Supplemental Material

CBE—Life Sciences Education

Aikens *et al.*

SUPPLEMENTAL MATERIAL

TABLE OF CONTENTS

APPENDIX A: METHODS

Characteristics of Undergraduate Participants

Table A1. Characteristics of all undergraduate participants (n = 842) overall and by triad type. For college GPA, we present the overall mean and the mean for each triad. Students may have chosen "Prefer not to respond" for some demographic information or for triad type. Therefore, the total number of students for each characteristic may be < 842, and the sum of students across triads may not equal the total number of students for a given characteristic.

Characteristics	Total	Triad I 	Triad II 	Triad III 	Triad IV 	Triad V 	Triad VI 	Triad VII 	Triad VIII 	Triad Other
Gender										
Male	311 (37%)	1	4	5	18	14	13	71	182	2
Female	525 (62%)	0	6	4	17	25	24	169	271	9
Other	2 (< 1%)	0	0	0	1	0	0	0	1	0
Race/Ethnicity										
White	381 (45%)	1	5	3	14	20	13	111	206	7
Asian	262 (31%)	0	3	5	13	7	15	93	125	1
Underserved	185 (22%)	0	2	1	9	11	9	34	116	3
Other	5 (< 1%)	0	0	0	0	1	0	0	4	0
First-generation college										
No	739 (88%)	1	8	7	31	33	36	218	395	9
Yes	80 (10%)	0	2	2	4	4	1	21	44	2
Prior research experience										
None	454 (54%)	1	7	5	17	20	24	134	241	5
1 experience	241 (29%)	0	2	4	12	13	9	61	137	3
2 experiences	100 (12%)	0	1	0	7	3	2	31	53	2
3+ experiences	47 (6%)	0	0	0	0	3	2	17	24	1
Duration of research experience										
1 semester	182 (22%)	0	2	3	7	13	4	60	90	3
2 semesters	192 (23%)	0	2	1	9	9	10	57	101	2
3 semesters	154 (18%)	1	1	1	8	5	12	43	81	2
4+ semesters	311 (37%)	0	5	4	11	12	11	83	181	4
Honors program										
No	513 (61%)	1	4	6	30	25	21	161	260	6
Yes	297 (35%)	0	5	3	4	14	15	73	176	5
Institution type										
Very high research	681 (81%)	1	8	8	32	26	26	220	349	10
High research	114 (14%)	0	2	1	0	11	6	12	82	0
Doctorate	13 (2%)	0	0	0	1	1	2	4	5	0
Masters	8 (1%)	0	0	0	0	1	0	0	6	1
Baccalaureate	1 (< 1%)	0	0	0	0	0	0	1	0	0
Research institute	23 (3%)	0	0	0	3	0	2	6	12	0
International	2 (< 1%)	0	0	0	0	0	1	0	1	0
College GPA	3.55	4.00	3.36	3.72	3.54	3.51	3.56	3.55	3.56	3.47

Scales

Table A2. The item stem, items, and response options for each outcome scale used in the analyses.

