www.sciencemag.org/content/360/3691/861/suppl/DC1

Supplementary Materials for

Disparities in science literacy

Nick Allum*, John Besley, Louis Gomez, Ian Brunton-Smith

*Corresponding author. Email: nallum@essex.ac.uk

Published 25 May 2018, *Science* **360**, 861 (2018)
DOI: 10.1126/science.aar8480

This PDF file includes:

Materials and Methods

Experimental design

This study uses data from the General Social Survey (GSS) spanning the years 2006-2016 to examine racial and ethnic disparities in science literacy. Combining multiple rounds of the survey is necessary to ensure a sufficient sample size to examine differences based on racial and ethnic groups. Our key objectives are to first describe disparities in science literacy between ethnic and racial groups, and secondly to use multivariate analyses to examine how these disparities are affected by accounting for potential confounders and relevant explanatory pathways.

GSS sample

The GSS uses primarily face-to-face data collection and achieves response rates of around 70%. The GSS presents the Science & Technology module to a subset of respondents on a biennial basis. Our analysis uses a pooled sample including all respondents presented with this module in the six waves fielded between 2006 and 2016 (available from: https://gssdataexplorer.norc.org/pages/show?page=gss%2Fgss_data). This yields analytic sample sizes of between two and six thousand respondents depending on the level of item non-response and the number of sample members asked each of the items.

Knowledge scale

The underlying questions were designed to capture both factual and procedural knowledge (including probability) (3) but research suggests combining all of the items together (*1*). The 13-item measure used here included 11 multiple choice questions, an open-ended question related to experimental design, and a self-report item related to perceived understanding of what it means to study something scientifically. The scale was computed by summing the number of correct responses for each respondent. "Don't know" responses and refusals to respond counted as incorrect.

Multiple choice questions

- *The center of the Earth is very hot.* (True)
- *All radioactivity is man-made.* (False)
- *It is the father's gene that decides whether the baby is a boy or a girl.* (True) or (in 2008) *It is the mother's gene that decides whether the baby is a boy or a girl.* (False) (Split ballot in 2008; 1,506 survey respondents were asked about "father's gene"; 515 survey respondents were asked about "mother's gene.")
- *Lasers work by focusing sound waves.* (False)
- *Electrons are smaller than atoms.* (True)
- *Antibiotics kill viruses as well as bacteria.* (False)
- *The continents on which we live have been moving their locations for millions of years and will continue to move in the future.* (True)
- *Does the Earth go around the Sun, or does the Sun go around the Earth?* (Earth around Sun)
- *How long does it take for the Earth to go around the Sun?* (One year) (Asked only if the respondent answered correctly that the Earth goes around the Sun.)

To be classified as understanding probability, the survey respondent had to answer correctly *A doctor tells a couple that their genetic makeup means that they've got one in four chances of having a child with an inherited illness.* (1) *Does this mean that if their first child has the illness, the next three will not have the illness?* (No); and (2) *Does this mean that each of the couple's children will have the same risk of suffering from the illness?* (Yes).

To be classified as understanding experiment, the survey respondent had to answer correctly (1) *Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1,000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?* and (2) *Why is it better to test the drug this way?* (The second way because a control group is used for comparison).

Open-ended question

To be classified as understanding scientific study, the survey respondent had to answer correctly (1) *When you read news stories, you see certain sets of words and terms. We are interested in how many people recognize certain kinds of terms. First, some articles refer to the results of a scientific study. When you read or hear the term scientific study, do you have a clear understanding of what it means, a general sense of what it means, or little understanding of what it means?* and (2) (If "clear understanding" or "general sense" response) *In your own words, could you tell me what it means to study something scientifically?* (Formulation of theories/test hypothesis, experiments/control group, or rigorous/systematic comparison).

The Cambridge Social Interaction and Stratification scale

We attach details from the Cambridge Social Interaction and Stratification (CAMSIS) scale score to each GSS respondent. CAMSIS scores are interpreted as representing an occupational unit's relative position within the national order of social interaction and stratification (or, more precisely, the average relative position held by the incumbents of the occupational unit). Members of groups that are socially more similar will tend to be more likely to interact socially than are members of groups that are socially less similar. Differences between groups in the relative frequencies of social interaction can be treated as reflecting the social distances between them. These relationships can be represented by a two-dimensional table, where the rows indicate the range of one partner's jobs, the columns the range of the other partner's jobs, and the cell frequencies the number of occurrences of each combination in the population. This table can be analysed to see if the distances are consistent with location in a social space of a limited number of dimensions. In particular, one would expect to find a major dimension relating to social inequality and stratification. Correspondence Analysis is used to reduce the dimensionality of partner occupation combinations and continuous score for each

occupational group can be estimated. This continuous score represents that occupation's location within the national stratification order. (see http://www.camsis.stir.ac.uk)

CAMSIS scores also take into account an individuals' current status in employment, and recognizes that the same occupation may have different social connections for female employees when compared to male employees. CAMSIS are therefore matched to all GSS respondents based on their census occupation code (2010), status in employment (distinguishing between self-employed, employees, and others), and gender.

Statistical Analysis

To correctly account for the multi-stage cluster sampling design adopted by the GSS, all reported analyses are estimated using a robust standard error adjustment. This incorporates information on the first- and second-stage sampling units used (identified using the variables VSTRAT and VPSU included in the GSS datafile). All estimates have also been weighted to correct for nonresponse (using GSS weight WTSSNR).

Decomposition analysis

We use the Blinder-Oaxaca method (*2*) to decompose the disparity in mean knowledge scores for blacks and Hispanics compared to whites. This procedure partitions the gap into that which is 'explained', i.e. attributable to inter-group differences in levels of our measured characteristics and that which is 'unexplained', by which we mean differences in group-specific 'returns' to these characteristics, along with any unobserved influences.

