

## Supplementary Materials for **Coming out in STEM: Factors affecting retention of sexual minority STEM students**

Bryce E. Hughes

Published 14 March 2018, *Sci. Adv.* **4**, eaao6373 (2018)

DOI: 10.1126/sciadv.aao6373

### **This PDF file includes:**

- Supplementary Text
- table S1. List of STEM majors.
- table S2. Hierarchical generalized linear model predicting likelihood of participation in an undergraduate research experience ( $n = 4162$ ).
- table S3. Hierarchical generalized linear model predicting likelihood of persistence in a STEM degree after 4 years ( $n = 4162$ ).
- table S4. List of all variables included in modeling and respective coding.
- table S5. Descriptive statistics for all variables included in HGLM.
- Reference (50)

## Supplementary Materials:

### Supplementary Text

#### Hierarchical Generalized Linear Model Equations

In logistic regression, the statistical model follows the form of the equation (49):

$$\ln \frac{\hat{p}}{1 - \hat{p}} = b_0 + b_1 X_{1i} + \dots + b_h X_{hi} + e_i$$

In this equation:

- $\hat{p}$  is the probability that the expected value of the dependent variable is 1
- $\ln \frac{\hat{p}}{1 - \hat{p}}$  is the logit link function for the model
- $b_0$  is the estimated value of the intercept
- $b_l$  is the estimated value of the slope for independent variable 1
- $X_{li}$  is the value for case  $i$  on independent variable 1
- $b_h$  is the estimated value of the slope for independent variable  $h$
- $X_{hi}$  is the value for case  $i$  on independent variable  $h$
- $e_i$  is the measurement error for case  $i$

Logistic regression analysis is only valid when several assumptions about the data being modeled are met (49). The dependent variable should be dichotomous in nature, the relationship between the variables and the logit function is linear, and the measurement error for each case is independent of the other cases. Data with a “nested” structure, where individual cases are grouped, such as education data where individual students are grouped within schools, the independence assumption is most likely not met (45). Students at the same school may be similar to each other in ways that differ by school.

To overcome this assumption violation, multilevel modeling is recommended to parse the variation within the dependent variable into that which can be attributed to individual differences within groups, or level-1 variance, and that which can be attributed to differences between groups, or level-2 variance (45). Multilevel modeling also helps address the problem that arises with respect the appropriate number of degrees of freedom used for significance testing of group-level variables. With single-level logistic regression, the value for group-level variables is assigned to individual-level cases, increasing the number of degrees of freedom used to determine significance. Multilevel modeling determines the number of degrees of freedom for group-level variables based on the number of groups, reducing the likelihood of a type-I error resulting from an inappropriate number of degrees of freedom.

The general form for the hierarchical generalized linear model (HGLM) is as follows (45):

Level 1 (student-level variables):

$$\ln \left( \frac{\varphi_{ij}}{1 - \varphi_{ij}} \right) = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \dots + \beta_{Qj} X_{Qij}$$

Level 2 (institution-level variables):

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}W_{1j} + \cdots \gamma_{0Z}W_{Zj} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ &\vdots \\ \beta_{Qj} &= \gamma_{Q0}\end{aligned}$$

- $\phi_{ij}$  is the probability of persisting in STEM for the  $i^{\text{th}}$  student at the  $j^{\text{th}}$  institution
- $\beta$ s represent the level-1 coefficients
- $\beta_{0j}$  is the intercept for the  $j^{\text{th}}$  institution
- $\gamma$ s represent the level-2 coefficients
- $\gamma_{00}$  is the overall intercept
- $X_{Qij}$  is the value of the  $Q^{\text{th}}$  level-1 variable for the  $i^{\text{th}}$  student at the  $j^{\text{th}}$  institution
- $W_{Zj}$  is the value of the  $Z^{\text{th}}$  level-2 variable for the  $j^{\text{th}}$  institution
- $u_{0j}$  is the level-2 error term

The mixed form of the equation is then developed by substituting the level-2 equations for the  $\beta$ s in the level-1 equation.

