The Promise and Peril for Universities Using Correlates of Race in Admissions in Response to the *Grutter* and *Fisher* Decisions

Mark Long
University of Washington

2015

The Civil Rights Project
Proyecto Derechos Civiles

The Civil Rights Project
Proyecto Derechos Civiles

Abstract

This paper examines using correlated indicators for a student's race, rather than using the student's actual race, in admission decisions for universities. The goal is to determine whether this method will permit universities to broaden admissions criteria to promote diversity without using race or ethnicity. Also addressed in this paper is whether the use of correlates of race is workable and race neutral from a legal, political, and university perspective.

Keywords: admission decisions; race; ethnicity; correlated indicators; diversity; Fisher decision; Grutter decision

The Civil Rights Project
Proyecto Derechos Civiles

Table of Contents

In the case of *Fisher v. University of Texas at Austin,* the U.S. Supreme Court concluded that the

> University must prove [to the reviewing court] that the means it chose to attain that diversity are narrowly tailored to its goal. On this point, the University receives no deference. . . . The reviewing court must ultimately be satisfied that no workable race-neutral alternatives would produce the educational benefits of diversity. (p. 10)

In this paper, I discuss the potential promise and peril for universities using correlated indicators for a student's race, rather than using the student's actual race, in their admission's decisions. Using a variety of data sources (including university admissions records and other more detailed surveys from the U.S. Census Bureau and U.S. Department of Education [ED]), I answer the following question: Can universities broaden admissions criteria to promote diversity without using race or ethnicity per se? More specifically, if university admissions offices used all of the information that they could obtain on an applicant (aside from the student's race), how well would that information correlate with the student's race (and thus promote diversity in admissions)? Subsequently, I address the following question: Is the use of correlates of race *workable* and *race neutral* from a legal, political, and university perspective?

Universities explicitly or implicitly score applicants based on observable characteristics and admit those students whose scores are above a threshold. For example, a university may observe a set of characteristics for student *i* (denoted by the vector $\boldsymbol{X}_i$) and give positive or negative weight to each characteristic (with the group of weights denoted by the vector $\boldsymbol{\omega}$), and admit student *i* if the weighted sum is above a given threshold (i.e., in matrix notation, the university would admit student *i* if $\boldsymbol{X}_i{}'\boldsymbol{\omega}$ > threshold). Prior to the *Gratz v. Bollinger* and *Grutter v. Bollinger* rulings, the University of Michigan (UM) used an explicit scoring system that gave a fixed number of points to underrepresented minority (URM) applicants. While the *Gratz* decision found UM's undergraduate admissions system to be unconstitutional, as it gave a *fixed* amount of weight to an applicant's race, the *Grutter* decision found UM's Law School admissions system to be constitutional, as it gave positive weight (but not a fixed weight) to the applicant's race. My assumption is that universities are complying with these rulings, and thus their scoring of applicants, at least as it pertains to race, is now implicit rather than explicit. That is to say, racial minorities are still more likely to be admitted than other observably similar applicants (Blume & Long, 2014), but the weight placed on race is no longer fixed.

I am assuming that universities would like to give positive weight for applicants who are URMs (i.e., Black, Hispanic, or Native American).[1] However, the *Grutter* and *Fisher* decisions make it more difficult for universities to directly give advantage to URMs in their admissions decisions. These decisions clarified the legality of the use of affirmative action in admissions but limited the mechanisms by which colleges could advantage minority applicants in

admissions decisions. The *Grutter* decision concluded that race can be used as one factor among many in an admissions decision so long as it is part of a holistic review of an applicant and so long as race does not have a uniform impact (as would follow from a point system). Moreover, the *Grutter* ruling held that for the use of an applicant's race to be permissible, the college needs to give "serious good faith" consideration to workable nonrace-based alternatives that increase diversity. (This question of serious good faith effort was at the heart of the *Fisher* case.) The Court's ruling on *Fisher* means universities now have a greater burden of proof that there was no workable race-neutral alternatives should their race-based admissions program be challenged in court.

Chan and Eyster (2003) developed a theoretical model that showed that in response to bans on affirmative action, universities that value both the academic qualifications and the ethnic and racial diversity of their student bodies will react by adopting "an admissions rule that partially ignores standardized-test scores and other traditional measures of academic ability" (p. 868). Long and Tienda (2008) showed that this prediction held true for the University of Texas at Austin (UT Austin) and Texas A&M University in the years after the 1996 *Hopwood v. Texas* decision. Long and Tienda found that the direct advantages given to Black and Hispanic applicants at these universities disappeared post-*Hopwood.* Furthermore, both universities changed the weights they placed on applicant characteristics in ways that increased the likelihood that URMs would be accepted. Yet, these changes in weights (combined with Texas's top 10% policy, which required the automatic admission of students who graduated in the top 10% of their high school classes) were insufficient to restore URMs' share of admitted students. Long and Tienda simulated the extent to which URMs' share of admitted students would have fallen if these universities had responded to the *Hopwood* ruling by holding the weights placed on applicant characteristics constant, set the weight on URM to zero, and did not have the top 10% policy. The URM share was then simulated with the new weights (and the new weights plus the top 10% policy) to see how much the URM share rebounds owing to these policy changes. The following rebounds were found: for UT Austin, 33% rebound due to the change in weights and 61% rebound due to the change in weights + the top 10% policy, and for Texas A&M, 19% rebound due to the change in weights and 29% rebound due to the change in weights + top 10% policy. Antonovics and Backes (2014) found similar results for the University of California (UC) campuses. They found that

UC campuses changed the weight given to SAT scores, high school [grade point average] and family background in response to California's ban on race-based affirmative action, and that these changes were able to substantially (though far from completely) offset the fall in minority admissions rates. (p. 295)

Furthermore, they found that "Berkeley was able to restore roughly 16–26 percent of the 31 percentage point direct fall in URMs' relative chances of admission" (p. 16). Thus, Long and Tienda (2008) and Antonovics and Backes (2014) each found that reweighting of applicant

The Civil Rights Project
Proyecto Derechos Civiles
Using Correlates of Race in Admissions

characteristics led to rebounds in URMs' share in the range of 16%–33% at various flagship campuses.

In response to the *Grutter* and *Fisher* decisions, I am assuming that universities will seek a variety of alternatives, one of which is to place more positive weight on observable characteristics that are positively correlated with URM. A more straightforward way to achieve the same result would be to replace the indicator for being a URM student (i.e., URM_i = 0 or 1) in a university's old admissions formula with the student's predicted likelihood of being a URM student ($\widehat{URM}_i$, which ranges from 0 to 1). If a university were able to *perfectly* predict a student's URM status (i.e., $\widehat{URM}_i = URM$), then the university would admit the same students that it would have admitted using traditional affirmative action, leading to the same diversity that would have been had under traditional affirmative action. Note that using an admissions system where URM status is replaced by the student's likelihood of being a minority might not be considered a race-neutral alternative. I discuss this issue in the last section of the paper. Before I get to that discussion, the goal in the next three sections of this paper is to empirically evaluate how close the university could come to such a perfect prediction.

I should emphasize that I am not suggesting that universities *should* give weight to the student's predicted likelihood of being a minority student. The intent is not to provide an admissions formula that universities could use to improve their diversity, nor to demonstrate how universities could circumvent the law. Rather, I am attempting to illustrate whether and to what extent such an admissions policy, if used, could be a successful alternative strategy for improving the diversity of the admitted class. As I illustrate in this paper, the efficacy of such an alternative policy has substantial limitations.

Methods

First, using existing data collected in university admissions records (from UT Austin), I estimate the probability that a student is a URM based on student characteristics that are already observable to UT Austin (Z) using a probit regression as shown in Equation 1:

$$\text{PROB}(URM_i = 1) = \Phi(Z_i'\boldsymbol{\beta} + \varepsilon_i). \tag{1}$$

Second, I reestimate Equation 1 using data from two additional data sources: (a) the American Community Survey (ACS) collected by the U.S. Census Bureau and (b) the Education Longitudinal Study of 2002 (ELS) collected by the ED. The purpose of using these data will be to assess the extent to which colleges could collect additional data that would most enhance the predictive power of Equation 1 (i.e., in more successfully finding correlates of race).