Supplementary Text

Figure S1 shows estimates for all covariates included in the multivariate models. Table S1 includes annual estimates of science knowledge for each ethnic group and associated robust standard errors, used to produce figure 1. Table S2 includes full model estimates (with robust standard errors) for the three models reported in Figure 2. For comparison, tables S3 and S4 include model estimates unweighted estimates assuming a simple random sample, based on complete cases and using listwise delete respectively. Our substantive conclusions remain unchanged in all cases.

Figure S1 (all covariates)

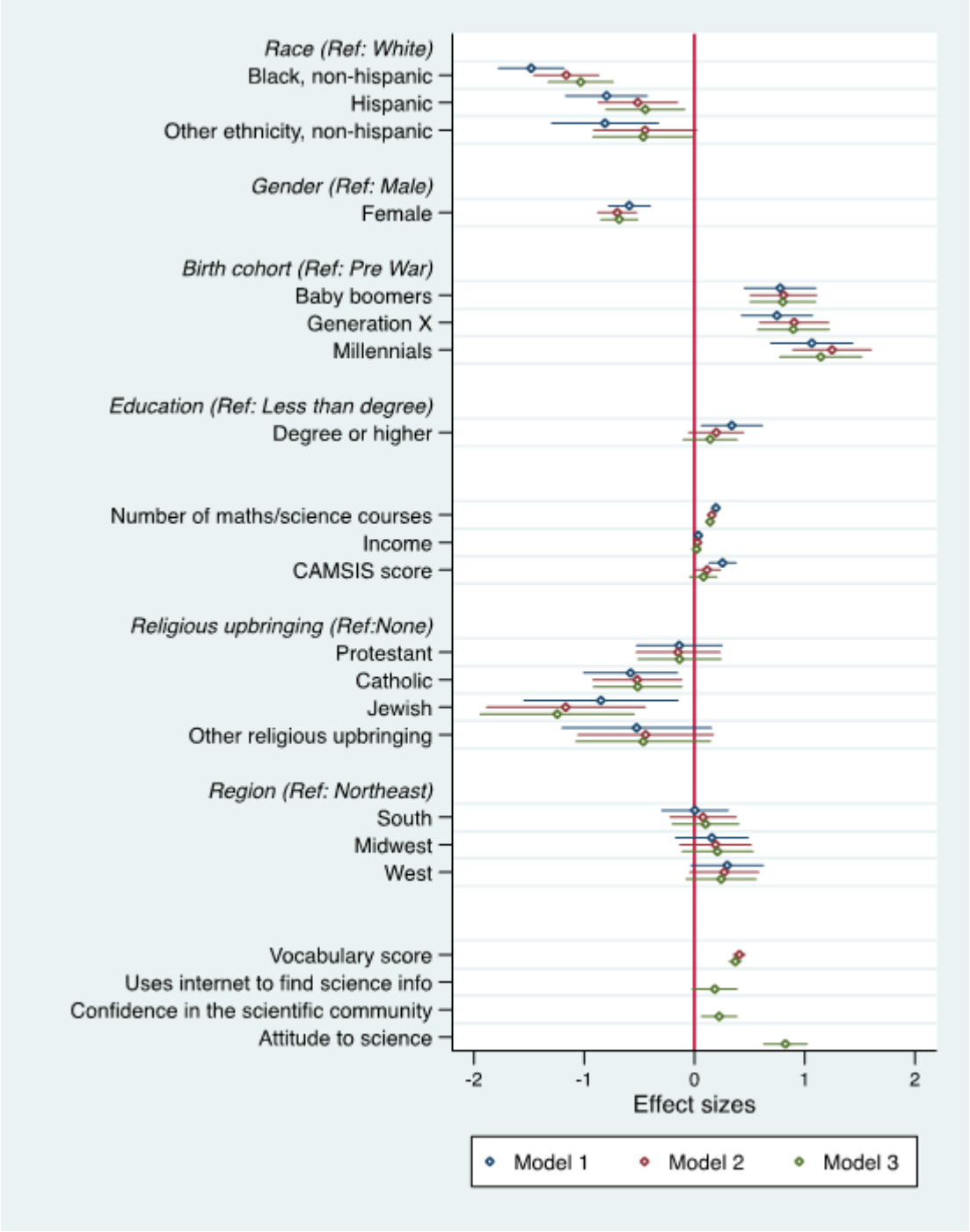

Table S1.

Annual science knowledge estimates by ethnicity (for figure 1)

(Robust standard errors)		**Mean**	**S.E**	**2.5%**	**97.5%**
White					
	2006	8.586	0.099	8.392	8.780
	2008	8.689	0.124	8.445	8.933
	2010	8.776	0.178	8.426	9.126
	2012	8.614	0.128	8.363	8.866
	2014	8.526	0.114	8.303	8.749
	2016	8.536	0.096	8.348	8.725
	Total	8.607	0.048	8.512	8.703
Black, non-Hispanic					
	2006	6.391	0.168	6.061	6.722
	2008	6.137	0.276	5.595	6.679
	2010	6.460	0.474	5.530	7.390
	2012	6.274	0.188	5.904	6.643
	2014	6.586	0.217	6.160	7.013
	2016	6.831	0.205	6.428	7.234
	Total	6.466	0.096	6.277	6.654
Hispanic					
	2006	6.873	0.287	6.310	7.436
	2008	6.776	0.292	6.203	7.350
	2010	6.230	0.335	5.572	6.888
	2012	7.295	0.261	6.781	7.808
	2014	7.057	0.216	6.632	7.482
	2016	6.469	0.243	5.992	6.947
	Total	6.813	0.113	6.591	7.034
Other ethnicity, non-Hispanic					
	2006	7.762	0.476	6.828	8.697
	2008	7.999	0.390	7.233	8.765
	2010	8.622	0.635	7.376	9.869
	2012	8.421	0.385	7.664	9.177
	2014	7.836	0.414	7.023	8.649
	2016	8.404	0.319	7.779	9.030
	Total	8.101	0.182	7.745	8.458

Table S2.