**table S1.**

**List of STEM majors.**

---

*Life Sciences*

Biology (general)  
Biochemistry or Biophysics  
Botany  
Environmental Science  
Marine (Life) Science  
Microbiology or Bacteriology  
Zoology  
Other Biological Science

*Engineering*

Aeronautical or Astronautical Engineering  
Civil Engineering  
Chemical Engineering  
Computer Engineering  
Electrical or Electronic Engineering  
Industrial Engineering  
Mechanical Engineering  
Other Engineering

*Physical Sciences and Mathematics*

Astronomy  
Atmospheric Science (incl. Meteorology)  
Chemistry  
Computer Science  
Earth Science  
Marine Science (incl. Oceanography)  
Mathematics  
Physics  
Other Physical Science

*Health-related Fields*

Health Technology (medical, dental,  
laboratory)  
Medicine, Dentistry, Veterinary Medicine  
Nursing  
Pharmacy

*Agriculture*

Agriculture

---

**table S2.****Hierarchical generalized linear model predicting likelihood of participation in an undergraduate research experience ( $n = 4162$ ).**

	B	Std. Err.	<i>t</i> -ratio	Sig
Intercept	-1.636	0.713	-2.294	*
Sexual minority	0.307	0.104	2.949	**
Sex: female	0.251	0.090	2.784	**
Student of color	0.090	0.062	1.46	
Either parent employed in STEM	0.146	0.092	1.599	
Low income (ref: middle)	0.112	0.106	1.057	
Middle-low income	-0.005	0.116	-0.044	
Middle-high income	-0.145	0.074	-1.961	
High income	-0.033	0.100	-0.327	
Mother achieved at least college degree	-0.038	0.079	-0.476	
High school GPA	0.163	0.037	4.416	***
SAT score, scaled by 100	0.162	0.028	5.798	***
Institutional type: four-year college (ref: university)	0.176	0.199	0.887	
Institutional control: private	0.005	0.344	0.016	
Selectivity (average SAT score, scaled by 100)	0.112	0.059	1.894	
Average student agreement with legality of same-sex marriage	0.117	0.188	0.626	
Academic self-concept construct score, scaled by 10	0.226	0.052	4.372	***
Social self-concept construct score, scaled by 10	-0.038	0.067	-0.562	
Goal: obtaining recognition from colleagues for contributions to my field	0.008	0.034	0.251	
Goal: making a theoretical contribution to science	0.549	0.026	21.452	***
Studied with other students	-0.031	0.045	-0.693	
Hours per week: studying or homework	0.097	0.024	4.034	***
Faculty interaction construct score, scaled by 10	0.436	0.079	5.533	***

Note: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**table S3.**

**Hierarchical generalized linear model predicting likelihood of persistence in a STEM degree after 4 years ( $n = 4162$ ).**

	B	Std. Err.	t-ratio	Sig	$\Delta p$
Intercept	-0.697	1.020	-0.683		
Sexual minority	-0.523	0.145	-3.604	***	-9.54%
<i>Demographic characteristics</i>					
Sex: female	-0.317	0.119	-2.668	**	-5.66%
Student of Color	0.082	0.126	0.651		
Either parent employed in STEM	0.190	0.082	2.336	*	4.23%
Low income (ref: middle)	0.005	0.170	0.031		
Middle-low income	-0.032	0.144	-0.225		
Middle-high income	0.134	0.094	1.435		
High income	-0.001	0.088	-0.009		
Mother achieved at least college degree	0.211	0.114	1.843		
<i>Precollege academic preparation</i>					
Average HS GPA	0.283	0.050	5.664	***	5.60%
SAT score or ACT equivalent, scaled by 100	0.200	0.046	4.331	***	4.02%
<i>Institutional characteristics (level-2)</i>					
Institutional selectivity (scaled by 100)	-0.172	0.106	-1.614		
Institutional type: four-year college	0.004	0.283	0.012		
Institutional control: private	0.724	0.431	1.682		
Average student agreement with legality of same-sex marriage	0.005	0.212	0.022		
<i>Expectations and self-concept</i>					
Academic self-concept	0.085	0.048	1.755		
Social self-concept	-0.356	0.045	-7.869	***	-7.96%
Goal: obtaining recognition from colleagues for contributions to my field	-0.277	0.056	-4.961	***	-6.12%
Goal: making a theoretical contribution to science	0.756	0.046	16.422	***	13.25%
<i>College experiences</i>					
Studied with other students	0.323	0.068	4.763	***	6.32%
Hours per week: studying or homework	0.218	0.032	6.755	***	4.37%
Participated in undergraduate research program	0.557	0.140	3.970	***	13.46%
Student-faculty interaction	-0.097	0.116	-0.837		
<i>Interaction term tests</i>					
Interaction term, sexual minority with undergraduate research	-0.388	0.283	-1.369		
Interaction term, sex with sexual minority status	0.701	0.238	2.952	**	
Interaction term, race with sexual minority status	0.191	0.302	0.630		