The Civil Rights Project
Proyecto Derechos Civiles

Prior to estimating Equation 1 with the full set of available variables, I use an iterative procedure to evaluate (a) the extent to which adding additional variables raises the predictive power of Equation 1 and (b) the nature of the most predictive variables. This iterative procedure is as follows:

Step 1: Find the variable with the largest correlation with URM.

Step 2: Enter this variable into the $\mathbf{Z}$ vector.

Step 3: Estimate Equation 1.

Step 4: Measure the predictive power of Equation 1 in two ways:

Method 1: Compute McFadden's pseudo-R^2.[2]

Method 2: Rank order the observations from the most to the least likely to be a URM based on $\widehat{URM}_i = \Phi(\mathbf{Z}_i'\hat{\beta})$. Assume that the top N observations are URMs, where N is the actual number of URMs in the data set. Compute the share of those who are tagged as a URM who are correctly identified as URMs. Compute the following ratio: Incorrectly tagged as a URM)/Correctly tagged as a URM.

Step 5: Find the prediction error: $\hat{\varepsilon}_i = URM_i - \widehat{URM}_i = URM_i - \Phi(\mathbf{Z}_i'\hat{\beta})$.

Step 6: Find the unused variable with the largest correlation with the prediction error.

Step 7: Repeat Steps 2–6 until all of the variables are used. Along the way, chart the increase in the predictive power of Equation 1 and note the order in which variables are added to $\mathbf{Z}$.

Data

The first source of data includes the census of applicants to UT Austin in the years 1998, 1999, and 2000. These data were compiled by the Texas Higher Education Opportunity Project (http://theop.princeton.edu) and include a total of 50,705 students.[3] Table 1 lists all of the variables included. Broadly speaking, these variables represent the kind of information that universities may readily have on hand under their current, typical admissions processes. Note that some variables are missing for some applicants. I impute all missing variables using the best available subset of nonmissing $\mathbf{Z}$ variables.[4]

Table 1 University of Texas at Austin (UT Austin) Data: List of Variables and Pre- and Postimputation Means

Variable	Number of nonmissing observations	Preimputation mean	Postimputation mean
Student is			
an underrepresented minority	50,705	19.0%	19.0%
Female	50,672	49.9%	49.9%
a U.S. citizen	50,705	89.1%	89.1%
Student's academic achievement measures			
SAT/ACT score (in SAT points)	48,907	1,192	1,191
Texas Academic Skills Program score	31,122	0.747	0.689
Class rank percentile (0–100, with 100 = top of class)	40,186	80.9	79.4
In top-10% of HS class	40,189	41.2%	37.5%
Took *Advanced Placement*® (*AP*®) test	50,705	24.4%	24.4%
Scored 3+ on AP math test	50,705	8.8%	8.8%
Scored 3+ on AP science test	50,705	4.3%	4.3%
Scored 3+ on AP foreign language test	50,705	4.1%	4.1%
Scored 3+ on AP social science test	50,705	7.2%	7.2%
Scored 3+ on other AP test	50,705	14.4%	14.4%
Parents' highest level of education is			
No high school	45,198	1.2%	1.3%
Some high school	45,198	1.4%	1.6%
High school graduate	45,198	5.7%	5.6%
Some college	45,198	15.2%	14.4%
College graduate	45,198	43.5%	44.6%
Father's income is			
$20,000 - $39,999	42,326	14.4%	14.7%
$40,000 - $59,999	42,326	15.9%	15.6%
$60,000 - $79,999	42,326	15.4%	15.1%
>= $80,000	42,326	47.4%	47.4%
Single parent family	50,705	14.5%	14.5%
Student's high school's characteristics			
Average SAT/ACT score of students (in ACT points)	47,864	21.8	21.9
Share of students who took the SAT or ACT	39,626	0.839	0.866
School is a UT-feeder (as defined by Tienda & Niu, 2006)	50,701	21.7%	21.7%
School is a recipient of Longhorn Opportunity scholarships	50,705	2.6%	2.6%

Variable	Number of nonmissing observations	Preimputation mean	Postimputation mean
School is a recipient of Century Scholars scholarships	50,705	2.2%	2.2%
School is a private high school	45,446	11.2%	12.7%
School is in Texas	50,577	81.2%	81.3%

Note. Squares of all continuous variables (i.e., those not listed above with "%" symbols) are included as well. HS = high school, UT = University of Texas.

The second source of data is the ACS of 2011 (public use microdata sample). ACS is nationally representative of the U.S. population. I restrict the analysis to those individuals aged 15–17 years who are not institutionalized. This analysis includes 119,879 youths. Table 2 lists all of the variables included. These variables mostly include measures of family resources, family structure (interacted with parents' nativity and parents' labor force participation), and youths' nativity. These variables could be easily collected by universities during their admissions processes (and many of these kinds of variables are collected by universities).[5]

The third source of data is the ELS (restricted use data), which is nationally representative of 10th graders in 2002 and includes 15,240 students.[6] ELS is a remarkably rich data set that includes students' academic records, achievement scores, attitudes and beliefs, and participation in sports and other activities; hours spent on TV, computers, video games, and homework; characteristics of the students' friends; family structure and resources; parents' expectations and other views; and characteristics of the students' school, administrator, and teachers' views. I again impute missing variables.[7] Because it may be more difficult to predict the URM status of students who apply to very competitive colleges than for all 10th graders, I evaluate the sensitivity of the results by further restricting the ELS data to the 4,770 students who submitted at least one application to one of the 456 colleges rated by Barron's (2004) as very, highly, or most selective. I further evaluate whether the set of most predictive variables is different for these students than for all 10th graders.

For each data set, I also include the squares of all continuous variables in Z (e.g., grade point average squared). The three data sets contain 33, 35, and 195 variables, respectively. For reasons I discuss later, some of the ELS variables would be difficult for colleges to obtain or use as they may be more likely to provoke legal challenge, be more politically unattractive, or be administratively infeasible to accurately obtain. I conduct a further analysis restricting myself to the 171 ELS variables that I expect would cause fewer such problems. See the appendix, where problematic variables, which are excluded from this analysis, are italicized.

The Civil Rights Project
Proyecto Derechos Civiles

Using Correlates of Race in Admissions

Table 2 American Community Survey (ACS) Data: List of Variables and Pre- and Postimputation Means

Variable	Number of observations	Mean
Youth is an underrepresented minority	119,879	34.7%
Household size	119,879	4.37
Household rent	119,879	$284
Household income	119,879	$84,721
Number of bedrooms in household/persons in household	119,879	0.854
Household receives SNAP	119,879	19.3%
Household receives SSI/AFDC/other welfare income	119,879	5.0%
Not living with parents (or missing)	119,879	6.8%
Living with two parents who are both U.S. natives	119,879	47.9%
Living with two parents; father only foreign born	119,879	2.1%
Living with two parents; mother only foreign born	119,879	2.2%
Living with two parents who are both foreign born	119,879	10.7%
Living with father only who is a U.S. native	119,879	5.4%
Living with father only who is foreign born	119,879	1.1%
Living with mother only who is a U.S. native	119,879	19.8%
Living with mother only who is a foreign born	119,879	3.9%
Living with two parents who are both in the labor force	119,879	43.3%
Living with two parents; father only in the labor force	119,879	14.9%
Living with two parents; mother only in the labor force	119,879	3.1%
Living with two parents who are neither in the labor force	119,879	1.6%
Living with father only who is in the labor force	119,879	5.6%
Living with father only who is not in the labor force	119,879	0.9%
Living with mother only who is in the labor force	119,879	19.0%
Living with mother only who is not in the labor force	119,879	4.7%
Born in the United States	119,879	92.8%
Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas	119,879	0.3%
Born abroad of American parent(s)	119,879	0.9%
U.S. citizen by naturalization	119,879	1.6%
Not a citizen of the United States	119,879	4.3%
Language other than English by adults	119,879	1.8%
Grandparents living with grandchildren present in the household	119,879	11.6%

Note. Squares of all continuous variables (i.e., those not listed above with "%" symbols) are included as well. AFDC = Aid to Families with Dependent Children, SNAP = Supplemental Nutrition Assistance Program, SSI = Supplemental Security Income.