Outcome	Stem	Item	Response options
Thinking and working like a scientist (Hunter *et al.*, 2009; Weston and Laursen, 2015)	Please indicate the extent of the gains you have made within each category.	Analyzing data for patterns.	1 = No gain; 2 = Little gain; 3 = Moderate gain; 4 = Good gain; 5 = Great gain; 6 = I don't know; 7 = Not applicable / No response
		Figuring out the next step in a research project.	
		Problem-solving in general.	
		Formulating a research question that could be answered with data.	
		Identifying limitations of research methods and designs.	
		Understanding the theory and concepts guiding my research project.	
		Understanding the connections among scientific disciplines.	
		Understanding the relevance of my research to my coursework.	
		Defending an argument when asked questions.	
Scientific self-efficacy (Estrada *et al.*, 2011)	Please indicate your level of confidence in your ability to…	Use technical science skills (use of tools, instruments, and/or techniques).	1 = Not confident; 2 = A little confident; 3 = Somewhat confident; 4 = Confident; 5 = Very confident; 6 = I don't know; 7 = Not applicable / No response
		Generate a research question to answer.	
		Figure out what data/observations to collect and how to collect them.	
		Create explanations for the results of the study.	
		Use scientific literature and/or reports to guide research.	
		Develop theories (integrate and coordinate results from multiple studies).	
Scientific identity (Estrada *et al.*, 2011)	Please indicate your level of agreement with the following statements.	I have a strong sense of belonging to the community of scientists.	1 = Strongly disagree; 2 = Disagree; 3 = Neither agree nor disagree; 4 = Agree; 5 = Strongly agree; 6 = I don't know; 7 = Not applicable / No response
		I have come to think of myself as a "scientist."	
		I feel like I belong in the field of science.	
		I derive great personal satisfaction from working on a team that is doing important research.	
		The daily work of a scientist is appealing to me.	
Research satisfaction (adapted from Volkwein and Carbone, 1994)		I am satisfied with my research experience in general.	1 = Strongly disagree; 2 = Disagree; 3 = Neither agree nor disagree; 4 = Agree; 5 = Strongly agree; 6 = I don't know; 7 = Not applicable / No response
		I am satisfied with the extent of my intellectual development during my research experience.	
		My research experience has had a positive influence on my intellectual growth.	
Career and education preparation (Hunter *et al.*, 2009)	Please indicate the extent to which you agree with the following statements.	My undergraduate research experience has prepared me to succeed in an academic career.	1 = Strongly disagree; 2 = Disagree; 3 = Neither agree nor disagree; 4 = Agree; 5 = Strongly agree; 6 = I don't know; 7 = Not applicable / No response
		My research experience has prepared me for a job.	
		Doing research has confirmed my interest in my field of study.	

		Doing research has clarified for me which field of study I want to pursue.	
		My research experience has prepared me for advanced coursework or thesis work.	
		My research experience has prepared me for graduate school.	
Scholarly productivity (self-authored)	Please indicate how many times you completed each of the following professional activities as a result of your research experience.	Presented a poster or talk as part of a local program or event	0; 1; 2; 3; 4; 5+
		Presented a poster at a regional, national, or international conference	
		Presented a talk at a regional, national, or international conference	
		Participated in writing a manuscript for publication in a peer-reviewed journal	
		Published an article in a peer-reviewed journal.	
Intentions to enroll in a Ph.D. program (Hunter *et al.*, 2009; Weston and Laursen, 2015)	Compared to your intentions before doing research, please indicate how likely are you now to…	Enroll in a Ph.D. program in science, mathematics, or engineering	1 = Not more likely; 2 = A little more likely; 3 = Somewhat more likely, 4 = Much more likely, 5 = Extremely more likely; 6 = I don't know; 7 = Not applicable / No response

Confirmatory Factor Analysis

We performed confirmatory factor analysis (lavaan package; Rosseel, 2012) on two scales to ensure that the addition of one item to each of these scales did not negatively impact the functioning of the scales. We used robust diagonally weighted least squares (WLSMV) as an estimator, which provides robust standard errors and a test statistic adjusted for the mean and variance that is appropriate for ordinal data (Finney and DiStefano, 2013). Confirmatory factor analysis provides several fit statistics to analyze; good fit is indicated by Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values above 0.95, a root mean square error of approximation (RMSEA) value less than 0.06, and a standardized root mean square residual (SRMR) value below 0.08 (Hu and Bentler, 1998, 1999). In practice, a holistic view of the fit indices is taken when analyzing model fit, such that failure of any one value to conform to these rules of thumb is not necessarily an indication that the model is a poor fit (Marsh *et al.*, 2004).

