Does Classroom Diversity Improve Academic Outcomes?*

Yan Lau†

December 2016

Abstract

This paper estimates the causal effect of racial diversity in the classroom on academic outcomes. I exploit a quasi-experimental setting where first-year students in a year-long mandatory humanities writing course at a U.S. liberal-arts college are assigned to discussion conference groups with varying levels of diversity in terms of classmate racial composition. This within-classroom diversity is effectively random conditional on students' scheduling availability, 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 find that a higher degree of racial diversity in the conference causes a statistically significant increase in the humanities course grade and the grade point average (GPA) at graduation. However, diversity has no statistically significant effect on GPA at the end of the first year. These results contribute to the debate over affirmative action in higher education, and offer modest justification for race-based admissions policies.

*I would like to thank Lars Lefgren for invaluable feedback on this project. I am indebted to John Colgrove and Mike Tamada for help with obtaining and understanding the data. Alicia Toshima provided excellent research assistance.

†Economics Department, Reed College, Portland, OR [Email: yanlau@reed.edu]

1 Introduction

Does being exposed to a racially-diverse classroom setting improve the academic outcomes of college students? One of the main rationales for affirmative action in higher education is to foster a diverse environment in which students are exposed to worldviews different from their own. Through interactions with one another, these students exchange unfamiliar ideas and opinions, thereby challenging prior belief systems and enriching the overall learning experience. Contenders argue that increasing diversity on college campuses potentially leads to better student outcomes in various ways. The most immediate educational benefit that comes to mind are improvements in grades while in college, but there are also other outcomes to consider. Graduates who possess this broadened perspective in a globalized world may possess job skills important to employers, thereby enjoying higher incomes following graduation. Society-at-large benefits from having democratically-engaged citizens who understand an increasingly multicultural society and who do not hold unfounded stereotypes.

Given that such positive effects exist, the policy implications are clear: if the marginal benefit to educational outcomes is positive for a marginal increase in diversity (net any marginal costs), then higher education administrators should increase the diversity of their student bodies. Advocates have focused on increasing what scholars in the field term structural diversity (diversity in terms of the racial composition of students) through affirmative action policies that give college admission preferences in favor of minority students. A host of legal cases over the years have reaffirmed the constitutionality and strengthened the application of affirmative action policies in U.S. colleges (Regents of the University of California v. Bakke, 1978; Grutter v. Bollinger, 2003; Fisher v. University of Texas, 2016), though not without challenge (Hopwood v. State of Texas, 1996; Gratz v. Bollinger, 2003). Many of these considerations are predicated on the premise that the marginal benefit of diversity is indeed positive. Thus, estimating the effect of diversity is crucial for justifying race-based admissions policies.

This paper contributes to the debate by focusing on the effects of racial diversity on academic outcomes. To answer the question of 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 humanities writing course at a liberal-arts college are assigned to discussion conference groups (sections). 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. Hence, this identification strategy estimates the causal effect of diversity on grades, avoiding selection bias arising in other situations where (say) better students may select themselves into more diverse environments.

I find that a higher degree of racial diversity in the conference causes a statistically significant increase in the humanities course grade and the cumulative grade point average (GPA) at graduation. However, diversity has no statistically significant effect on GPA at the end of the first year. I do not detect heterogeneous effects between male and female students, or between white and minority students. On the other hand, higher ability students benefit less from racial diversity in the classroom.

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; Terenzini et al., 2001; Gurin et al., 2002; Hu and Kuh, 2003; Umbach and Kuh, 2006; Denson and Chang, 2009), as well as on attitudes and perceptions (Rothman et al., 2003; Umbach and Kuh, 2006; Denson and Chang, 2009). Beyond tertiary education, several papers have quantified the effects of desegregation on educational outcomes in the United States, both at specific diversity-inducing programs (Angrist and Lang, 2004) and at a more macro level (Guryan, 2004). Related literature also considers the impact of race-blind (as opposed to race-conscious) college admission policies (Chan and Eyster, 2003; Epple et al., 2008; Fryer et al., 2008), as well as the interaction between affirmative action and school choice (Hafalir et al., 2013; Alcalde and Subiza, 2014; Ehlers et al., 2014).[1]

The main contribution of this paper to the literature is the estimation of causal effects of diversity on college grades. 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). While results in some experimental studies that randomize diversity in lab settings do estimate causal effects (e.g. Antonio et al., 2004), these suffer from generalizability issues. The quasi-experimental approach this paper adopts involves a real-world setting that offers both causal estimates with internal validity and a greater degree of external validity. Furthermore, many studies use survey data with self-reported measures of both outcomes and diversity, which may be unreliable and imprecise (e.g. Gurin et al., 2002; Hu and Kuh, 2003; Umbach and Kuh, 2006; Denson and Chang, 2009). My use of administrative student data and focus on quantitative grading outcomes sidesteps these issues.

The remainder of the paper proceeds as follows. In Section 2, I present a human capital framework for understanding why racial diversity in the classroom might lead to better grades. In Section 3, I

[1] A somewhat related strand of the literature examines the effects of racial diversity at the workplace on firm outcomes such as productivity. See Kahane et al. (2013); Ozgen et al. (2013); Parrotta et al. (2014); and Trax et al. (2015).

discuss the institutional background at Reed College used as part of my identification strategy. Section 4 details the reduced-form empirical analyses conducted and presents the results obtained. Section 5 provides further discussion and concludes.