Results

Figure 1 shows the results using the UT Austin data. Nineteen percent of applicants to UT Austin (9,644 out of 50,705) in these 3 years were URMs. Thus, if one were to randomly tag 9,644 applicants as being URMs, one would get it right by dumb luck 19% of the time. As we add variables to Z, our ability to correctly identify students as URMs increases. However, even with all 33 variables included, one can correctly identify students as URMs only 53.9% of the time. That is, out of the 9,644 applicants with the highest likelihood of being URMs (i.e., the highest value of $\widehat{URM_i}$), only 53.9% of them are in fact URMs. Another way to understand this inability to accurately predict is shown in Figure 2. Here with 33 variables in the model, 0.9 students who are not URMs were incorrectly tagged for every 1 student who was correctly tagged as being a URM.[8]

Figure 3 shows the results using the ACS data. URMs comprise a higher share of youths in the ACS data (34.7%; 41,568 out of 119,879) than among the UT Austin applicants. As a result, at base, one can correctly tag (by dumb luck) a higher share of students as URMs. As the number of variables included in the prediction model are increased, the share that is correctly identified as URMs rises from 34.7% to 64.1%. (Correspondingly, one would incorrectly tag 0.6 non-URMs for every 1 URM correctly identified.) Although one can correctly identify a higher share of students, this better performance is due to the higher baseline share of URMs in the ACS data. In fact, the pseudo-R^2 measure of how well one has predicted URM status is lower for the ACS data (.20 with all 35 variables included) than for the UT Austin data (.23 with all 33 variables included).

In contrast, one can do *remarkably* better at predicting URM status for the ELS data using the full set of 195 variables, as shown in Figure 4. The share that is correctly identified as URMs rises from 28.6% at base to 82.3% with all variables included (and, correspondingly, only 0.22 non-URMs are incorrectly tagged for every URM correctly identified). Amazingly, almost all of this gain in predictive power comes from the first four variables that are entered into the probit regression.

To understand what variables are leading our ability to predict URM status, Table 3 lists the first 10 variables added to the models. I begin my discussion with the third column for the ELS data. Note that the first four variables entered are indicators for whether the student's three best friends are URMs and whether the student's native language is Spanish. Just using the first of these indicators (whether the best friend is a URM) is able to raise the share correctly identified as URMs from 28.6% at base to 74.5%. (Appallingly, but not surprisingly, 75% of URM students' best friends are URMs, while only 9% of non-URM students' best friends are URMs.)
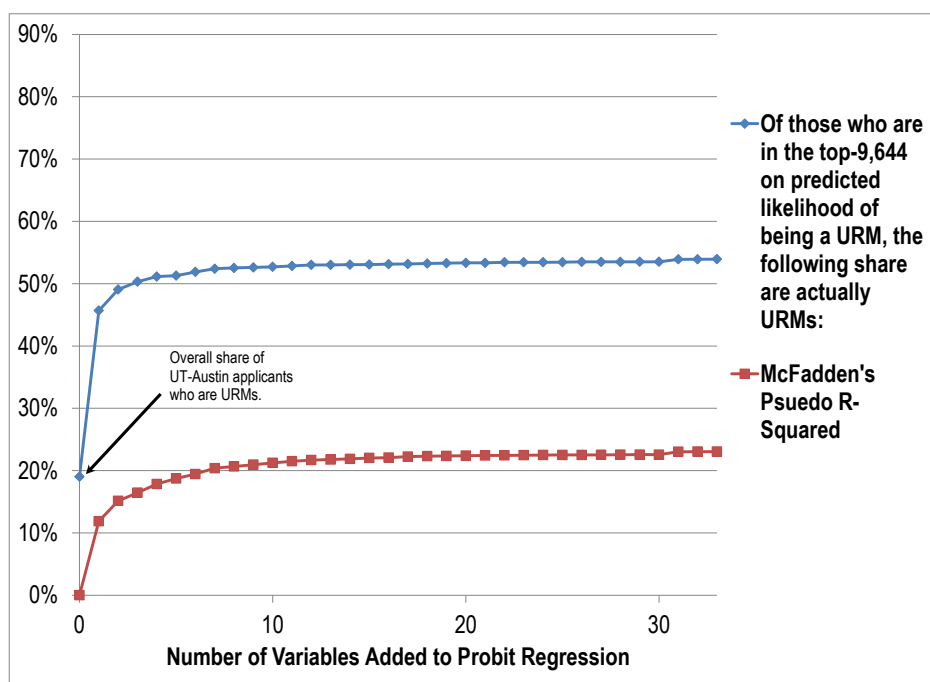
Figure 1. University of Texas (UT) at Austin data: Correspondence of students' actual underrepresented minority (URM) status with predicted URM status.

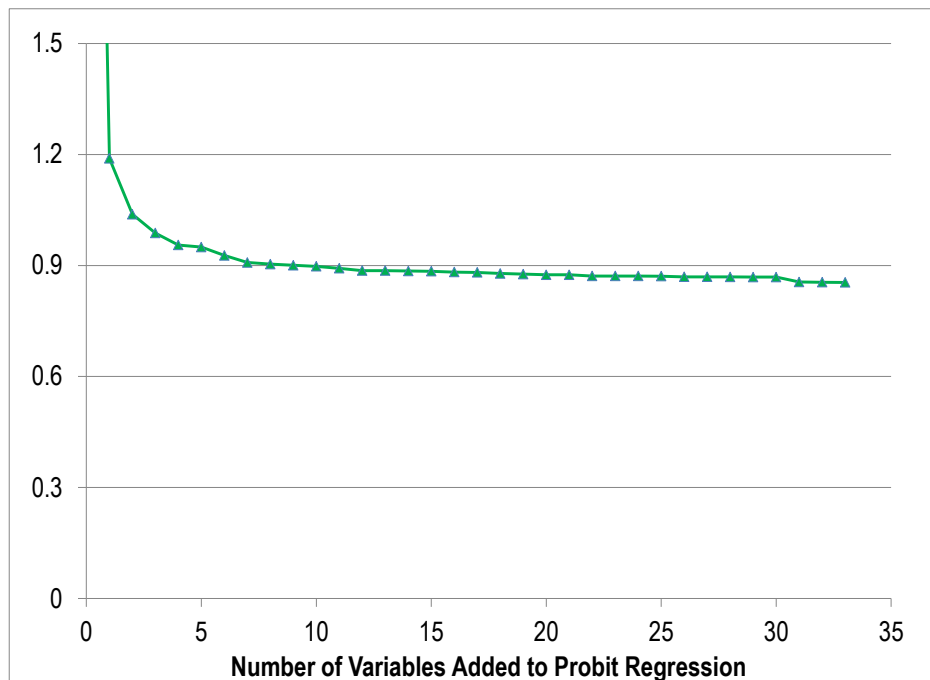*Figure 2. University of Texas (UT) at Austin data: Ratio of students incorrectly identified as underrepresented minorities (URMs) to those correctly identified as URMs.*
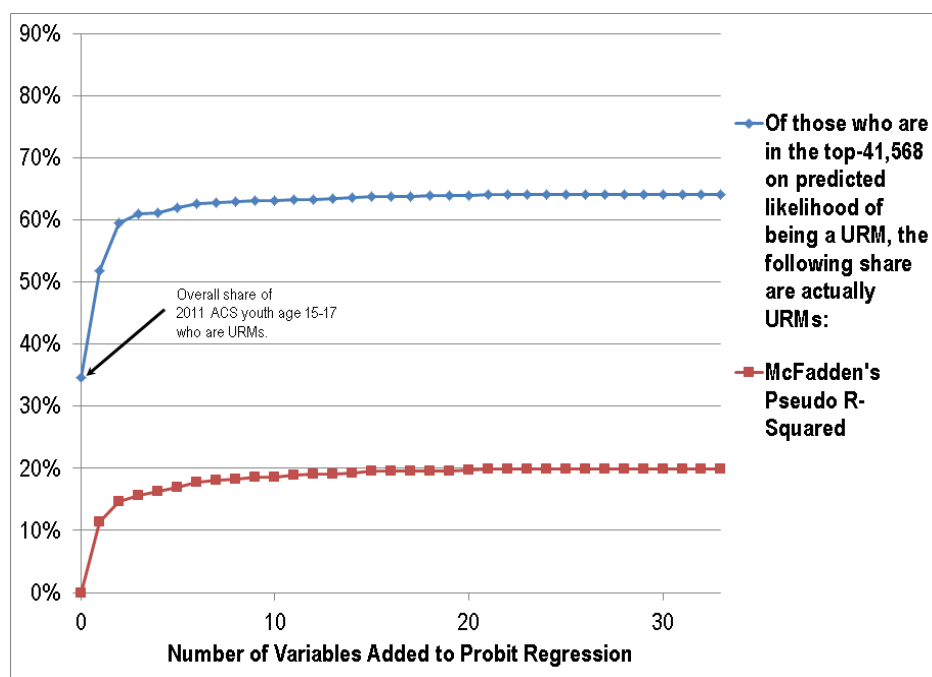
Figure 3. Same as Figure 1, but for American Community Survey (ACS) data. URM = underrepresented minority.
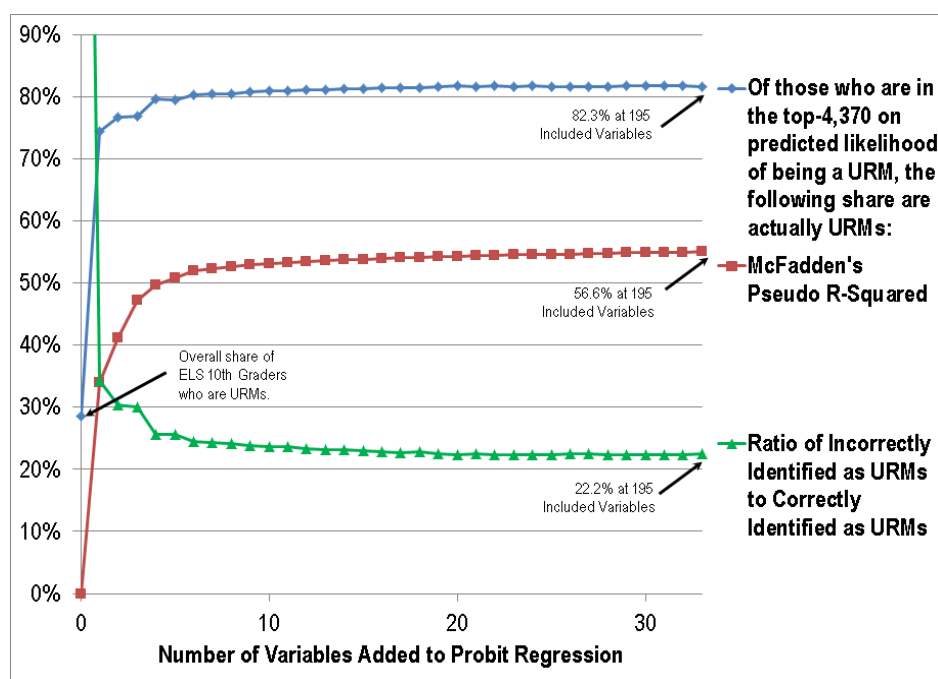

