the degree of diversity. Unlike the original diversity index, Shannon entropy is not bounded between 0 and 1. Given this change in units, the magnitudes of the regression coefficient estimates in Panel (C) are not directly comparable to those of Panel (A). Nonetheless, the estimates are all in the positive direction, and the relative magnitudes of the estimates between different dependent variables exhibit a similar pattern as the original estimates. As before, only the estimates in the last three columns are statistically significant. Reinterpreting these estimates in terms of hypothetical student replacements as in Section 4.2 yield similar findings.

Panel (D) of Table 9 replaces the diversity index with the proportion of non-White students. This is a common proxy for diversity used in the literature, where a higher proportion of non-White students is considered a more diverse setting. Again, given the change in units, the magnitudes of the regression coefficient estimates in Panel (D) are not directly comparable to those of Panel (A). Nonetheless, the estimates are all in the positive direction, and relative magnitudes exhibit a similar pattern compared to the original estimates. As before, only the estimates in the last three columns are statistically significant. It should be noted that proportion non-White is highly correlated with the diversity index, with a correlation coefficient of 0.9455.