Thinking and Working Like a Scientist Scale. We added one item to the Thinking and Working Like a Scientist scale: "defending an argument when asked questions." This item had moderate correlations with items from the Thinking and Working Like a Scientist scale in our undergraduate population (r_s = 0.440 – 0.534). The original scale had adequate fit ($\chi^2(20, n=796) = 162.457$, $p < 0.001$; CFI = 0.864; TLI = 0.809; RMSEA = 0.095 [0.081-0.108]; SRMR = 0.040), but adding the extra item improved model fit ($\chi^2(27, n=785) = 160.635$, $p < 0.001$; CFI =

0.892; TLI = 0.856; RMSEA = 0.079 [0.068-0.092]; SRMR = 0.038) as demonstrated by the higher CFI and TLI values and the lower RMSEA and SRMR values. Cronbach's α for the nine-item scale was 0.90, indicating high internal consistency. Thus, we incorporated the extra item into our measure of thinking and working like a scientist.

Career and Education Preparation Scale. We added one item to the Career and Education Preparation scale: "My undergraduate research experience has prepared me to succeed in an academic career." This item had moderately high correlations with items from the Career and Education Preparation scale in our undergraduate population (r_s = 0.470 – 0.667). The original scale had adequate fit ($\chi^2(5, n=710) = 51.122$, $p < 0.001$; CFI = 0.926; TLI = 0.853; RMSEA = 0.114 [0.087-0.143]; SRMR = 0.038). Adding the extra item caused a comparable model fit ($\chi^2(9, n=708) = 70.860$, $p < 0.001$; CFI = 0.915; TLI = 0.858; RMSEA = 0.099 [0.078-0.121]; SRMR = 0.036). Cronbach's α for the six-item scale was 0.89, indicating high internal consistency. Thus, we incorporated the extra item into our measure of career and education preparation.

Likelihood Ratio Tests

We tested whether to include a race/ethnicity x gender interaction term and a race/ethnicity x first-generation interaction term in the regression models using a likelihood ratio test. Likelihood ratio tests are used to compare a full model and a nested model, which contains one variable less than the full model. The null hypothesis is that the models fit equivalently well; in other words, the parameter estimate of the dropped variable is zero. The alternative hypothesis is that the full model is a significantly better fit to the data than the nested model, and thus, the variable needs to be retained. Therefore, if the likelihood ratio statistic is not significant ($p < 0.10$, a conservative value), the variable can be dropped from the regression, but if it is significant, the variable must be retained in the regression. We compared a full regression model that contained all main effects and both interaction terms to two different nested models: one without the race/ethnicity x gender interaction and one without the race/ethnicity x first-generation interaction. In all likelihood ratio tests, the nested model without the race/ethnicity x gender interaction term was not significantly different from the full model, so we dropped this interaction term from all regression models. The nested model with the race/ethnicity x first-generation interaction term was significantly different from the full model for the linear

regressions predicting thinking and working like a scientist and career and education preparation. Thus, we retained the race/ethnicity x first-generation interaction term for these models.

References

Estrada M, Woodcock A, Hernandez PR, Schultz PW (2011). Toward a model of social influence that explains minority student integration into the scientific community. J Educ Psych *103*, 206-222.

Finney SJ, DiStefano C (2013). Non-normal and categorical data in structural equation modeling. In: Structural Equation Modeling: A Second Course (2nd ed.), ed. GR Hancock and RO Mueller, Greenwich, CT: Information Age.

Hu L, Bentler PM (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol Methods *3*, 424-453.

Hu L, Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling *6*, 1-55.

Hunter A-B, Weston TJ, Laursen SL, Thiry H (2009). URSSA: Evaluating student gains from undergraduate research in the sciences. CUR Quarterly *29*, 15-19.

Marsh HW, Hau K-T, Wen Z (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct Equ Modeling *11*, 320-341.

Rosseel Y (2012). lavaan: An R package for structural equation modeling. J Stat Software *48*, 1-36.

Volkwein JF, Carbone DA (1994). The impact of departmental research and teaching climates on undergraduate growth and satisfaction. J High Educ *65*, 147-167.

Weston TJ, Laursen SL (2015). The undergraduate research student self-assessment (URSSA): validation for use in program evaluation. CBE Life Sci Educ *14*, ar33.