2 Human Capital Model of Diversity

In this section, I develop a simple theoretical framework for understanding why racial diversity in the classroom leads to better grades. This human capital model is for illustrative purposes and serves to motivate, but is nevertheless separate from, the reduced-form empirical analyses to follow. The model assumes that there are two distinct types of human capital, denoted g and h, that contribute to education production.[2] Without loss of generality, suppose there are only two students in the classroom, student i and their peer $-i$.[3] The education production function is given by

$$Y_i = F\left(\underbrace{g_i + \lambda_g g_{-i}}_{g}, \underbrace{h_i + \lambda_h h_{-i}}_{h}\right)$$

where

- Y_i is the output of the education production at the end of the production process for student i,
- the function $F(.)$ describes the production technology using two types of human capital g and h,
- g_i and h_i are student i's human capital levels of the two types,
- g_{-i} and h_{-i} are peer $-i$'s human capital levels of the two types,
- λ_g and λ_h are positive parameters determining the degree to which peer $-i$'s human capital (of each respective type) substitutes for student i's own human capital (of that same type) in the production of Y_i

Y_i can be thought of as any measurable output for student i at the end of the production process, such as a course grade or the GPA at some point in time. The higher the parameter values of λ, the more peer human capital matters in the production of Y_i. A value of $\lambda_g = 1$ implies that peer human

[2]Without loss of generality, assume that both types of human capital are measured in the same units.

[3]The model can be easily generalized to settings with more than one peer. In such settings, the variables attributed to peer $-i$ would be for a representative peer aggregated across multiple peers.

capital of type g is a perfect substitute for student i's own human capital of type g. A value of $\lambda_g = 0$ implies that there are no peer effects for human capital of type g.

In this model, diversity can be thought of as having peers whose human capital combinations (g_{-i}, h_{-i}) are different from one's own (g_i, h_i). Greater diversity is productively beneficial to the education production process if the two types of human capital g and h are complements in production. This can be shown mathematically after imposing the following assumptions.

Assumption 1: Diminishing Returns Let $\frac{\partial F}{\partial g} > 0$ and $\frac{\partial F}{\partial h} > 0$, but $\frac{\partial^2 F}{\partial g^2} < 0$ and $\frac{\partial^2 F}{\partial h^2} < 0$.

Assumption 2: Complementary Inputs The marginal product of one type of human capital is increasing in the other type. That is, $\frac{\partial^2 F}{\partial g \partial h} > 0$.

Under these assumptions, the following proposition holds.

Proposition 1. *Suppose Assumptions 1 and 2 hold. Given student i with human capital combination* (g_i, h_i), *it is output-maximizing to match this student with peer* $-i$ *with human capital combination* (g_{-i}, h_{-i}) *such that*

1) for students with higher g_i, *it is optimal to match them to peers with higher* h_{-i} *and lower* g_{-i};

2) for students with higher h_i, *it is optimal to match them to peers with higher* g_{-i} *and lower* h_{-i};

—conditional on the ceteris paribus constraint that the total human capital level of each peer being considered for the match $(g_{-i} + h_{-i})$ *remain constant.*

Proof. See Appendix A. □

This proposition states that if an administrator were matching student i with human capital combination (g_i, h_i) to some peer, and had to choose between peers with the same total amount of human capital $(g_{-i} + h_{-i})$ but different combinations (g_{-i}, h_{-i}), then it is output maximizing to choose the peer with (g_{-i}, h_{-i}) that is as different as possible from (g_i, h_i). In this way, the peer-effect-inclusive human capital of two types experienced by student i (i.e. $g_i + \lambda_g g_{-i}$ and $h_i + \lambda_h h_{-i}$, the arguments going into the education production function) are more evenly spread between the two types, thus taking full advantage of human capital complementarities.

Insofar as racial differences are correlated with differences in human capital combinations, greater racial diversity will lead to higher levels of output. For instance, suppose white students have high levels of g and low levels of h, while black students have low levels of g and high levels of h, and that the total human capital levels of the peers being considered remain the same. Then placing students of different races into the same classroom will lead to greater output levels for all students, compared to placing students of the same race in the same classroom, because of positive peer effects from having additional human capital of the complementary type relative to one's own.

