sociological science

Supplement to:

Williams-Baron, Emma, Aliya Saperstein. 2026. “A Roadmap for Inequality Research: Transparency, Intersectionality, and Multiple Measures of Race” Sociological Science 13: 825-863.

Table of Contents

Appendix A. Approaches to Specifying Race and Ethnicity

Appendix A Section 1. Self-Identified Race and Ethnicity Measures

The 1979 wave of the National Longitudinal Survey of Youth asked respondents "What is your origin or descent?" and presented a list of more than two dozen responses, of which they could select up to six. The list of responses is shown in Panel A of Table A1. All but 107 respondents answered this question; the nonresponses include one person who refused to answer, 11 invalid skips, and 98 people who responded they did not know their origin or descent.

In the 2002 survey wave, following updated federal government standards, respondents were first asked whether they were of Hispanic, Latino, or Spanish origin (yes/no), and then asked in a separate question to indicate "what race or races do you consider yourself to be?" by marking all that applied from a list of six options, with a seventh option that interviewers could use to code "respondent refuses to classify race except as Hispanic/Latino." These data are available for download from the NLSY Investigator site as a dichotomous variable for the first Hispanic/Latino question and a series of seven dichotomous variables for each race category. Sample sizes for these groups by gender are shown in Appendix Table A1. Sixty-one percent of the original respondents were interviewed in 2002 (7,723 of 12,686).

As all analysts working with these data do, we faced a series of decisions about how to operationalize the self-identification 1979 data. The raw data are available through the NLS Investigator site as six separate variables: the first variable contains respondents' first self-identified category from the list of 30 options, the second variable contains the second self-identified category, and so on, for up to six possible self-identified categories per person. From these variables, we created thirty category-level dichotomous variables coding for "respondent self-identified as [category] in any of the six possible responses." After generating these

variables, we faced the empirical and theoretical decision points that we discuss in the body of the paper, namely: multiracial categorization; Hispanic categorization; aggregation (due to small cell sizes); and choosing the variable metric. We also faced decisions about the number of parameters; the number of model degrees of freedom matters for its relative comparison to other measures because we evaluate the performance of each race/ethnicity measure using penalized fit statistics (AIC and BIC). Unlike most researchers, we did not have to choose just one approach to making these decisions, as our project is concerned with evaluating the relative performance of many measures. When researchers do have to choose one approach, explicit justification of their preferred strategy is important given that different operationalization schemes can produce quite different results in terms of statistical fit, substantive conclusions, or both.

Appendix Table A1. Sample Sizes by Self-Identification Category

Panel A 1979 Self-Identification			Panel B 1979 Self-Identification Responses in Corresponding 2002 Categories			Panel C 2002 Self-Identification		
	Women	Men		Women	Men		Women	Men
Chicano	34	43	Hispanic, Latino, or Spanish Origin	1,016	973	Hispanic, Latino, or Spanish Origin	729	685
Cuban or Cubano	62	77						
Mexican or Mexicano	218	215						
Mexican-American	393	368						
Other Latino, Hispano, or Latin-American Descent	69	75						
Other Spanish Descent	121	68						
Puerto Rican, Puertorriqueno, or Borincano	178	178						
English	1,275	1,226	White	3,209	3,203	White	2,322	2,209
French	541	441						
German	1,506	1,380						
Greek	27	27						
Irish	1,280	1,128						
Italian	361	363						
Polish	228	230						
Portuguese	65	66						
Russian	68	70						
Scottish	373	341						
Welsh	131	112						
Black, Afro-American, or Negro	1,521	1,561	Black or African American	1,521	1,561	Black or African American	1,200	1,133
Chinese	16	14	Asian	96	87	Asian	16	14
Filipino or Philipino	29	29						
Indian-Asian	28	22						
Japanese	17	15						
Korean	5	4						
Vietnamese	2	3						
Hawaiian or Pacific Islander	16	18	Native Hawaiian or Other Pacific Islander	16	18	Native Hawaiian or Other Pacific Islander	10	5
Indian-American or Native American	720	597	American Indian or Alaska Native	720	597	American Indian or Alaska Native	64	46
American	382	396	Some Other Race	1,164	1,173	Some Other Race	176	166
None	37	28						
Other	838	841						

Note: Columns do not sum to total survey sample size because respondents could select multiple categories. See user's guide for information on the artificially inflated number of respondents self-identifying as "Native American" in 1979 because they believed this category was intended to include anyone who was born in America. Category labels shown are those presented to respondents in the survey.

We developed approaches to span different possible responses to the key decision points, resulting in the ten unique approaches shown in Table 2, Panel A in the main text. We developed approaches that use different metrics (e.g, a mutually exclusive categorical variable vs. a series of non-mutually exclusive dichotomous indicators), different approaches to multiraciality (e.g., one "multiracial" category vs. counting each response with non-mutually exclusive indicators), and different approaches to Hispanic origin responses (e.g., including a Hispanic category as mutually exclusive from race categories vs. a dichotomous indicator for Hispanic origin alongside the race categories).

Although we tested many different specifications, even more approaches are possible: we quickly realized that these decision points are interlocking and that changing the "order of operations" in which we applied decision rules could result in exponential different possibilities. Table A2 depicts the specific decisions we confronted using the NLSY79 data, which intertwine and affect each other. For example, one could aggregate categories with small cell sizes either before or after creating an indicator for multiracial responses; if you first aggregate categories and then tally multiracial responses as only those that cross the aggregated categories, you will have fewer multiracial respondents than if you first count all multiple responses and only then aggregate small categories into the residual for remaining participants.

Appendix Table A2. Intertwined Decision Rules for Coding Multiple Responses

Multiracial Determination	Hispanic	Aggregation	Metric
Any two or more responses	Hispanic responses not considered in the "two or more responses" multiracial determination	Aggregate categories if under a sample size minimum	Mutually exclusive multiracial category
Two or more responses in multiple 2000 Census categories	Hispanic responses considered in the "two or more responses" multiracial determination only in 1979	Aggregate by theoretically relevant factor (e.g., geography)	Series of non-exclusive category indicators
	Hispanic responses considered in the "two or more responses" multiracial determination in 1979 and 2002	Any two or more responses within a geographic region (e.g., "multiple European origins")	Series of non-exclusive category indicators and a multi flag

Note: Columns are not shown in any particular order.

Given the number of potential combinations implied by the various category options shown in Table A1 and decisions points in Table A2, it is not feasible for us to empirically evaluate *every* possible approach, but we do test a wide range of approaches, each of which represents the consequences of changing one particular decision compared to another nearly identical approach. In the interests of contributing to transparent science, we describe in detail each approach we tested below. Further, we share Stata code for others to adapt for their own projects; each of these approaches corresponds to code in the replication file.

The ten approaches to operationalizing self-identification in 1979 are grouped into three "families" of approaches. Across all approaches, we established a minimum sample size of 25 women and 25 men. Categories with fewer than 25 men or women were deemed too small for analysis and are incorporated into a residual category, as described below. Also across all approaches, if a respondent was recorded as "none" but also had at least one response recorded among the other (six) indicator variables, we recoded their observation to drop the "none" identification in the interests of logical consistency (we suspect these cases were due to data entry errors).

The first two families of approaches have corresponding 2002 versions, allowing for comparison between the different years and formats of self-identification questions. The first family includes three approaches that approximate the 1978 screener variable. We generated these approaches to match the number of parameters and types of categories used by the screener measure, so that any differences in AIC or BIC are due exclusively to differences in how the questions were asked (e.g. a calculated variable vs. self-identification and 30 vs. three categories), rather than the number of terms in the model. Table A3 shows sample sizes for the original screener measure (Panel A) and both the 1979 and 2002 self-identification variables recoded to mimic the screener categories (Panels B and C, respectively).

Appendix Table A3. Sample Sizes by Screener 1978 Category

	Panel A: 1978 Screener		**Panel B: 1979 in Screener Categories**		**Panel C: 2002 in Screener Categories**	
	Women	Men	Women	Men	Women	Men
Hispanic	1,002	1,000	1,016	973	729	685
Black	1,561	1,613	1,512	1,558	1,181	1,126
Non-Black, Non-Hispanic, and All Remaining Respondents	3,720	3,790	3,755	3,872	4,373	4,592

Note: Forty percent of respondents were not interviewed in 2002.

In Approach 1a, to approximate the three categories available in the screener variable, we aggregate the 1979 responses into a variable with the categories "Hispanic," "non-Hispanic Black," and "all remaining respondents." The Hispanic category includes respondents who also identified multiple origins, including Black or African American ancestry. The non-Hispanic Black category includes those who selected multiple origins, except Hispanic origins. The residual category includes those who did not report Black ancestry or an Hispanic origin, including those who reported multiple ancestries (except Black and/or Hispanic origins), and missing data. We follow a similar process to aggregate the 2002 responses.

The other two approaches in this first family highlight dichotomous contrasts with the two main screener categories. In Approach 1b, we aggregated the groups into a dichotomous variable of "Hispanic alone" and "all other respondents," the latter of which includes multi-origin responses and missing data. In Approach 1c, we similarly aggregated the variable into "Black alone" and "all other respondents." For equal parameter comparisons, we recoded versions of the 1978 screener variable (dichotomizing to either Black or Hispanic) to match Approaches 1b and 1c as well.

The second family of approaches approximates the federal government race/ethnicity categories used in the 2000 Census. The 2002 self-identification question uses this format, so by applying this format to the 1979 self-identification data, we can compare the 1979 and 2002 measures holding the number of parameters and analogous responses constant. The four approaches within this family differ in how multiracial and Hispanic identities are incorporated. In generating these approaches, we faced several intertwined decision points that quickly multiplied the number of potential operationalizations, including how to include Hispanic and multiracial responses (and whether to count Hispanic origins for the purposes of defining a

multiracial response) and when to aggregate categories with small cell sizes. We strove to make consistent decisions across measures, but this was not always appropriate, as we describe below.

Approach 2a is a mutually exclusive categorical variable that aggregates the 30 categories from the 1979 self-identification measure into six categories corresponding to the 2002 self-identification measure. This approach treats aggregated Hispanic origins as equivalent to a race category by including it in the same variable. This approach also includes a separate mutually exclusive category for people who gave two or more responses. The categories for this variable include: White alone; Black alone; Hispanic alone or in combination with other responses; Multiracial; All remaining responses, including "other" alone or small sample size in 2002; and Missing data. The multiracial category is defined as respondents who selected two or more responses from the 2002 categories, aside from Hispanic (e.g., reporting Black and an Asian identity in 1979 is treated as multiracial; reporting Black and a Hispanic origin is not). Reporting a Hispanic origin and another race category or categories is not treated as multiorigin or multiracial because the 2002 question formatting asked about Hispanic origin in a separate question from the race categories, consistent with the 2000 Census standards. Further, in the interests of keeping the 1979 and 2002 categories as consistent as possible, we checked the cell sizes by gender for both the 1979 and 2002 responses. In the 2002 self-identification measure, several categories fell below the cell size minimum, including the categories of Asian alone, Native Hawaiian or Pacific Islander alone, and American Indian or Alaska Native alone. We tested two versions of the 1979 self-identification approach 2a: one version that matched only those categories that met the minimum in 2002, and one version that used all categories that met the minimum in 1979. We tested both versions by gender for each outcome variable. We found that the former version generated better BIC values for both genders on all outcomes, and better

AIC values in five gender-by-outcome tests (see Table A4). We therefore aggregate the 30 categories in 1979 to the 2002 categories as described at the start of this section.

Appendix Table A4. Model Fit Statistics for Self-Identification 1979 Approach 2a Alternative Specifications

	Women				Men			
	AIC		BIC		AIC		BIC	
	Match 2002 categories that surpass sample size criteria	All categories above sample size criteria in 1979	Match 2002 categories that surpass sample size criteria	All categories above sample size criteria in 1979	Match 2002 categories that surpass sample size criteria	All categories above sample size criteria in 1979	Match 2002 categories that surpass sample size criteria	All categories above sample size criteria in 1979
Wages	5,794	5,787	5,831	5,837	5,735	5,731	5,772	5,781
Salary	8,410	8,410	8,447	8,460	8,437	8,437	8,474	8,487
Unemployment	44,068	44,071	44,109	44,125	46,593	46,596	46,634	46,650
Depression	24,832	24,834	24,871	24,885	22,737	22,739	22,775	22,790
School Discipline	5,609	5,608	5,649	5,662	7,585	7,588	7,625	7,641

Approach 2b modifies Approach 2a by removing the Hispanic category from the categorical race variable and including a separate dichotomous indicator. This allows respondents who report both a Hispanic origin and an additional race category to be represented in both variables, rather than coding them in only the Hispanic category and effectively erasing their race responses as in Approach 2a.

Approach 2c further modifies Approach 2b by converting all responses into a series of non-exclusive dichotomous indicators and eliminates the standalone multiracial category. Here, a 1979 respondent who reported a European origin, an Asian origin, and a Hispanic origin would be coded one on all three dichotomous indicators for "White," "Hispanic," and "All remaining responses" (because the Asian category in 2002 falls below the sample size minimum). We also relaxed the minimum cell size criteria we used elsewhere in order to include American Indian and Alaska Native among the responses to aid in comparisons across measures (see, e.g., Figure 1 in the main text). The sample sizes shown in Table A1 for 2002 reflect American Indian or Alaska Native responses alone or in combination; when considering only monoracial responses, as is necessary in the mutually exclusive categorical approaches, there are 21 men and 38 women

who self-identified as American Indian or Alaska Native alone in 2002. Thus, Approaches 2a and 2b do not include a separate American Indian or Alaska Native category but Approach 2c does.

Approach 2d adapts Approach 2c by adding an additional dichotomous multiracial indicator. This means that multiple responses are represented in two ways: each response is counted in the dichotomous variables for each category they select, and there is an additional flag for multiple responses, therefore measuring whether accounting for a shared "multiracial" experience adds explanatory power to the model, in addition to the experience of identifying with each particular race or origin category. Here, "multiracial" for the 2002 specification or "multiorigin" for 1979 is defined as selecting responses that cross more than one 2000 Census race category, aside from Hispanic origin (e.g., reporting Italian and Black would be coded as "multiracial," but Italian and Cuban would not be coded as "multiracial" in this approach).

The third family of approaches consists of the most disaggregated versions of the 1979 self-identification responses. A corresponding version using 2002 is not possible due to the 2002 question format; the most disaggregated version possible for 2002 corresponds to the second family of approaches. In this third family, we aim to use as many categories as possible given sample size constraints, allowing this high level of detail to capture nuance in respondents' reported ancestries or origins beyond what is possible with the 2002 self-identification measure.

Approach 3a is a mutually exclusive categorical variable that includes the 15 categories that surpass the cell size criteria as "ancestry/origin category alone;"[1] all responses below the cell size minimum are aggregated into a 16th category for "All remaining responses," those selecting multiple ancestries or origins are coded as the 17th category of "multiorigin," and there is one

[1] These groups are: Black; English; French; German; Native American; Irish; Italian; Cuban; Mexican; Mexican-American; Puerto Rican; Other Hispanic; Polish; Portuguese; and American.

final category for missing data. Because this variable uses a mutually exclusive categories, it was technically necessary to define "multiple categories" as those selecting two or more responses of any kind, including two or more responses that might not intuitively "seem" multiracial, such as two European origins. However, while we might not feel that a respondent who selected English and Irish is multiracial, for example, they cannot be coded in both categories simultaneously in a mutually exclusive categorical variable. One alternative approach would be to create multiple multiorigin categories, such as "multiple European categories," "multiple Asian categories," and "multiple Hispanic categories," and "multiple origins spanning 2000 Census categories," which we considered. Ultimately, we did not pursue this alternative for several reasons. First, it would add several additional parameters to the approach, which is already at a disadvantage in the penalized model fit statistics because it is one of the least parsimonious approaches amongst all the measures by a wide margin. Second, it departs substantially from the question format as originally presented to respondents. Third, it would be less consistent with the decision rules we apply to the 2002 self-identification measure, which cannot produce different "multiorigin" categories. Finally, the alternative approach would implicitly suggest that some groups "should" be aggregated together, when the degree of intergroup heterogeneity might be high; indeed some categories that are not aggregated together have more similar outcomes on our dependent variables than the categories they would be aggregated with based on typical assumptions. Of course, many of these considerations inspire their own empirical questions, and future research should further explore these issues.

Approach 3b adapts Approach 3a by changing the metric from a mutually exclusive categorical variable to a series of dichotomous indicators, and by accounting for multiple origins differently. Each dichotomous variable indicates whether respondents selected the category in

any of their up to six ancestries or origins. A final dichotomous variable aggregates "all remaining respondents," including those who selected "Other" alone or selected a category that falls below the sample size minimum. When generating this approach, we tested the relative performances of using the 15 categories that appear in Approach 3a (those with sample sizes large enough when excluding respondents who selected multiple categories) compared to using the 21 categories that surpass the sample size criteria regardless of other selections (including those who selected multiple categories). We found that for both women and men, the version matching the categories in Approach 3a fit better on every outcome for both women and men on BIC (see Table A5) and fit the data substantially better on AIC (i.e., by three to eleven points) for three of the five outcomes. Because of the overwhelmingly better fit of this approach, and because it allows comparison to Approach 3a, we used this specification. When compared to Approach 3a, this approach accounts for multiraciality using the same distinction as between Approaches 2a/2b and Approaches 2c/2d: rather than creating one "multiorigin" category that is mutually exclusive from all other categories, the series of dichotomous variables enables each respondent's data to be represented in each category they selected.

Appendix Table A5. Model Fit Statistics for Self-Identification 1979 Approach 3b Alternative Specifications

	Women				Men			
	AIC		BIC		AIC		BIC	
	Matching Approach 3a	All categories above sample size criteria	Matching Approach 3a	All categories above sample size criteria	Matching Approach 3a	All categories above sample size criteria	Matching Approach 3a	All categories above sample size criteria
Wages	5,743	5,738	5,848	5,880	5,688	5,677	5,794	5,820
Salary	8,394	8,396	8,499	8,538	8,397	8,381	8,503	8,524
Unemployment	44,089	44,092	44,203	44,247	46,603	46,611	46,717	46,767
Depression	24,808	24,810	24,916	24,956	22,741	22,749	22,848	22,894
School Discipline	5,569	5,566	5,683	5,714	7,533	7,529	7,647	7,683

That said, we faced another difficult decision point with Approach 3b: whether or not to code respondents who select a Hispanic origin and one other non-Hispanic origin as multiorigin. We did not treat Hispanic origins as a separate "kind" of category from the other

ancestries/origins, because the question format as originally presented to respondents offered Hispanic origin options in the same list, rather than in a separate question as in the 2002 Census-style format. Therefore, in this approach, selecting both a Hispanic origin and one other origin is classified as "multiorigin." However, our approaches for operationalizing self-identification measures collected in 2002 do treat Hispanic selections as a separate "origin" variable that is different from the "race" variables, following the format of those questions. Given ongoing debates about how to collect data on Hispanic origins and identities, it is not surprising that our approaches vary based on how the data were originally collected, which changed between 1979 and 2002. Future research should evaluate how other approaches to classifying Hispanic categories in the 1979 data might change the empirical performance of these measures.