Full regression estimates with robust standard errors (for figure 2)

	B	SE	B	SE	B	SE
Race (ref: White)						
Black, non-Hispanic	-1.481***	0.151	-1.162***	0.150	-1.031***	0.150
Hispanic	-0.797***	0.188	-0.513**	0.183	-0.446*	0.182
Other ethnicity, non-Hispanic	-0.811**	0.247	-0.447	0.239	-0.465*	0.232
Female	-0.590***	0.097	-0.700***	0.090	-0.682***	0.086
Generation (ref: pre-War)						
Baby boomers	0.778***	0.166	0.809***	0.153	0.803***	0.152
Generation X	0.749***	0.164	0.905***	0.16	0.898***	0.165
Millenials	1.065***	0.190	1.247***	0.181	1.146***	0.189
Degree or higher	0.340*	0.141	0.198	0.127	0.144	0.125
Number of maths/science courses	0.195***	0.013	0.159***	0.013	0.141***	0.012
Income	0.036***	0.010	0.028**	0.010	0.021*	0.010
CAMSIS score	0.255***	0.063	0.115	0.062	0.082	0.063
Religious upbringing (ref: None)						
Protestant upbringing	-0.137	0.197	-0.147	0.194	-0.134	0.192
Catholic upbringing	-0.579**	0.217	-0.519*	0.205	-0.514*	0.205
Jewish upbringing	-0.848*	0.355	-1.167**	0.366	-1.244***	0.355
Other religious upbringing	-0.523	0.345	-0.441	0.313	-0.464	0.311
Region (Ref: Northeast)						
South	0.006	0.154	0.079	0.153	0.101	0.154
Midwest	0.158	0.168	0.191	0.165	0.210	0.163
West	0.298	0.168	0.271	0.159	0.243	0.162
Vocabulary score			0.407***	0.028	0.371***	0.028
Uses internet to find science info					0.184	0.104
Confidence in scientific community					0.225**	0.083
Attitude to science					0.825***	0.100
Constant	6.483***	0.307	4.193***	0.335	2.290***	0.390
N	2339		2339		2339	

*p<.05, *p<.01, ***p<.001

Table S3.

Full regression estimates using complete cases for each model to preserve maximum sample size

	B	**SE**	**B**	**SE**	**B**	**SE**
Race (ref: White)						
Black, non-Hispanic	-1.612***	0.086	-1.283***	0.103	-1.015***	0.128
Hispanic	-1.176***	0.104	-0.707***	0.123	-0.385*	0.15
Other ethnicity, non-Hispanic	-1.036***	0.145	-0.516**	0.175	-0.365	0.212
Female	-0.504***	0.055	-0.693***	0.0653	-0.696***	0.080
Generation (ref: pre-War)						
Baby boomers	0.835***	0.084	0.981***	0.100	0.960***	0.123
Generation X	0.960***	0.089	1.190***	0.107	1.076***	0.133
Millenials	1.180***	0.101	1.450***	0.119	1.263***	0.153
Degree or higher	0.599***	0.079	0.349***	0.0924	0.243*	0.114
Number of maths/science courses	0.194***	0.008	0.158***	0.00922	0.130***	0.0112
Income	0.024***	0.005	0.018**	0.00645	0.0208*	0.008
CAMSIS score	0.266***	0.036	0.144***	0.0428	0.112*	0.053
Religious upbringing (ref: None)						
Protestant upbringing	-0.030	0.106	0.024	0.123	-0.026	0.150
Catholic upbringing	-0.249*	0.111	-0.282*	0.129	-0.443**	0.157
Jewish upbringing	-0.561**	0.213	-0.647*	0.261	-1.119***	0.330
Other religious upbringing	-0.168	0.172	-0.197	0.207	-0.378	0.256
Region (Ref: Northeast)						
South	-0.104	0.084	0.030	0.099	0.042	0.122
Midwest	0.205*	0.089	0.241*	0.105	0.207	0.129
West	0.427***	0.092	0.350**	0.107	0.252	0.129
Vocabulary score			0.418***	0.020	0.386***	0.025
Uses internet to find science info					0.268**	0.091
Confidence in scientific community					0.270***	0.072
Attitude to science					0.716***	0.095
Constant	6.289***	0.169	3.847***	0.229	2.068***	0.337
N	5808		3667		2339	

*p<.05, *p<.01, ***p<.001

Table S4.

Full regression estimates using listwise delete

	B	SE	B	SE	B	SE
Race (ref: White)						
Black, non-Hispanic	-1.480***	0.137	-1.140***	0.131	-1.015***	0.128
Hispanic	-0.853***	0.161	-0.482**	0.153	-0.385*	0.150
Other ethnicity, non-Hispanic	-0.780***	0.227	-0.370	0.216	-0.365	0.212
Female	-0.630***	0.086	-0.735***	0.082	-0.696***	0.080
Generation (ref: pre-War)						
Baby boomers	0.985***	0.131	1.002***	0.123	0.960***	0.123
Generation X	1.003***	0.139	1.152***	0.131	1.076***	0.133
Millenials	1.278***	0.158	1.427***	0.150	1.263***	0.153
Degree or higher	0.499***	0.122	0.309**	0.116	0.243*	0.114
Number of maths/science courses	0.182***	0.012	0.147***	0.011	0.130***	0.011
Income	0.036***	0.009	0.027**	0.008	0.021*	0.008
CAMSIS score	0.286***	0.057	0.147**	0.054	0.112*	0.053
Religious upbringing (ref: None)						
Protestant upbringing	-0.054	0.162	-0.032	0.153	-0.026	0.150
Catholic upbringing	-0.507**	0.169	-0.416**	0.160	-0.443**	0.157
Jewish upbringing	-0.850*	0.356	-1.066**	0.337	-1.119***	0.330
Other religious upbringing	-0.502	0.277	-0.378	0.262	-0.378	0.256
Region (Ref: Northeast)						
South	-0.0272	0.131	0.044	0.124	0.042	0.122
Midwest	0.214	0.139	0.211	0.132	0.207	0.129
West	0.364**	0.139	0.311*	0.131	0.252	0.129
Vocabulary score			0.417***	0.025	0.386***	0.025
Uses internet to find science info					0.268**	0.091
Confidence in scientific community					0.270***	0.072
Attitude to science					0.716***	0.095
Constant	6.238***	0.262	3.895***	0.284	2.068***	0.337
N	2339		2339		2339	

*p<.05, *p<.01, ***p<.001

Table S5.