Figure 4. Same as Figure 2, but for Education Longitudinal Study (ELS) data. URM = underrepresented minority.

Table 3 First 10 Variables Entered Into Probit Regression

No.	UT-Austin	ACS	ELS
1	High school's average SAT/ACT score	Living with two parents who are both U.S. natives	Best friend #1 is a URM
2	Student's SAT/ACT score	Number of bedrooms in household/persons in household	Native language is Spanish
3	Father's income is >= $80,000	Household receives SNAP	Best friend #2 is a URM
4	U.S. citizen	Living with father only who is a U.S. native	Best friend #3 is a URM
5	Longhorn Opportunity Scholarship high school	Household rent	ELS 10th grade math test score squared
6	Passed AP foreign language test	Household income	High school's percent of students who are minorities
7	Private high school	Living with mother only who is a foreign born	Live with two parents/guardians
8	Neither parent attended high school	Living with two parents who are both foreign born	Parents expect student to earn a graduate degree
9	$60,000 <= father's income < $80,000	Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas	Born in the United States
10	$40,000 <= father's income < $60,000	Living with father only who is foreign born	Student participated in interscholastic football

Note. ACS = American Community Survey, AP = advanced placement, ELS = Education Longitudinal Study, SNAP = Supplemental Nutrition Assistance Program, URM = underrepresented minority, UT = University of Texas.

Further note that school attributes show up prominently in these lists. For UT Austin, a high school's average score on the *SAT®* test or ACT test, being a Longhorn Opportunity School, and being a private school are each highly predictive. After the *Hopwood* decision led to a ban on affirmative action in Texas, UT Austin named a set of schools to be Longhorn Opportunity Schools. This set of schools includes those with high shares of low-income students and that have historically low rates of sending students to UT Austin.[9] These schools, as one would expect, also have high shares of students who are URMs. Although not included in the model for UT Austin, it would be easy to add the direct measure of the high schools' percentages of students who are URMs (derived from the ED's Common Core of Data and Private School Survey), as I have done in the third column using the ELS data. Note that *high school's percent of students who are minorities* was the sixth most important variable in predicting URM status for the ELS data, and this suggests that the power of prediction might have been somewhat improved for UT Austin had I included this measure.

When I restrict the ELS data to the students who submitted at least one application to a very competitive college, I find similar results. Using the full set of 195 variables, the pseudo-R^2 is

54.4% (as compared to 56.6% using all 10th graders), and the rate of correctly tagging URMs rises from 18.7% at base using blind luck to 76.5% for the full model (as compared to 28.6% [blind luck] rising to 82.3% [full model] using all 10th graders).

As discussed in the next section, the use of some applicant characteristics (such as the race or ethnicity of the applicant's friends and participation in a college preparation program for disadvantaged students, many of which are specifically targeted to racial minorities) might prompt legal challenge, because such characteristics would appear to be too nakedly an attempt to create a proxy for the applicant's race. Other variables that are included in ELS based on parental surveys would be problematic. *How often the parent knows 10th grader's whereabouts* would be nearly impossible to accurately collect from the applicant, whereas the parents' frequency of attending religious services with the 10th grader would draw political and legal challenge. Most of the variables that are included in ELS based on administrator surveys would be unavailable in administrative databases or in other surveys (e.g., ED's Common Core of Data and Private School Survey). In Figure 5, I show the effect of eliminating problematic variables from Z. As shown, using the reduced set of variables modestly reduces the accuracy of the prediction of the student's URM status. When using the full set of 171 nonproblematic variables, the rate of correctly tagging URMs is 75.7% (as opposed to 82.3% when using all 195 variables), and correspondingly, one incorrectly tags 0.32 students who are not underrepresented minorities for every 1 student whom one correctly tags as a URM (rather than 0.22 when using all 195 variables).[10]
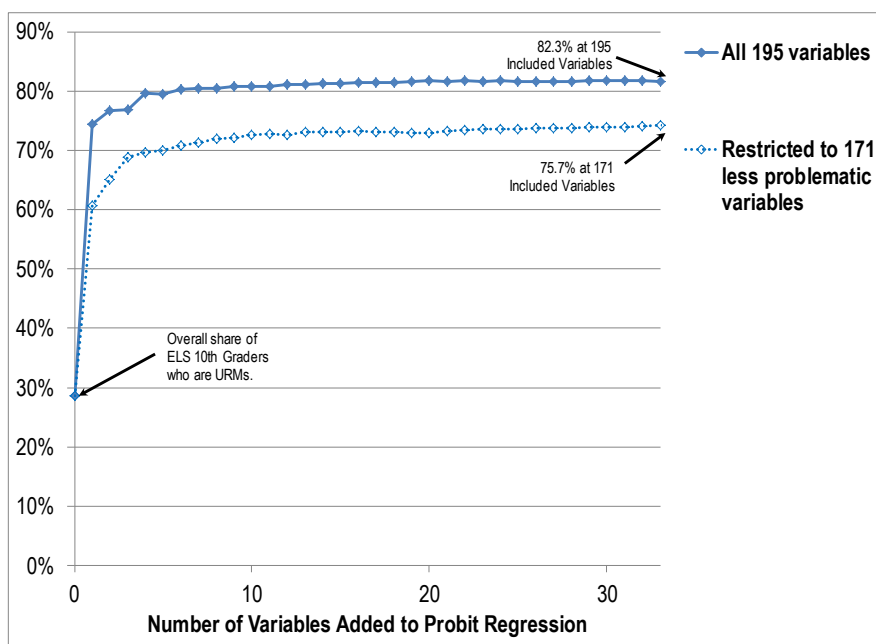

Figure 5. Education Longitudinal Study of 2002 data: Effect of excluding more problematic variables. ELS = English as a second language, URM = underrepresented minority.