So far, Y_i has been thought of as an output measured immediately after the education production process (e.g. GPA at the end of the first year). However, the framework can be generalized to explain potential longer term effects of having a diverse classroom in the first year (e.g. effects on GPA at graduation). Suppose a similar setup describes the production of the human capital of the two types itself. In this case, Y_i is replaced with variables representing subsequent values of g_i and h_i at the end of a production period. Adding time superscripts, such a dynamic human capital production process for student i can be described by

$$g_i^{t+1} = G\left(\underbrace{g_i^t + \lambda_g^G g_{-i}^t}_{g}, \underbrace{h_i^t + \lambda_h^G h_{-i}^t}_{h}\right)$$

$$h_i^{t+1} = H\left(\underbrace{g_i^t + \lambda_g^H g_{-i}^t}_{g}, \underbrace{h_i^t + \lambda_h^H h_{-i}^t}_{h}\right)$$

Presumably, baseline human capital of one type would be more effective at producing human capital of the same type. Thus, experiencing more diversity in the classroom (i.e. exposure to higher levels of human capital of the type different from one's own) would make student i more "well-rounded" with a more balanced human capital combination $\left(g_i^{t+1}, h_i^{t+1}\right)$ at the end of period t, compared to another student exposed to a lower-diversity classroom. Then, longer term outputs produced in subsequent periods $(t+1)$ would be functions of these new values of own human capital. For instance,

$$Y_i^{t+1} = F\left(\underbrace{g_i^{t+1} + \lambda_g g_{-i}^{t+1}}_{g}, \underbrace{h_i^{t+1} + \lambda_h h_{-i}^{t+1}}_{h}\right)$$

In this way, initial exposure to diversity in period t (e.g. in a first-year class) affects longer-term educational outcomes in later periods (e.g. GPA at graduation) through the accumulation of varying combinations of the human capital types and their subsequent use in education production of these outputs in period $t+1$. Such a dynamic interpretation of the model with long-term effects is especially relevant to the empirical setting at hand.

3 Institutional Background

Introduction to Humanities (henceforth Hum 110) is a year-long mandatory course that is taken by all first-year students at Reed College.[4] The syllabus comprise readings of classical texts from antiquity and the course serves as the foundational writing and critical thinking component of the liberal arts education at Reed.[5] All students attend trice-weekly lectures before breaking up into smaller conferences (sections) of approximately 15 students per classroom. Conferences meet either trice-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. Faculty members 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 others say." 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 understand how racial diversity plays a role here, it is helpful to think about diversity as exhibited in two forms: structural diversity and curricular diversity.[6] Structural diversity refers to the numerical makeup of the different racial groups within a student body. This form of diversity is directly related to generating the peer effects of different human capital types described in the model in Section 2. On the other hand, curricular diversity refers to students learning content that includes subject matters relating to different peoples and worldviews. One does not necessarily follow the other. A racially-homogeneous classroom can be taught a diverse curriculum; on the other hand, a classroom comprising students of many different ethnicities can be taught a culturally-uniform curriculum. Given that the Hum 110 curriculum is identical across conference classrooms, and its content has been more-or-less constant across years, it is the between-classroom variation in structural diversity that identifies our estimated treatment effect.

Scholars have also 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 (structural or curricular) within formalized settings established by the

[4] An exception is that incoming transfer students may opt to take another humanities course in lieu of Hum 110.

[5] Appendix B contains an excerpt of the 2015-2016 syllabus.

[6] This distinction, as well as other classifications to follow, has been made by many in the existing literature, including Gurin et al. (2002) and Denson and Chang (2009).

institution, such as in lectures, conferences, or seminars. The second route (also known as 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. While the effect on grades estimated in this paper acts directly through the formal setting of Hum 110 conferences, it is also possible for there to be diversity effects through informal channels, insofar as the conference peer groups are correlated with social groupings developed beyond the classroom. Moreover, diversity within these informal interactions could lead not only to better academic outcomes—if, for example, the formation of diverse study groups results in better grades—but also to better non-academic outcomes—if, say, having a diverse social network results in better job opportunities.

To estimate an internally-valid causal effect of racial diversity, it is necessary to understand where the variation in diversity comes from. In other words: How are students assigned to their Hum110 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. In theory, it is possible for students to 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. If students do not have specific friends to coordinate with to enter the same conference, 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, Reed's Registrar Office will balance students randomly across conferences held at the same 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 Registrar's Office post-actively 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 in which students are allowed access to the course sign-up system is

idiosyncratic. Students are only allowed to sign up for courses after having met with their advisors and obtained a PIN code, and appointments with advisors (which all occur on a single day before the start of classes) are essentially random. 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. Students will only have been on campus for a few days prior to signing up. It is unlikely they will base Hum 110 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 each grouping of friends is small (two or three students), the students within the friends group are still subjected to the randomness in diversity generated by conference peers outside the friends 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 Hum 110 conferences

Given these reasons, I argue that the racial composition of students in any one particular Hum 110 conference is effectively random.

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 present the results obtained.

4.1 Data

This study uses student-level administrative data obtained from Reed College for first-year and transfer students who took Hum 110 between academic years 1995-1996 and 2011-2012. Each year is a cross-section of Hum 110 students, and I combine these together into a pooled dataset. The analysis sample is restricted to students who subsequently graduated from Reed College, for whom a final cumulative GPA at graduation is observed. In addition to cumulative GPA at graduation, the other dependent variables of interest in the data are cumulative GPA at the end of the first year and the Hum 110 grade. All grades and GPAs are numerically scaled using a 4-point grade-point scale, 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 1.00 and zero points respectively.[7] The course registration data allows us to identify exactly which students are in which Hum 110 conference with which other students, as well as the full schedule of other enrolled courses during the semester for any given student. Other transcript 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.[8] The data also contain student demographic characteristics: most importantly 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.[9]

Table 1 reports summary statistics for most of these variables. Mean grades are slightly above the B letter grade. The average student takes just enough courses to meet the graduation requirement of 30 units. 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.