Oaxaca-Blinder decompositions comparing whites and Blacks, and whites and Hispanics (for model 1)

	Whites vs Blacks		Whites vs Hispanics	
	B	**S.E**	**B**	**S.E**
Whites	8.754**	0.025	8.755**	0.025
Non-whites	6.720**	0.062	7.168**	0.061
Difference	2.034*	0.086	1.587*	0.085
Decomposition				
% Explained	21%		23%	
% Unexplained	79%		77%	
Explained				
Female	0.002	0.000	-0.027	0.011
Generation (ref: pre-War)				
Baby boomers	0.031	0.022	0.120	0.041
Generation X	-0.037	0.011	-0.077	0.021
Millenials	-0.102	0.020	-0.182	0.024
Degree or higher	0.089	0.013	0.115	0.013
Number of maths/science courses	0.213	0.037	0.249*	0.018
Income	0.063	0.007	0.037	0.011
CAMSIS score	0.128	0.026	0.105*	0.002
Religious upbringing (ref: None)				
Protestant upbringing	0.010	0.031	-0.029	0.065
Catholic upbringing	-0.042	0.061	0.161	0.159
Jewish upbringing	-0.019	0.002	-0.020	0.002
Other religious upbringing	0.009	0.006	-0.002	0.004
Region (Ref: Northeast)				
South	0.027	0.005	-0.001	0.002
Midwest	0.017	0.008	0.037	0.007
West	0.048	0.004	-0.126	0.021
Total	0.437	0.080	0.358	0.079
Unexplained				
Female	0.189	0.114	0.336	0.147
Generation (ref: pre-War)				
Baby boomers	-0.025	0.078	0.364	0.357
Generation X	-0.099	0.013	0.455	0.382
Millenials	-0.074*	0.002	0.517	0.339
Degree or higher	-0.105	0.080	-0.084	0.064
Number of maths/science courses	0.105	0.022	-0.179	0.055
Income	0.005	0.387	0.367	0.103
CAMSIS score	-0.009	0.017	-0.034	0.019
Religious upbringing (ref: None)				
Protestant upbringing	0.053	0.550	0.024	0.087
Catholic upbringing	0.017	0.090	0.840	0.382
Jewish upbringing	0.002	0.004	0.002	0.002
Other religious upbringing	0.001	0.079	-0.016	0.009

Region (Ref: Northeast)				
South	0.182	0.219	0.056	0.150
Midwest	0.091	0.086	0.012	0.023
West	0.020	0.106	-0.098	0.177
Constant	1.243	1.801	-1.334	0.339
Total	1.598*	0.112	1.229*	0.038

*p<.05, *p<.01, ***p<.001

Table S6.

Oaxaca-Blinder decompositions comparing whites and Blacks, and whites and Hispanics (for model 2)

	Whites vs Blacks		Whites vs Hispanics	
	B	**S.E**	**B**	**S.E**
Whites	8.714**	0.043	8.714**	0.043
Non-whites	6.827**	0.043	7.287**	0.106
Difference	1.888*	0.061	1.427	0.145
Decomposition				
% Explained	35%		40%	
% Unexplained	65%		60%	
Explained				
Female	0.010	0.003	-0.050	0.019
Generation (ref: pre-War)				
Baby boomers	0.052	0.012	0.129	0.047
Generation X	-0.069	0.017	-0.113	0.027
Millenials	-0.120*	0.006	-0.221	0.044
Degree or higher	0.035*	0.002	0.050	0.007
Number of maths/science courses	0.169	0.029	0.156	0.060
Income	0.062	0.043	0.022	0.029
CAMSIS score	0.065	0.021	0.053	0.016
Religious upbringing (ref: None)				
Protestant upbringing	0.012	0.026	-0.013	0.023
Catholic upbringing	-0.053	0.071	0.157	0.130
Jewish upbringing	-0.017	0.008	-0.018	0.008
Other religious upbringing	0.006	0.003	-0.004	0.005
Region (Ref: Northeast)				
South	0.001	0.040	-0.003	0.006
Midwest	0.017	0.008	0.042	0.015
West	0.036	0.015	-0.100	0.021
Vocabulary score	0.457*	0.017	0.479*	0.020
Total	0.664	0.074	0.565	0.113
Unexplained				
Female	0.257	0.117	0.187	0.255
Generation (ref: pre-War)				
Baby boomers	-0.138	0.104	0.423	0.194
Generation X	-0.008	0.069	0.576	0.250
Millenials	0.040	0.093	0.628	0.215
Degree or higher	-0.072	0.056	-0.185	0.033
Number of maths/science courses	-0.369	0.176	-0.080	0.157
Income	0.043	0.229	0.806	0.225
CAMSIS score	-0.056	0.022	-0.031	0.024
Religious upbringing (ref: None)				
Protestant upbringing	0.381	0.091	-0.018	0.111
Catholic upbringing	0.058	0.035	0.778	0.523
Jewish upbringing	0.003	0.003	0.002	0.002

Other religious upbringing	0.012	0.029	-0.011	0.022
Region (Ref: Northeast)				
South	0.065	0.151	0.112	0.196
Midwest	0.129	0.020	0.030	0.047
West	0.013	0.068	-0.070	0.408
Vocabulary score	0.311	0.951	0.219	0.182
Constant	0.553	1.350	-2.503	0.951
Total	1.224*	0.054	0.862*	0.042

*p<.05, *p<.01, ***p<.001

Table S7.