Finally, Approach 3c is identical to Approach 3b with the addition of a dichotomous flag for those selecting more than one category, mirroring the difference between Approach 2c and Approach 2d. Note that here, the "multiple categories" variable includes all instances of multiple categories, not just those spanning 2000 Census categories. We did test both approaches, with the "multiple categories" flag being limited to only those respondents who reported, for example, a European origin and an Asian origin. We found that the version with a dichotomous indicator for "multiple origins, regardless of whether they span 2000 Census categories" performs better on AIC and BIC for both women and men on all outcomes aside from women's unemployment and men's depression (see Table A6). Further, the AIC and BIC results using the 2000 Census version of multiraciality are not lower than the best-performing overall measures. Thus, we include the version that identifies all multiple-response observations, and either way, the final results do not change.

Appendix Table A6. Model Fit Statistics for Self-Identification 1979 Approach 3c Alternative Specifications

	Women				Men			
	AIC		BIC		AIC		BIC	
	Any multiple origins	Multiple 2000 Census categories only	Any multiple origins	Multiple 2000 Census categories only	Any multiple origins	Multiple 2000 Census categories only	Any multiple origins	Multiple 2000 Census categories only
Wages	5,732	5,734	5,850	5,852	5,672	5,673	5,790	5,791
Salary	8,385	8,387	8,502	8,504	8,388	8,391	8,506	8,509
Unemployment	44,093	44,092	44,221	44,220	46,604	46,606	46,732	46,735
Depression	24,812	24,812	24,933	24,933	22,741	22,739	22,861	22,860
School Discipline	5,570	5,571	5,697	5,698	7,529	7,532	7,657	7,660

Appendix A Section 2. Interviewer-Observed Race

In the 17 survey waves of 1979-1986 and 1988-1998, interviewers recorded their perception of the respondents' race and ethnicity, in three categories: Black, White, and Other. Interviewers could select only one response per participant per year. No classifications were recorded for the 1987 wave, when all interviews were conducted by telephone (though classifications were recorded for some interviews that were completed by phone in other years). In our analyses, we test three approaches to operationalizing these data for each category, resulting in nine total specifications. Each approach aggregates the previous approach into fewer categories, as described below.

The first observed race approach is a continuous variable for the percent of survey waves in which interviewers recorded their classification of the respondent as Black, Other, or White, divided by ten to aid interpretation of regression coefficients (i.e. a value of 1 indicates 10 percent of survey waves). We considered using a simple count variable, either in addition or instead of this percentage approach, but opted not to for reasons we describe below.

The second observed race approach is a categorical variable that aggregates the percent of survey waves approach into deciles. For example, a value of seven indicates participants were observed to be Black, Other, or White in 70-79 percent of survey waves. This specification

allows for non-linearities in the relationship between how respondents were perceived racially by interviewers and the outcome of interest. We also considered aggregating the percentages into smaller buckets, such as quintiles, but wanted to avoid interpretations that resembled biological attributions of blood quantum (e.g., one-quarter Black, one-half White, etc.).

The third observed race approach consists of a categorical variable with three categories corresponding to "always, ever, or never observed by interviewers as [Black/Other/White]." This approach adapts the second approach, above, but aggregates the middle categories into one. For example, respondents who interviewers did not classify as Black in any survey wave are "never observed as Black," respondents classified as Black in at least one survey wave are "ever observed as Black," and respondents who interviewers classified as Black in all survey waves are "always observed as Black" in this scheme. Theoretically, this approach is useful for determining whether a "one-drop rule" for classification is in effect; if so, then estimates for the "ever" category will be indistinguishable from estimates for the "always" category. Alternatively, outcomes for people with ambiguous or inconsistent racial classification could be better, worse, or in between their consistently classified counterparts. See Guluma and Saperstein (2022) for a similar analysis using interviewer classifications in Add Health. We show the cross-classification of this observed race approach with 1979 Approach 3b in Table A7, which demonstrates considerable variation by self-identification in how respondents were perceived by interviewers and motivates our dual-measure approaches shown in the main text.

Appendix Table A7. Percent of Respondents Interviewers Always, Ever, and Never Observed to be Black, Other, and White, by Self-Identification 1979 Approach 3b

Panel A. Women

	Black			Other			White		
	Always	Ever	Never	Always	Ever	Never	Always	Ever	Never
Black	86.2	13.6	0.2	0.0	3.0	97.0	0.1	12.0	87.9
Portuguese	3.1	18.5	78.5	0.0	33.8	66.2	60.0	32.3	7.7
Puerto Rican	1.1	23.0	75.8	3.4	88.2	8.4	5.6	88.8	5.6
Native American	6.4	3.3	90.3	0.1	10.0	89.9	81.3	12.1	6.7
Mexican-American	0.5	2.0	97.5	0.3	89.3	10.4	9.9	89.3	0.8
Italian	0.6	1.4	98.1	0.0	6.9	93.1	92.2	6.9	0.8
Polish	0.0	2.6	97.4	0.0	1.3	98.7	96.1	3.9	0.0
French	1.3	2.6	96.1	0.0	4.8	95.2	92.2	6.3	1.5
English	2.4	2.0	95.6	0.0	3.0	97.0	92.7	4.9	2.4
Irish	0.3	1.9	97.8	0.0	2.5	97.5	95.5	4.2	0.3
American	1.3	1.0	97.6	0.0	5.5	94.5	92.4	6.3	1.3
Cuban	0.0	6.5	93.5	0.0	69.4	30.6	29.0	69.4	1.6
German	1.0	3.3	95.7	0.4	13.6	86.0	83.3	15.0	1.7
Other Hispanic	0.3	1.5	98.2	0.0	2.7	97.3	95.7	4.0	0.3
Mexican	1.4	15.9	82.6	1.4	82.6	15.9	14.5	78.3	7.2
All Remaining Responses	0.0	2.3	97.7	0.0	84.9	15.1	14.7	85.3	0.0

Panel B. Men

	Black			Other			White		
	Always	Ever	Never	Always	Ever	Never	Always	Ever	Never
Black	88.1	11.7	0.1	0.0	2.1	97.9	0.1	10.6	89.4
Portuguese	3.0	19.7	77.3	0.0	37.9	62.1	56.1	31.8	12.1
Puerto Rican	0.0	23.6	76.4	2.2	93.3	4.5	4.5	92.1	3.4
Native American	5.5	4.2	90.3	0.5	10.9	88.6	79.7	13.9	6.4
Mexican-American	0.3	0.5	99.2	0.5	90.5	9.0	8.7	90.5	0.8
Italian	0.8	2.5	96.7	0.0	6.1	93.9	90.9	8.3	0.8
Polish	0.0	2.6	97.4	0.0	2.2	97.8	95.7	4.3	0.0
French	1.4	3.2	95.5	0.2	5.2	94.6	90.7	7.7	1.6
English	2.9	2.3	94.9	0.1	3.1	96.8	91.8	5.1	3.0
Irish	0.2	2.5	97.3	0.1	3.4	96.5	94.3	5.2	0.4
American	1.5	2.3	96.2	0.0	5.8	94.2	90.7	7.8	1.5
Cuban	0.0	5.2	94.8	0.0	77.9	22.1	20.8	79.2	0.0
German	0.2	1.7	98.1	0.1	2.5	97.5	95.7	3.9	0.4
Other Hispanic	0.0	16.0	84.0	0.0	80.0	20.0	18.7	81.3	0.0
Mexican	0.9	0.9	98.1	0.5	88.8	10.7	9.8	88.8	1.4
All Remaining Responses	1.2	2.3	96.5	0.7	14.9	84.5	82.4	15.3	2.3

Note: Values are row percentages within Black, Other, and White categories. Rows are arranged in the order shown in Figure 2 to facilitate ease of comparison.

A key decision point facing analysts using the interviewer-observed race data is whether to use all 17 waves of data, as we did, or to use just one wave, such as 1998, the last year for which observed race is available (and the closest in time to when our outcomes are measured). We decided to use all 17 years because the ability to detect fluctuations in how interviewers assigned race to each respondent is a unique aspect of these measures that could matter for inequality outcomes (see, e.g., Saperstein and Penner 2012). However, using just one wave of data could be justifiable if it is closest to the year of data collection to the measured outcome(s);

a single year of classification data also allows models to include interviewer fixed effects, which could improve model performance.

We tested the performance of one approach to including the 1998 interviewer-observed race data (one dichotomous variable each for "observed [Black/Other/White] in 1998", with each model containing one of the three), both with and without interviewer fixed effects, across all our outcome variables and split by gender, resulting in 60 tests (two fixed effects options * two genders * three race measures * five outcome variables). Comparing fit statistics for the versions with and without fixed effects, we find that the versions without fixed effects perform better on AIC in 21 out of 30 comparisons, and perform better on BIC in all 30 comparisons (Table A8). This suggests that accounting for variation between interviewers (i.e., in how they judge racial classification) generally does not sufficiently improve model fit for our outcomes of interest to warrant using this approach. Future research could also evaluate the performance of a categorical version of these data with three categories (Black, Other, and White) for the 1998 interviewer observations, rather than dichotomous variables for each individual category.

Appendix Table A8. Model Fit Statistics for Alternative Observed Race Measure Specifications: 1998 With and Without Interviewer Fixed Effects

Panel A. Observed Black

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects
Wages	5,402	5,455	6,835	5,455	5,157	5,212	6,554	5,224
Salary	8,042	7,965	9,475	7,965	7,573	7,611	8,966	7,623
Unemployment	30,741	30,545	32,239	30,545	30,406	30,210	31,867	30,223
Depression	23,545	23,358	25,031	23,358	20,655	20,485	22,086	20,497
School Discipline	3,873	3,710	5,105	3,710	4,832	4,673	6,104	4,685

Panel B. Observed Other

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects
Wages	5,442	5,524	6,875	5,524	5,328	5,435	6,725	5,448
Salary	8,055	7,982	9,487	7,982	7,738	7,844	9,131	7,856
Unemployment	30,755	30,580	32,254	30,580	30,434	30,251	31,894	30,264
Depression	23,557	23,383	25,042	23,383	20,651	20,499	22,081	20,512
School Discipline	3,949	3,834	5,181	3,834	4,905	4,814	6,177	4,826

Panel C. Observed White

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects	1998 With Interviewer Fixed Effects	1998 Without Interviewer Fixed Effects
Wages	5,401	5,471	6,835	5,471	5,135	5,191	6,532	5,203
Salary	8,040	7,969	9,472	7,969	7,553	7,597	8,946	7,610
Unemployment	30,747	30,552	32,246	30,552	30,408	30,210	31,868	30,222
Depression	23,541	23,356	25,026	23,356	20,644	20,475	22,075	20,488
School Discipline	3,848	3,681	5,081	3,681	4,845	4,687	6,117	4,700

We also compared fit statistics from the 1998 alternative without interviewer fixed effects to the "percent of waves observed by interviewer as [Black/Other/White]" approach in the main analyses, which uses the same number of parameters (one). The approach we use in the main analyses performs better in 16 out of 60 comparisons on AIC and 17 instances on BIC (across gender) and the two measures tie in four instances on AIC and two instances on BIC (Table A9). Many of these comparisons suggest the two options provide equivalent results (i.e., within one or two points of each other). However, when the "percent of waves observed [Black/Other/White]" approach does perform best, it is by a larger magnitude (up to 31 points vs. up to 7 points). These results support our decision to use multiple waves of observed race data in our main analyses.

Appendix Table A9. Model Fit Statistics for Alternative Observed Race Measure Specifications: 1998 Without Interviewer Fixed Effects

Panel A. Observed Black

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves
Wages	5,477	5,480	5,490	5,492	5,287	5,290	5,300	5,303
Salary	8,002	8,003	8,014	8,015	7,718	7,724	7,730	7,737
Unemployment	30,828	30,825	30,841	30,838	30,724	30,726	30,737	30,738
Depression	23,560	23,558	23,572	23,571	20,786	20,787	20,798	20,799
School Discipline	3,854	3,855	3,866	3,868	4,823	4,823	4,836	4,835

Panel B. Observed Other

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves
Wages	5,547	5,544	5,559	5,556	5,517	5,516	5,529	5,528
Salary	8,019	8,017	8,031	8,029	7,961	7,955	7,973	7,967
Unemployment	30,864	30,865	30,877	30,878	30,768	30,764	30,781	30,777
Depression	23,585	23,585	23,597	23,598	20,803	20,801	20,815	20,814
School Discipline	3,991	3,996	4,004	4,009	4,967	4,962	4,979	4,974

Panel C. Observed White

	Women				Men			
	AIC		BIC		AIC		BIC	
	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves	1998 Without Interviewer Fixed Effects	Continuous Percent of Waves
Wages	5,493	5,488	5,506	5,500	5,266	5,244	5,278	5,256
Salary	8,006	8,006	8,018	8,018	7,703	7,672	7,716	7,685
Unemployment	30,836	30,822	30,848	30,835	30,722	30,713	30,735	30,726
Depression	23,558	23,556	23,571	23,569	20,777	20,777	20,789	20,789
School Discipline	3,832	3,837	3,844	3,850	4,838	4,833	4,850	4,846

Finally, researchers might consider using a straightforward count of the number of survey waves interviewers classified participants as Black, Other, or White. However, this approach does not account for missing waves of data. For example, a respondent could have only five years of interviewer observations recorded and in all five the interviewers classified them as White. In terms of how many times they were classified as White, this respondent would be empirically indistinguishable from a respondent with 17 waves of data, in five of which interviewers classified them as White, and in 12 of which interviewers classified them as Black. Thus, we do not encourage researchers to use a simple count measure for these data.

Appendix B. Fit Statistics by Outcome and Gender

Appendix B Section 1. AIC and BIC comparisons

Appendix Table B1. Wage Analyses Fit Statistics, by Gender

	Self-Id 1979			Self-Id 2002		Observed race				Screener		
	Approach	*AIC*	*BIC*	*AIC*	*BIC*	*Group*	*Approach*	*AIC*	*BIC*	*Approach*	*AIC*	*BIC*
Women	1a	+65	+27	+48	+09	Black	Percentages	+55	+10	Default	+52	+14
	1b	+116	+71	+116	+71	Black	Categorical Deciles	+51	+62	Hispanic	+116	+71
	1c	+67	+22	+52	+08	Black	Always/Ever/Never	+38	**5,787**	Black	+56	+12
	2a	+64	+44	+50	+30	Other	Percentages	+112	+68			
	2b	+65	+46	+52	+32	Other	Categorical Deciles	+118	+130			
	2c	+54	+34	+52	+32	Other	Always/Ever/Never	+116	+78			
	2d	+56	+43	+54	+40	White	Percentages	+62	+17			
	3a	+25	+80	-	-	White	Categorical Deciles	+62	+73			
	3b	+01	+49	-	-	White	Always/Ever/Never	+42	+04			
	3c	**5,730**	+55	-	-							
Men	1a	+69	+33	+101	+65	Black	Percentages	+98	+56	Default	+56	+21
	1b	+344	+302	+346	+305	Black	Categorical Deciles	+97	+112	Hispanic	+345	+304
	1c	+119	+77	+132	+90	Black	Always/Ever/Never	+98	+62	Black	+108	+67
	2a	+60	+43	+64	+48	Other	Percentages	+339	+298			
	2b	+53	+37	+62	+46	Other	Categorical Deciles	+347	+361			
	2c	+39	+22	+43	+26	Other	Always/Ever/Never	+344	+309			
	2d	+33	+23	+44	+34	White	Percentages	+47	+6			
	3a	+41	+99	-	-	White	Categorical Deciles	+36	+51			
	3b	+2	+54	-	-	White	Always/Ever/Never	+36	**5,729**			
	3c	**5,675**	+58	-	-							

Note: Actual values shown for best fit statistics, bolded, by gender and fit statistic. All other fit statistics are shown as distance from the lowest fit statistic.

Appendix Table B2. Salary Analyses Fit Statistics, by Gender

	Self-Id 1979			Self-Id 2002		Observed race				Screener		
	Approach	*AIC*	*BIC*	*AIC*	*BIC*	*Group*	*Approach*	*AIC*	*BIC*	*Approach*	*AIC*	*BIC*
Women	1a	+23	+20	+15	+12	Black	Percentages	+15	+06	Default	+18	+16
	1b	+29	+20	+29	+20	Black	Categorical Deciles	+12	+59	Hispanic	+29	+20
	1c	+21	+12	+14	+05	Black	Always/Ever/Never	+02	**8,406**	Black	+17	+09
	2a	+25	+41	+15	+31	Other	Percentages	+27	+18			
	2b	+24	+40	+17	+33	Other	Categorical Deciles	+38	+85			
	2c	+25	+41	+19	+35	Other	Always/Ever/Never	+29	+27			
	2d	+26	+48	+18	+40	White	Percentages	+18	+09			
	3a	+07	+97	-	-	White	Categorical Deciles	+26	+72			
	3b	+02	+86	-	-	White	Always/Ever/Never	+08	+06			
	3c	**8,385**	+90	-	-							
Men	1a	+59	+45	+98	+83	Black	Percentages	+83	+63	Default	+42	+28
	1b	+325	+304	+329	+308	Black	Categorical Deciles	+83	+112	Hispanic	+324	+304
	1c	+107	+87	+122	+101	Black	Always/Ever/Never	+76	+62	Black	+100	+79
	2a	+48	+52	+41	+45	Other	Percentages	+314	+293			
	2b	+47	+51	+40	+44	Other	Categorical Deciles	+326	+361			
	2c	+30	+34	+15	+19	Other	Always/Ever/Never	+324	+309			
	2d	+28	+39	+17	+28	White	Percentages	+22	+2			
	3a	+56	+135	-	-	White	Categorical Deciles	+21	+56			
	3b	+02	+74	-	-	White	Always/Ever/Never	+15	**8,422**			
	3c	**8,389**	+79	-	-							

Note: Actual values shown for best fit statistics, bolded, by gender and fit statistic. All other fit statistics are shown as distance from the lowest fit statistic.

Appendix Table B3. Unemployment Analyses Fit Statistics, by Gender

	Self-Id 1979			Self-Id 2002		Observed race				Screener		
	Approach	*AIC*	*BIC*	*AIC*	*BIC*	*Group*	*Approach*	*AIC*	*BIC*	*Approach*	*AIC*	*BIC*
Women	1a	+51	+31	+45	+24	Black	Percentages	+77	+50	Default	+52	+31
	1b	+167	+139	+162	+135	Black	Categorical Deciles	+93	+126	Hispanic	+165	+138
	1c	+69	+42	+64	+37	Black	Always/Ever/Never	+84	+63	Black	+73	+45
	2a	+49	+49	+05	+05	Other	Percentages	+169	+142			
	2b	+50	+50	+01	**44,060**	Other	Categorical Deciles	+180	+213			
	2c	+57	+57	**44,019**	**44,060**	Other	Always/Ever/Never	+165	+144			
	2d	+59	+66	**44,019**	+07	White	Percentages	+70	+43			
	3a	+62	+143	-	-	White	Categorical Deciles	+75	+109			
	3b	+70	+144	-	-	White	Always/Ever/Never	+63	+42			
	3c	+72	+153	-	-							
Men	1a	+84	+63	+38	+17	Black	Percentages	+110	+82	Default	+71	+50
	1b	+215	+188	+215	+188	Black	Categorical Deciles	+116	+150	Hispanic	+214	+186
	1c	+110	+82	+58	+31	Black	Always/Ever/Never	+115	+95	Black	+102	+75
	2a	+88	+88	+08	+08	Other	Percentages	+212	+184			
	2b	+89	+88	+10	+10	Other	Categorical Deciles	+216	+250			
	2c	+90	+89	**46,505**	**46,546**	Other	Always/Ever/Never	+216	+196			
	2d	+91	+97	+02	+09	White	Percentages	+82	+55			
	3a	+92	+173	-	-	White	Categorical Deciles	+85	+118			
	3b	+98	+172	-	-	White	Always/Ever/Never	+98	+77			
	3c	+97	+178	-	-							

Note: Actual values shown for best fit statistics, bolded, by gender and fit statistic. All other fit statistics are shown as distance from the lowest fit statistic.