Note: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**table S4.****List of all variables included in modeling and respective coding.**


---

Variables	Coding
<u>Dependent variable</u>	
STEM major in 2015	1 Non-STEM major 2 STEM major
<u>Independent variables</u>	
<i>Student background characteristics</i>	
Sexual minority	1 No 2 Yes
Your sex:	1 Male 2 Female
Student of Color	1 No 2 Yes
Either parent employed in STEM	1 No 2 Yes
Family income	
Low income (less than \$25,000)	1 No
Middle-low income (\$25,000 to \$49,999)	2 Yes
Middle income (\$50,000 to \$99,999)	
Middle-high income (\$100,000 to \$199,999)	
High income (\$200,000 or higher)	
Mother completed at least a college degree	1 No 2 Yes
<i>Pre-college academic preparation</i>	
What was your average grade in high school?	1 D 2 C 3 C+ 4 B- 5 B 6 B+ 7 A- 8 A or A+
SAT score (or ACT equivalent)	
<i>Institutional characteristics</i>	
Institutional selectivity	Avg. SAT score
Institution type	1 University 2 4-year
Institution control	1 Public 2 Private

Agg support for same-sex marriage

Same-sex couples should have the right to legal marital status

- 1 Disagree Strongly
- 2 Disagree Somewhat
- 3 Agree Somewhat
- 4 Agree Strongly

*Expectations and self-concept*

Academic self-concept construct score

Social self-concept construct score

Faculty interaction construct score

Goal: Obtaining recognition from my colleagues for contributions to my special field

(for construct technical information, see 50)

- 1 Not important
- 2 Somewhat important
- 3 Very important
- 4 Essential

Goal: Making a theoretical contribution to science

- 1 Not important
- 2 Somewhat important
- 3 Very important
- 4 Essential

*STEM-related college experiences*

Frequency studied with other students

- 1 Not at all
- 2 Occasionally
- 3 Frequently

HPW: studying or homework

- 1 None
- 2 1 to 5 hours
- 3 6 to 10 hours
- 4 11 to 15 hours
- 5 16 to 20 hours
- 6 Over 20 hours

Participated in an undergraduate research program

- 1 No
  - 2 Yes
-



**table S5.****Descriptive statistics for all variables included in HGLM.**

	Mean	SD	Min	Max
Retained in STEM	1.70	0.46	1	2
Sexual minority	1.08	0.27	1	2
Sex: female	1.58	0.49	1	2
Student of Color	1.27	0.45	1	2
Either parent employed in STEM	1.31	0.46	1	2
Low income (less than \$25,000)	1.07	0.25	1	2
Middle-low income (\$25,000 to \$49,999)	1.12	0.32	1	2
Middle-high income (\$100,000 to \$199,999)	1.32	0.46	1	2
High income (\$200,000 or higher)	1.20	0.40	1	2
Mother completed at least a college degree	1.72	0.45	1	2
What was your average grade in high school?	7.19	0.99	2	8
SAT (or ACT equivalent) scaled by 100	12.68	1.75	4	16
Institutional Selectivity scaled by 100	12.45	1.44	8.55	15
Institution Type	1.65	0.48	1	2
Institution Control	1.94	0.23	1	2
Agg support for same-sex marriage	3.44	0.31	1.94	4
Academic self-concept score, scaled by 10	5.24	0.98	1.96	7.97
Social self-concept score, scaled by 10	5.24	0.85	2.23	7.36
Goal: Obtaining recognition from my colleagues for contributions to my special field	2.57	0.84	1	4
Goal: Making a theoretical contribution to science	2.31	0.98	1	4
Act: Studied with other students	2.42	0.59	1	3
Hours per week: studying or homework (rescaled)	3.99	1.39	1	6
Act in College: Participated in an undergraduate research program	1.42	0.49	1	2
Faculty interaction score, scaled by 10	5.17	0.56	3.59	6.51
Act in College: Taken a women's studies course	1.22	0.42	1	2
Act in College: Participated in an LGBTQ student organization	1.07	0.26	1	2
Sense of belonging score, scaled by 10	5.10	0.91	2.42	7.55
Worked off campus	1.38	0.48	1	2
Commuter student	1.49	0.50	1	2
Act in Past Year: Sought personal counseling	1.38	0.61	1	3