Oaxaca-Blinder decompositions comparing whites and Blacks, and whites and Hispanics (for model 3)

	Whites vs Blacks		Whites vs Hispanics	
	B	**S.E**	**B**	**S.E**
Whites	8.735**	0.037	8.735**	0.037
Non-whites	6.826**	0.107	7.587*	0.240
Difference	1.909*	0.125	1.149	0.250
Decomposition				
% Explained	46%		61%	
% Unexplained	54%		39%	
Explained				
Female	0.042	0.009	-0.021	0.007
Generation (ref: pre-War)				
Baby boomers	0.043	0.024	0.100	0.026
Generation X	-0.049	0.016	-0.096	0.022
Millenials	-0.109	0.054	-0.158	0.030
Degree or higher	0.016	0.011	0.023	0.009
Number of maths/science courses	0.162	0.021	0.099	0.025
Income	0.095*	0.007	0.043	0.017
CAMSIS score	0.041	0.049	0.036	0.048
Religious upbringing (ref: None)				
Protestant upbringing	0.041	0.079	-0.044	0.114
Catholic upbringing	-0.102	0.091	0.219	0.178
Jewish upbringing	-0.030	0.007	-0.033	0.008
Other religious upbringing	0.016	0.024	0.003	0.009
Region (Ref: Northeast)				
South	-0.028	0.047	0.002	0.008
Midwest	0.016	0.007	0.036	0.008
West	0.029	0.017	-0.078	0.034
Vocabulary score	0.450*	0.027	0.427	0.079
Uses internet to find science info	0.012	0.006	0.000	0.002
Confidence in scientific community	0.041	0.005	-0.002	0.001
Attitude to science	0.199	0.073	0.144	0.039
Total	0.884	0.089	0.698	0.112
Unexplained				
Female	0.314	0.136	0.276	0.025
Generation (ref: pre-War)				
Baby boomers	-0.158	0.052	0.423	0.291
Generation X	-0.093*	0.005	0.606	0.400
Millenials	0.040	0.024	0.587	0.325
Degree or higher	-0.103	0.032	-0.213	0.038
Number of maths/science courses	-0.186	0.108	0.133	0.274
Income	0.010	0.380	0.691	0.198
CAMSIS score	-0.057	0.020	-0.040	0.025
Religious upbringing (ref: None)				

Protestant upbringing	0.216	0.460	-0.135	0.121
Catholic upbringing	0.010	0.097	0.174	0.589
Jewish upbringing	0.002	0.006	-0.001	0.002
Other religious upbringing	-0.005	0.070	-0.031	0.069
Region (Ref: Northeast)				
South	-0.126	0.086	-0.027	0.064
Midwest	0.061	0.046	0.004	0.098
West	-0.030	0.099	-0.300	0.313
Vocabulary score	0.178	0.735	0.098	0.138
Uses internet to find science info	-0.006	0.013	-0.222	0.381
Confidence in scientific community	-0.648	0.332	-0.800	1.407
Attitude to science	1.112	0.776	0.496	0.919
Constant	0.493	1.294	-1.270	2.569
Total	1.025	0.045	0.451	0.218

*p<.05, *p<.01, ***p<.001

References

1. D. M. Kahan, 'Ordinary science intelligence': A science-comprehension measure for study of risk and science communication, with notes on evolution and climate change. *J. Risk Res.* **20**, 995–1016 (2017). doi:10.1080/13669877.2016.1148067
2. O. O'Donnell, E. Van Doorslaer, A. Wagstaff, M. Lindelow, *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation* (World Bank Publications, 2008).