Appendix Table B4. Depression Symptoms Analyses Fit Statistics, by Gender

	Self-Id 1979			Self-Id 2002		Observed race				Screener		
	Approach	*AIC*	*BIC*	*AIC*	*BIC*	*Group*	*Approach*	*AIC*	*BIC*	*Approach*	*AIC*	*BIC*
Women	1a	+31	+22	+15	+06	Black	Percentages	+22	+06	Default	+27	+18
	1b	+48	+32	+48	+33	Black	Categorical Deciles	+25	+67	Hispanic	+47	+32
	1c	+29	+14	+15	**24,836**	Black	Always/Ever/Never	+18	+09	Black	+25	+10
	2a	+24	+35	+16	+26	Other	Percentages	+48	+33			
	2b	+22	+32	+22	+32	Other	Categorical Deciles	+56	+98			
	2c	+10	+21	+17	+27	Other	Always/Ever/Never	+49	+40			
	2d	+12	+28	+19	+35	White	Percentages	+18	+03			
	3a	+38	+124	-	-	White	Categorical Deciles	+31	+73			
	3b	**24,808**	+80	-	-	White	Always/Ever/Never	+18	+09			
	3c	+02	+88	-	-							
Men	1a	+32	+19	+30	+17	Black	Percentages	+31	+12	Default	+28	+15
	1b	+47	+28	+49	+29	Black	Categorical Deciles	+43	+80	Hispanic	+49	+29
	1c	+37	+18	+32	+13	Black	Always/Ever/Never	+32	+19	Black	+32	+12
	2a	+17	+23	+15	+21	Other	Percentages	+41	+21			
	2b	+34	+40	+10	+16	Other	Categorical Deciles	+45	+82			
	2c	+26	+32	**22,720**	+06	Other	Always/Ever/Never	+48	+35			
	2d	+24	+37	+02	+14	White	Percentages	+20	**22,752**			
	3a	+16	+97	-	-	White	Categorical Deciles	+20	+57			
	3b	+19	+95	-	-	White	Always/Ever/Never	+23	+10			
	3c	+19	+101	-	-							

Note: Actual values shown for best fit statistics, bolded, by gender and fit statistic. All other fit statistics are shown as distance from the lowest fit statistic.

Appendix Table B5. School Discipline Analyses Fit Statistics, by Gender

	Self-Id 1979		Self-Id 2002		Observed race				Screener		
	Approach	*AIC*	*BIC*	*AIC*	*BIC*	*Group*	*Approach*	*AIC*	*BIC*	*Approach*	*AIC*	*BIC*
Women	1a	+28	+14	+43	+29	Black	Percentages	+21	**5,602**	Default	+25	+11
	1b	+159	+138	+161	+140	Black	Categorical Deciles	+19	+45	Hispanic	+161	+140
	1c	+27	+06	+41	+20	Black	Always/Ever/Never	+19	+05	Black	+26	+05
	2a	+41	+47	+28	+34	Other	Percentages	+163	+143			
	2b	+42	+48	+31	+37	Other	Categorical Deciles	+168	+208			
	2c	+17	+23	+34	+40	Other	Always/Ever/Never	+163	+149			
	2d	+19	+32	+32	+45	White	Percentages	+22	+02			
	3a	+28	+115	-	-	White	Categorical Deciles	+29	+69			
	3b	**5,568**	+80	-	-	White	Always/Ever/Never	+21	+07			
	3c	**5,568**	+87	-	-							
Men	1a	+45	+10	+88	+54	Black	Percentages	+42	**7,583**	Default	+50	+15
	1b	+206	+164	+207	+166	Black	Categorical Deciles	+44	+49	Hispanic	+206	+165
	1c	+43	+01	+87	+46	Black	Always/Ever/Never	+51	+16	Black	+48	+06
	2a	+57	+42	+69	+54	Other	Percentages	+203	+161			
	2b	+57	+43	+76	+61	Other	Categorical Deciles	+207	+219			
	2c	+33	+18	+74	+60	Other	Always/Ever/Never	+194	+153			
	2d	+35	+27	+75	+67	White	Percentages	+70	+29			
	3a	+37	+103	-	-	White	Categorical Deciles	+58	+77			
	3b	+04	+63	-	-	White	Always/Ever/Never	+66	+31			
	3c	**7,528**	+66	-	-							

Note: Actual values shown for best fit statistics, bolded, by gender and fit statistic. All other fit statistics are shown as distance from the lowest fit statistic.

Appendix B Section 2. Cross-Validation Benchmarks

Appendix Table B6. Wage Analyses Fit Statistics, Cross-Validation, by Gender

	Self-Id 1979		Self-Id 2002	Observed race			Screener	
	Approach	*Improvement*	*Improvement*	*Group*	*Approach*	*Improvement*	*Approach*	*Improvement*
Women	1a	0.680%	0.925%	Black	Percentages	0.827%	Default	0.873%
	1b	-0.064%	-0.029%	Black	Categorical Deciles	0.649%	Hispanic	-0.027%
	1c	0.666%	0.876%	Black	Always/Ever/Never	1.064%	Black	0.752%
	2a	0.541%	0.736%	Other	Percentages	0.033%		
	2b	0.665%	0.863%	Other	Categorical Deciles	0.044%		
	2c	0.818%	0.923%	Other	Always/Ever/Never	0.056%		
	2d	0.794%	0.798%	White	Percentages	0.788%		
	3a	1.423%	-	White	Categorical Deciles	0.615%		
	3b	1.393%	-	White	Always/Ever/Never	0.997%		
	3c	**1.528%**	-					
Men	1a	3.735%	3.244%	Black	Percentages	3.257%	Default	3.734%
	1b	0.046%	-0.017%	Black	Categorical Deciles	3.156%	Hispanic	-0.014%
	1c	3.020%	2.817%	Black	Always/Ever/Never	3.219%	Black	3.214%
	2a	3.628%	3.624%	Other	Percentages	0.060%		
	2b	3.844%	3.708%	Other	Categorical Deciles	0.088%		
	2c	4.012%	3.979%	Other	Always/Ever/Never	-0.057%		
	2d	4.099%	3.911%	White	Percentages	3.935%		
	3a	3.968%	-	White	Categorical Deciles	3.999%		
	3b	4.472%	-	White	Always/Ever/Never	4.122%		
	3c	**4.480%**	-					

Note: Values are the percent improvement of the RMSE from five-fold cross-validation compared to the constant-only benchmark. The largest improvement within gender is bolded.

Appendix Table B7. Salary Analyses Fit Statistics, Cross-Validation, by Gender

	Self-Id 1979	Self-Id 2002	Observed race			Screener	
	Approach	*Improvement*	*Improvement*	*Group*	*Approach*	*Improvement*	*Approach*	*Improvement*
Women	1a	0.123%	0.162%	Black	Percentages	0.173%	Default	0.120%
	1b	-0.030%	-0.041%	Black	Categorical Deciles	0.168%	Hispanic	-0.078%
	1c	0.072%	0.118%	Black	Always/Ever/Never	0.242%	Black	0.131%
	2a	0.007%	0.080%	Other	Percentages	-0.009%		
	2b	-0.043%	0.138%	Other	Categorical Deciles	-0.163%		
	2c	-0.076%	0.151%	Other	Always/Ever/Never	-0.109%		
	2d	0.057%	0.015%	White	Percentages	0.034%		
	3a	0.152%	-	White	Categorical Deciles	-0.122%		
	3b	0.143%	-	White	Always/Ever/Never	0.160%		
	3c	**0.415%**	-					
Men	1a	3.629%	3.088%	Black	Percentages	3.291%	Default	3.905%
	1b	0.028%	0.016%	Black	Categorical Deciles	3.331%	Hispanic	0.020%
	1c	2.955%	2.755%	Black	Always/Ever/Never	3.354%	Black	2.963%
	2a	3.767%	3.901%	Other	Percentages	0.311%		
	2b	3.853%	3.933%	Other	Categorical Deciles	-0.005%		
	2c	4.028%	4.311%	Other	Always/Ever/Never	-0.011%		
	2d	4.034%	4.152%	White	Percentages	4.121%		
	3a	3.682%	-	White	Categorical Deciles	4.190%		
	3b	4.310%	-	White	Always/Ever/Never	4.129%		
	3c	**4.428%**	-					

Note: Values are the percent improvement of the RMSE from five-fold cross-validation compared to the constant-only benchmark. The largest improvement within gender is bolded.

Appendix Table B8. Unemployment Analyses Fit Statistics, Cross-Validation, by Gender

	Self-Id 1979	Self-Id 2002	Observed race			Screener	
	Approach	*Improvement*	*Improvement*	*Group*	*Approach*	*Improvement*	*Approach*	*Improvement*
Women	1a	0.922%	1.071%	Black	Percentages	0.966%	Default	0.683%
	1b	0.028%	-0.020%	Black	Categorical Deciles	0.452%	Hispanic	0.222%
	1c	1.000%	0.598%	Black	Always/Ever/Never	0.481%	Black	0.882%
	2a	0.879%	1.251%	Other	Percentages	0.013%		
	2b	0.885%	1.214%	Other	Categorical Deciles	-0.237%		
	2c	0.785%	1.146%	Other	Always/Ever/Never	0.229%		
	2d	**1.618%**	1.277%	White	Percentages	0.679%		
	3a	0.876%	-	White	Categorical Deciles	0.636%		
	3b	0.732%	-	White	Always/Ever/Never	0.923%		
	3c	1.035%	-					
Men	1a	0.839%	1.215%	Black	Percentages	1.195%	Default	1.014%
	1b	-0.008%	-0.201%	Black	Categorical Deciles	0.768%	Hispanic	0.092%
	1c	0.644%	1.141%	Black	Always/Ever/Never	0.730%	Black	0.726%
	2a	1.057%	1.494%	Other	Percentages	0.498%		
	2b	0.780%	1.404%	Other	Categorical Deciles	0.249%		
	2c	1.055%	1.548%	Other	Always/Ever/Never	0.124%		
	2d	0.768%	**1.560%**	White	Percentages	1.183%		
	3a	1.174%	-	White	Categorical Deciles	1.055%		
	3b	0.852%	-	White	Always/Ever/Never	0.821%		
	3c	0.762%	-					

Note: Values are the percent improvement of the RMSE from five-fold cross-validation compared to the constant-only benchmark. The largest improvement within gender is bolded.

Appendix Table B9. Depression Symptoms Analyses Fit Statistics, Cross-Validation, by Gender

		Self-Id 1979	Self-Id 2002		Observed race		Screener	
	Approach	*Improvement*	*Improvement*	*Group*	*Approach*	*Improvement*	*Approach*	*Improvement*
Women	1a	0.057%	0.227%	Black	Percentages	0.196%	Default	0.141%
	1b	-0.105%	-0.075%	Black	Categorical Deciles	0.151%	Hispanic	-0.119%
	1c	0.166%	0.279%	Black	Always/Ever/Never	0.313%	Black	0.181%
	2a	0.248%	0.356%	Other	Percentages	-0.051%		
	2b	0.186%	0.222%	Other	Categorical Deciles	-0.131%		
	2c	0.312%	0.205%	Other	Always/Ever/Never	-0.162%		
	2d	0.312%	0.218%	White	Percentages	0.275%		
	3a	0.274%	-	White	Categorical Deciles	0.144%		
	3b	**0.533%**	-	White	Always/Ever/Never	0.366%		
	3c	0.382%	-					
Men	1a	0.156%	0.276%	Black	Percentages	0.160%	Default	0.274%
	1b	0.037%	0.087%	Black	Categorical Deciles	0.168%	Hispanic	-0.154%
	1c	0.186%	0.148%	Black	Always/Ever/Never	0.282%	Black	0.160%
	2a	0.345%	0.341%	Other	Percentages	0.055%		
	2b	0.135%	0.354%	Other	Categorical Deciles	-0.218%		
	2c	0.314%	**0.546%**	Other	Always/Ever/Never	0.061%		
	2d	0.264%	0.368%	White	Percentages	0.407%		
	3a	0.118%	-	White	Categorical Deciles	0.341%		
	3b	0.343%	-	White	Always/Ever/Never	0.301%		
	3c	0.352%	-					

Note: Values are the percent improvement of the RMSE from five-fold cross-validation compared to the constant-only benchmark. The largest improvement within gender is bolded.

Appendix Table B10. School Discipline Analyses Fit Statistics, Cross-Validation, by Gender

		Self-Id 1979	Self-Id 2002		Observed race		Screener	
	Approach	*Improvement*	*Improvement*	*Group*	*Approach*	*Improvement*	*Approach*	*Improvement*
Women	1a	1.173%	1.045%	Black	Percentages	1.277%	Default	1.211%
	1b	0.016%	0.001%	Black	Categorical Deciles	1.304%	Hispanic	-0.021%
	1c	1.193%	1.090%	Black	Always/Ever/Never	1.300%	Black	1.203%
	2a	1.006%	1.224%	Other	Percentages	-0.009%		
	2b	1.008%	1.174%	Other	Categorical Deciles	-0.102%		
	2c	1.252%	1.078%	Other	Always/Ever/Never	-0.023%		
	2d	1.170%	1.179%	White	Percentages	1.244%		
	3a	1.107%	-	White	Categorical Deciles	1.079%		
	3b	1.317%	-	White	Always/Ever/Never	1.205%		
	3c	**1.445%**	-					
Men	1a	1.405%	1.071%	Black	Percentages	1.482%	Default	1.367%
	1b	0.071%	0.051%	Black	Categorical Deciles	1.363%	Hispanic	0.026%
	1c	1.451%	1.071%	Black	Always/Ever/Never	1.413%	Black	1.401%
	2a	1.298%	1.148%	Other	Percentages	0.087%		
	2b	1.340%	1.038%	Other	Categorical Deciles	-0.120%		
	2c	1.506%	1.158%	Other	Always/Ever/Never	-0.006%		
	2d	1.471%	1.191%	White	Percentages	1.239%		
	3a	1.456%	-	White	Categorical Deciles	1.292%		
	3b	1.693%	-	White	Always/Ever/Never	1.252%		
	3c	**1.787%**	-					

Note: Values are the percent improvement of the RMSE from five-fold cross-validation compared to the constant-only benchmark. The largest improvement within gender is bolded.

Appendix C. Full Regression Tables for Single Measure Analyses

Appendix C Section 1. Self-identification 1979 Models

Appendix Table C1a. Women's Wage Regression Analyses by Specifications of Self-Identification 1979

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.0101		Reference Category: Black		Reference Category: Black		Black	-0.0685*	-0.0760
Hispanic	0.104***			(0.0226)		Hispanic alone	0.167***	Hispanic	0.00884		(0.0344)	(0.0489)
	(0.0259)		Black		-0.135***		(0.0236)		(0.0489)	Hispanic	0.0248	0.0185
Non-Black, non-Hispanic	0.147***				(0.0192)	European alone	0.107**	European alone	0.163***		(0.0341)	(0.0449)
	(0.0202)						(0.0344)		(0.0234)	European	0.0848**	0.0767
						Other alone	0.109***	Other alone	0.107**		(0.0290)	(0.0471)
							(0.0264)		(0.0338)	American Indian	-0.127***	-0.128***
						Multiorigin	0.155***	Multiorigin	0.154***	or Alaska Native	(0.0361)	(0.0364)
							(0.0266)		(0.0266)	Other	0.0590*	0.0490
						Missing data	0.00444	Missing data	0.0916		(0.0276)	(0.0536)
							(0.117)		(0.0544)	Multiorigin		0.0127
												(0.0584)
Constant	2.480***		Constant	2.573***	2.615***	Constant	2.473***	Constant	2.475***	Constant	2.549***	2.556***
	(0.0162)			(0.00979)	(0.0104)		(0.0165)		(0.0165)		(0.0314)	(0.0468)
Observations	3,680		Observations	3,680	3,680	Observations	3,680	Observations	3,680	Observations	3,680	3,680
AIC	5,795		AIC	5,846	5,797	AIC	5,794	AIC	5,795	AIC	5,784	5,786
BIC	5,814		BIC	5,858	5,809	BIC	5,831	BIC	5,833	BIC	5,821	5,830
R^2	0.014		R^2	0.000	0.013	R^2	0.016	R^2	0.016	R^2	0.019	0.019
Adjusted R^2	0.014		Adjusted R^2	0.000	0.013	Adjusted R^2	0.015	Adjusted R^2	0.014	Adjusted R^2	0.017	0.017
5-fold CV RMSE	0.532		5-fold CV RM	0.536	0.532	5-fold CV RMSE	0.532	5-fold CV RMSE	0.532	5-fold CV RMSE	0.531	0.531

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C1b. Women's Wage Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	-0.0681*	-0.0785**
Cuban	0.401***		(0.0282)	(0.0289)
	(0.0993)	Cuban	0.287***	0.269**
Mexican	-0.0709		(0.0822)	(0.0829)
	(0.0519)	Mexican	-0.120*	-0.135**
Mexican-American	0.0994**		(0.0473)	(0.0482)
	(0.0372)	Mexican-American	0.00639	-0.00610
Puerto Rican	0.187**		(0.0378)	(0.0386)
	(0.0568)	Puerto Rican	0.123*	0.111*
Other Hispanic	0.0647		(0.0546)	(0.0551)
	(0.0907)	Other Hispanic	0.0159	0.000243
English	-0.0162		(0.0782)	(0.0788)
	(0.0406)	English	-0.0291	-0.0445
French	0.248		(0.0251)	(0.0268)
	(0.129)	French	-0.0316	-0.0470
German	0.189***		(0.0349)	(0.0362)
	(0.0480)	German	0.119***	0.0933**
Native American	0.0873		(0.0241)	(0.0287)
	(0.0695)	Native American	-0.0791*	-0.0997**
Irish	0.113		(0.0310)	(0.0335)
	(0.0685)	Irish	0.00547	-0.0148
Italian	0.144		(0.0258)	(0.0287)
	(0.0812)	Italian	0.0904*	0.0710
Polish	0.159		(0.0403)	(0.0420)
	(0.142)	Polish	0.0924	0.0692
Portuguese	0.0486		(0.0490)	(0.0511)
	(0.200)	Portuguese	0.233	0.209
All remaining responses	0.215***		(0.133)	(0.133)
	(0.0440)	All remaining responses	0.101***	0.0780**
American	0.0260		(0.0238)	(0.0278)
	(0.0491)	American	-0.0145	-0.0291
Multiracial	0.189***		(0.0445)	(0.0454)
	(0.0221)	Multiracial		0.0563
Missing	0.00444			(0.0348)
	(0.116)			
Constant	2.473***	Constant	2.549***	2.558***
	(0.0164)		(0.0238)	(0.0244)
Observations	3,680	Observations	3,680	3,680
AIC	5,755	AIC	5,731	5,730
BIC	5,867	BIC	5,836	5,842
R^2	0.033	R^2	0.039	0.039
Adjusted R^2	0.028	Adjusted R^2	0.034	0.035
5-fold CV RMSE	0.528	5-fold CV RMSE	0.528	0.527

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C2a. Women's Salary Regression Analyses by Specifications of Self-Identification 1979