To measure the diversity within a particular Hum 110 conference, I construct a diversity index based on the probability that two randomly selected students from a given Hum 110 conference are of different racial ethnicities.

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

where $proportion\left(g\right)$ is the proportion of students in the conference belonging to race group g. A higher diversity measure represents Hum 110 conferences that are more diverse in terms of racial composition.[10] The summation term $\sum_g \left(proportion\left(g\right)\right)^2$ in the above formulation is commonly

[7] Grades of D-plus and D-minus are not given at Reed.

[8] Majors at Reed are grouped into five divisions: 1) Arts; 2) History and Social Sciences; 3) Literature and Languages; 4) Mathematics and Natural Sciences; and 5) Philosophy, Religion, Psychology, and Linguistics. A sixth category in this variable accounts for students with inter-disciplinary majors or double majors.

[9] 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 College Board (2009).

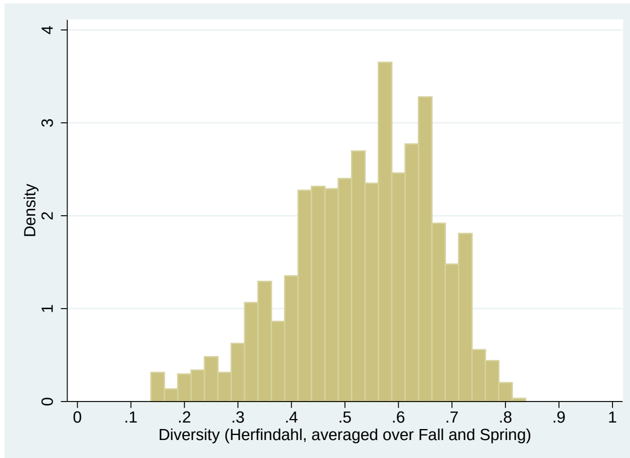
[10] It should be noted that diversity is not necessarily the same as the proportion of minority students in a classroom. For example, a classroom with 100% white students is equally diverse as a classroom with 100% black students, both of which are at the opposite end of the diversity spectrum compared to a classroom with a 50-50 split of white and black students.

Table 1: Summary Statistics

Variable / Indicator	Mean (Standard Deviation)
Cumulative GPA at Graduation	3.187
	(0.390)
Cumulative GPA at End of First Year	3.091
	(0.475)
Hum 110 Grade	3.197
	(0.517)
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)
(Normalized) Diversity	0.532
	(0.133)
N	4733

Notes: Grades and GPAs are measured in grade-points. SAT scores are expressed in thousands. Standard deviations reported in parentheses.

Figure 1: Histogram of Diversity Index Values



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.[11]

At Reed, the mean diversity within Hum 110 conferences is 0.532 (reported in Table 1).[12] 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 Hum 110 conferences, and suggests that there is sufficient sample variation in the index to identify diversity effects.

[11] Appendix C considers using alternative measures of diversity in the subsequent analyses.

[12] As mentioned previously, the regression sample is restricted to students who subsequently graduated from Reed College. However, when calculating the diversity index for a conference, I also include students in the conference who were not observed to have graduated.

Table 2: Example 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 Hisp. 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. "Asian/PI" refers to the Asian / Pacific Islander category; "other/mult." refers to other / multiple race category.

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 Table 1. The resulting measure of 0.668 for this hypothetical conference is somewhat 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 number 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.

4.2 Effects of Diversity

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

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

where

- $outcome_{icty}$ is one of the academic outcomes of interest (Hum 110 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;[13]
- $race\left(g\right)_{icty}$ 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 (in thousands) of student i;

[13] This exposition is a slight simplification of the actual situation. While uncommon, students are allowed to petition to change conference groups 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 Hum 110 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;[14]

- $free\left(\hat{t}\hat{y}\right)_{icty}$ 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 $\hat{y}$;[15]

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

- μ_y are year fixed effects (across all students who took Hum 110 in year y); and

- μ_t are time slot fixed effects (across all students who took Hum 110 in time slot t).[16]

It is important to distinguish between a Hum 110 conference c (which is the groupings of students within which they experience the structural diversity) and a Hum 110 time slot t (in 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.

The estimate of the coefficient β measures the effect of an (index) unit increase in the racial diversity of a student's Hum 110 conference on the academic outcome of interest, in grade-point units. As argued previously, this effect estimate has a causal interpretation because assignment of students to Hum 110 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 Hum 110 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 grade-points caused

[14]Similar to the situation described in Footnote 13, $meanSAT_{cty}$ 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.

[15]Because of the situation described in Footnote 13, 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 Hum 110 conference is offered, rather than the full universe of time slots available for scheduling at Reed College.

[16]Because of the situation described in Footnote 13, the full set of fixed effects includes fixed effects for time slots in both the Fall and Spring semesters. 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 Hum 110 conferences are scheduled are consistent across years).

by an increase in racial diversity in the Hum 110 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 $race\,(g)_{icty}$ are included as controls 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. Lastly, 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 trice-weekly Monday-Wednesday-Friday morning conferences, as opposed to late owls in twice-weekly Tuesday-Thursday afternoon conferences), even after conditioning on scheduling availability.