The Civil Rights Project
Proyecto Derechos Civiles
Using Correlates of Race in Admissions

Discussion: Is the Use of Correlates of Race Workable and Race Neutral?

Replacing URM status with predicted likelihood of being a URM in the university's admissions decisions has serious limitations that challenge its workability and might not be deemed race neutral by the courts.

First, let us evaluate whether it is workable. If URM status cannot be perfectly predicted, then to maintain the same racial composition of admitted students, universities would need to put more weight on $\widehat{URM}_i$ than they formerly put on URM_i. This shifting weight will likely reduce the quality of admitted students, where *quality* means the multidimensional desirability of the class of admitted students from the perspective of the university. To illustrate this concept, consider two admissions systems:

$$\text{Admit} = 1 \text{ if } [\mu_1 URM_i + X_i'\omega > \text{threshold}]. \tag{2}$$

$$\text{Admit} = 1 \text{ if } [\mu_2 \widehat{URM}_i + X_i'\omega = \mu_2 \Phi(Z_i'\hat{\beta}) + X_i'\omega > \text{threshold}]. \tag{3}$$

Suppose that Equation 2 represents the university's ideal admissions system (i.e., the admissions system that will yield the most desirable class of admitted students). If the university could perfectly predict URM status, then $URM_i = \widehat{URM}_i$, and Equation 3 would yield the exact same group of admitted students. Given the challenges in perfectly predicting URM status that this paper has illustrated, μ_2 would need to be made larger than μ_1 to yield the same number of admitted URMs, as a certain number of non-URMs would incorrectly receive the advantage conferred by $\mu_2 \widehat{URM}_i$ and thus would be more likely to be admitted than they would have been under Equation 2. Making $\mu_2 > \mu_1$ results in a shift of emphasis in admissions from $X_i'\omega$ toward $Z_i'\hat{\beta}$ and thus distorts the university's optimal admitted class. Simulations conducted by Fryer, Loury, and Yuret (2008) and Long (2015) have demonstrated that such alterative admissions systems are inefficient relative to traditional affirmative action.

To ward off the adverse effects of using an imperfect predictor of race, the university could seek to obtain additional information on students to help predict their URM status. The university may want to go to great lengths and follow the path of private businesses that try to predict the characteristics of their customers. For example, as Duhigg (2012) noted,

> Target can buy data about your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought (or lost) your house, where you went to college, what kinds of topics you talk about online, whether you prefer certain brands of coffee, paper towels, cereal or applesauce, your political leanings, reading habits, charitable giving and the number of cars you own.

Although universities may want to go down this path, they may be thwarted by the monetary cost of purchasing such information, the political challenge that would be likely to follow from such privacy invasion, and the distaste it would engender in applicants.

An additional concern is that using correlates of minority status will invite adverse behavioral responses on the part of students (and possibly parents and school administrators). If students become aware that predicted likelihood of being a URM is a factor in admissions, they will be incentivized to "invest" in those characteristics in Z that are positively weighted. For example, Cullen, Long, and Reback (2013) investigated whether Texas's top 10% policy led students to strategically enroll in high schools that would increase their chances of being in the top 10% by choosing a high school with lower achieving peers. Cullen et al. found that

> among the subset of students with both motive and opportunity for strategic high school choice, at least 5% enroll in a different high school to improve the chances of being in the top 10%. These students tend to choose the neighborhood high school in lieu of transferring to more competitive schools and, regardless of own race, typically displace minority students from the top 10% pool. (p. 32)

Thus students are engaging in costly behavior that has no productive rationale other than to effectively undo the reweighting of applicant characteristics that is generated by the top 10% policy. One should expect such undesirable responses in reaction to the use of a URM proxy. Additionally, one should expect some degree of lying on admissions forms with respect to questions that are hard to verify. For example, if universities followed the ELS lead in asking for the race or ethnicity of the applicant's best friend, there would be no way to confirm or refute the applicant's claim. Conversely, if universities only used existing information from their current application forms to create $\widehat{URM}_i$, doing so would reduce adverse behavioral responses, yet the poor results using the UT Austin data suggest that such a strategy is not workable given the high number of non-URMs who would be incorrectly tagged and thus given weight in the admissions decision.

Next, let us evaluate whether using Equation 3 is race neutral from a legal perspective. The short answer is probably not. A cogent discussion of the definition of *race neutral* is found in a College Board report by Coleman, Palmer, and Winnick (2008). They noted that "policies that are neutral on their face but that are motivated by a racially discriminatory purpose, resulting in racially discriminatory effects," are deemed "race-conscious policies" and "trigger 'strict scrutiny' review" (p. 4). They further noted that

> as the term [*proxy*] is frequently used, it can literally refer to a substitute for a race-conscious policy or program, where the racial diversity goal is precisely the same. In that circumstance, the mere shift to a potentially viable "proxy" for race may mitigate some risk of being sued (nothing on the face of the policy would indicate

that race was a factor in relevant decisions)—but, with evidence that racial goals were driving the development and implementation of the policy, such a shift would not likely insulate the "proxy" policy from strict scrutiny review. (p. 7)

Thus, if the *intent* of the university's new admissions system, as given in Equation 3, is to confer advantage to URMs, it might not be considered race neutral based on the analysis by Coleman et al. (2008). To further illustrate this point, Coleman et al. cited a case decided by the ED's Office of Civil Rights:

> If the evidence shows a deliberate use of race-neutral criteria as proxies for race . . . OCR would then apply Title VI strict scrutiny standards. . . . Proxy allegations raise issues of intentional discrimination, [for which certain] . . . factors may be evidence of intent to discriminate, [including] the impact of the official action (i.e., whether it impacts more heavily upon one racial group than another); a pattern of discrimination unexplainable on grounds other than race; the historical background of a decision, particularly the specific sequence of events leading to the challenged policy; departure from the normal procedural sequence; and the legislative or administrative history, particularly contemporaneous statements of members of the decision-making body. (p. 7)

To make this idea concrete, suppose a university had a simple admissions rule before the *Gratz* and *Grutter* rulings:

$$\text{Admit} = 1 \text{ if } [URM_i + GPA_i > 4]. \tag{4}$$

After these rulings, suppose the university instituted the following new admissions rule, where SES reflects the student's socioeconomic status (measured continuously from 0 = highest SES to 1 = lowest SES) and FRPL is the percentage of the student's high school's students who received free or reduced-price lunch:

$$\text{Admit} = 1 \text{ if } [0.5 \times SES_i + 0.5 \times FRPL_i + GPA_i > 4]. \tag{5}$$

Whether courts deem Equation 5 to be race neutral may depend on the intent of the university in using this admissions system. If the university became motivated to care about low-income students and students from low-income backgrounds, and it instituted Equation 5 for the purpose of utilizing so-called class-based affirmative action, then Equation 5 would likely be deemed race neutral, even if such an admissions system yielded the side benefit of increasing racial diversity of the admitted class. The U.S. Supreme Court appears to be hoping that universities will become inspired in this manner and that utilizing such admissions systems will serve as one possible workable race-neutral alternative that would produce the educational benefits of diversity.

Now, suppose instead that prior to implementing Equation 5, the university had undertaken a linear probability model approach to estimate Equation 1 using only SES and FRPL as predictors of URM status and found that $\beta_{SES} = \beta_{FRPL} = 0.5$, and thus $\widehat{URM}_i =$ $0.5 \times SES_i + 0.5 \times FRPL_i$, and thus used $0.5 \times SES_i + 0.5 \times FRPL_i$ as a proxy for URM_i, yielding Equation 5. The development of this proxy indicator might not be deemed race neutral by a reviewing court, as the intent is to confer advantage to URMs.[11] That is, it is not the form of the admissions system that determines whether it is race neutral; rather, it is the intent of university officials that may matter.[12]

So, if running a regression to predict URM status and then using the resulting prediction in an admissions system like 2b may not be deemed by the courts to be race neutral, then what is to be gained by the exercise conducted in this paper? The aim of this paper is not to show universities how to implement such a system; rather, the aim is to illustrate both the promise that is suggested by using correlates of race and the legal, political, and educational peril in doing so. The promise is illustrated by the relatively successful results illustrated when using 195 characteristics of ELS students and correctly predicting the student's URM status 82% of the time. If one were to obtain additional information on an applicant (perhaps by buying information on household purchasing habits), one could probably do even better in correctly predicting URM status. However, doing so nakedly may not be deemed race neutral, would add cost to the university's applicant review, would engender adverse behavioral responses, and would invite political challenge. Thus, instead, the university is invited by the courts to think about other goals (e.g., improving access for low-income students) and hope that such efforts will yield side benefits to racial and ethnic diversity. Pursuing these approaches would not yield as many admitted URMs as would be achieved by the direct approach shown in Equation 3, which draws into question the extent to which such indirect so-called race-neutral approaches are workable.