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.0125		Reference Category: Black		Reference Category: Black		Black	-0.0579	-0.0129
Hispanic	0.0690			(0.0339)		Hispanic alone	0.117**	Hispanic	0.0229		(0.0526)	(0.0748)
	(0.0392)		Black		-0.0828**		(0.0356)		(0.0738)	Hispanic	0.00849	0.0462
Non-Black, non-Hispanic	0.0878**				(0.0292)	European alone	0.113*	European alone	0.118***		(0.0519)	(0.0684)
	(0.0307)						(0.0525)		(0.0353)	European	0.0374	0.0851
						Other alone	0.0760	Other alone	0.108*		(0.0443)	(0.0717)
							(0.0399)		(0.0515)	American Indian	-0.0921	-0.0859
						Multiorigin	0.0846*	Multiorigin	0.0845*	or Alaska Native	(0.0549)	(0.0554)
							(0.0403)		(0.0404)	Other	0.0329	0.0924
						Missing data	0.102	Missing data	0.0505		(0.0417)	(0.0818)
							(0.183)		(0.0823)	Multiorigin		-0.0751
												(0.0886)
Constant	10.01***		Constant	10.07***	10.09***	Constant	10.000***	Constant	9.999***	Constant	10.07***	10.02***
	(0.0247)			(0.0147)	(0.0157)		(0.0252)		(0.0252)		(0.0480)	(0.0715)
Observations	3,554		Observations	3,554	3,554	Observations	3,554	Observations	3,554	Observations	3,554	3,554
AIC	8,408		AIC	8,414	8,406	AIC	8,410	AIC	8,409	AIC	8,410	8,411
BIC	8,426		BIC	8,426	8,418	BIC	8,447	BIC	8,446	BIC	8,447	8,454
R^2	0.002		R^2	0.000	0.002	R^2	0.003	R^2	0.004	R^2	0.003	0.004
Adjusted R^2	0.002		Adjusted R^2	0.000	0.002	Adjusted R^2	0.002	Adjusted R^2	0.002	Adjusted R^2	0.002	0.002
5-fold CV RMSE	0.789		5-fold CV RM	0.79	0.789	5-fold CV RMSE	0.79	5-fold CV RMSE	0.79	5-fold CV RMSE	0.79	0.789

*p < .05, **p < .01, ***p < .001 (two-tailed tests)

Appendix Table C2b. Women's Salary Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	-0.0648	-0.0843
Cuban	0.531***		(0.0431)	(0.0441)
	(0.151)	Cuban	0.445***	0.412**
Mexican	-0.153		(0.126)	(0.127)
	(0.0793)	Mexican	-0.182*	-0.211**
Mexican-American	0.0695		(0.0725)	(0.0739)
	(0.0561)	Mexican-American	-0.0229	-0.0464
Puerto Rican	0.144		(0.0573)	(0.0584)
	(0.0866)	Puerto Rican	0.0654	0.0416
Other Hispanic	-0.0532		(0.0831)	(0.0838)
	(0.139)	Other Hispanic	-0.0843	-0.115
English	-0.0957		(0.121)	(0.121)
	(0.0618)	English	-0.0670	-0.0968*
French	-0.104		(0.0381)	(0.0408)
	(0.192)	French	-0.0586	-0.0903
German	0.124		(0.0527)	(0.0549)
	(0.0715)	German	0.0786*	0.0301
Native American	0.147		(0.0364)	(0.0435)
	(0.105)	Native American	-0.0621	-0.101*
Irish	0.143		(0.0475)	(0.0511)
	(0.105)	Irish	-0.0336	-0.0717
Italian	0.0945		(0.0391)	(0.0434)
	(0.125)	Italian	0.0956	0.0589
Polish	0.0173		(0.0614)	(0.0639)
	(0.212)	Polish	0.0342	-0.00845
Portuguese	0.0575		(0.0738)	(0.0767)
	(0.298)	Portuguese	0.292	0.247
All remaining responses	0.229***		(0.198)	(0.199)
	(0.0670)	All remaining responses	0.0962**	0.0519
American	0.0269		(0.0360)	(0.0420)
	(0.0752)	American	-0.0379	-0.0664
Multiracial	0.134***		(0.0677)	(0.0691)
	(0.0336)	Multiracial		0.107*
Missing	0.102			(0.0525)
	(0.182)			
Constant	10.000***	Constant	10.08***	10.09***
	(0.0251)		(0.0362)	(0.0371)
Observations	3,554	Observations	3,554	3,554
AIC	8,392	AIC	8,387	8,385
BIC	8,503	BIC	8,492	8,496
R^2	0.015	R^2	0.016	0.017
Adjusted R^2	0.01	Adjusted R^2	0.011	0.012
5-fold CV RMSE	0.788	5-fold CV RMSE	0.789	0.786

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C3a. Women's Unemployment Regression Analyses by Specifications of Self-Identification 1979

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black				Reference Category: Black	Reference Category: Black		
Hispanic	-1.395***	Hispanic	0.511	Hispanic alone	-2.821***	Hispanic	1.309*	Black	2.654***	2.531***
	(0.327)		(0.279)		(0.280)		(0.543)		(0.391)	(0.558)
Non-Black, non-Hispanic	-2.674***	Black		2.402***	European alone	-3.118***	European alone	-2.851***	Hispanic	1.257**	1.158*
	(0.246)			(0.238)		(0.381)		(0.278)		(0.386)	(0.503)
					Other alone	-1.520***	Other alone	-2.986***	European	0.0358	-0.0946
						(0.336)		(0.376)		(0.320)	(0.530)
					Multiorigin	-2.466***	Multiorigin	-2.509***	American Indian	0.500	0.480
						(0.315)		(0.316)	or Alaska Native	(0.393)	(0.398)
					Missing data	-3.511**	Missing data	-2.837***	Other	-0.198	-0.355
						(1.208)		(0.619)		(0.314)	(0.598)
									Multiorigin		0.202
											(0.653)
Constant	4.099***	Constant	2.193***	1.698***	Constant	4.221***	Constant	4.210***	Constant	1.413***	1.536**
	(0.207)		(0.112)	(0.117)		(0.213)		(0.213)		(0.345)	(0.526)
Observations	6,283	Observations	6,283	6,283	Observations	6,283	Observations	6,283	Observations	6,283	6,283
AIC	44,070	AIC	44,186	44,088	AIC	44,068	AIC	44,069	AIC	44,076	44,078
BIC	44,091	BIC	44,199	44,102	BIC	44,109	BIC	44,110	BIC	44,117	44,126
R^2	0.019	R^2	0.001	0.016	R^2	0.020	R^2	0.020	R^2	0.019	0.019
Adjusted R^2	0.019	Adjusted R^2	0.000	0.016	Adjusted R^2	0.019	Adjusted R^2	0.019	Adjusted R^2	0.018	0.018
5-fold CV RMSE	8.051	5-fold CV RM	8.12	8.044	5-fold CV RMSE	8.05	5-fold CV RMSE	8.05	5-fold CV RMSE	8.062	7.994

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C3b. Women's Unemployment Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	2.552***	2.547***
Cuban	-3.230*		(0.330)	(0.337)
	(1.263)	Cuban	-0.788	-0.797
Mexican	-2.265***		(1.045)	(1.052)
	(0.660)	Mexican	0.0317	0.0246
Mexican-American	-0.661		(0.588)	(0.597)
	(0.491)	Mexican-American	1.884***	1.878***
Puerto Rican	-2.022**		(0.470)	(0.478)
	(0.699)	Puerto Rican	0.529	0.522
Other Hispanic	-0.814		(0.647)	(0.654)
	(1.172)	Other Hispanic	1.981*	1.973*
English	-3.066***		(0.994)	(1.001)
	(0.461)	English	-0.0750	-0.0828
French	-2.715*		(0.280)	(0.300)
	(1.326)	French	0.496	0.487
German	-3.485***		(0.376)	(0.394)
	(0.558)	German	-0.147	-0.159
Native American	-2.830***		(0.270)	(0.319)
	(0.739)	Native American	0.233	0.223
Irish	-3.367***		(0.334)	(0.363)
	(0.736)	Irish	0.0364	0.0260
Italian	-3.208***		(0.282)	(0.319)
	(0.927)	Italian	-0.313	-0.323
Polish	-0.849		(0.450)	(0.473)
	(1.488)	Polish	0.278	0.268
Portuguese	-2.243		(0.554)	(0.574)
	(1.380)	Portuguese	-0.411	-0.420
All remaining responses	-2.526***		(1.020)	(1.028)
	(0.499)	All remaining responses	0.0279	0.0173
American	-3.267***		(0.269)	(0.309)
	(0.522)	American	-0.482	-0.489
Multiorigin	-2.488***		(0.469)	(0.479)
	(0.270)	Multiorigin		0.0275
Missing	-3.511**			(0.390)
	(1.208)			
Constant	4.221***	Constant	1.519***	1.523***
	(0.213)		(0.264)	(0.271)
Observations	6,283	Observations	6,283	6,283
AIC	44,081	AIC	44,089	44,091
BIC	44,203	BIC	44,204	44,213
R^2	0.022	R^2	0.020	0.020
Adjusted R^2	0.019	Adjusted R^2	0.018	0.018
5-fold CV RMSE	8.054	5-fold CV RMSE	8.066	8.041

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C4a. Women's Depression Regression Analyses by Specifications of Self-Identification 1979

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	-0.144		Reference Category: Black		Reference Category: Black		Black	0.995***	1.196**
Hispanic	-0.582**			(0.175)		Hispanic alone	-0.983***	Hispanic	0.795*		(0.266)	(0.378)
	(0.202)		Black		0.662***		(0.185)		(0.373)	Hispanic	0.395	0.563
Non-Black, non-Hispanic	-0.691***				(0.150)	European alone	-0.575*	European alone	-0.949***		(0.263)	(0.346)
	(0.159)						(0.266)		(0.184)	European	0.0982	0.312
						Other alone	-0.589**	Other alone	-0.509		(0.224)	(0.363)
							(0.206)		(0.261)	American Indian	0.999***	1.025***
						Multiorigin	-0.187	Multiorigin	-0.219	or Alaska Native	(0.282)	(0.284)
							(0.209)		(0.209)	Other	0.262	0.527
						Missing data	-0.478	Missing data	-1.493***		(0.218)	(0.416)
							(0.904)		(0.417)	Multiorigin		-0.340
												(0.455)
Constant	4.328***		Constant	3.889***	3.666***	Constant	4.318***	Constant	4.314***	Constant	3.280***	3.080***
	(0.126)			(0.0765)	(0.0818)		(0.129)		(0.129)		(0.243)	(0.361)
Observations	4,255		Observations	4,255	4,255	Observations	4,255	Observations	4,255	Observations	4,255	4,255
AIC	24,839		AIC	24,856	24,837	AIC	24,832	AIC	24,830	AIC	24,818	24,820
BIC	24,858		BIC	24,868	24,850	BIC	24,871	BIC	24,868	BIC	24,857	24,864
R^2	0.005		R^2	0.000	0.005	R^2	0.007	R^2	0.008	R^2	0.011	0.011
Adjusted R^2	0.004		Adjusted R^2	0.000	0.004	Adjusted R^2	0.006	Adjusted R^2	0.007	Adjusted R^2	0.010	0.009
5-fold CV RMSE	4.481		5-fold CV RM	4.49	4.477	5-fold CV RMSE	4.47	5-fold CV RMSE	4.48	5-fold CV RMSE	4.470	4.470

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C4b. Women's Depression Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	0.823***	0.790***
Cuban	-1.936*		(0.222)	(0.228)
	(0.778)	Cuban	-0.494	-0.546
Mexican	-1.398***		(0.659)	(0.665)
	(0.402)	Mexican	-0.441	-0.487
Mexican-American	-0.788**		(0.365)	(0.372)
	(0.297)	Mexican-American	0.175	0.136
Puerto Rican	0.287		(0.300)	(0.306)
	(0.430)	Puerto Rican	1.389***	1.351**
Other Hispanic	-1.413*		(0.416)	(0.420)
	(0.703)	Other Hispanic	-0.574	-0.622
English	-0.440		(0.609)	(0.614)
	(0.322)	English	0.0578	0.00916
French	-0.468		(0.200)	(0.215)
	(1.009)	French	0.626*	0.579*
German	-0.890*		(0.274)	(0.284)
	(0.385)	German	-0.439*	-0.515*
Native American	-0.00292		(0.192)	(0.228)
	(0.540)	Native American	1.164***	1.101***
Irish	-0.924		(0.241)	(0.262)
	(0.547)	Irish	0.191	0.130
Italian	-2.024**		(0.205)	(0.227)
	(0.640)	Italian	-0.228	-0.288
Polish	-0.880		(0.316)	(0.330)
	(1.126)	Polish	0.426	0.356
Portuguese	-2.032		(0.388)	(0.404)
	(1.696)	Portuguese	-0.393	-0.469
All remaining responses	-1.000**		(1.059)	(1.066)
	(0.341)	All remaining responses	-0.0473	-0.116
American	-0.273		(0.188)	(0.219)
	(0.381)	American	0.803*	0.759*
Multiorigin	-0.512**		(0.345)	(0.353)
	(0.175)	Multiorigin		0.170
Missing	-0.478			(0.275)
	(0.904)			
Constant	4.318***	Constant	3.457***	3.484***
	(0.129)		(0.187)	(0.193)
Observations	4,255	Observations	4,255	4,255
AIC	24,846	AIC	24,808	24,810
BIC	24,960	BIC	24,916	24,924
R^2	0.010	R^2	0.018	0.018
Adjusted R^2	0.006	Adjusted R^2	0.014	0.014
5-fold CV RMSE	4.472	5-fold CV RMSE	4.460	4.467

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C5a. Women's School Discipline Regression Analyses by Specifications of Self-Identification 1979

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	-0.212*		Reference Category: Black		Reference Category: Black		Black	0.803***	0.844***
Hispanic	-0.782***			(0.0954)		Hispanic alone	-0.923***	Hispanic	0.179		(0.128)	(0.184)
	(0.105)		Black		0.851***		(0.0891)		(0.184)	Hispanic	0.0275	0.0599
Non-Black, non-Hispanic	-0.871***				(0.0712)	European alone	-0.832***	European alone	-0.920***		(0.134)	(0.170)
	(0.0746)						(0.127)		(0.0882)	European	-0.107	-0.0634
						Other alone	-0.794***	Other alone	-0.792***		(0.111)	(0.179)
							(0.109)		(0.123)	American	-0.145	-0.0927
						Multiorigin	-0.718***	Multiorigin	-0.721***	Indian or	(0.115)	(0.204)
							(0.0989)		(0.0992)	Other	-0.145	-0.0927
						Missing data	-0.758	Missing data	-0.988***		(0.115)	(0.204)
							(0.416)		(0.209)	Multiorigin		-0.0687
												(0.223)
Constant	-0.905***		Constant	-1.475***	-1.756***	Constant	-0.907***	Constant	-0.911***	Constant	-1.718***	-1.759***
	(0.0577)			(0.0361)	(0.0417)		(0.0593)		(0.0593)		(0.119)	(0.177)
Observations	6,049		Observations	6,049	6,049	Observations	6,049	Observations	6,049	Observations	6,049	6,049
AIC	5,596		AIC	5,727	5,595	AIC	5,609	AIC	5,610	AIC	5,585	5,587
BIC	5,616		BIC	5,740	5,608	BIC	5,649	BIC	5,650	BIC	5,625	5,634
Pseudo R^2	0.024		Pseudo R^2	0.001	0.024	Pseudo R^2	0.023	Pseudo R^2	0.023	Pseudo R^2	0.027	0.027
5-fold CV RMSE	0.381		5-fold CV RMSE	0.385	0.381	5-fold CV RMSE	0.381	5-fold CV RMSE	0.381	5-fold CV RMSE	0.380	0.381

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C5b. Women's School Discipline Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	0.892***	0.860***
Cuban	-1.233*		(0.108)	(0.111)
	(0.532)	Cuban	-0.0673	-0.127
Mexican	-1.693***		(0.389)	(0.392)
	(0.318)	Mexican	-0.731**	-0.779**
Mexican-American	-0.624***		(0.271)	(0.273)
	(0.156)	Mexican-American	0.367*	0.330*
Puerto Rican	-0.327		(0.155)	(0.159)
	(0.216)	Puerto Rican	0.496*	0.455*
Other Hispanic	-1.268**		(0.208)	(0.210)
	(0.476)	Other Hispanic	-0.593	-0.644
English	-0.898***		(0.435)	(0.437)
	(0.160)	English	0.0720	0.0233
French	-0.415		(0.0995)	(0.107)
	(0.402)	French	0.163	0.111
German	-1.552***		(0.129)	(0.136)
	(0.246)	German	-0.233*	-0.304**
Native American	-0.480*		(0.0990)	(0.114)
	(0.231)	Native American	0.429***	0.360**
Irish	-1.095***		(0.108)	(0.121)
	(0.281)	Irish	-0.0435	-0.107
Italian	-0.381		(0.102)	(0.114)
	(0.289)	Italian	0.302*	0.236
Polish	-0.798		(0.150)	(0.159)
	(0.547)	Polish	0.166	0.0945
Portuguese	-0.634		(0.195)	(0.203)
	(0.454)	Portuguese	0.539	0.480
All remaining responses	-1.140***		(0.313)	(0.316)
	(0.189)	All remaining responses	-0.0984	-0.165
American	-0.844***		(0.0979)	(0.111)
	(0.179)	American	-0.0364	-0.0813
Multiorigin	-0.770***		(0.170)	(0.174)
	(0.0825)	Multiorigin		0.171
Missing	-0.758			(0.135)
	(0.416)			
Constant	-0.907***	Constant	-1.814***	-1.788***
	(0.0593)		(0.0936)	(0.0967)
Observations	6,049	Observations	6,049	6,049
AIC	5,596	AIC	5,568	5,568
BIC	5,717	BIC	5,682	5,689
Pseudo R^2	0.029	Pseudo R^2	0.034	0.034
5-fold CV RMSE	0.381	5-fold CV RMSE	0.380	0.380

*p < .05, **p < .01, ***p < .001 (two-tailed tests)

Appendix Table C6a. Men's Wage Regression Analyses by Specifications of Self-Identification 1979

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black				Reference Category: Black	Reference Category: Black		
Hispanic	0.169***	Hispanic	-0.0486*	Hispanic alone	0.382***	Hispanic	0.0117	Black	-0.228***	-0.129**
	(0.0255)		(0.0229)		(0.0226)		(0.0481)		(0.0336)	(0.0490)
Non-Black, non-Hispanic	0.337***	Black	-0.293***	European alone	0.262***	European alone	0.378***	Hispanic	-0.0721*	0.0112
	(0.0198)		(0.0190)		(0.0322)		(0.0224)		(0.0334)	(0.0450)
				Other alone	0.168***	Other alone	0.271***	European	0.135***	0.238***
					(0.0259)		(0.0316)		(0.0284)	(0.0470)
				Multiorigin	0.289***	Multiorigin	0.288***	American Indian	-0.162***	-0.149***
					(0.0276)		(0.0276)	or Alaska Native	(0.0380)	(0.0383)
				Missing data	0.102	Missing data	0.131*	Other	0.0436	0.167**
					(0.110)		(0.0534)		(0.0277)	(0.0526)
								Multiorigin		-0.160**
										(0.0580)
Constant	2.587***	Constant	2.805***	2.880***	Constant	2.584***	Constant	2.585***	Constant	2.816***	2.718***
	(0.0160)		(0.00985)	(0.0102)		(0.0163)		(0.0163)		(0.0306)	(0.0468)
Observations	3,726	Observations	3,726	3,726	Observations	3,726	Observations	3,726	Observations	3,726	3,726
AIC	5,744	AIC	6,019	5,794	AIC	5,735	AIC	5,728	AIC	5,714	5,708
BIC	5,762	BIC	6,031	5,806	BIC	5,772	BIC	5,766	BIC	5,751	5,752
R^2	0.073	R^2	0.001	0.060	R^2	0.077	R^2	0.078	R^2	0.082	0.084
Adjusted R^2	0.072	Adjusted R^2	0.001	0.06	Adjusted R^2	0.075	Adjusted R^2	0.077	Adjusted R^2	0.080	0.082
5-fold CV RMSE	0.522	5-fold CV RM	0.542	0.526	5-fold CV RMSE	0.523	5-fold CV RMSE	0.522	5-fold CV RMSE	0.521	0.520