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. 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.[17] 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 Hum 110 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 Hum 110 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 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

[17] See Footnote 8 for the categories used.

them.

The effect of diversity on Hum 110 grade is positive, but statistically significant only at the 10% level (columns (1) through (3)). The estimates suggest that a one-unit increase in the diversity index (i.e. moving from a single-race class to an evenly-balanced class) increases Hum 110 grade by approximately 0.15 grade-points. This is equivalent to moving up half-way between grades B and B-plus. An alternative interpretation is that replacing one white student with one minority student in a "typical" conference (a 0.08 unit increase in diversity) increases Hum 110 grade by 0.012 grade-points. In this student-replacement scenario, the change represents a 0.023 standard deviation increase in Hum 110 grade.

There does not appear to be any diversity effect on cumulative GPA at the end of the first year (columns (4) through (6)). The point estimates across the different specifications are all positive and around 0.08, but are not statistically significant. This implies that diversity in Hum 110 conferences does not have contemporaneous effects on grades during the first year while the students are in the course. However, the next set of estimates suggests it 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 Hum 110 conference increases graduation GPA by approximately 0.10 grade-points. Alternatively, replacing one white student with one minority student in a "typical" conference increases graduation GPA by 0.008 grade-points. In this student-replacement scenario, the change represents a 0.021 standard deviation 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 does not affect dependent variables from the first year, but does have an impact on cumulative GPA at graduation. This is not as surprising as it seems, since the Hum 110 grade depends largely on submitted written work, and the content and quality of a student's writing may be affected more by diversity of ideas rather than by peer academic quality. On the other hand, social networks formed through Hum 110 conference groupings may lead to longer-term peer effects in other classes, and 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 Hum 110 grade; this may be due to language barrier issues. Moreover, taking a greater

Table 3: Effect of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Hum 110 Grade			Cum. GPA at End of First Year			Cum. GPA at Graduation		
Diversity	0.147*	0.150*	0.151*	0.076	0.081	0.079	0.103**	0.096**	0.101**
	(0.080)	(0.080)	(0.080)	(0.062)	(0.063)	(0.062)	(0.047)	(0.047)	(0.047)
Own SAT	0.910***	0.889***	0.890***	1.111***	1.093***	1.053***	0.747***	0.833***	0.827***
	(0.070)	(0.070)	(0.070)	(0.062)	(0.062)	(0.061)	(0.053)	(0.053)	(0.053)
Mean Conf. SAT	-0.123	-0.182	-0.139	0.306	0.243	0.242	0.543***	0.445**	0.460**
	(0.328)	(0.327)	(0.326)	(0.235)	(0.235)	(0.234)	(0.200)	(0.197)	(0.197)
Male		-0.092***	-0.091***		-0.089***	-0.094***		-0.103***	-0.101***
		(0.016)	(0.016)		(0.014)	(0.015)		(0.012)	(0.012)
International		0.007	0.002		0.122***	0.117***		0.101***	0.095***
		(0.033)	(0.033)		(0.032)	(0.031)		(0.027)	(0.027)
First Year Fall Units		0.043**	0.041**		0.077***	0.073***			
		(0.021)	(0.021)		(0.019)	(0.019)			
First Year Spring Units		0.182***	0.174***		0.173***	0.178***			
		(0.017)	(0.018)		(0.015)	(0.015)			
Units at Graduation								0.013***	0.011***
								(0.002)	(0.002)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year & Time Slot FEs	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	0.149	0.185	0.194	0.177	0.224	0.236	0.156	0.178	0.188

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

Notes: Grades and GPAs are measured in grade-points. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the Hum 110 conference level. All regressions include race indicators, free indicators, year and time slot fixed effects.

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 interact 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 heterogeneity regressions, I include the controls X_{icty} (male indicator, international status, and number of course units), except for 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 Hum 110 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. Surprisingly, inclusion of the interaction term actually increases the magnitude and statistical significance of the diversity effect estimates.

Lower ability students (where ability is proxied for by own SAT score) benefit more academically from the positive effects of diversity; however, this heterogeneous effect is observed to be statistically significant only for Hum 110 grade and cumulative GPA at the end of first year (columns (2) and (5)). For a student with the mean own SAT score of 1.354 (recall that SAT scores are expressed in thousands), the effect of diversity on Hum 110 grade is 0.149 grade-points. Consider a 0.1 or 100-point increase in own SAT score, which is approximately 1 standard deviation in the own SAT score distribution. On average, moving from a lower-scoring student to a higher-scoring student reduces the diversity effect on Hum 110 grade by 0.101, or about 68% of the effect magnitude at the mean SAT score. For cumulative GPA at the end of first year, the equivalent reduction in the diversity effect is by 0.082, though this estimate is only statistically significant at the 10% level. For cumulative GPA at graduation, student ability does not have an impact on the diversity effect's size, as seen by the statistically insignificant coefficient estimate on the interaction term (column (8)).

That higher-ability students benefit less from racial diversity in the classroom—at least with respect to certain academic outcomes—is consistent with diminishing returns in the human capital model in Section 2. If these higher-ability students already have high levels of human capital of both types—and in particular high levels of the type not typical of their own race—then they would benefit less from having diverse peers given their already "well-rounded" combination of own human capital of the two types.