References

Antonovics, K., & Backes, B. (2014). The effect of banning affirmative action on college admissions rules and student quality. *Journal of Human Resources, 49*(2), 295–322.

Barron's. (2004). *Profiles of American colleges: 2005* (26th ed.). New York, NY: Baron's Education Series.

Blume, G. H., & Long, M. C. (2014). Changes in levels of affirmative action in college admissions in response to statewide bans and judicial rulings. *Educational Evaluation and Policy Analysis*, *36*(2), 228–252.

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

Chan, J., & Eyster, E. (2003). Does banning affirmative action lower college student quality? *American Economic Review, 93*(3), 858–872.

Coleman, A. L., Palmer, S. R., & Winnick, S. Y. (2008). *Race-neutral policies in higher education: from theory to action*. Retrieved from http://advocacy.collegeboard.org/sites/default/files/Race-Neutral_Policies_in_Higher_Education.pdf

Cullen, J. B., Long, M., & Reback, R. (2013). Jockeying for position: High school student mobility and Texas' top-ten percent rule. *Journal of Public Economics, 97*(1), 32–48.

Dickson, L. M. (2006). Does ending affirmative action in college admissions lower the percent of minority students applying to college? *Economics of Education Review, 25*(1), 109–119.

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times Magazine*. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all

Fisher v. University of Texas, 570 U.S. __ (2013).

Fryer, R. G., Jr., Loury, G. C., & Yuret, T. (2008). An economic analysis of color-blind affirmative action. *Journal of Law, Economics & Organization, 24*, 319–355.

Gratz v. Bollinger, 539 U.S. 244 (2003).

Grutter v. Bollinger, 539 U.S. 306 (2003).

Hopwood v. Texas, 78 F.3d 932 (5th Cir. 1996), cert. denied, 518 U.S. 1033 (1996).

Jaschik, S. (2006, December 6). Proxies for race. *Inside Higher Ed*. Retrieved from
http://www.insidehighered.com/news/2006/12/06/wayne

Laird, B. (2005). *The case for affirmative action in university admissions.* Berkeley, CA: Bay
Tree.

Long, M. C. (2004). Race and college admission: An alternative to affirmative action? *Review
of Economics and Statistics, 86*(4), 1020–1033.

Long, M. C. (2015). Is there a "workable" race-neutral alternative to affirmative action in
college admissions? *Journal of Policy Analysis and* Management, *34*(1), 162–183.

Long, M. C., & Tienda, M. (2008). Winners and losers: Changes in Texas University
admissions post-*Hopwood. Educational Evaluation and Policy Analysis*, *30*(3), 255–
280.

Tienda, M., & Niu, S. (2006). Flagships, feeders, and the Texas top 10% plan. *Journal of
Higher Education, 77*(4), 712–739.

The Civil Rights Project
Proyecto Derechos Civiles

Using Correlates of Race in Admissions

Notes

[1] I group Asian American students and White students together because affirmative action practices at colleges and universities in the United States have historically not given preference to Asian students (Bowen & Bok, 1998; Laird, 2005). Long (2004) and Long and Tienda (2008) found no significant advantage or disadvantage given to Asian applicants relative to White applicants using, respectively, national college admissions data from 1992 and institutional admissions data from the University of Texas in pre-*Hopwood* years, yet Long and Tienda found a modest disadvantage given to Asian applicants at Texas A&M University in pre-*Hopwood* years.

[2] McFadden's pseudo-R^2 is defined as $1 -$ log-likelihood of a model predicting URM with $\widehat{URM_i}$/log-likelihood of a model predicting URM with $\overline{URM}$. If one could perfectly predict URM status, then McFadden's pseudo-R^2 would equal 1. If one had no ability to improve on $\overline{URM}$ as a predictor of URM$_i$ for all students, then McFadden's pseudo-R^2 would equal 0.

[3] I drop all individuals who have missing indicators for race or ethnicity. The same restriction is placed on the other two sources of data.

[4] This procedure is done using Stata's impute command. Note that multiple imputation is unneeded in this context because I am not using Equation 1 for hypothesis testing and thus have no need to add uncertainty to our prediction of missing variables. Rather, I am simply trying to predict URM using available information as efficiently as possible.

[5] Missing values in the ACS data were imputed by the U.S. Census Bureau.

[6] Sample sizes are rounded to the nearest 10 owing to requirements of the ED for using restricted-access data.

[7] Unlike for the UT Austin data, I impute each missing variable using only variables that have no missing values (instead of the best available subset of available nonmissing information using all variables), again using Stata's impute command. I do this because Stata's impute command has a limit of 31 variables in each imputation. The right-hand-side variables in these imputations included family's socioeconomic status composite (Version 2); scores on ED-administered 10th-grade math and reading tests; 10th-grade enrollment; student lived with two parents; a set of indicators for parent's highest level of education; and indicators for parent's income (which was itself imputed by ED if missing).

[8] Furthermore, even this low level of predictive validity might be somewhat optimistic if the university were to maintain its set of weights constant across years. What might be highly predictive of minority status for one cohort may not be as predictive for another cohort. To cross-validate the results and to get a sense of

how accurate the predictions would be if the obtained equations were applied to another data set, I ran the probit regression on the 1998 cohort of applicants and applied the resulting weights to the 2000 cohort. In doing so, I found that with 33 variables in the model, I got close to the same performance as found using the full data (e.g., out of the 3,230 applicants in the 2000 cohort with the highest likelihood of being a URM, 52.4% of them are in fact URMs). I get similar performance splitting the full sample in half, running the probit on one half of the data, and applying the results to the other half of the data.

[9] Dickson (2006) described the early years of the program as follows: "This program offers scholarships of $4000 each year to students who graduate from qualifying high schools. High schools qualify for the program if the average parental income at the high school is less than $35,000 and if less than 35% of the high school graduates sent college admission test scores to the University of Texas in the previous year" (p. 110). Dickson listed 68 high schools as program recipients by 2000.

[10] Following is the list of the top 10 variables entered using the nonproblematic variables, in order: percentage of the school's students who are minority students, native language is Spanish, score on ED-administered 10th-grade math test, two parents living with 10th grader, student participated in interscholastic basketball, parents expect student to earn a graduate degree, grade point average, born in the United States, hours per day spent watching TV/DVD on weekends, and family has access to the Internet.

[11] See Jaschik (2006) for a discussion of a debate about whether a new admissions system at the law school at Wayne State University was serving as a proxy for race.

[12] However, the university may be able to argue that it is intending to benefit all of its enrolled students by providing the educational benefits that flow from having a diverse student body. In this sense, the university's intent is not to be racially discriminatory. There is doubt regarding how a court would view this argument. (I would like to thank Liliana Garces for suggesting this argument to me in a personal correspondence. I would also like to thank Judith Winston for providing very helpful feedback and references that helped me develop this section of the paper.)