Code for replication from GSS public files (2006 – 2016)

```
***SPSS syntax for generation of full database***

//FREQUENCIES WTSSNR VSTRAT VPSU.
//FREQUENCIES wordsum.

//FREQUENCIES sex.
//RECODE sex (1=0)(2=1) INTO female.
//FREQUENCIES female.

//FREQUENCIES degree.

*For income, use the following has constant dollar income,
though many missing values..
//FREQUENCIES CONINC.

*RECONSTRUCTIING THE RACE VARIALBES.
//FREQUENCIES race hispanic.

//RECODE hispanic (1=0)(2 thru 50=100)(ELSE=100) INTO
hispanic100.
//FREQUENCIES hispanic100.

//COMPUTE blackhispanic = SUM.2(race, hispanic100).
//FREQUENCIES blackhispanic.
//RECODE blackhispanic (1=1)(2=2)(3=4)(101 thru 103=3) INTO
race_eth.
*note that the n = 24 who said DK or NA on the hispanic
question are not marked as hispanic.

//VALUE LABELS race_eth
//1 'white, non-hispanic'
//2 'black, non-hispanic'
//3 'hispanic'
//4 'other'.

//FREQUENCIES race_eth.

//RECODE race_eth (1=1)(ELSE=0) INTO white_nonH.
//RECODE race_eth (2=1)(ELSE=0) INTO black_nonH.
//RECODE race_eth (3=1)(ELSE=0) INTO hispanic_r.
//RECODE race_eth (4=1)(ELSE=0) INTO other_r.

//FREQUENCIES white_nonH black_nonH hispanic2 race_other.
```

```
*RECONSTRUCTING THE COHORT VARIABLES.

//USE ALL.
//COMPUTE filter_$=( year = 2006).
//VARIABLE LABELS filter_$ ' year = 2006 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//COMPUTE filter_$=( year = 2008).
//VARIABLE LABELS filter_$ ' year = 2008 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//COMPUTE filter_$=( year = 2010).
//VARIABLE LABELS filter_$ ' year = 2010 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//COMPUTE filter_$=( year = 2012).
//VARIABLE LABELS filter_$ ' year = 2012 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//COMPUTE filter_$=( year = 2014).
//VARIABLE LABELS filter_$ ' year = 2014 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
```

```
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//COMPUTE filter_$=( year = 2016).
//VARIABLE LABELS filter_$ ' year = 2016 (FILTER)'.
//VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
//FORMATS filter_$ (f1.0).
//FILTER BY filter_$.

//COMPUTE yearborn = year-age.
//FREQUENCIES yearborn.

//USE ALL.
//FREQUENCIES yearborn.

//RECODE yearborn (1860 thru 1945=1)(ELSE=0) INTO greatest.
//RECODE yearborn (1946 thru 1964=1)(ELSE=0) INTO boomers.
//RECODE yearborn (1965 thru 1980=1)(ELSE=0) INTO genx.
//RECODE yearborn (1981 thru 1996=1)(ELSE=0) INTO geny.

//FREQUENCIES greatest boomers genx geny.


*INTERNET.
//FREQUENCIES seeksci scifrom.
//RECODE seeksci (3=1)(ELSE=0) INTO seek_int.
//RECODE scifrom (3=1)(ELSE=0) INTO scifrom_int.
//FREQUENCIES seek_int scifrom_int.
//COMPUTE internet_sciX = SUM( seek_int, scifrom_int).
//FREQUENCIES internet_sciX.
//RECODE internet_sciX (0=0)(1 thru 2=1) INTO internet_sci.
//FREQUENCIES internet_sci.

*ATTITUDES.

*Be careful with consci as it doesn't overlap with other
science questions in many years.
//FREQUENCIES consci.
//RECODE consci (1=3)(2=2)(3=1)(ELSE=SYSIS) INTO consci_r.
//FREQUENCIES consci_r.


*SCIENCE KNOWLEDGE.
```

```
//FREQUENCIES VARIABLES=hotcore radioact boyorgrl lasers
electron viruses condrift earthsun solarrev
  //STATISTICS=STDDEV SEMEAN MEAN
  //ORDER=ANALYSIS.

//RECODE hotcore boyorgrl electron condrift earthsun
(1=1)(2=0)(8=0)(9=0)(0=SYSMIS)
//INTO hotcore_cat boyorgrl_cat electron_cat condrift_cat
earthsun_cat.

//RECODE radioact lasers viruses
(1=0)(2=1)(8=0)(9=0)(0=SYSMIS) INTO radioact_cat lasers_cat
viruses_cat.

//FREQUENCIES earthsun solarrev.
//RECODE earthsun (1=1)(2=10)(ELSE=SYSMIS) INTO earthsun10.
//FREQUENCIES earthsun10.

//RECODE solarrev (3=1)(1=0)(2=0)(8=0)(9=0)(0=100) INTO
solarrev_cat100.
//FREQUENCIES solarrev_cat100.
//COMPUTE solar2 = earthsun10+solarrev_cat100.
//FREQUENCIES solar2.
//RECODE solar2 (1=0)(2=1)(110=0) INTO solarrev_cat.
//FREQUENCIES solarrev_cat.


//FREQUENCIES scistudy scitext expdesgn exptext odds1
odds2.
//RECODE scistudy (1=1)(2=0)(3=0)(8=0)(9=0)(ELSE=SYSMIS)
INTO scistudy_cat.
//RECODE exptext (1=1)(2=0)(3 thru 99=0)(ELSE=SYSMIS) INTO
exptext_cat.
*THE 2-CATEGORY COULD ALSO BE CODED FOR A .5 POINT.

//RECODE odds1 (1=0)(2=1)(8=0)(9=0)(ELSE=SYSMIS) INTO
odds1_cat.
//RECODE odds2 (1=1)(2=0)(8=0)(9=0)(ELSE=SYSMIS) INTO
odds2_cat.

//FREQUENCIES hotcore_cat boyorgrl_cat radioact_cat
lasers_cat electron_cat viruses_cat condrift_cat
earthsun_cat  solarrev_cat scistudy_cat exptext_cat
odds1_cat odds2_cat.
```

```
*.70 ... could be slightly better (.71) if boyorgrl_cat
were dropped.
//RELIABILITY
  //VARIABLES=hotcore_cat boyorgrl_cat radioact_cat
lasers_cat electron_cat viruses_cat condrift_cat
earthsun_cat  solarrev_cat scistudy_cat exptext_cat
odds1_cat odds2_cat
  //SCALE('ALL VARIABLES') ALL
  //MODEL=ALPHA
  //STATISTICS=DESCRIPTIVE SCALE CORR
  //SUMMARY=TOTAL MEANS.

//COMPUTE sciscore13 = SUM.13(hotcore_cat, boyorgrl_cat,
radioact_cat, lasers_cat, electron_cat, viruses_cat,
//condrift_cat, earthsun_cat, solarrev_cat, scistudy_cat,
exptext_cat, odds1_cat, odds2_cat).

//FREQUENCIES sciscore13.

*RECONSTRUCTING NUMMATHSCI.
//RECODE HSBIO HSCHEM  HSPHYS (1=1)(2=0)(8 thru
9=0)(ELSE=SYSMIS) INTO HSBIO_dich HSCHEM_dich  HSPHYS_dich.

//COMPUTE HSsci_cat = SUM.3 (HSBIO_dich, HSCHEM_dich,
HSPHYS_dich).

//RECODE COLSCI (1=0)(2=0)(8 thru 9=0)(ELSE=SYSMIS) INTO
colsci_dich.
//FREQUENCIES colsci_dich.

//FREQUENCIES COLSCINM.
//RECODE COLSCINM (1 thru 10=COPY)(10 thru
90=10)(ELSE=SYSMIS) INTO colsci10.
//FREQUENCIES colsci10.
//COMPUTE colsci_cat = SUM(colsci10, colsci_dich).
//FREQUENCIES colsci_cat.

//FREQUENCIES HSmath.
//RECODE HSmath (0 thru 2=0)(9 thru 10=0)(3=1)(5=2)(4=3)(6