APPENDIX B: RESULTS

Correlation Matrix of Variables

Table B1. Correlations between the predictor and outcome variables. Correlations between continuous variables and between a continuous and dichotomous variable are based on Pearson's *r*. The latter is equivalent to using point biserial correlation coefficients. Correlations between a continuous and an ordinal variable are based on Spearman's rho. Correlations between an ordinal and a dichotomous variable are based on Somer's D, which is equivalent to the rank biserial correlation coefficient. Correlations between dichotomous variables are based on phi. Gender: 1=male, 2=female; First-generation: 1=non-first-generation, 2=first-generation; Honors program 1=non-Honors, 2= Honors; Intent to enroll in a Ph.D. program in STEM: 1=not more likely, 2=more likely. †continuous variable; ‡ordinal variable; §dichotomous variable

	(1)§	(2)§	(3)§	(4)†	(5)†	(6)†	(7)†	(8)†	(9)†	(10)‡
(1) Gender§	-									
(2) First-generation§	-0.003	-								
(3) Honors program§	0.033	-0.070	-							
(4) College GPA†	-0.051	-0.045	0.336	-						
(5) Thinking and working like a scientist†	-0.022	0.085	0.064	-0.011	-					
(6) Scientific self-efficacy†	-0.098	0.022	0.088	-0.042	0.647	-				
(7) Scientific identity†	-0.116	0.017	0.077	-0.167	0.444	0.465	-			
(8) Research satisfaction†	-0.067	0.043	0.032	0.017	0.614	0.486	0.517	-		
(9) Career and education preparation†	-0.023	0.018	0.102	-0.021	0.566	0.499	0.643	0.659	-	
(10) Scholarly productivity‡	-0.094	0.112	0.173	0.034	0.319	0.312	0.276	0.246	0.305	-
(11) Intent to enroll in a Ph.D. program in STEM§	0.024	0.007	0.049	-0.157	0.151	0.108	0.323	0.194	0.246	0.073

Linear Regression Table

Table B2. Linear regression results for continuous outcomes. Regression coefficients (± heteroskedastic-consistent SE) are presented. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Variable	Thinking and working like a scientist	Scientific self-efficacy	Scientific identity	Research satisfaction	Career and education preparation
Intercept	4.357 (0.076)***	4.115 (0.075)***	4.351 (0.073)***	4.657 (0.074)***	4.358 (0.078)***
Triad type (reference level: Triad VIII)					
Triad IV	-0.553 (0.193)**	-0.373 (0.170)*	-0.454 (0.149)**	-0.407 (0.197)*	-0.446 (0.191)*
Triad V	-0.553 (0.129)***	-0.401 (0.131)**	-0.195 (0.130)	-0.344 (0.156)*	-0.259 (0.149)
Triad III/VI	-0.270 (0.120)*	-0.160 (0.111)	-0.079 (0.105)	-0.193 (0.123)	-0.113 (0.116)
Triad VII	-0.355 (0.064)***	-0.274 (0.064)***	-0.302 (0.062)***	-0.331 (0.072)***	-0.342 (0.075)***
Gender (reference level: Male)					
Female	-0.012 (0.054)	-0.151 (0.056)**	-0.144 (0.052)**	-0.036 (0.060)	0.008 (0.061)
Race/ethnicity (reference level: White)					
Asian	0.025 (0.060)	-0.086 (0.063)	-0.160 (0.061)**	-0.140 (0.070)*	-0.136 (0.070)
Underserved	0.084 (0.082)	0.073 (0.075)	0.015 (0.069)	0.144 (0.075)	0.167 (0.089)
First-generation status (reference level: Non-first-generation)					
First-gen	0.324 (0.149)*	-0.040 (0.096)	-0.029 (0.103)	-0.024 (0.091)	0.302 (0.141)*
Prior research experience (reference level: 0 prior experiences)					
1 prior experience	0.043 (0.059)	0.174 (0.062)**	0.066 (0.060)	-0.009 (0.067)	-0.009 (0.074)
2+ prior experiences	0.118 (0.073)	0.256 (0.076)***	0.200 (0.070)**	0.038 (0.080)	-0.188 (0.081)*
Duration of research experience (reference level: 4+ semesters)					
1 semester	-0.346 (0.075)***	-0.300 (0.077)***	-0.212 (0.071)**	-0.300 (0.081)***	-0.325 (0.084)***
2 semesters	-0.226 (0.068)***	-0.186 (0.075)*	-0.166 (0.073)*	-0.233 (0.075)**	-0.279 (0.085)**
3 semesters	-0.190 (0.072)**	-0.261 (0.076)***	-0.170 (0.071)*	-0.257 (0.083)**	-0.236 (0.079)**
College GPA	-0.101 (0.087)	-0.218 (0.085)*	-0.374 (0.081)***	0.055 (0.102)	-0.039 (0.095)
Honors program (reference: Not Honors)					
Honors	0.039 (0.055)	0.047 (0.060)	0.113 (0.056)*	-0.026 (0.065)	0.122 (0.066)
Race/ethnicity x first-generation status (reference level: White non-first-generation)					
Asian first-gen	-0.660 (0.268)*	NA	NA	NA	-0.616 (0.278)*
Underserved first-gen	-0.199 (0.195)	NA	NA	NA	-0.403 (0.229)
R^2	0.140	0.111	0.125	0.090	0.131
Cohen's f^2	0.087	0.041	0.048	0.044	0.050
$F_{(\text{num df, denom df})}$	$6.213_{(17,680)}$***	$5.736_{(15,716)}$***	$7.300_{(15,705)}$***	$4.924_{(15,727)}$***	$5.733_{(17,608)}$***