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C6b. Men's Wage Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	-0.235***	-0.254***
Cuban	0.427***		(0.0279)	(0.0295)
	(0.0948)	Cuban	0.207**	0.181*
Mexican	0.0798		(0.0747)	(0.0759)
	(0.0481)	Mexican	-0.116*	-0.137**
Mexican-American	0.124***		(0.0459)	(0.0472)
	(0.0369)	Mexican-American	-0.130***	-0.150***
Puerto Rican	0.163**		(0.0379)	(0.0393)
	(0.0550)	Puerto Rican	-0.0811	-0.104*
Other Hispanic	0.224*		(0.0512)	(0.0525)
	(0.0948)	Other Hispanic	0.0371	0.0111
English	0.211***		(0.0802)	(0.0813)
	(0.0380)	English	0.0254	-0.000857
French	0.295*		(0.0253)	(0.0286)
	(0.115)	French	0.0368	0.0164
German	0.339***		(0.0373)	(0.0387)
	(0.0433)	German	0.0987***	0.0668*
Native American	0.117		(0.0243)	(0.0293)
	(0.0702)	Native American	-0.139***	-0.165***
Irish	0.332***		(0.0329)	(0.0356)
	(0.0614)	Irish	0.0631*	0.0334
Italian	0.427***		(0.0260)	(0.0301)
	(0.0696)	Italian	0.162***	0.130**
Polish	0.527***		(0.0389)	(0.0422)
	(0.105)	Polish	0.0896	0.0584
Portuguese	-0.0630		(0.0465)	(0.0492)
	(0.197)	Portuguese	0.0226	0.0102
All remaining responses	0.350***		(0.119)	(0.120)
	(0.0414)	All remaining responses	0.0814***	0.0485
American	0.231***		(0.0238)	(0.0292)
	(0.0456)	American	-0.0123	-0.0353
Multiorigin	0.371***		(0.0430)	(0.0446)
	(0.0223)	Multiorigin		0.0686
Missing	0.102			(0.0353)
	(0.110)			
Constant	2.584***	Constant	2.825***	2.842***
	(0.0162)		(0.0236)	(0.0252)
Observations	3,726	Observations	3,726	3,726
AIC	5,716	AIC	5,677	5,675
BIC	5,828	BIC	5,783	5,787
R^2	0.087	R^2	0.096	0.097
Adjusted R^2	0.083	Adjusted R^2	0.092	0.093
5-fold CV RMSE	0.521	5-fold CV RMSE	0.518	0.518

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C7a. Men's Salary Regression Analyses by Specifications of Self-Identification 1979

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	0.249***
	(0.0384)
Non-Black, non-Hispanic	0.495***
	(0.0297)
Constant	10.24***
	(0.0241)
Observations	3,649
AIC	8,448
BIC	8,467
R^2	0.072
Adjusted R^2	0.072
5-fold CV RMSE	0.770

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.0773*	
	(0.0343)	
Black		-0.433***
		(0.0286)
Constant	10.56***	10.67***
	(0.0146)	(0.0151)
Observations	3,649	3,649
AIC	8,714	8,496
BIC	8,726	8,509
R^2	0.001	0.059
Adjusted R^2	0.001	0.059
5-fold CV RM	0.799	0.775

	2a
	One Categorical Variable
Reference Category: Black	
Hispanic alone	0.564***
	(0.0337)
European alone	0.379***
	(0.0479)
Other alone	0.245***
	(0.0391)
Multiorigin	0.433***
	(0.0411)
Missing data	0.258
	(0.156)
Constant	10.23***
	(0.0246)
Observations	3,649
AIC	8,437
BIC	8,474
R^2	0.077
Adjusted R^2	0.075
5-fold CV RMSE	0.769

	2b
	One Categorical Variable + Hispanic Dichotomous Variable
Reference Category: Black	
Hispanic	-0.0490
	(0.0698)
European alone	0.557***
	(0.0334)
Other alone	0.386***
	(0.0470)
Multiorigin	0.434***
	(0.0412)
Missing data	0.269***
	(0.0779)
Constant	10.23***
	(0.0246)
Observations	3,649
AIC	8,436
BIC	8,473
R^2	0.077
Adjusted R^2	0.075
5-fold CV RMSE	0.768

	2c	2d
	Series of Dichotomous Variables	
Black	-0.345***	-0.246***
	(0.0498)	(0.0721)
Hispanic	-0.115*	-0.0324
	(0.0493)	(0.0659)
European	0.191***	0.295***
	(0.0419)	(0.0690)
American	-0.227***	-0.215***
Indian or	(0.0563)	(0.0567)
Other	0.0492	0.174*
	(0.0407)	(0.0774)
Multiorigin		-0.162
		(0.0851)
Constant	10.58***	10.48***
	(0.0451)	(0.0687)
Observation:	3,649	3,649
AIC	8,419	8,417
BIC	8,456	8,461
R^2	0.081	0.082
Adjusted R^2	0.080	0.08
5-fold CV RM	0.767	0.767

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C7b. Men's Salary Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	-0.383***	-0.410***
Cuban	0.684***		(0.0415)	(0.0437)
	(0.138)	Cuban	0.339**	0.300**
Mexican	0.134		(0.108)	(0.110)
	(0.0722)	Mexican	-0.220**	-0.250***
Mexican-American	0.225***		(0.0686)	(0.0703)
	(0.0562)	Mexican-American	-0.190***	-0.220***
Puerto Rican	0.211*		(0.0569)	(0.0590)
	(0.0859)	Puerto Rican	-0.189*	-0.223**
Other Hispanic	0.182		(0.0784)	(0.0803)
	(0.138)	Other Hispanic	-0.112	-0.149
English	0.370***		(0.117)	(0.119)
	(0.0568)	English	0.0114	-0.0269
French	0.339*		(0.0371)	(0.0420)
	(0.169)	French	-0.0392	-0.0696
German	0.574***		(0.0548)	(0.0570)
	(0.0630)	German	0.159***	0.113**
Native American	0.220*		(0.0357)	(0.0429)
	(0.104)	Native American	-0.208***	-0.246***
Irish	0.445***		(0.0489)	(0.0527)
	(0.0932)	Irish	0.0766*	0.0335
Italian	0.515***		(0.0382)	(0.0442)
	(0.104)	Italian	0.167**	0.120
Polish	0.607***		(0.0570)	(0.0619)
	(0.159)	Polish	0.0619	0.0163
Portuguese	0.320		(0.0691)	(0.0730)
	(0.291)	Portuguese	0.0732	0.0548
All remaining responses	0.491***		(0.176)	(0.176)
	(0.0607)	All remaining responses	0.0971**	0.0498
American	0.323***		(0.0350)	(0.0427)
	(0.0681)	American	-0.0737	-0.107
Multiorigin	0.530***		(0.0641)	(0.0663)
	(0.0333)	Multiorigin		0.101
Missing	0.258			(0.0520)
	(0.156)			
Constant	10.23***	Constant	10.62***	10.65***
	(0.0246)		(0.0348)	(0.0370)
Observations	3,649	Observations	3,649	3,649
AIC	8,445	AIC	8,391	8,389
BIC	8,557	BIC	8,496	8,501
R^2	0.080	R^2	0.094	0.094
Adjusted R^2	0.076	Adjusted R^2	0.09	0.09
5-fold CV RMSE	0.769	5-fold CV RMSE	0.764	0.763

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C8a. Men's Unemployment Regression Analyses by Specifications of Self-Identification 1979

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.816*		Reference Category: Black		Reference Category: Black		Black	2.912***	2.472***
Hispanic	-1.468***			(0.324)		Hispanic alone	-3.386***	Hispanic	1.203		(0.452)	(0.663)
	(0.376)		Black		2.854***		(0.310)		(0.661)	Hispanic	1.479**	1.106
Non-Black, non-Hispanic	-3.203***				(0.268)	European alone	-3.017***	European alone	-3.393***		(0.459)	(0.616)
	(0.276)						(0.418)		(0.308)	European	-0.372	-0.834
						Other alone	-1.525***	Other alone	-3.116***		(0.372)	(0.631)
							(0.383)		(0.413)	American	-0.153	-0.225
						Multiorigin	-2.990***	Multiorigin	-3.019***	Indian or	(0.472)	(0.479)
							(0.372)		(0.372)	Other	0.0805	-0.459
						Missing data	-4.365***	Missing data	-2.650***		(0.364)	(0.698)
							(1.192)		(0.727)	Multiorigin		0.691
												(0.763)
Constant	4.821***		Constant	2.537***	1.967***	Constant	4.888***	Constant	4.884***	Constant	1.905***	2.343***
	(0.233)			(0.126)	(0.132)		(0.238)		(0.238)		(0.400)	(0.627)
Observations	6,403		Observations	6,403	6,403	Observations	6,403	Observations	6,403	Observations	6,403	6,403
AIC	46,589		AIC	46,720	46,615	AIC	46,593	AIC	46,594	AIC	46,595	46,596
BIC	46,609		BIC	46,734	46,628	BIC	46,634	BIC	46,634	BIC	46,635	46,643
R^2	0.022		R^2	0.001	0.017	R^2	0.022	R^2	0.022	R^2	0.022	0.022
Adjusted R^2	0.021		Adjusted R^2	0.001	0.017	Adjusted R^2	0.021	Adjusted R^2	0.021	Adjusted R^2	0.021	0.021
5-fold CV RMSE	9.191		5-fold CV RMSE	9.269	9.209	5-fold CV RMSE	9.171	5-fold CV RMSE	9.196	5-fold CV RM	9.171	9.198

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C8b. Men's Unemployment Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	3.107***	3.300***
Cuban	-2.836*		(0.373)	(0.389)
	(1.362)	Cuban	-0.532	-0.227
Mexican	-1.987**		(1.072)	(1.086)
	(0.727)	Mexican	1.669*	1.893**
Mexican-American	-1.465**		(0.677)	(0.689)
	(0.563)	Mexican-American	1.702**	1.917***
Puerto Rican	-0.288		(0.551)	(0.565)
	(0.792)	Puerto Rican	2.472***	2.710***
Other Hispanic	1.063		(0.740)	(0.752)
	(1.297)	Other Hispanic	2.601*	2.890**
English	-3.186***		(1.088)	(1.101)
	(0.503)	English	0.000952	0.283
French	-3.860**		(0.324)	(0.363)
	(1.362)	French	0.503	0.774
German	-3.259***		(0.467)	(0.493)
	(0.567)	German	-0.222	0.120
Native American	-3.856***		(0.314)	(0.372)
	(0.832)	Native American	-0.00814	0.289
Irish	-3.762***		(0.408)	(0.443)
	(0.754)	Irish	-0.377	-0.0398
Italian	-1.116		(0.328)	(0.382)
	(0.893)	Italian	0.467	0.796
Polish	-4.506**		(0.515)	(0.550)
	(1.376)	Polish	0.106	0.454
Portuguese	-4.624**		(0.630)	(0.662)
	(1.594)	Portuguese	-1.362	-1.097
All remaining responses	-2.896***		(1.150)	(1.160)
	(0.539)	All remaining responses	0.0622	0.417
American	-3.047***		(0.309)	(0.371)
	(0.580)	American	0.0958	0.337
Multiracial	-3.247***		(0.526)	(0.545)
	(0.310)	Multiracial		-0.784
Missing	-4.384***			(0.458)
	(1.191)			
Constant	4.907***	Constant	1.702***	1.526***
	(0.238)		(0.298)	(0.315)
Observations	6,403	Observations	6,403	6,403
AIC	46,597	AIC	46,603	46,602
BIC	46,719	BIC	46,718	46,724
R^2	0.025	R^2	0.024	0.024
Adjusted R^2	0.022	Adjusted R^2	0.021	0.022
5-fold CV RMSE	9.160	5-fold CV RMSE	9.190	9.198

*p < .05, **p < .01, ***p < .001 (two-tailed tests)

Appendix Table C9a. Men's Depression Regression Analyses by Specifications of Self-Identification 1979

	1a	1b		1c	2a		2b		2c		2d
	One Categorical Variable	Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables		
Reference Category: Black		Hispanic	0.225		Reference Category: Black		Reference Category: Black		Black	0.354	-0.131
Hispanic	-0.147		(0.155)		Hispanic alone	-0.695***	Hispanic	-0.220		(0.238)	(0.343)
	(0.178)	Black		0.466***		(0.159)		(0.337)	Hispanic	0.231	-0.184
Non-Black, non-Hispanic	-0.581***			(0.132)	European alone	-0.448*	European alone	-0.661***		(0.237)	(0.318)
	(0.139)					(0.227)		(0.158)	European	-0.310	-0.822*
					Other alone	-0.121	Other alone	-0.385		(0.201)	(0.330)
						(0.181)		(0.224)	American Indian	0.818**	0.750**
					Multiorigin	-0.301	Multiorigin	-0.290	or Alaska Native	(0.270)	(0.272)
						(0.194)		(0.195)	Other	-0.178	-0.795*
					Missing data	2.576***	Missing data	0.215		(0.196)	(0.371)
						(0.679)		(0.370)	Multiorigin		0.800
											(0.408)
Constant	3.191***	Constant	2.819***	2.725***	Constant	3.151***	Constant	3.148***	Constant	2.832***	3.315***
	(0.111)		(0.0668)	(0.0716)		(0.113)		(0.113)		(0.216)	(0.328)
Observations	4,109	Observations	4,109	4,109	Observations	4,109	Observations	4,109	Observations	4,109	4,109
AIC	22,752	AIC	22,767	22,757	AIC	22,737	AIC	22,754	AIC	22,746	22,744
BIC	22,771	BIC	22,780	22,770	BIC	22,775	BIC	22,792	BIC	22,784	22,789
R^2	0.005	R^2	0.001	0.003	R^2	0.010	R^2	0.006	R^2	0.008	0.009
Adjusted R^2	0.004	Adjusted R^2	0.000	0.003	Adjusted R^2	0.009	Adjusted R^2	0.005	Adjusted R^2	0.006	0.007
5-fold CV RMSE	3.855	5-fold CV RM	3.860	3.854	5-fold CV RMSE	3.848	5-fold CV RMSE	3.856	5-fold CV RMSE	3.849	3.851

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C9b. Men's Depression Regression Analyses by Specifications of Self-Identification 1979

	3a		3b	3c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Black	0.326	0.424*
Cuban	-0.280		(0.197)	(0.208)
	(0.633)	Cuban	-0.451	-0.309
Mexican	-0.770*		(0.513)	(0.522)
	(0.340)	Mexican	-0.198	-0.0837
Mexican-American	-0.345		(0.326)	(0.335)
	(0.260)	Mexican-American	-0.0229	0.0890
Puerto Rican	1.185**		(0.269)	(0.280)
	(0.377)	Puerto Rican	1.412***	1.529***
Other Hispanic	0.315		(0.360)	(0.369)
	(0.610)	Other Hispanic	0.233	0.362
English	-0.102		(0.541)	(0.548)
	(0.274)	English	0.118	0.260
French	-1.053		(0.180)	(0.204)
	(0.846)	French	-0.251	-0.139
German	-0.678*		(0.267)	(0.277)
	(0.306)	German	-0.439*	-0.271
Native American	0.217		(0.173)	(0.207)
	(0.497)	Native American	0.736**	0.881***
Irish	-0.148		(0.235)	(0.254)
	(0.425)	Irish	-0.139	0.0210
Italian	0.0609		(0.184)	(0.213)
	(0.483)	Italian	-0.140	0.0328
Polish	0.198		(0.275)	(0.299)
	(0.761)	Polish	-0.271	-0.0996
Portuguese	-2.005		(0.333)	(0.352)
	(1.456)	Portuguese	-1.123	-1.056
All remaining responses	-0.774**		(0.887)	(0.888)
	(0.292)	All remaining responses	-0.261	-0.0837
American	-0.314		(0.169)	(0.207)
	(0.321)	American	-0.0740	0.0447
Multiorigin	-0.648***		(0.304)	(0.315)
	(0.156)	Multiorigin		-0.372
Missing	2.579***			(0.252)
	(0.678)			
Constant	3.148***	Constant	2.856***	2.765***
	(0.113)		(0.167)	(0.178)
Observations	4,109	Observations	4,109	4,109
AIC	22,736	AIC	22,739	22,739
BIC	22,849	BIC	22,847	22,853
R^2	0.016	R^2	0.015	0.015
Adjusted R^2	0.012	Adjusted R^2	0.011	0.011
5-fold CV RMSE	3.857	5-fold CV RMSE	3.848	3.847

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C10a. Men's School Discipline Regression Analyses by Specifications of Self-Identification 1979

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	-0.802***
	(0.0899)
Non-Black, non-Hispanic	-0.812***
	(0.0634)
Constant	-0.106*
	(0.0519)
Observations	6,091
AIC	7,573
BIC	7,593
Pseudo R^2	0.022
5-fold CV RMSE	0.464

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.242**	
	(0.0791)	
Black		0.810***
		(0.0613)
Constant	-0.666***	-0.916***
	(0.0293)	(0.0326)
Observations	6,091	6,091
AIC	7,734	7,571
BIC	7,747	7,584
Pseudo R^2	0.001	0.022
5-fold CV RM	0.470	0.464

	2a
	One Categorical Variable
Reference Category: Black	
Hispanic alone	-0.870***
	(0.0727)
European alone	-0.764***
	(0.0996)
Other alone	-0.805***
	(0.0918)
Multiorigin	-0.599***
	(0.0859)
Missing data	-0.729*
	(0.287)
Constant	-0.118*
	(0.0531)
Observations	6,091
AIC	7,585
BIC	7,625
Pseudo R^2	0.022
5-fold CV RMSE	0.465

	2b
	One Categorical Variable + Hispanic Dichotomous Variable
Reference Category: Black	
Hispanic	0.129
	(0.160)
European alone	-0.862***
	(0.0722)
Other alone	-0.765***
	(0.0983)
Multiorigin	-0.602***
	(0.0860)
Missing data	-0.944***
	(0.176)
Constant	-0.120*
	(0.0531)
Observations	6,091
AIC	7,585
BIC	7,626
Pseudo R^2	0.021
5-fold CV RMSE	0.465

	2c	2d
	Series of Dichotomous Variables	
Black	0.860***	0.837***
	(0.107)	(0.160)
Hispanic	0.0532	0.0345
	(0.112)	(0.150)
European	-0.0136	-0.0372
	(0.0899)	(0.154)
American Indian	0.0357	0.00825
or Alaska Native	(0.0893)	(0.170)
Other	0.0357	0.00825
	(0.0893)	(0.170)
Multiorigin		0.0351
		(0.185)
Constant	-0.977***	-0.955***
	(0.0969)	(0.152)
Observations	6,091	6,091
AIC	7,561	7,563
BIC	7,601	7,610
Pseudo R^2	0.025	0.025
5-fold CV RMSE	0.464	0.464