thru 7=4)(8=5) INTO HSmath_cat.
//FREQUENCIES HSmath_cat.

//FREQUENCIES HSmath_cat, colsci_cat, hssci_cat.
//COMPUTE NumMathSci = SUM.3(HSmath_cat, colsci_cat,
hssci_cat).
//FREQUENCIES NumMathSci.
```

```
//FREQUENCIES year.
//FREQUENCIES year female degree white_nonH black_nonH
hispanic_r other_r greatest boomers genx millennials
consci_r internet_sci wordsum sciscore13 nummathsci.
//DESCRIPTIVES  female degree white_nonH black_nonH
hispanic_r other_r greatest boomers genx millennials
consci_r internet_sci sciscore13 wordsum nummathsci.

//FREQUENCIES  year.
//FREQUENCIES female degree white_nonH black_nonH
hispanic_r race_other greatest boomers genx geny consci_r
internet_sci wordsum sciscore13 nummathsci.
//DESCRIPTIVES female degree white_nonH black_nonH
hispanic_r race_other greatest boomers genx geny consci_r
internet_sci sciscore13 wordsum nummathsci.


//FREQUENCIES white_nonH black_nonH hispanic_r other_r.


//Scales – to be constructed as previously. I do not have
this syntax
//sciscore13 (science literacy score)
//NumMathSci (total number of maths courses)
//wordsum (foundational literacy)


//Other controls in original file
//ethnicity (Black_nonH Other_nonH Hispanic). SPSS syntax
refs race_eth
//sex - female
//generation? (boomers genx millennials), SPSS syntax
references birth_year
//education? - degree
//income
//consci_r (confidence in science)
//internet_sci (uses internet for seeksci or scifrom)


*allow missing values to be used in computation.
//missing values nextgen advfront toofast scibnfts balpos
balneg ().
//fre nextgen advfront toofast scibnfts balpos balneg.

*compute composite benefit/harm variable (I have re-
allocated people who said dk to scibnfts to the middle
category
```

```
//I also reallocated those who said benefit or harm to that
question but then said dk to the followup to the 'slightly'
category
//on the basis that they already said which direction their
attitude was and if they then can't choose between slightly
and strongly
//we can reasonably think they are not having strong views.

//if (scibnfts=2 or scibnfts=8) benharm=3.
//if (balpos=1) benharm=5.
//if (balpos=2 or balpos=8) benharm=4.
//if (balneg=1) benharm=1.
//if (balneg=2 or balpos=8) benharm=2.
//fre benharm.

*rescale this variable so it runs from 1-4 like the others.
//compute benharm2=.75*(benharm-5)+4.

*reverse code it so that positive to science = low score,
like the others.
//compute benharm2r=5-benharm2.
//fre benharm2r.

*reinstate missing values.
//missing values nextgen advfront toofast scibnfts balpos
balneg (0,8,9).

*reverse code the neg worded item.
//compute toofastr=5-toofast.

*compute final attitude variable so it runs from 0 to 3
with high score = positive attitude.
//compute sciatt=4-
(mean(nextgen,advfront,toofastr,benharm2r)).
//fre sciatt.


*****************************
***Stata code for analyses***
*****************************

***Generating the appropriate CAMSIS measures for inclusion
in GSS.***

*GSS uses census occupation codes from 2010, but CAMSIS is
constructed using 2000 codes.
*Steps*
```

```
*1. Go from the 2010 census codes to Standard Occupation
Coding (SOC).  The 2010 to 2002 SOC is here:
https://www.census.gov/people/eeotabulation/data/2010_OccCo
deswithCrosswalkfrom2002-2011nov04.xls
*2. Go from SOC to the 2000 Census Occupation Codes. The
2002 SOC to OCC is here:
http://www.workforceinfodb.org/ftp/DOWNLOAD/xwalks/cen02soc
.zip
*In most cases, this means a slight loss of granularity
(e.g. a small number of 2010 categories are merged into
general categories in 2000)*
*797 cases not merged. Of these: 745 coded in GSS as IAP or
NA (inapplicable/not applicable). 52 cases included in
lookup, but code not used in GSS.

*Generate appropriate matching variable to include details
of occupation status (employed, self employed, unknown)
when matching from camsis)
gen stdempst = 0
replace stdempst = 1 if WRKSLF ==1
replace stdempst = 6 if WRKSLF ==2
tab stdempst

*formatting CAMSIS for use in analysis - ensures it is on
correct scale
drop if soc2000>=10000
gen soc2000_new = soc2000*10

*Merge CAMSIS with GSS data**

merge m:1 stdempst soc2000_new using "CAMSIS file.dta"
//745 cases with missing information for occupation status
drop if _merge==2
drop _merge
gen finalCAM = mcamsis
replace finalCAM = fcamsis if SEX==2
egen zcam = std(finalCAM)



*****************************
**Data recoding for models**
*****************************

*Region*
tab REGION
tab REGION, nol
```

```
gen south_r = 0
replace south_r = 1 if REGION==5 | REGION==6 | REGION==7
gen midwest_r = 0
replace midwest_r = 1 if REGION==3 | REGION==4
gen west_r = 0
replace west_r = 1 if REGION==8 | REGION==9

*Religion*
tab RELIG16
tab RELIG16, nol
gen protestant16_r = 0
replace protestant16_r = 1 if RELIG16==1
replace protestant16_r = . if RELIG16==0 | RELIG16==98 |
RELIG16==99
gen catholic16_r = 0
replace catholic16_r = 1 if RELIG16==2
replace catholic16_r = . if RELIG16==0 | RELIG16==98 |
RELIG16==99
gen jewish16_r = 0
replace jewish16_r = 1 if RELIG16==3
replace jewish16_r = . if RELIG16==0 | RELIG16==98 |
RELIG16==99
gen otherrelig16_r = 0
replace otherrelig16_r = 1 if RELIG16>=5 & RELIG16<=13
replace otherrelig16_r = . if RELIG16==0 | RELIG16==98 |
RELIG16==99

*Respondent has a degree*
tab DEGREE
tab DEGREE, nol
gen degree = 0
replace degree = 1 if DEGREE==3 | DEGREE==4
replace degree = . if DEGREE==8 | DEGREE==9

*WordSUM*
tab WORDSUM
tab WORDSUM, nol
gen wordsum = WORDSUM
replace wordsum = . if WORDSUM ==-1 | WORDSUM==99

*Income*
gen income = INCOME06
replace income = INCOME16 if INCOME06 ==0
replace income = 26 if income==27