Non-white minorities do not experience larger diversity effects compared to their white counterparts (columns (3), (6), and (9)). 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 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 regress

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

where

- x_{icty} is one of the predetermined covariates in vector X_{icty}; and
- all other variables are as before.

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

Table 4: Heterogeneous Effects of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Hum 110 Grade			Cum. GPA at End of First Year			Cum. GPA at Graduation		
Diversity	0.190**	1.521**	0.153*	0.129*	1.185**	0.090	0.118**	0.717	0.087*
	(0.096)	(0.669)	(0.086)	(0.078)	(0.574)	(0.070)	(0.060)	(0.499)	(0.052)
Diversity×Male	-0.093			-0.110			-0.050		
	(0.119)			(0.110)			(0.089)		
Diversity×Own SAT		-1.013**			-0.815*			-0.459	
		(0.494)			(0.428)			(0.371)	
Diversity×Minority			-0.014			-0.037			0.035
			(0.125)			(0.115)			(0.094)
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year & Time Slot FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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	0.185	0.186	0.185	0.224	0.225	0.224	0.178	0.179	0.178

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

Notes: Grades and GPAs are measured in grade-points. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the Hum 110 conference level. All regressions include race indicators, free indicators, year and time slot fixed effects, as well as other controls (own SAT score, mean SAT score in conference, male indicator, international status, number of course units).

Table 5: Randomization Checks

	(1)	(2)	(3)
Dep. Var.:	Male	International	Own SAT
Diversity	-0.043	0.012	-0.016
	(0.056)	(0.030)	(0.016)
Race Indicators	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes
Year & Time Slot FEs	Yes	Yes	Yes
N	4733	4733	4733
R-square	0.128	0.181	0.218

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

Notes: SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the Hum 110 conference level. All regressions include race indicators, free indicators, year and time slot fixed effects.

Table 5 presents the results from these randomization checks for three dependent variables: an indicator for being male, an indicator for being an international student, and own SAT score.[18] None of the estimates of the coefficient on $diversity_{cty}$ are statistically significant, consistent with racial diversity in Hum 110 conferences being exogenous.

5 Discussion & Conclusion

The findings in this paper suggest that a greater degree of racial diversity in the Hum 110 classroom causes a statistically significant increase in the cumulative grade point average (GPA) at graduation. Hum 110 course grades also increase, but this estimated effect of diversity is statistically significant only at the 10% level. On the other hand, diversity has no statistically significant effect on cumulative GPA at the end of the first year. I do not detect heterogeneous effects between male and female students, or between white and minority students. However, lower ability students benefit more from racial diversity in the classroom. These results suggest that experiencing diverse human capital types in the classroom can have both short- and long-term effects on academic outcomes.

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 the graduation GPA by 0.008 grade-points. Jones and Jackson (1990) report that a 1 grade-point increase in GPA is associated with a 9% increase in earnings.[19] Assuming linearity, in this scenario, replacing one white

[18]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.

[19]Loury and Garman (1993) report similar estimates separately for white and black males, and find that white male

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 (US Department of Education, 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.

Furthermore, these results highlight the possibility of costless yet 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 compensating 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 of 30.6 students, this costless intervention represents a $4,600 net increase in aggregate annual earnings.[20]

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 dispels the anecdotal notion that white students are somehow hurt academically by being in more racially-diverse classrooms, all else being equal. Second, insofar as administrators are concerned with Hum 110 grade and cumulative GPA at the end of the first year, that lower ability students benefit more from racially-diverse classrooms means that they should receive priority when being assigned to high-diversity conferences. Undoubtedly, implementing an efficient conference-assignment policy based

students with higher GPAs earn 6% more per grade-point, while black male students earn 27% more per grade-point.

[20]That is, 30.6 students × 0.50 net diversity gain per student × 0.10 grade-point effect per diversity unit × 9% increase in earnings per grade point × $33,400 annual earnings. This calculation again uses the coefficient estimates for cumulative GPA at graduation and assumes linearity.

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. 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 then wholly segregated into classrooms by race post-matriculation.

This paper has two main contributions. First, I exploit a quasi-experimental identification strategy which generates internally-valid causal effects, given the exogenous variation in racial diversity. Second, the Hum 110 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.

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 human capital framework of diversity developed here could be applicable to on-the-job production of human capital, leading to diversity effects on worker productivity or earnings stemming 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 role racial diversity plays in higher education and 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.

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.

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.

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 Psychometric.

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.

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, 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.

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, March.

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.

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.

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

Appendices

A Proof of Proposition 1

Suppose an administrator wants to maximize the outputs of student i and peer $-i$ and has the social welfare function

$$Y_i + Y_{-i} = F\left(g_i + \lambda_g g_{-i},\, h_i + \lambda_h h_{-i}\right) + f\left(g_{-i} + \pi_g g_i,\, h_{-i} + \pi_h h_i\right)$$

where $f\left(.\right)$ is the analogous education production function for peer $-i$, with corresponding parameters π_g and π_h. Let both production functions exhibit properties stated in Assumptions 1 and 2.