Logistic Regression Table

Table B3. Ordinal and logistic regression results for non-continuous outcomes. Regression coefficients (±SE) are presented. e^{β} is the odds-ratio. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

	Scholarly productivity		Intent to enroll in Ph.D. program	
Variable	β (± SE)	e^{β}	β (± SE)	e^{β}
Intercept 0/1	-2.019 (0.207)	0.133	1.008 (0.240)	2.740
Intercept 1/2	-0.565 (0.195)	0.568	NA	NA
Intercept 2/3	0.160 (0.194)	1.174	NA	NA
Triad type (reference level: Triad VIII)				
Triad IV	-1.155 (0.353)**	0.315	-1.255 (0.394)**	0.285
Triad V	-0.394 (0.323)	0.674	0.175 (0.429)	1.191
Triad III/VI	-0.238 (0.307)	0.788	-0.160 (0.355)	0.852
Triad VII	-0.395 (0.158)*	0.674	-0.493 (0.189)**	0.611
Gender (reference level: Male)				
Female	-0.359 (0.142)*	0.699	0.064 (0.174)	1.066
Race/ethnicity (reference level: White)				
Asian	0.107 (0.159)	1.113	-0.216 (0.188)	0.806
Underserved	0.225 (0.182)	1.252	0.399 (0.249)	1.491
First-generation status (reference level: Non-first-generation)				
First-gen	0.296 (0.229)	1.345	-0.014 (0.303)	0.986
Prior research experience (reference level: 0 prior experiences)				
1 prior experience	0.372 (0.158)*	1.450	-0.087 (0.194)	0.917
2+ prior experiences	0.669 (0.194)***	1.952	0.031 (0.235)	1.032
Duration of research experience (reference level: 4+ semesters)				
1 semester	-1.539 (0.190)***	0.215	0.068 (0.236)	1.070
2 semesters	-1.146 (0.185)***	0.318	-0.297 (0.225)	0.743
3 semesters	-0.755 (0.195)***	0.470	-0.357 (0.235)	0.700
College GPA	-0.118 (0.211)	0.889	-1.386 (0.290)***	0.250
Honors program (reference: Not Honors)				
Honors	0.397 (0.153)**	1.487	0.515 (0.190)**	1.674