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C10b. Men's School Discipline Regression Analyses by Specifications of Self-Identification 1979

	3a
One Categorical Variable	
Reference Category: Black	
Cuban	-1.935***
	(0.478)
Mexican	-1.222***
	(0.200)
Mexican-American	-0.715***
	(0.136)
Puerto Rican	-0.621**
	(0.197)
Other Hispanic	-0.899**
	(0.328)
English	-0.686***
	(0.120)
French	-0.892**
	(0.341)
German	-1.312***
	(0.154)
Native American	-0.646**
	(0.199)
Irish	-0.488**
	(0.176)
Italian	-0.929***
	(0.230)
Polish	-0.830*
	(0.344)
Portuguese	-0.574
	(0.373)
All remaining responses	-1.181***
	(0.142)
American	-0.413**
	(0.132)
Multiorigin	-0.717***
	(0.0715)
Missing	-0.728*
	(0.287)
Constant	-0.119*
	(0.0531)
Observations	6,091
AIC	7,565
BIC	7,686
Pseudo R^2	0.027
5-fold CV RMSE	0.464

	3b	3c
Series of Dichotomous Variables		
Black	0.861***	0.796***
	(0.0887)	(0.0933)
Cuban	-0.796*	-0.908**
	(0.332)	(0.335)
Mexican	-0.272	-0.351
	(0.181)	(0.184)
Mexican-American	0.205	0.131
	(0.134)	(0.138)
Puerto Rican	0.267	0.182
	(0.182)	(0.186)
Other Hispanic	-0.125	-0.229
	(0.281)	(0.285)
English	0.0202	-0.0795
	(0.0802)	(0.0902)
French	0.111	0.0166
	(0.114)	(0.120)
German	-0.202*	-0.321***
	(0.0790)	(0.0930)
Native American	0.422***	0.317**
	(0.0962)	(0.106)
Irish	0.123	0.00334
	(0.0805)	(0.0943)
Italian	0.193	0.0763
	(0.126)	(0.135)
Polish	0.126	0.00238
	(0.155)	(0.164)
Portuguese	-0.111	-0.205
	(0.288)	(0.291)
All remaining responses	-0.156*	-0.280**
	(0.0776)	(0.0931)
American	0.468***	0.386**
	(0.123)	(0.128)
Multiorigin		0.276*
		(0.113)
Constant	-0.978***	-0.918***
	(0.0737)	(0.0784)
Observations	6,091	6,091
AIC	7,532	7528
BIC	7,646	7649
Pseudo R^2	0.031	0.032
5-fold CV RMSE	0.463	0.462

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix C Section 2. Self-identification 2002 Models

Appendix Table C11. Women's Wage Regression Analyses by Specifications of Self-Identification 2002

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	-0.00373		Reference Category: Black		Reference Category: Black		Black	-0.150***	-0.153***
Hispanic	0.110***			(0.0235)		Hispanic	0.113***	Hispanic	-0.0600*		(0.0310)	(0.0321)
	(0.0269)		Black		-0.156***		(0.0270)		(0.0262)	Hispanic	-0.0631*	-0.0634*
Non-Black, non-Hispanic	0.171***				(0.0195)	White	0.170***	White	0.171***		(0.0264)	(0.0264)
	(0.0203)						(0.0211)		(0.0207)	White	0.0194	0.0167
						Some other race	0.292***	Some other race	0.228***		(0.0277)	(0.0288)
							(0.0747)		(0.0473)	American Indian	-0.00567	-0.0223
						Multiracial	0.124	Multiracial	0.146*	or Alaska Native	(0.0724)	(0.0872)
							(0.0845)		(0.0737)	Some other race	0.0819	0.0761
						Noninterview	0.177***	Noninterview	0.148***		(0.0465)	(0.0495)
							(0.0387)		(0.0327)	Multiracial		0.0313
												(0.0914)
Constant	2.462***		Constant	2.576***	2.619***	Constant	2.459***	Constant	2.462***	Constant	2.614***	2.616***
	(0.0166)			(0.00969)	(0.0103)		(0.0168)		(0.0167)		(0.0268)	(0.0278)
Observations	3,680		Observations	3,680	3,680	Observations	3,680	Observations	3,680	Observations	3,680	3,680
AIC	5,778		AIC	5,846	5,782	AIC	5,780	AIC	5,782	AIC	5,782	5,784
BIC	5,796		BIC	5,858	5,795	BIC	5,817	BIC	5,819	BIC	5,819	5,827
R^2	0.019		R^2	0.000	0.017	R^2	0.020	R^2	0.019	R^2	0.019	0.019
Adjusted R^2	0.018		Adjusted R^2	0.000	0.017	Adjusted R^2	0.019	Adjusted R^2	0.018	Adjusted R^2	0.018	0.018
5-fold CV RMSE	0.530		5-fold CV RM	0.535	0.531	5-fold CV RMSE	0.531	5-fold CV RMSE	0.531	5-fold CV RMSE	0.530	0.531

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C12. Women's Salary Regression Analyses by Specifications of Self-Identification 2002

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black		Hispanic	0.00547	Reference Category: Black	Reference Category: Black	Black	-0.163***	-0.184***
Hispanic	0.0883*		(0.0350)	Hispanic	0.0926*	Hispanic	-0.0490		(0.0476)	(0.0493)
	(0.0407)	Black		(0.0408)	(0.0394)		-0.0466	-0.0490
Non-Black, non-Hispanic	0.124***		-0.115***	White	0.112***	White	0.122***		(0.0397)	(0.0397)
	(0.0309)		(0.0297)		(0.0320)		(0.0314)		-0.0432	-0.0626
				Some other race	0.239*	Some other race	0.142*		(0.0425)	(0.0441)
					(0.113)		(0.0716)	American Indian	0.0403	-0.0811
				Multiracial	0.217	Multiracial	0.219*	or Alaska Native	(0.111)	(0.133)
					(0.127)		(0.111)	Some other race	-0.0138	-0.0550
				Noninterview	0.218***	Noninterview	0.178***		(0.0704)	(0.0747)
					(0.0606)		(0.0505)	Multiracial		0.224
										(0.137)
Constant	9.987***	Constant	10.07***	10.10***	Constant	9.983***	Constant	9.981***	Constant	10.15***	10.17***
	(0.0253)		(0.0146)	(0.0155)		(0.0256)		(0.0255)		(0.0414)	(0.0427)
Observations	3,554	Observations	3,554	3,554	Observations	3,554	Observations	3,554	Observations	3,554	3,554
AIC	8,400	AIC	8,414	8,399	AIC	8,400	AIC	8,402	AIC	8,404	8,403
BIC	8,418	BIC	8,426	8,411	BIC	8,437	BIC	8,439	BIC	8,441	8,446
R^2	0.004	R^2	0.000	0.004	R^2	0.006	R^2	0.006	R^2	0.005	0.006
Adjusted R^2	0.004	Adjusted R^2	0.000	0.004	Adjusted R^2	0.005	Adjusted R^2	0.004	Adjusted R^2	0.004	0.004
5-fold CV RMSE	0.788	5-fold CV RM	0.790	0.789	5-fold CV RMSE	0.789	5-fold CV RMSE	0.789	5-fold CV RMSE	0.788	0.790

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C13. Women's Unemployment Regression Analyses by Specifications of Self-Identification 2002

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.890**		Reference Category: Black		Reference Category: Black		Black	3.614***	3.666***
Hispanic	-1.398***			(0.321)		Hispanic	-1.370***	Hispanic	1.143***		(0.280)	(0.283)
	(0.380)		Black		2.690***		(0.380)		(0.339)	Hispanic	1.206***	1.176***
Non-Black, non-Hispanic	-2.906***				(0.261)	White	-1.936***	White	-2.107***		(0.342)	(0.343)
	(0.264)						(0.298)		(0.292)	White	1.480***	1.528***
						Some other race	-2.676*	Some other race	-2.272***		(0.228)	(0.231)
							(1.040)		(0.650)	American Indian	1.005	1.911
						Multiracial	-0.992	Multiracial	-1.147	or Alaska Native	(1.011)	(1.202)
							(1.172)		(1.041)	Some other race	1.038	1.276*
						Noninterview	-3.666***	Noninterview	-3.666***		(0.623)	(0.646)
							(0.289)		(0.285)	Multiracial		-1.746
												(1.255)
Constant	4.460***		Constant	2.172***	1.770***	Constant	4.432***	Constant	4.489***	Constant	0.855***	0.831***
	(0.235)			(0.109)	(0.113)		(0.237)		(0.235)		(0.160)	(0.161)
Observations	6,283		Observations	6,283	6,283	Observations	6,283	Observations	6,283	Observations	6,283	6,283
AIC	44,064		AIC	44,181	44,083	AIC	44,024	AIC	44,020	AIC	44,019	44,019
BIC	44,084		BIC	44,195	44,097	BIC	44,065	BIC	44,060	BIC	44,060	44,067
R^2	0.020		R^2	0.001	0.017	R^2	0.027	R^2	0.028	R^2	0.028	0.028
Adjusted R^2	0.02		Adjusted R^2	0.001	0.016	Adjusted R^2	0.026	Adjusted R^2	0.027	Adjusted R^2	0.027	0.027
5-fold CV RMSE	8.038		5-fold CV RM	8.127	8.077	5-fold CV RMSE	8.024	5-fold CV RMSE	8.027	5-fold CV RMSE	8.032	8.022

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C14. Women's Depression Regression Analyses by Specifications of Self-Identification 2002

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black		Hispanic	-0.0566	Reference Category: Black	Reference Category: Black	Black	0.860***	0.902***
Hispanic	-0.693**		(0.184)	Hispanic	-0.669**	Hispanic	0.168		(0.227)	(0.234)
	(0.213)	Black		0.887***		(0.214)		(0.204)	Hispanic	0.221	0.223
Non-Black, non-Hispanic	-0.946***			(0.154)	White	-0.940***	White	-0.941***		(0.206)	(0.206)
	(0.160)					(0.167)		(0.164)	White	-0.101	-0.0617
					Some other race	0.754	Some other race	-0.746*		(0.202)	(0.208)
						(0.597)		(0.371)	American Indian	0.986	1.291
					Multiracial	-0.524	Multiracial	-0.383	or Alaska Native	(0.578)	(0.697)
						(0.652)		(0.585)	Some other race	-0.154	-0.0631
					Noninterview	-1.001***	Noninterview	-0.863***		(0.363)	(0.381)
						(0.268)		(0.238)	Multiracial		-0.561
											(0.719)
Constant	4.507***	Constant	3.871***	3.620***	Constant	4.484***	Constant	4.493***	Constant	3.634***	3.601***
	(0.131)		(0.0755)	(0.0803)		(0.133)		(0.132)		(0.191)	(0.195)
Observations	4,255	Observations	4,255	4,255	Observations	4,255	Observations	4,255	Observations	4,255	4,255
AIC	24,823	AIC	24,856	24,823	AIC	24,824	AIC	24,830	AIC	24,825	24,827
BIC	24,842	BIC	24,869	24,836	BIC	24,862	BIC	24,868	BIC	24,863	24,871
R^2	0.008	R^2	0.000	0.008	R^2	0.009	R^2	0.008	R^2	0.009	0.009
Adjusted R^2	0.008	Adjusted R^2	0.000	0.008	Adjusted R^2	0.008	Adjusted R^2	0.007	Adjusted R^2	0.008	0.008
5-fold CV RMSE	4.474	5-fold CV RM	4.487	4.471	5-fold CV RMSE	4.468	5-fold CV RMSE	4.474	5-fold CV RMSE	4.475	4.474

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C15. Women's School Discipline Regression Analyses by Specifications of Self-Identification 2002

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black		Hispanic	-0.174	Reference Category: Black	Reference Category: Black	Black	0.707***	0.730***
Hispanic	-0.816***		(0.109)	Hispanic	-0.830***	Hispanic	-0.0366		(0.0834)	(0.0844)
	(0.121)	Black		0.863***		(0.122)		(0.120)	Hispanic	-0.00749	-0.0227
Non-Black, non-Hispanic	-0.871***			(0.0755)	White	-1.096***	White	-1.045***		(0.121)	(0.122)
	(0.0773)					(0.0945)		(0.0914)	White	-0.336***	-0.313***
					Some other race	-0.496	Some other race	-0.438*		(0.0816)	(0.0827)
						(0.319)		(0.201)	American Indian	0.0507	0.465
					Multiracial	-0.961*	Multiracial	-0.939*	or Alaska Native	(0.317)	(0.379)
						(0.413)		(0.367)	Some other race	0.157	0.253
					Noninterview	-0.732***	Noninterview	-0.737***		(0.200)	(0.207)
						(0.0859)		(0.0848)	Multiracial		-0.811
											(0.450)
Constant	-0.846***	Constant	-1.488***	-1.708***	Constant	-0.831***	Constant	-0.844***	Constant	-1.567***	-1.579***
	(0.0643)		(0.0353)	(0.0397)		(0.0649)		(0.0647)		(0.0543)	(0.0549)
Observations	6,049	Observations	6,049	6,049	Observations	6,049	Observations	6,049	Observations	6,049	6,049
AIC	5,611	AIC	5,729	5,609	AIC	5,596	AIC	5,599	AIC	5,602	5,600
BIC	5,631	BIC	5,742	5,622	BIC	5,636	BIC	5,639	BIC	5,642	5,647
Pseudo R^2	0.021	Pseudo R^2	0.000	0.021	Pseudo R^2	0.025	Pseudo R^2	0.025	Pseudo R^2	0.024	0.025
5-fold CV RMSE	0.381	5-fold CV RM	0.385	0.381	5-fold CV RMSE	0.381	5-fold CV RMSE	0.381	5-fold CV RMSE	0.381	0.381

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C16. Men's Wage Regression Analyses by Specifications of Self-Identification 2002

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	0.185***
	(0.0267)
Non-Black, non-Hispanic	0.323***
	(0.0202)
Constant	2.583***
	(0.0166)
Observations	3,726
AIC	5,776
BIC	5,794
R^2	0.065
Adjusted R^2	0.064
5-fold CV RMSE	0.525

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.0339	
	(0.0238)	
Black		-0.291***
		(0.0195)
Constant	2.801***	2.874***
	(0.00975)	(0.0101)
Observations	3,726	3,726
AIC	6,021	5,807
BIC	6,034	5,819
R^2	0.001	0.056
Adjusted R^2	0.000	0.056
5-fold CV RM	0.543	0.527

	2a
	One Categorical Variable
Reference Category: Black	
Hispanic	0.184***
	(0.0267)
White	0.355***
	(0.0208)
Some other race	0.273***
	(0.0805)
Multiracial	0.160
	(0.0953)
Noninterview	0.131***
	(0.0351)
Constant	2.583***
	(0.0166)
Observations	3,726
AIC	5,739
BIC	5,777
R^2	0.075
Adjusted R^2	0.074
5-fold CV RMSE	0.523

	2b
	One Categorical Variable + Hispanic Dichotomous Variable
Reference Category: Black	
Hispanic	-0.0878***
	(0.0261)
White	0.349***
	(0.0205)
Some other race	0.311***
	(0.0473)
Multiracial	0.206*
	(0.0844)
Noninterview	0.169***
	(0.0307)
Constant	2.582***
	(0.0166)
Observations	3,726
AIC	5,737
BIC	5,775
R^2	0.076
Adjusted R^2	0.075
5-fold CV RMSE	0.523

	2c	2d
	Series of Dichotomous Variables	
Black	-0.172***	-0.175***
	(0.0299)	(0.0304)
Hispanic	-0.0983***	-0.0985***
	(0.0262)	(0.0263)
White	0.180***	0.177***
	(0.0263)	(0.0268)
American Indian	-0.216**	-0.252*
or Alaska Native	(0.0820)	(0.104)
Some other race	0.184***	0.178***
	(0.0462)	(0.0477)
Multiracial		0.0603
		(0.106)
Constant	2.754***	2.757***
	(0.0252)	(0.0256)
Observations	3,726	3,726
AIC	5,718	5,719
BIC	5,755	5,763
R^2	0.081	0.081
Adjusted R^2	0.08	0.079
5-fold CV RMSE	0.521	0.522

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C17. Men's Salary Regression Analyses by Specifications of Self-Identification 2002

	1a	1b	1c	2a	2b	2c	2d
	One Categorical Variable	Series of Dichotomous Variables		One Categorical Variable	One Categorical Variable + Hispanic Dichotomous Variable	Series of Dichotomous Variables	
Reference Category: Black		Hispanic	-0.0359	Reference Category: Black	Reference Category: Black	Black	-0.218***	-0.220***
Hispanic	0.286***		(0.0355)	Hispanic	0.284***	Hispanic	-0.0741		(0.0454)	(0.0461)
	(0.0402)	Black		-0.428***		(0.0400)	(0.0390)	Hispanic	-0.0933*	-0.0935*
Non-Black, non-Hispanic	0.469***			(0.0292)	White	0.525***	White	0.518***		(0.0392)	(0.0392)
	(0.0302)					(0.0310)		(0.0306)	White	0.302***	0.301***
					Some other race	0.142	Some other race	0.331***		(0.0398)	(0.0407)
						(0.116)		(0.0703)	American Indian	-0.502***	-0.515***
					Multiracial	0.0510	Multiracial	0.147	or Alaska Native	(0.121)	(0.150)
						(0.142)		(0.126)	Some other race	0.196**	0.193**
					Noninterview	0.150**	Noninterview	0.210***		(0.0691)	(0.0714)
						(0.0539)		(0.0467)	Multiracial		0.0230
											(0.156)
Constant	10.23***	Constant	10.55***	10.66***	Constant	10.23***	Constant	10.23***	Constant	10.45***	10.45***
	(0.0250)		(0.0145)	(0.0150)		(0.0250)		(0.0249)		(0.0384)	(0.0391)
Observations	3,649	Observations	3,649	3,649	Observations	3,649	Observations	3,649	Observations	3,649	3,649
AIC	8,487	AIC	8,718	8,511	AIC	8,430	AIC	8,429	AIC	8,404	8,406
BIC	8,505	BIC	8,730	8,523	BIC	8,467	BIC	8,466	BIC	8,441	8,450
R^2	0.062	R^2	0.000	0.055	R^2	0.078	R^2	0.078	R^2	0.085	0.085
Adjusted R^2	0.062	Adjusted R^2	0.000	0.055	Adjusted R^2	0.077	Adjusted R^2	0.077	Adjusted R^2	0.083	0.083
5-fold CV RMSE	0.774	5-fold CV RM	0.799	0.777	5-fold CV RMSE	0.768	5-fold CV RMSE	0.767	5-fold CV RMSE	0.764	0.766

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C18. Men's Unemployment Regression Analyses by Specifications of Self-Identification 2002