To the administrator, student i has a fixed human capital combination of $\left(\overline{g_i}, \overline{h_i}\right)$, and a peer with some human capital combination $\left(g_{-i}, h_{-i}\right)$ must be chosen to maximize the above social welfare function.[21] To make a ceteris paribus comparison, the administrator is to choose among peers whose total human capital $g_{-i} + h_{-i}$ is constant. Thus, the administrator's maximization problem is

$$\max_{g_{-i}, h_{-i}} F\left(\overline{g_i} + \lambda_g g_{-i},\, \overline{h_i} + \lambda_h h_{-i}\right) + f\left(g_{-i} + \pi_g \overline{g_i},\, h_{-i} + \pi_h \overline{h_i}\right)$$

subject to the constraint

$$g_{-i} + h_{-i} = C$$

The first order condition (FOC) is given by

$$FOC \equiv \frac{\partial F}{\partial g}\lambda_g + \frac{\partial f}{\partial g} - \frac{\partial F}{\partial h}\lambda_h - \frac{\partial f}{\partial h} = 0$$

(The second order conditions confirm this solution to be a maximum.) We can do comparative statics by applying the implicit function theorem on the FOC to obtain

$$\frac{dg_{-i}}{d\overline{h_i}} = -\frac{\frac{dFOC}{d\overline{h_i}}}{\frac{dFOC}{dg_{-i}}} = -\frac{\frac{\partial^2 F}{\partial g \partial h}\lambda_g + \frac{\partial^2 f}{\partial g \partial h}\pi_h - \frac{\partial^2 F}{\partial h^2}\lambda_h - \frac{\partial^2 f}{\partial h^2}\pi_h}{\frac{\partial^2 F}{\partial g^2}\lambda_g^2 + \frac{\partial^2 f}{\partial g^2} - \frac{\partial^2 F}{\partial h \partial g}\lambda_h \lambda_g - \frac{\partial^2 f}{\partial h \partial g}}$$

Under Assumptions 1 and 2,

$$\frac{dg_{-i}}{d\overline{h_i}} > 0$$

[21] Without loss of generality, outputs Y_i and Y_{-i} are weighted equally.

Using similar algebra, it can be shown that

$$\frac{dh_{-i}}{d\overline{g_i}} > 0$$

This means that at the optimum output-maximizing solution (on the FOC), when faced with a student i with higher $\overline{h_i}$ ($\overline{g_i}$), the administrator should choose a peer with a higher g_{-i} (h_{-i}), the other human capital type.

Furthermore, applying similar techniques reveals that

$$\frac{dg_{-i}}{d\overline{g_i}} < 0$$

and

$$\frac{dh_{-i}}{d\overline{h_i}} > 0$$

This means that at the optimum output-maximizing solution (on the FOC), when faced with a student i with higher $\overline{g_i}$ ($\overline{h_i}$), the administrator should choose a peer with a lower g_{-i} (h_{-i}).

B Hum 110 Syllabus (2015-2016)

The following is an excerpt from the Hum 110 syllabus, downloaded in June 2016 from the website http://www.reed.edu/humanities/hum110/.

> "The humanities"—literae humaniores in Latin—referred originally to the study of texts written by human, rather than divine, hands. In modern education, the humanities have come to include the study of literature, history, philosophy, religion, politics, and the arts. From these perspectives, students of the humanities consider some of the ways in which people have represented and reflected on the physical, social, psychological, and ideological features of the world in which they live. They investigate the various materials that form the basis of a culture and that simultaneously provide key terms for its critique and transformation. Typical questions include: How have different cultures distinguished appearance from reality, nature from culture, particular from universal? How have they made sense of the connection between the individual and the various groups in which individuals claim membership? How can we understand the relations between reason and desire, word

and deed, the worldly and the transcendent? In pursuing such questions, moreover, humanistic study seeks to employ a set of analytic perspectives – literary, aesthetic, historical, philosophical, social scientific – that have helped shape our intellectual tradition and now compose the foundations of a liberal arts education.

Humanities 110 introduces students to humanistic inquiry by considering a range of artistic, intellectual, political, and religious strategies that emerged in ancient Greece and in the larger Mediterranean world of which it was a part. The course examines how varieties of human thought interact to produce a culture's distinctive way of life. Recognizing that no culture is self-contained, we seek as well to interpret ancient sources as artifacts of cultural exchange, influence, and differentiation. For example, we might consider how materials from ancient Athens intersect and diverge from one another in their reflections on democracy, empire, gender, race, or class, while also considering how these materials compare with those of Mesopotamian, Egyptian, Persian, Israelite, Hellenistic, Jewish, Roman, or early Christian cultures. In doing so, we will encounter issues of continuing relevance pertaining to ideals of truth, beauty, virtue, justice, happiness, and freedom, as well as challenges posed by social inequality, war, power, and prejudice.

As the only course required of all first-year students at Reed, Humanities 110 serves as the College's foundational writing course and introduces students to the skills and habits of mind necessary for academic inquiry in their future work at Reed. Over the course of the year students should become more practiced and adept at:

Framing questions that elicit deeper analysis;

Cultivating intellectual curiosity;

Crafting, analyzing, critiquing, and defending arguments using evidence;

Expressing ideas in writing and speech clearly, persuasively, and honestly;

Participating productively and respectfully in a Reed conference discussion;

Interpreting primary sources in a range of media and genres;

Practicing the basic methods of various disciplines in the humanities and social sciences.