Appendix: Education Longitudinal Study of 2002 (ELS) Data: List of Variables and Pre- and Postimputation Means

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Student is						
An underrepresented minority	15,240	28.6%	28.6%	4,770	18.7%	18.7%
Black (Non-Hispanic)	15,240	13.3%	13.3%	4,770	9.6%	9.6%
Native American (Non-Hispanic)	15,240	0.9%	0.9%	4,770	0.4%	0.4%
Hispanic	15,240	14.5%	14.5%	4,770	8.7%	8.7%
Student's academic achievement measures						
Score on ED-administered 10th grade math test	15,240	50.72	50.72	4,770	57.22	57.22
Score on ED-administered 10th grade reading test	15,240	50.55	50.55	4,770	56.56	56.56
Grade point average	13,980	2.72	2.71	4,430	3.23	3.21
Number of AP/IB courses by 10th grade	13,990	0.75	0.75	4,440	1.79	1.76
Number of AP exams with a 3+ score by 10th grade	15,240	0.28	0.28	4,770	0.77	0.77
How far student thinks he/she will get in school						
Some college	13,790	10.5%	10.8%	4,520	2.5%	2.7%
Bachelor's degree (but not higher)	13,790	39.3%	39.3%	4,520	33.9%	34.1%
Graduate degree	13,790	42.2%	41.5%	4,520	62.7%	62.1%
Student's views (1 = strongly agree, …, 5 = strongly disagree):						
Students at my school get along well with teachers	14,550	2.20	2.20	4,550	2.08	2.09

The Civil Rights Project
Proyecto Derechos Civiles

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Students at my school are friendly with other racial groups	14,490	1.80	1.80	4,530	1.76	1.76
Other students at my school often disrupt class	14,470	2.11	2.11	4,530	2.22	2.22
The teaching is good at my school	14,390	2.06	2.06	4,500	1.95	1.95
Do not feel safe at my school	14,390	3.27	3.27	4,510	3.43	3.43
There are gangs in my school	14,340	2.94	2.94	4,490	3.12	3.11
The rules at my school are strictly enforced	14,420	2.23	2.23	4,500	2.22	2.22
Had something stolen at my school	14,560	1.46	1.46	4,540	1.42	1.42
Someone offered drugs at my school	14,540	1.32	1.32	4,550	1.22	1.23
Classes at my school are interesting and challenging	14,520	2.40	2.40	4,540	2.30	2.30
Education is important to get a job later	14,470	1.42	1.42	4,530	1.31	1.32
Teachers at my school expect success in school	14,460	2.31	2.31	4,520	2.26	2.27
Parents expect success in school	14,490	1.54	1.54	4,530	1.49	1.49
How much likes school	14,680	2.13	2.13	4,630	2.21	2.21
Thinks reading is fun	11,810	2.49	2.51	4,010	2.31	2.33
Thinks math is fun	11,670	2.79	2.79	3,970	2.66	2.68
Important to friends to go to parties	10,340	2.14	2.14	3,550	2.13	2.13
Student's reported importance of (1 = not; 2 = somewhat; 3 = very important)						
Having lots of money	14,590	2.33	2.33	4,650	2.27	2.27

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Being able to find steady work	14,480	2.83	2.83	4,620	2.86	2.86
Working to correct inequalities	14,500	1.92	1.92	4,640	1.91	1.91
Student's reported hours/week spent on homework						
In school	14,670	4.72	4.71	4,670	4.69	4.69
Out of school	14,790	6.08	6.03	4,710	8.21	8.20
Student's reported hours/day on						
Computer for school work	13,960	1.18	1.19	4,460	1.31	1.30
Computer other than for school	13,990	2.20	2.20	4,470	2.19	2.19
Watching TV/DVD on weekdays	14,050	2.97	2.98	4,460	2.51	2.52
Watching TV/DVD on weekends	14,070	4.00	4.01	4,480	3.85	3.85
Video/computer games on weekdays	14,000	1.09	1.09	4,450	0.80	0.81
Video/computer games on weekends	14,040	1.76	1.76	4,470	1.53	1.54
Student ever in/participated in						
Science/math fair at my school	14,490	14.0%	14.0%	4,530	17.9%	17.8%
Voc/tech skills competition at my school	14,470	7.9%	7.9%	4,530	6.3%	6.3%
ESL program	14,240	8.6%	8.7%	4,490	5.4%	5.4%
Dropout prevention program	14,330	2.9%	2.9%	4,510	1.2%	1.2%
Special education program	14,300	7.5%	7.6%	4,510	3.7%	3.8%
Program to help prepare for college	14,310	22.5%	22.5%	4,500	25.3%	25.1%
School band or chorus	14,890	22.0%	21.9%	4,700	27.2%	27.1%
School play or musical	14,890	12.2%	12.2%	4,700	15.1%	15.1%
Student government	14,840	7.2%	7.1%	4,680	12.1%	12.1%
Academic honor society	14,870	9.6%	9.5%	4,690	17.4%	17.4%

The Civil Rights Project
Proyecto Derechos Civiles

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
College preparation program for disadvantaged [a]	10,250	4.8%	5.3%	3,840	4.3%	4.2%
Student's interscholastic participation in						
Baseball	14,030	7.1%	7.1%	4,410	6.7%	6.7%
Softball	14,020	9.9%	9.9%	4,420	10.4%	10.4%
Basketball	14,030	13.8%	13.8%	4,440	15.3%	15.2%
Football	14,060	16.6%	16.7%	4,420	15.7%	15.7%
Soccer	13,900	9.7%	9.7%	4,380	13.4%	13.3%
Other team sport	13,940	24.6%	24.6%	4,400	33.4%	33.1%
Solo sport	13,980	14.2%	14.1%	4,420	20.4%	20.3%
Cheerleading	14,020	12.8%	12.8%	4,410	13.3%	13.3%
Student reports that family has						
Daily newspaper	13,390	64.7%	64.3%	4,430	71.6%	71.2%
Computer	13,390	88.8%	88.0%	4,430	96.7%	96.3%
Access to the Internet	13,420	82.8%	81.9%	4,440	93.2%	92.7%
More than 50 books	13,380	82.1%	81.2%	4,420	91.4%	90.9%
Other yes/no questions for student						
Recognized for good attendance at my school	14,480	21.9%	21.9%	4,530	21.1%	21.0%
School has library media/resource center	14,230	95.6%	95.6%	4,500	97.4%	97.3%
Did not participate in work-based learning experiences	13,000	57.4%	57.4%	4,290	50.7%	51.2%
Ever worked for pay not around house	13,110	58.7%	58.5%	4,300	58.0%	58.0%

	All 10th graders			10th graders who apply to very competitive or higher colleges		
Variable	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Has own room	13,330	86.5%	86.0%	4,410	89.3%	89.1%
Observed students betting on sports	9,740	49.1%	49.1%	3,380	49.7%	49.6%
Native language is Spanish	14,850	7.0%	7.1%	4,710	3.5%	3.6%
Has an individualized education plan	7,980	11.9%	11.8%	2,480	2.9%	3.4%
1st friend is an underrepresented minority [a]	14,130	27.8%	28.1%	4,450	18.3%	18.3%
2nd friend is an underrepresented minority [a]	13,840	28.8%	29.2%	4,380	19.3%	19.3%
3rd friend is an underrepresented minority [a]	13,160	28.9%	29.5%	4,190	20.1%	20.2%
Student held job for pay during 2001–2002 school year	12,840	37.7%	37.6%	4,230	35.9%	36.0%
Took or plans to take commercial SAT/ACT preparation course	10,150	13.9%	13.5%	3,840	22.5%	21.3%
Received or plans to receive private tutoring for SAT/ACT	10,130	9.7%	9.9%	3,820	12.6%	12.1%
Parent reported:						
# of dependents	12,420	2.67	2.69	4,200	2.59	2.60
How often know 10th grader's whereabouts (1 = never; 2 = seldom; 3 = usually; 4 = always) [a]	12,280	3.80	3.80	4,140	3.84	3.83
Provide 10th grader advice about applying to college/school after HS (1 = never; 2 = sometimes; 3 = often)	12,220	2.12	2.11	4,130	2.27	2.25

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Provide 10th grader advice about jobs to apply for after high school (1 = never; 2 = sometimes; 3 = often)	12,230	2.15	2.15	4,140	2.10	2.11
Attended religious services with 10th grader (1 = never; 2 = rarely; 3 = sometimes; 4 = frequently) [a]	12,270	3.05	3.03	4,150	3.18	3.16
Went to restaurants with 10th grader (1 = never; 2 = rarely; 3 = sometimes; 4 = frequently)	12,310	3.39	3.38	4,160	3.45	3.44
Level of crime in neighborhood (1 = high; 2 = moderate; 3 = low)	12,240	2.87	2.86	4,140	2.91	2.90
How safe is neighborhood (1= very safe; 2 = somewhat safe; 3 = somewhat unsafe; 4 = very unsafe)	12,250	1.37	1.38	4,130	1.28	1.29
Away from home while attending post-sec important to parent (1 = somewhat important; 2 = very important; 3 = not important)	11,650	2.12	2.12	4,060	2.05	2.06
Parents' highest level of education is						
HS dropout	15,240	6.1%	6.1%	4,770	1.9%	1.9%
HS graduate	15,240	19.8%	19.8%	4,770	10.0%	10.0%
Some college	15,240	32.9%	32.9%	4,770	24.5%	24.5%
Bachelor's degree (but not higher)	15,240	22.7%	22.7%	4,770	30.6%	30.6%
Family income is						
<= $25,000	15,240	20.8%	20.8%	4,770	10.9%	10.9%
$25,001-$50,000	15,240	30.5%	30.5%	4,770	22.3%	22.3%