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.952*		Reference Category: Black		Reference Category: Black		Black	4.694***	4.699***
Hispanic	-2.345***			(0.376)		Hispanic	-2.243***	Hispanic	1.071**		(0.321)	(0.322)
	(0.444)		Black		3.878***		(0.444)		(0.401)	Hispanic	1.096**	1.092**
Non-Black, non-Hispanic	-4.106***				(0.301)	White	-3.171***	White	-3.152***		(0.402)	(0.403)
	(0.305)						(0.345)		(0.340)	White	1.452***	1.458***
						Some other race	-1.253	Some other race	-1.787*		(0.258)	(0.260)
							(1.348)		(0.778)	American Indian	2.564	2.763
						Multiracial	2.240	Multiracial	0.239	or Alaska Native	(1.352)	(1.701)
							(1.591)		(1.408)	Some other race	2.870***	2.901***
						Noninterview	-4.667***	Noninterview	-4.610***		(0.743)	(0.760)
							(0.327)		(0.324)	Multiracial		-0.342
												(1.771)
Constant	5.857***		Constant	2.559***	1.979***	Constant	5.755***	Constant	5.739***	Constant	1.111***	1.109***
	(0.273)			(0.123)	(0.126)		(0.274)		(0.273)		(0.173)	(0.173)
Observations	6,403		Observations	6,403	6,403	Observations	6,403	Observations	6,403	Observations	6,403	6,403
AIC	46,543		AIC	46,720	46,563	AIC	46,513	AIC	46,515	AIC	46,505	46,507
BIC	46,563		BIC	46,734	46,577	BIC	46,554	BIC	46,556	BIC	46,546	46,555
R^2	0.029		R^2	0.001	0.025	R^2	0.034	R^2	0.034	R^2	0.035	0.035
Adjusted R^2	0.028		Adjusted R^2	0.001	0.025	Adjusted R^2	0.033	Adjusted R^2	0.033	Adjusted R^2	0.034	0.034
5-fold CV RMSE	9.156		5-fold CV RM	9.287	9.163	5-fold CV RMSE	9.130	5-fold CV RMSE	9.139	5-fold CV RMSE	9.125	9.124

*p < .05, **p < .01, ***p < .001 (two-tailed tests)

Appendix Table C19. Men's Depression Regression Analyses by Specifications of Self-Identification 2002

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	0.151		Reference Category: Black		Reference Category: Black		Black	0.267	0.266
Hispanic	-0.287			(0.164)		Hispanic	-0.256	Hispanic	0.0207		(0.195)	(0.198)
	(0.190)		Black		0.564***		(0.190)		(0.183)	Hispanic	0.0724	0.0723
Non-Black, non-Hispanic	-0.642***				(0.136)	White	-0.770***	White	-0.749***		(0.184)	(0.184)
	(0.141)						(0.147)		(0.145)	White	-0.501**	-0.502**
						Some other race	0.444	Some other race	0.445		(0.171)	(0.174)
							(0.567)		(0.334)	American Indian	2.526***	2.513***
						Multiracial	1.492*	Multiracial	1.224*	or Alaska Native	(0.570)	(0.723)
							(0.670)		(0.592)	Some other race	0.516	0.514
						Noninterview	-0.115	Noninterview	-0.222		(0.325)	(0.335)
							(0.218)		(0.199)	Multiracial		0.0218
												(0.752)
Constant	3.275***		Constant	2.837***	2.711***	Constant	3.244***	Constant	3.248***	Constant	2.990***	2.991***
	(0.117)			(0.0658)	(0.0702)		(0.117)		(0.117)		(0.159)	(0.161)
Observations	4,109		Observations	4,109	4,109	Observations	4,109	Observations	4,109	Observations	4,109	4,109
AIC	22,750		AIC	22,769	22,752	AIC	22,735	AIC	22,730	AIC	22,720	22,722
BIC	22,769		BIC	22,781	22,765	BIC	22,773	BIC	22,768	BIC	22,758	22,766
R^2	0.005		R^2	0.000	0.004	R^2	0.010	R^2	0.011	R^2	0.014	0.014
Adjusted R^2	0.005		Adjusted R^2	0.000	0.004	Adjusted R^2	0.009	Adjusted R^2	0.01	Adjusted R^2	0.013	0.013
5-fold CV RMSE	3.850		5-fold CV RM	3.858	3.855	5-fold CV RMSE	3.848	5-fold CV RMSE	3.847	5-fold CV RMSE	3.840	3.847

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C20. Men's School Discipline Regression Analyses by Specifications of Self-Identification 2002

	1a		1b	1c		2a		2b		2c	2d
	One Categorical Variable		Series of Dichotomous Variables			One Categorical Variable		One Categorical Variable + Hispanic Dichotomous Variable		Series of Dichotomous Variables	
Reference Category: Black			Hispanic	-0.258**		Reference Category: Black		Reference Category: Black		Black	0.641***	0.634***
Hispanic	-0.857***			(0.0919)		Hispanic	-0.851***	Hispanic	-0.102		(0.0735)	(0.0737)
	(0.106)		Black		0.776***		(0.106)		(0.100)	Hispanic	-0.0888	-0.0817
Non-Black, non-Hispanic	-0.765***				(0.0681)	White	-0.960***	White	-0.922***		(0.101)	(0.101)
	(0.0691)						(0.0808)		(0.0792)	White	-0.285***	-0.295***
						Some other race	-0.713*	Some other race	-0.661***		(0.0637)	(0.0643)
							(0.328)		(0.189)	American Indian	0.313	0.0412
						Multiracial	0.0212	Multiracial	-0.180	or Alaska Native	(0.309)	(0.397)
							(0.354)		(0.318)	Some other race	-0.0533	-0.100
						Noninterview	-0.621***	Noninterview	-0.633***		(0.184)	(0.189)
							(0.0746)		(0.0740)	Multiracial		0.472
												(0.416)
Constant	-0.0753		Constant	-0.674***	-0.852***	Constant	-0.0818	Constant	-0.0865	Constant	-0.722***	-0.718***
	(0.0606)			(0.0287)	(0.0309)		(0.0610)		(0.0609)		(0.0419)	(0.0421)
Observations	6,091		Observations	6,091	6,091	Observations	6,091	Observations	6,091	Observations	6,091	6,091
AIC	7,616		AIC	7,735	7,615	AIC	7,597	AIC	7,604	AIC	7,602	7,603
BIC	7,637		BIC	7,749	7,629	BIC	7,637	BIC	7,644	BIC	7,643	7,650
Pseudo R^2	0.017		Pseudo R^2	0.001	0.017	Pseudo R^2	0.02	Pseudo R^2	0.019	Pseudo R^2	0.019	0.019
5-fold CV RMSE	0.466		5-fold CV RM	0.471	0.466	5-fold CV RMSE	0.465	5-fold CV RMSE	0.466	5-fold CV RMSE	0.465	0.465

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix C Section 3. Observed Race Models

Appendix Table C21. Women's Wage Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	-0.0151***		
	(0.00193)		
Percent Observed Other		0.0107	
		(0.00584)	
Percent Observed White			0.0147***
			(0.00198)
Constant	2.621***	2.569***	2.481***
	(0.0105)	(0.00947)	(0.0154)
Observations	3,680	3,680	3,680
AIC	5,785	5,842	5,792
BIC	5,797	5,855	5,804
R^2	0.016	0.001	0.015
Adjusted R^2	0.016	0.001	0.014
5-fold CV RMSE	0.531	0.535	0.531

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	0.0145	-0.0181	0.225*
	(0.133)	(0.0407)	(0.0952)
20-29% Observed	-0.0671	-0.0322	0.206*
	(0.306)	(0.0442)	(0.0844)
30-39% Observed	-0.0224	0.0469	0.265**
	(0.265)	(0.0642)	(0.0982)
40-49% Observed	-0.247	-0.0576	0.153*
	(0.306)	(0.0633)	(0.0775)
50-59% Observed	0.0750	0.00379	0.0265
	(0.217)	(0.0728)	(0.0659)
60-69% Observed	0.378	0.138	0.128*
	(0.265)	(0.102)	(0.0622)
70-79% Observed	-0.107	0.123	0.0849
	(0.306)	(0.102)	(0.0473)
80-89% Observed	0.0514	0.467**	0.0721
	(0.113)	(0.155)	(0.0427)
90-99% Observed	0.0270	0.0704	0.103
	(0.0473)	(0.239)	(0.0526)
100% Observed	-0.177***	0.263	0.161***
	(0.0202)	(0.309)	(0.0202)
Constant	2.619***	2.574***	2.473***
	(0.0105)	(0.00966)	(0.0160)
Observations	3,680	3,680	3,680
AIC	5,781	5,848	5,792
BIC	5,849	5,917	5,860
R^2	0.022	0.004	0.019
Adjusted R^2	0.02	0.001	0.017
5-fold CV RMSE	0.532	0.535	0.532

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	0.192***	-0.246	-0.0438*
	(0.0375)	(0.310)	(0.0219)
Never Observed	0.178***	-0.266	-0.183***
	(0.0202)	(0.309)	(0.0210)
Constant	2.442***	2.837***	2.634***
	(0.0172)	(0.309)	(0.0123)
Observations	3,680	3,680	3,680
AIC	5,768	5,846	5,772
BIC	5,787	5,865	5,791
R^2	0.021	0.000	0.020
Adjusted R^2	0.021	0.000	0.020
5-fold CV RMSE	0.530	0.535	0.530

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C22. Women's Salary Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	-0.0111***		
	(0.00294)		
Percent Observed Other		0.0130	
		(0.00881)	
Percent Observed White			0.0101***
			(0.00301)
Constant	10.10***	10.06***	10.01***
	(0.0158)	(0.0142)	(0.0235)
Observations	3,554	3,554	3,554
AIC	8,400	8,412	8,403
BIC	8,412	8,424	8,415
R^2	0.004	0.001	0.003
Adjusted R^2	0.004	0.000	0.003
5-fold CV RMSE	0.788	0.790	0.789

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	-0.139	-0.000639	0.150
	(0.211)	(0.0615)	(0.142)
20-29% Observed	0.107	-0.0433	0.149
	(0.455)	(0.0661)	(0.129)
30-39% Observed	-0.178	-6.82e-05	0.396**
	(0.394)	(0.0962)	(0.151)
40-49% Observed	-0.226	-0.0281	0.190
	(0.455)	(0.0949)	(0.120)
50-59% Observed	0.301	0.00202	-0.00921
	(0.322)	(0.113)	(0.0987)
60-69% Observed	0.538	0.261	0.0307
	(0.394)	(0.153)	(0.0943)
70-79% Observed	-0.157	0.0968	0.0575
	(0.455)	(0.156)	(0.0717)
80-89% Observed	0.270	0.482*	0.0402
	(0.169)	(0.229)	(0.0651)
90-99% Observed	0.113	0.0974	0.0745
	(0.0714)	(0.354)	(0.0791)
100% Observed	-0.149***	0.395	0.114***
	(0.0307)	(0.456)	(0.0307)
Constant	10.10***	10.07***	9.998***
	(0.0158)	(0.0145)	(0.0244)
Observations	3,554	3,554	3,554
AIC	8,397	8,423	8,411
BIC	8,465	8,491	8,478
R^2	0.010	0.003	0.006
Adjusted R^2	0.007	0.000	0.003
5-fold CV RMSE	0.788	0.791	0.791

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	0.226***	-0.369	-0.0102
	(0.0569)	(0.457)	(0.0331)
Never Observed	0.149***	-0.399	-0.147***
	(0.0308)	(0.456)	(0.0319)
Constant	9.954***	10.46***	10.11***
	(0.0263)	(0.456)	(0.0185)
Observations	3,554	3,554	3,554
AIC	8,387	8,414	8,393
BIC	8,406	8,433	8,412
R^2	0.008	0.000	0.006
Adjusted R^2	0.007	0.000	0.006
5-fold CV RMSE	0.788	0.791	0.788

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C23. Women's Unemployment Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.231***		
	(0.0239)		
Percent Observed Other		0.0620	
		(0.0670)	
Percent Observed White			-0.240***
			(0.0240)
Constant	1.701***	2.242***	3.949***
	(0.118)	(0.109)	(0.196)
Observations	6,283	6,283	6,283
AIC	44,096	44,188	44,089
BIC	44,110	44,202	44,103
R^2	0.015	0.000	0.016
Adjusted R^2	0.014	0.000	0.016
5-fold CV RMSE	8.047	8.124	8.070

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	0.293	0.838	-1.716
	(1.413)	(0.523)	(1.089)
20-29% Observed	1.357	1.060	-0.859
	(2.699)	(0.572)	(1.099)
30-39% Observed	0.377	0.364	-1.446
	(4.047)	(0.768)	(1.139)
40-49% Observed	0.986	0.660	-1.593
	(3.620)	(0.835)	(0.949)
50-59% Observed	1.907	0.393	-0.697
	(2.442)	(0.885)	(0.826)
60-69% Observed	-0.851	0.790	-1.258
	(3.305)	(1.146)	(0.762)
70-79% Observed	-1.706	-0.543	-0.592
	(4.047)	(1.344)	(0.621)
80-89% Observed	0.555	-0.573	-1.098*
	(1.353)	(1.543)	(0.544)
90-99% Observed	2.385***	-0.610	-1.472*
	(0.652)	(2.040)	(0.627)
100% Observed	2.323***	-1.685	-2.513***
	(0.250)	(2.180)	(0.245)
Constant	1.706***	2.188***	3.923***
	(0.118)	(0.111)	(0.205)
Observations	6,283	6,283	6,283
AIC	44,112	44,199	44,094
BIC	44,186	44,273	44,169
R^2	0.015	0.001	0.018
Adjusted R^2	0.013	0.000	0.016
5-fold CV RMSE	8.089	8.145	8.074

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-1.101*	2.361	1.498***
	(0.481)	(2.190)	(0.265)
Never Observed	-2.323***	1.656	2.540***
	(0.250)	(2.179)	(0.254)
Constant	4.029***	0.504	1.410***
	(0.220)	(2.176)	(0.134)
Observations	6,283	6,283	6,283
AIC	44,103	44,184	44,082
BIC	44,123	44,204	44,102
R^2	0.014	0.001	0.017
Adjusted R^2	0.014	0.001	0.017
5-fold CV RMSE	8.086	8.107	8.051

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C24. Women's Depression Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.0782***		
	(0.0151)		
Percent Observed Other		0.0205	
		(0.0447)	
Percent Observed White			-0.0847***
			(0.0155)
Constant	3.624***	3.849***	4.400***
	(0.0825)	(0.0739)	(0.120)
Observations	4,255	4,255	4,255
AIC	24,830	24,856	24,826
BIC	24,842	24,869	24,839
R^2	0.006	0.000	0.007
Adjusted R^2	0.006	0.000	0.007
5-fold CV RMSE	4.475	4.486	4.472

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	1.788	-0.514	0.528
	(1.088)	(0.322)	(0.756)
20-29% Observed	-0.624	-0.279	-0.736
	(2.583)	(0.348)	(0.672)
30-39% Observed	-1.624	0.339	-0.145
	(2.238)	(0.493)	(0.746)
40-49% Observed	0.626	-0.425	-0.422
	(2.238)	(0.493)	(0.596)
50-59% Observed	0.543	0.594	-0.458
	(1.828)	(0.546)	(0.498)
60-69% Observed	-2.874	0.332	-0.238
	(2.238)	(0.742)	(0.481)
70-79% Observed	-2.957	-0.529	-1.084**
	(2.583)	(0.810)	(0.374)
80-89% Observed	1.492	1.973	-0.935**
	(0.881)	(1.202)	(0.338)
90-99% Observed	-0.146	1.116	-0.133
	(0.371)	(1.698)	(0.408)
100% Observed	0.892***	-2.134	-0.829***
	(0.158)	(2.245)	(0.158)
Constant	3.624***	3.884***	4.388***
	(0.0827)	(0.0754)	(0.125)
Observations	4,255	4,255	4,255
AIC	24,833	24,864	24,839
BIC	24,903	24,934	24,909
R^2	0.010	0.002	0.008
Adjusted R^2	0.007	0.000	0.006
5-fold CV RMSE	4.477	4.490	4.478

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-0.725*	2.113	0.219
	(0.295)	(2.250)	(0.172)
Never Observed	-0.900***	2.113	0.934***
	(0.158)	(2.246)	(0.164)
Constant	4.515***	1.750	3.560***
	(0.134)	(2.245)	(0.0972)
Observations	4,255	4,255	4,255
AIC	24,826	24,857	24,826
BIC	24,845	24,876	24,845
R^2	0.008	0.000	0.008
Adjusted R^2	0.007	0.000	0.007
5-fold CV RMSE	4.470	4.491	4.468

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C25. Women's School Discipline Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.0875***		
	(0.00717)		
Percent Observed Other		-0.0154	
		(0.0226)	
Percent Observed White			-0.0889***
			(0.00739)
Constant	-1.770***	-1.499***	-0.934***
	(0.0422)	(0.0354)	(0.0553)
Observations	6049	6049	6049
AIC	5,589	5,731	5,590
BIC	5,602	5,745	5,604
Pseudo R^2	0.025	0.000	0.025
5-fold CV RMSE	0.380	0.385	0.380

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	-1.569	-0.283	-0.877*
	(1.019)	(0.184)	(0.386)
20-29% Observed	0.510	-0.180	-0.210
	(0.803)	(0.196)	(0.324)
30-39% Observed	-	0.294	-0.744
	-	(0.227)	(0.390)
40-49% Observed	0.377	-0.184	-0.564
	(1.119)	(0.284)	(0.313)
50-59% Observed	1.763**	-0.230	-0.577*
	(0.634)	(0.316)	(0.267)
60-69% Observed	0.153	-0.00168	-0.461
	(1.096)	(0.371)	(0.235)
70-79% Observed	-	0.391	-0.755***
	-	(0.387)	(0.210)
80-89% Observed	0.153	-1.729	-0.939***
	(0.449)	(1.020)	(0.193)
90-99% Observed	0.559**	-0.456	-0.409*
	(0.195)	(0.757)	(0.194)
100% Observed	0.903***	0.643	-0.914***
	(0.0738)	(0.691)	(0.0752)
Constant	-1.763***	-1.490***	-0.914***
	(0.0422)	(0.0359)	(0.0571)
Observations	6043	6049	6049
AIC	5,587	5,736	5,597
BIC	5,647	5,810	5,671
Pseudo R^2	0.027	0.002	0.027
5-fold CV RMSE	0.380	0.386	0.381

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-0.483***	-0.755	0.321***
	(0.146)	(0.695)	(0.0893)
Never Observed	-0.915***	-0.643	0.945***
	(0.0742)	(0.691)	(0.0773)
Constant	-0.860***	-0.847	-1.829***
	(0.0606)	(0.690)	(0.0490)
Observations	6049	6049	6049
AIC	5,587	5,731	5,589
BIC	5,607	5,751	5,609
Pseudo R^2	0.026	0.000	0.025
5-fold CV RMSE	0.380	0.385	0.381

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C26. Men's Wage Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	-0.0306***		
	(0.00190)		
Percent Observed Other		-0.0172**	
		(0.00569)	
Percent Observed White			0.0342***
			(0.00193)
Constant	2.887***	2.806***	2.576***
	(0.0103)	(0.00954)	(0.0150)
Observations	3,726	3,726	3,726
AIC	5,773	6,014	5,722
BIC	5,785	6,027	5,735
R^2	0.065	0.002	0.078
Adjusted R^2	0.065	0.002	0.077
5-fold CV RMSE	0.525	0.542	0.521