Conference Assignments:

The Registrar makes initial assignments to conferences in this course that continue through the year. Students who subsequently find it necessary to change conferences due to time conflicts must petition the chair of Humanities 110.

Required Texts (Fall):

Aeschylus, The Oresteia, trans. Fagles (Penguin)

Anonymous, The Epic of Gilgamesh, trans. George (Penguin)

New Oxford Annotated Bible with the Apocrypha, trans. Coogan, et al. (Oxford)

Curd, ed., Presocratics Reader: Selected Fragments and Testimonia, trans. McKirahan (Hackett)

Freeman, Egypt, Greece and Rome (Oxford)

Herodotus, The Histories, trans. Selincourt (Penguin)

Hesiod, Theogony and Works and Days, trans. Lombardo (Hackett)

Homer, The Iliad, trans. Lattimore (Chicago)

Miller, Greek Lyric: An Anthology in Translation (Hackett)

The Tale of Sinuhe and Other Ancient Egyptian Poems, trans. Parkinson (Oxford)

Sophocles, Sophocles I: Oedipus the King, Oedipus at Colonus, Antigone, trans. Grene and Lattimore (Chicago)

Thucydides, The Peloponnesian War, trans. Warner (Penguin)

Required Texts (Spring):

Apuleius, The Golden Ass, trans. Lindsay (Indiana University Press)

Aristophanes, Three Comedies: The Birds, The Clouds, The Wasps, trans. Arrowsmith (University of Michigan Press)

Aristotle, The Nicomachean Ethics, trans. Irwin (Hackett)

Cicero, Selected Works, trans. M. Grant (Penguin)

Epictetus, The Handbook, trans. Nicholas P. White (Hackett)

Euripides, Euripides I, ed. David Grene and Richmond Lattimore (University of Chicago Press)

Freeman, Egypt, Greece, and Rome, 2nd ed. (Oxford) Josephus, The Jewish War, trans. Williamson (Penguin)

Livy, The Rise of Rome; Ab Urbe Condita, trans. Luce (Oxford)

Lucretius, On the Nature of Things, trans. Englert (Focus Philosophical Library)

The New Oxford Annotated Bible with the Apocrypha New Revised Standard Version: College Edition (Oxford)

Ovid, Metamorphoses, trans. Melville (Oxford) Plato, Republic, trans. Reeve (Hackett)

Plato, Trial and Death of Socrates, trans. Grube (Hackett)

Seneca, The Stoic Philosophy of Seneca: Essays and Letters, trans. Hadas (Norton)

Theocritus, Idylls, trans. Verity (Oxford) Virgil, The Aeneid, trans. Mandelbaum (Bantam Doubleday Dell)

Recommended Texts:

Harvey, The Nuts and Bolts of College Writing (Hackett)

Williams, The Craft of Argument (Univ. of Chicago Press)

C 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 6 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.

Panel (B) of Table 6 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

Table 6: Effect of Alternate Measures of Diversity on Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	Hum 110 Grade			Cum. GPA at End of 1st Year			Cum. GPA at Graduation		
Panel (A)									
Diversity	0.147*	0.150*	0.151*	0.076	0.081	0.079	0.103**	0.096**	0.101**
	(0.080)	(0.080)	(0.080)	(0.062)	(0.063)	(0.062)	(0.047)	(0.047)	(0.047)
N	4733	4733	4733	4733	4733	4733	4733	4733	4733
R-square	0.149	0.185	0.194	0.177	0.224	0.236	0.156	0.178	0.188
Panel (B)									
Peer Diversity	0.131*	0.132*	0.131*	0.070	0.073	0.071	0.100**	0.092**	0.095**
	(0.074)	(0.074)	(0.074)	(0.057)	(0.058)	(0.057)	(0.044)	(0.044)	(0.044)
N	4733	4733	4733	4733	4733	4733	4733	4733	4733
R-square	0.149	0.185	0.194	0.177	0.224	0.236	0.156	0.178	0.188
Panel (C)									
Shannon Entropy	0.083*	0.088*	0.089*	0.036	0.042	0.038	0.052*	0.049*	0.052*
	(0.045)	(0.045)	(0.045)	(0.035)	(0.035)	(0.035)	(0.028)	(0.027)	(0.027)
N	4733	4733	4733	4733	4733	4733	4733	4733	4733
R-square	0.149	0.185	0.194	0.177	0.224	0.235	0.156	0.178	0.187
Race Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Free Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year & Time Slot FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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 measured in grade-points. SAT scores are expressed in thousands. Standard errors in parentheses are clustered at the Hum 110 conference level. All regressions include race indicators, free indicators, year and time slot fixed effects.

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. From the standpoint of the human capital model of diversity presented in Section 2, the production of education output depends on one's own human capital combination (and thus, one's own race) as well as the combinations 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 6 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, they 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. Reinterpreting these estimates in terms of hypothetical student replacements as in Section 4.2 yield similar findings.