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
>=$100,001	15,240	14.8%	14.8%	4,770	28.0%	28.0%
Family's socioeconomic status composite (Version 2)	15,240	0.04	0.04	4,770	0.45	0.45
How far in school wants 10th grader to go						
Some college	13,090	8.4%	8.8%	4,330	1.0%	1.4%
Bachelor's degree (but not higher)	13,090	44.1%	44.3%	4,330	37.0%	37.5%
Graduate degree	13,090	44.1%	43.3%	4,330	61.6%	60.4%
Other yes/no questions for parent						
English is parent respondent's native language	13,270	81.9%	81.6%	4,390	80.7%	80.9%
Belong to parent-teacher organization	12,190	26.2%	25.4%	4,110	38.4%	37.1%
Attend parent-teacher organization meetings	12,210	36.6%	36.7%	4,120	39.7%	39.1%
Take part in parent-teacher organization activities	12,100	31.0%	30.3%	4,080	39.8%	38.6%
Act as a volunteer at the school	12,130	31.7%	30.7%	4,090	43.2%	41.8%
10th grader has biological/adoptive parent living outside home	12,230	29.4%	30.4%	4,150	19.9%	21.3%
Done something specific to have some money for 10th grader's education after HS	11,310	55.0%	52.5%	4,090	66.7%	65.1%
Two parents living with 10th grader	15,240	74.9%	74.9%	4,770	81.6%	81.6%
Mother was born in the United States	13,250	78.1%	77.8%	4,380	75.7%	75.9%
Father was born in the United States	13,110	78.1%	77.9%	4,330	76.5%	76.7%

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
10th grader was born in the United States	13,340	89.5%	89.4%	4,420	89.5%	89.6%
School is a						
Public magnet school	14,920	7.5%	7.5%	4,670	7.4%	7.4%
Area vocational school/center	14,940	38.8%	38.9%	4,670	31.6%	31.7%
Full-time technical/vocational school	15,000	4.4%	4.4%	4,700	3.4%	3.4%
Other technical or vocational school	14,960	3.0%	3.0%	4,690	3.0%	3.0%
Catholic diocesan school	14,890	7.8%	7.7%	4,660	12.9%	12.7%
Catholic parish school	14,920	1.0%	1.0%	4,660	1.3%	1.3%
Catholic religious order school	14,990	4.3%	4.3%	4,680	8.5%	8.5%
Catholic independent school	14,940	2.2%	2.2%	4,680	4.3%	4.3%
Other private school with religious affiliation	15,000	5.4%	5.3%	4,700	7.5%	7.4%
Private school without religious affiliation	14,930	4.2%	4.1%	4,670	7.6%	7.5%
Boarding school	14,920	1.8%	1.8%	4,670	3.5%	3.5%
Indian reservation school	14,940	0.2%	0.2%	4,680	0.0%	0.0%
Military academy	14,950	0.1%	0.1%	4,690	0.1%	0.1%
Alternative/dropout prevention/continuation school	14,940	2.9%	2.9%	4,690	2.9%	2.9%
Charter school	14,940	1.1%	1.1%	4,660	0.4%	0.4%
Grade 10 enrollment-2001–2002 school roster	15,240	318	318	4,770	302	302
Student/teacher ratio	14,790	16.56	16.55	4,670	15.84	15.85

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Percent of the school's students that are minority students	14,960	33.78	33.78	4,720	28.99	28.98
Percent of your student body (or 10th graders) participated in/are						
Academic counseling program	12,690	79.18	79.31	4,030	81.95	81.77
Vocational counseling program	10,030	47.81	47.60	2,760	46.30	46.30
Dropout prevention program	7,300	19.49	19.14	1,820	18.13	17.85
Gang prevention program	3,510	23.43	24.05	940	22.62	22.29
Alcohol/drug prevention program	10,910	43.90	43.70	3,450	48.41	47.72
AIDS education program	9,310	42.53	42.54	2,940	46.42	45.61
Crisis prevention program	7,760	41.25	42.17	2,370	45.15	45.62
General HS program	12,360	45.03	45.48	3,880	34.93	36.13
College prep program	12,770	61.99	61.47	4,080	73.22	72.10
Other specialized programs	11,950	12.58	12.72	3,840	12.24	12.35
Voc/tech/business program	12,370	15.12	15.33	3,910	10.92	11.38
Special ed program	14,250	9.26	9.29	4,450	7.28	7.37
Alternative program	12,180	1.92	1.97	3,850	1.47	1.52
Receive bilingual education	12,130	3.65	3.62	3,810	3.29	3.24
Receive ESL	12,380	3.60	3.81	3,900	3.16	3.30
Receive remedial reading	12,450	4.23	4.33	3,900	3.21	3.34
Receive remedial math	12,240	5.79	5.91	3,850	4.37	4.55
After school/summer outreach	12,140	3.81	3.89	3,850	2.95	3.05
LEP or non-English proficient	14,490	4.06	4.18	4,570	3.16	3.27
Receive free/reduced-price lunch	13,990	23.48	23.65	4,430	15.02	15.32

The Civil Rights Project
Proyecto Derechos Civiles

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Tardiness [a]	12,720	1.15	1.14	4,060	1.22	1.21
Absenteeism	12,690	1.20	1.19	4,040	1.29	1.28
Physical conflicts [a]	12,720	3.60	3.57	4,070	3.76	3.73
Robbery/theft [a]	12,740	3.73	3.72	4,070	3.74	3.73
Vandalism [a]	12,730	3.89	3.88	4,070	3.95	3.93
Alcohol [a]	12,730	3.93	3.92	4,070	3.95	3.94
Illegal drugs [a]	12,710	3.87	3.86	4,060	3.90	3.89
Gang activity [a]	12,740	4.54	4.52	4,070	4.65	4.63
Learning hindered by poor condition of buildings (1 = not at all; 2 = very little; 3 = to some extent; 4 = a lot) [a]	12,690	1.54	1.54	4,050	1.45	1.46
How accurate is the following (1 = not at all; 2 = very little; 3 = to some extent; 4 = a lot)						
Student morale is high [a]	12,680	3.98	3.97	4,030	4.12	4.11
Teacher morale is high [a]	12,690	3.80	3.79	4,050	3.93	3.92
Lowest salary paid to full-time teachers	12,120	28,300	28,421	3,820	28,381	28,413
Percentage of your full-time teachers who						
Are certified	14,670	91.99	92.00	4,520	89.35	89.39
Teach out of field	12,420	4.25	4.18	3,900	4.11	4.11
Other yes/no questions for administrator						
Good teachers receive higher pay [a]	13,170	5.4%	5.3%	4,190	7.5%	7.3%
Students must pass a test for high school diploma [a]	13,320	57.0%	58.0%	4,200	48.1%	49.3%
Require students pass through metal detector [a]	13,400	2.4%	2.5%	4,230	1.1%	1.2%

Variable	All 10th graders			10th graders who apply to very competitive or higher colleges		
	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean	Number of nonmissing observations	Pre-imputation mean	Post-imputation mean
Require drug testing for any students [a]	13,310	16.0%	15.9%	4,190	15.9%	15.8%
Require students to wear uniforms [a]	13,370	15.9%	15.4%	4,230	23.1%	22.1%
Enforce strict dress code [a]	13,260	56.7%	56.5%	4,180	57.1%	56.9%
Emergency call button in classrooms [a]	13,310	53.8%	54.2%	4,200	48.4%	49.2%
Use paid security at any time during school hours [a]	13,370	64.2%	65.7%	4,220	59.5%	60.8%

Note. Squares of all continuous variables (i.e., those not listed above with % symbols) are included as well. Sample sizes are rounded to the nearest 10 due to requirements of the U.S. Department of Education for using restricted-access data. AP = Advanced Placement, ED = U.S. Department of Education, ESL = English as a second language, HS = high school, IB = International Baccalaureate, LEP = limited English proficiency, post-sec = postsecondary; tech = technical, Voc = vocational.

[a] Italicized variables are those that would either invite legal or political challenge, or be difficult to obtain in an admissions application.

The Civil Rights Project
Proyecto Derechos Civiles