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	-0.295	-0.0442	0.201*
	(0.152)	(0.0426)	(0.0918)
20-29% Observed	0.641	0.0803	0.211*
	(0.371)	(0.0483)	(0.0978)
30-39% Observed	0.162	-0.111	0.257*
	(0.303)	(0.0570)	(0.117)
40-49% Observed	-0.561	-0.0882	-0.000689
	(0.524)	(0.0603)	(0.0718)
50-59% Observed	0.0217	-0.174*	0.181**
	(0.262)	(0.0685)	(0.0582)
60-69% Observed	0.669	0.0513	0.0948
	(0.524)	(0.116)	(0.0545)
70-79% Observed	-0.118	-0.104	0.280***
	(0.235)	(0.109)	(0.0503)
80-89% Observed	-0.0220	-0.169	0.175***
	(0.131)	(0.125)	(0.0430)
90-99% Observed	-0.203***	-0.0903	0.211***
	(0.0500)	(0.222)	(0.0466)
100% Observed	-0.319***	-0.267	0.355***
	(0.0197)	(0.313)	(0.0197)
Constant	2.886***	2.804***	2.582***
	(0.0103)	(0.00972)	(0.0156)
Observations	3,726	3,726	3,726
AIC	5,772	6,022	5,711
BIC	5,841	6,090	5,780
R^2	0.070	0.005	0.085
Adjusted R^2	0.067	0.003	0.082
5-fold CV RMSE	0.526	0.542	0.521

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	0.201***	0.216	-0.188***
	(0.0392)	(0.314)	(0.0214)
Never Observed	0.319***	0.269	-0.365***
	(0.0198)	(0.313)	(0.0204)
Constant	2.567***	2.537***	2.937***
	(0.0168)	(0.313)	(0.0121)
Observations	3,726	3,726	3,726
AIC	5,773	6,019	5,711
BIC	5,791	6,038	5,729
R^2	0.066	0.002	0.081
Adjusted R^2	0.065	0.001	0.081
5-fold CV RMSE	0.525	0.543	0.520

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C27. Men's Salary Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	-0.0456***		
	(0.00285)		
Percent Observed Other		-0.0351***	
		(0.00861)	
Percent Observed White			0.0517***
			(0.00289)
Constant	10.68***	10.57***	10.21***
	(0.0152)	(0.0141)	(0.0227)
Observations	3,649	3,649	3,649
AIC	8,472	8,703	8,411
BIC	8,485	8,715	8,424
R^2	0.065	0.005	0.081
Adjusted R^2	0.065	0.004	0.081
5-fold CV RMSE	0.773	0.796	0.766

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	-0.525*	-0.0592	0.178
	(0.233)	(0.0645)	(0.139)
20-29% Observed	0.697	-0.0147	0.267
	(0.546)	(0.0725)	(0.152)
30-39% Observed	-0.0376	-0.135	0.335
	(0.546)	(0.0892)	(0.177)
40-49% Observed	-0.290	-0.211*	0.0394
	(0.772)	(0.0903)	(0.108)
50-59% Observed	-0.218	-0.231*	0.233**
	(0.386)	(0.101)	(0.0862)
60-69% Observed	-	0.0626	0.144
	-	(0.179)	(0.0849)
70-79% Observed	-0.307	-0.275	0.312***
	(0.345)	(0.160)	(0.0751)
80-89% Observed	0.118	-0.502*	0.297***
	(0.188)	(0.200)	(0.0654)
90-99% Observed	-0.293***	-0.255	0.344***
	(0.0761)	(0.326)	(0.0688)
100% Observed	-0.477***	-0.218	0.530***
	(0.0296)	(0.564)	(0.0294)
Constant	10.68***	10.57***	10.22***
	(0.0152)	(0.0144)	(0.0235)
Observations	3,649	3,649	3,649
AIC	8,472	8,715	8,410
BIC	8,534	8,783	8,478
R^2	0.070	0.006	0.086
Adjusted R^2	0.067	0.003	0.083
5-fold CV RMSE	0.772	0.799	0.765

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	0.263***	0.121	-0.291***
	(0.0586)	(0.565)	(0.0320)
Never Observed	0.481***	0.218	-0.544***
	(0.0297)	(0.565)	(0.0305)
Constant	10.20***	10.35***	10.75***
	(0.0254)	(0.564)	(0.0176)
Observations	3,649	3,649	3,649
AIC	8,465	8,713	8,404
BIC	8,484	8,731	8,422
R^2	0.068	0.002	0.083
Adjusted R^2	0.067	0.002	0.083
5-fold CV RMSE	0.772	0.799	0.766

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C28. Men's Unemployment Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.286***		
	(0.0269)		
Percent Observed Other		0.238**	
		(0.0745)	
Percent Observed White			-0.319***
			(0.0269)
Constant	1.942***	2.529***	4.870***
	(0.134)	(0.123)	(0.219)
Observations	6,403	6,403	6,403
AIC	46,615	46,717	46,587
BIC	46,628	46,730	46,601
R^2	0.017	0.002	0.022
Adjusted R^2	0.017	0.001	0.021
5-fold CV RMSE	9.158	9.223	9.159

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	0.599	0.348	-2.276
	(1.563)	(0.609)	(1.209)
20-29% Observed	0.538	0.636	-1.157
	(3.261)	(0.684)	(1.218)
30-39% Observed	2.221	1.144	-0.841
	(3.764)	(0.787)	(1.593)
40-49% Observed	2.457	0.737	1.833
	(5.322)	(0.890)	(1.100)
50-59% Observed	-0.218	3.022**	-1.191
	(3.485)	(0.961)	(0.860)
60-69% Observed	2.114	3.673*	-0.981
	(3.764)	(1.438)	(0.775)
70-79% Observed	-0.392	2.290	-1.795*
	(3.074)	(1.492)	(0.751)
80-89% Observed	-1.082	-1.339	-1.890**
	(1.660)	(1.575)	(0.623)
90-99% Observed	4.508***	1.850	-3.054***
	(0.777)	(2.684)	(0.687)
100% Observed	2.751***	-1.649	-3.190***
	(0.279)	(2.256)	(0.274)
Constant	1.946***	2.516***	4.766***
	(0.134)	(0.126)	(0.229)
Observations	6,403	6,403	6,403
AIC	46,621	46,721	46,590
BIC	46,696	46,796	46,664
R^2	0.019	0.004	0.024
Adjusted R^2	0.018	0.002	0.022
5-fold CV RMSE	9.198	9.246	9.171

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-0.441	2.465	1.995***
	(0.557)	(2.272)	(0.300)
Never Observed	-2.774***	1.664	3.008***
	(0.280)	(2.257)	(0.284)
Constant	4.696***	0.867	1.576***
	(0.245)	(2.254)	(0.152)
Observations	6,403	6,403	6,403
AIC	46,620	46,721	46,603
BIC	46,641	46,742	46,623
R^2	0.017	0.001	0.020
Adjusted R^2	0.016	0.001	0.019
5-fold CV RMSE	9.201	9.257	9.193

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C29. Men's Depression Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.0573***		
	(0.0132)		
Percent Observed Other		0.115**	
		(0.0387)	
Percent Observed White			-0.0745***
			(0.0136)
Constant	2.688***	2.791***	3.335***
	(0.0723)	(0.0647)	(0.105)
Observations	4,109	4,109	4,109
AIC	22,751	22,761	22,740
BIC	22,764	22,773	22,752
R^2	0.005	0.002	0.007
Adjusted R^2	0.004	0.002	0.007
5-fold CV RMSE	3.855	3.859	3.845

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	0.123	-0.0473	0.765
	(0.967)	(0.289)	(0.633)
20-29% Observed	1.311	0.329	-0.456
	(2.227)	(0.330)	(0.650)
30-39% Observed	2.978	0.943*	-0.761
	(2.227)	(0.377)	(0.867)
40-49% Observed	-0.689	0.00914	0.932
	(3.856)	(0.405)	(0.500)
50-59% Observed	-1.689	0.751	-0.631
	(1.929)	(0.456)	(0.400)
60-69% Observed	0.311	0.323	0.591
	(3.856)	(0.774)	(0.375)
70-79% Observed	-1.261	-0.226	-0.333
	(1.459)	(0.732)	(0.360)
80-89% Observed	-0.589	3.097***	-0.407
	(0.865)	(0.887)	(0.303)
90-99% Observed	0.787*	-0.131	-0.746*
	(0.348)	(1.576)	(0.322)
100% Observed	0.569***	2.536	-0.724***
	(0.137)	(2.228)	(0.139)
Constant	2.689***	2.797***	3.261***
	(0.0724)	(0.0659)	(0.109)
Observations	4,109	4,109	4,109
AIC	22,763	22,765	22,740
BIC	22,832	22,834	22,809
R^2	0.006	0.006	0.012
Adjusted R^2	0.004	0.003	0.009
5-fold CV RMSE	3.855	3.869	3.848

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-0.0999	-2.288	0.568***
	(0.269)	(2.234)	(0.150)
Never Observed	-0.580***	-2.520	0.700***
	(0.138)	(2.231)	(0.144)
Constant	3.258***	5.333*	2.538***
	(0.117)	(2.230)	(0.0855)
Observations	4,109	4,109	4,109
AIC	22,752	22,768	22,743
BIC	22,771	22,787	22,762
R^2	0.005	0.001	0.007
Adjusted R^2	0.004	0.000	0.006
5-fold CV RMSE	3.850	3.859	3.849

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C30. Men's School Discipline Regression Analyses by Specifications of Observed Race

	1a Black	1b Other	1c White
	Continuous Variable		
Percent Observed Black	0.0815***		
	(0.00615)		
Percent Observed Other		-0.0673***	
		(0.0193)	
Percent Observed White			-0.0753***
			(0.00623)
Constant	-0.924***	-0.667***	-0.194***
	(0.0330)	(0.0288)	(0.0492)
Observations	6,091	6,091	6,091
AIC	7,570	7,731	7,598
BIC	7,583	7,744	7,612
Pseudo R^2	0.022	0.002	0.019
5-fold CV RMSE	0.464	0.470	0.465

	2a Black	2b Other	2c White
	One Categorical Variable		
0-9% Observed (*reference*)			
10-19% Observed	-0.181	-0.00911	-0.416
	(0.410)	(0.142)	(0.273)
20-29% Observed	0.00156	-0.382*	-0.894**
	(0.837)	(0.174)	(0.316)
30-39% Observed	-	-0.466*	-1.336**
	-	(0.209)	(0.456)
40-49% Observed	0.225	-0.527*	-0.563*
	(1.225)	(0.235)	(0.255)
50-59% Observed	0.00156	-0.172	-1.105***
	(0.837)	(0.237)	(0.233)
60-69% Observed	-	-0.434	-0.929***
	-	(0.366)	(0.203)
70-79% Observed	-0.181	-0.0693	-0.908***
	(0.817)	(0.352)	(0.191)
80-89% Observed	0.710	-0.116	-0.550***
	(0.375)	(0.365)	(0.146)
90-99% Observed	0.704***	0.105	-0.782***
	(0.173)	(0.627)	(0.168)
100% Observed	0.823***	-	-0.788***
	(0.0636)	-	(0.0633)
Constant	-0.918***	-0.665***	-0.130*
	(0.0330)	(0.0292)	(0.0511)
Observations	6,080	6,081	6,091
AIC	7,572	7,735	7,586
BIC	7,632	7,802	7,660
Pseudo R^2	0.023	0.002	0.023
5-fold CV RMSE	0.464	0.471	0.465

	3a Black	3b Other	3c White
	One Categorical Variable		
Always Observed (*reference*)			
Ever Observed	-0.475***	-0.275***	0.103
	(0.128)	(0.0760)	(0.0731)
Never Observed	-0.825***	-	0.796***
	(0.0638)	-	(0.0654)
Constant	-0.0946	-0.655***	-0.918***
	(0.0544)	(0.0296)	(0.0374)
Observations	6,091	6,081	6,091
AIC	7,579	7,722	7,594
BIC	7,599	7,736	7,614
Pseudo R^2	0.022	0.002	0.020
5-fold CV RMSE	0.464	0.471	0.465

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix C Section 4. Screener Models

Appendix Table C31. Women's Wage Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
One Categorical Variable		Series of Dichotomous Variables		
Reference Category: Black		Hispanic	0.00358	
Hispanic	0.105***		(0.0225)	
	(0.0256)	Black		-0.147***
Non-Black, non-Hispanic	0.163***			(0.0190)
	(0.0201)			
Constant	2.473***	Constant	2.574***	2.620***
	(0.0159)		(0.00981)	(0.0105)
Observations	3,680	Observations	3,680	3,680
AIC	5,782	AIC	5,846	5,786
BIC	5,801	BIC	5,858	5,799
R^2	0.018	R^2	0.000	0.016
Adjusted R^2	0.017	Adjusted R^2	0.000	0.016
5-fold CV RMSE	0.531	5-fold CV RMSE	0.535	0.531

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C32. Women's Salary Regression Analyses by Specifications of Screener 1978

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	0.0734
	(0.0388)
Non-Black, non-Hispanic	0.108***
	(0.0305)
Constant	10.00***
	(0.0242)
Observations	3,554
AIC	8,403
BIC	8,422
R^2	0.004
Adjusted R^2	0.003
5-fold CV RMSE	0.789

	1b	1c
	Series of Dichotomous Variables	
Hispanic	0.00503	
	(0.0338)	
Black		-0.0987***
		(0.0289)
Constant	10.07***	10.10***
	(0.0147)	(0.0158)
Observations	3,554	3,554
AIC	8,414	8,402
BIC	8,426	8,415
R^2	0.000	0.003
Adjusted R^2	0.000	0.003
5-fold CV RMSE	0.790	0.789

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C33. Women's Unemployment Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
	One Categorical Variable		Series of Dichotomous Variables	
Reference Category: Black		Hispanic	0.605*	
Hispanic	-1.246***		(0.281)	
	(0.327)	Black		2.335***
Non-Black, non-Hispanic	-2.628***			(0.236)
	(0.243)			
Constant	4.030***	Constant	2.179***	1.695***
	(0.204)		(0.112)	(0.118)
Observations	6,283	Observations	6,283	6,283
AIC	44,071	AIC	44,184	44,092
BIC	44,091	BIC	44,198	44,105
R^2	0.019	R^2	0.001	0.015
Adjusted R^2	0.019	Adjusted R^2	0.001	0.015
5-fold CV RMSE	8.070	5-fold CV RMSE	8.107	8.054

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C34. Women's Depression Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
	One Categorical Variable		Series of Dichotomous Variables	
Reference Category: Black		Hispanic	-0.188	
Hispanic	-0.652**		(0.174)	
	(0.199)	Black		0.721***
Non-Black, non-Hispanic	-0.748***			(0.149)
	(0.158)			
Constant	4.362***	Constant	3.898***	3.641***
	(0.124)		(0.0766)	(0.0824)
Observations	4,255	Observations	4,255	4,255
AIC	24,835	AIC	24,855	24,833
BIC	24,854	BIC	24,868	24,846
R^2	0.006	R^2	0.000	0.005
Adjusted R^2	0.005	Adjusted R^2	0.000	0.005
5-fold CV RMSE	4.478	5-fold CV RM	4.489	4.476

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C35. Women's School Discipline Regression Analyses by Specifications of Screener 1978

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	-0.723***
	(0.104)
Non-Black, non-Hispanic	-0.883***
	(0.0744)
Constant	-0.914***
	(0.0569)
Observations	6,049
AIC	5,593
BIC	5,613
Pseudo R^2	0.025
5-fold CV RMSE	0.381

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.153	
	(0.0947)	
Black		0.848***
		(0.0707)
Constant	-1.484***	-1.762***
	(0.0361)	(0.0420)
Observations	6,049	6,049
AIC	5,729	5,594
BIC	5,742	5,607
Pseudo R^2	0.000	0.024
5-fold CV RMSI	0.385	0.381

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C36. Men's Wage Regression Analyses by Specifications of Screener 1978

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	0.175***
	(0.0248)
Non-Black, non-Hispanic	0.343***
	(0.0197)
Constant	2.589***
	(0.0156)
Observations	3,726
AIC	5,731
BIC	5,750
R^2	0.076
Adjusted R^2	0.075
5-fold CV RMSE	0.522

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.0394	
	(0.0224)	
Black		-0.296***
		(0.0188)
Constant	2.804***	2.885***
	(0.00992)	(0.0103)
Observations	3,726	3,726
AIC	6,020	5,783
BIC	6,033	5,796
R^2	0.001	0.062
Adjusted R^2	0.001	0.062
5-fold CV RM	0.543	0.525

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C37. Men's Salary Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
	One Categorical Variable		Series of Dichotomous Variables	
Reference Category: Black		Hispanic	-0.0811*	
Hispanic	0.243***		(0.0336)	
	(0.0374)	Black		-0.435***
Non-Black, non-Hispanic	0.506***			(0.0282)
	(0.0295)			
Constant	10.24***	Constant	10.56***	10.67***
	(0.0236)		(0.0147)	(0.0152)
Observations	3,649	Observations	3,649	3,649
AIC	8,431	AIC	8,713	8,489
BIC	8,450	BIC	8,726	8,501
R^2	0.076	R^2	0.002	0.061
Adjusted R^2	0.076	Adjusted R^2	0.001	0.061
5-fold CV RMSE	0.768	5-fold CV RM	0.799	0.775

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C38. Men's Unemployment Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
	One Categorical Variable		Series of Dichotomous Variables	
Reference Category: Black		Hispanic	0.908**	
Hispanic	-1.410***		(0.320)	
	(0.370)	Black		2.910***
Non-Black, non-Hispanic	-3.305***			(0.265)
	(0.273)			
Constant	4.838***	Constant	2.520***	1.928***
	(0.229)		(0.126)	(0.133)
Observations	6,403	Observations	6,403	6,403
AIC	46,576	AIC	46,719	46,607
BIC	46,596	BIC	46,732	46,621
R^2	0.024	R^2	0.001	0.018
Adjusted R^2	0.023	Adjusted R^2	0.001	0.018
5-fold CV RMSE	9.175	5-fold CV RM	9.260	9.201

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C39. Men's Depression Regression Analyses by Specifications of Screener 1978

	1a		1b	1c
	One Categorical Variable		Series of Dichotomous Variables	
Reference Category: Black		Hispanic	0.144	
Hispanic	-0.268		(0.151)	
	(0.174)	Black		0.552***
Non-Black, non-Hispanic	-0.665***			(0.131)
	(0.138)			
Constant	3.245***	Constant	2.833***	2.693***
	(0.109)		(0.0673)	(0.0721)
Observations	4,109	Observations	4,109	4,109
AIC	22,748	AIC	22,769	22,752
BIC	22,767	BIC	22,781	22,764
R^2	0.006	R^2	0.000	0.004
Adjusted R^2	0.005	Adjusted R^2	0.000	0.004
5-fold CV RMSE	3.850	5-fold CV RM	3.867	3.855

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)

Appendix Table C40. Men's School Discipline Regression Analyses by Specifications of Screener 1978

	1a
	One Categorical Variable
Reference Category: Black	
Hispanic	-0.773***
	(0.0885)
Non-Black, non-Hispanic	-0.795***
	(0.0630)
Constant	-0.127*
	(0.0511)
Observations	6,091
AIC	7,578
BIC	7,598
Pseudo R^2	0.022
5-fold CV RMSE	0.464

	1b	1c
	Series of Dichotomous Variables	
Hispanic	-0.234**	
	(0.0781)	
Black		0.790***
		(0.0607)
Constant	-0.666***	-0.918***
	(0.0294)	(0.0328)
Observations	6,091	6,091
AIC	7,734	7,576
BIC	7,748	7,589
Pseudo R^2	0.001	0.022
5-fold CV RMSE	0.471	0.464

*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed tests)