replace income = . if income==98
tab income
```

```
gen sub=1
replace sub=0 if sciscore13==.


********************
**Empirical models**
********************

**Science scores by year (Supplementary table S1)**
svyset VPSU [weight=WTSSNR], strata (VSTRAT)
svy, subpop(if sub==1): mean sciscore13, over(ethnic_r
YEAR)
svy, subpop(if sub==1): mean sciscore13, over(ethnic_r)


**Models including robust standard errors (Supplementary
table S2)**
reg sciscore13 black_nonH hispanic_r race_other female
boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r wordsum internet_sci consci_r
sciatt
generate sample=e(sample)
svyset VPSU [weight=WTSSNR], strata (VSTRAT)
svy, subpop(if sample==1): reg sciscore13 black_nonH
hispanic_r race_other female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r
estimates store est1
svy, subpop(if sample==1): reg sciscore13 black_nonH
hispanic_r race_other female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum
estimates store est2
svy, subpop(if sample==1): reg sciscore13 black_nonH
hispanic_r race_other female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum
internet_sci consci_r sciatt
estimates store est3
esttab est1 est2 est3, se wide nopa //Nb. Sample sizes
taken from models above.

**Reduced coefficient plot (Figure 1)**
coefplot est1 est2 est3, drop(_cons female boomers genx
geny south_r midwest_r west_r income protestant16_r
catholic16_r jewish16_r otherrelig16_r) msymbol(d)
```

```
msize(1.2) ciopts(recast(rspike) lwidth(thin)) ysize(4)
xsize(5) mfcolor(white) xline(0) xlabel(-2 -1 0 1 2,
labsize(2)) xtitle("Effect sizes", size(3)) base omit
coeflabels(black_nonH = "Black, non-hispanic" ///
race_other = "Other ethnicity, non-hispanic" ///
hispanic_r = "Hispanic" ///
degree = "Degree or higher" ///
NumMathSci = "Number of maths/science courses" ///
wordsum = "Vocabulary score" ///
internet_sci = "Uses internet to find science info" ///
sciatt = "Attitude to science" ///
zcam = "CAMSIS score" ///
consci_r = "Confidence in the scientific community",
labsize(2.6)) ///
headings(black_nonH = "{it:Race (Ref: White)}" ///
degree="{it:Education (Ref: Less than degree)}" ///
wordsum="{it:}") ///
legend(rows(1) label(2 "Model 1")label(4 "Model 2")label(6
"Model 3") size(2.2))
eststo clear
svyset, clear

**Full coefficient plot (Supplementary Figure S1)**
coefplot est1 est2 est3, drop(_cons) msymbol(d) msize(0.8)
ciopts(recast(rspike) lwidth(thin)) ysize(7) xsize(5.5)
mfcolor(white) xline(0) xlabel(-2 -1 0 1 2, labsize(2))
xtitle("Effect sizes", size(2.5)) base omit
coeflabels(black_nonH = "Black, non-hispanic" ///
race_other = "Other ethnicity, non-hispanic" ///
hispanic_r = "Hispanic" ///
female = "Female" ///
boomers = "Baby boomers" ///
genx = "Generation X" ///
geny = "Millennials" ///
degree = "Degree or higher" ///
south_r = "South" ///
midwest_r= "Midwest" ///
west_r = "West" ///
NumMathSci = "Number of maths/science courses" ///
wordsum = "Vocabulary score" ///
internet_sci = "Uses internet to find science info" ///
income = "Income" ///
sciatt = "Attitude to science" ///
zcam = "CAMSIS score" ///
protestant16_r = "Protestant" ///
catholic16_r  = "Catholic" ///
jewish16_r = "Jewish" ///
```

```
otherrelig16_r = "Other religious upbringing" ///
consci_r = "Confidence in the scientific community",
labsize(2.2)) ///
headings(black_nonH = "{it:Race (Ref: White)}" ///
female="{it:Gender (Ref: Male)}" ///
boomers="{it:Birth cohort (Ref: Pre War)}" ///
degree="{it:Education (Ref: Less than degree)}" ///
NumMathSci="{it:}" ///
wordsum="{it:}" ///
south_r="{it:Region (Ref: Northeast)}" ///
protestant16_r="{it:Religious upbringing (Ref:None)}") ///
legend(rows(1) label(2 "Model 1")label(4 "Model 2")label(6
"Model 3") size(2.2))

**Models estimated using full sample size (Supplementary
table S3)**
eststo clear
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r wordsum
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r wordsum internet_sci consci_r
sciatt
esttab, se wide nopa
eststo clear

**Models estimated using listwise delete (Supplementary
table S4)**
eststo clear
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r if sample==1
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
south_r midwest_r west_r wordsum if sample==1
eststo:reg sciscore13 black_nonH hispanic_r race_other
female boomers genx geny degree NumMathSci income zcam
protestant16_r catholic16_r jewish16_r otherrelig16_r
```

```
south_r midwest_r west_r wordsum internet_sci consci_r
sciatt if sample==1
esttab, se wide nopa
eststo clear

**OAXACA differentials (Supplementary Tables S5-S7)**
ssc install oaxaca

gen black_nonH_solo = black_nonH
replace black_nonH_solo =. if hispanic_r==1 | race_other==1

gen hispanic_r_solo = hispanic_r
replace hispanic_r_solo =. if black_nonH==1 | race_other==1

gen race_other_solo = race_other
replace race_other_solo =. if hispanic_r==1 | black_nonH==1

svyset VPSU [weight=WTSSNR]
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r,
by(black_nonH_solo)  pooled detail svy
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum,
by(black_nonH_solo) pooled detail svy
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum
internet_sci consci_r sciatt, by(black_nonH_solo) pooled
detail svy
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r,
by(hispanic_r_solo) pooled  detail svy
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum,
by(hispanic_r_solo)  pooled detail svy
oaxaca sciscore13 female boomers genx geny degree
NumMathSci income zcam protestant16_r catholic16_r
jewish16_r otherrelig16_r south_r midwest_r west_r wordsum
internet_sci consci_r sciatt, by(hispanic_r_solo) pooled
detail svy
```