

**Online Appendix for “Diverging Pathways: First-Year College Courses and the Gender
Gap in STEM Degree Attainment”**

The appendices present results from several sensitivity analyses and auxiliary analyses to complement our main findings. The table below details the analyses:

Appendix #	Content
A	Explanation of the attitude indices
B	Percentage distributions of STEM disciplines in which students took courses in their first year, by plans and gender
C	Replication of the models in Table 4 with alternative definitions of STEM curricula immersion (number of courses and proportion of STEM credits)
D	Model estimates for Figure 2
E	Model estimates for Table 5
F	Replication of Figure 2 with alternative model and variable specification
G	Replication of the main results for students with undecided plans
H	Assessment of fit for the models in Table 5

Appendix A: Indices for attitudes

Math and English self-assessments (10th grade): These are composite scores of two-factor solutions from factor analyses (with varimax rotation) estimated on the following items (response scale: almost never, sometimes, often, almost always). (1) Math self-assessment (Cronbach's alpha=0.94, Factor eigen value=3.67): Can do excellent job on math tests; Can understand difficult math texts; Can understand difficult math class; Can do excellent job on math assignments; Can master math class skills. (2) English self-assessment (Cronbach's alpha=0.928, Factor eigen value=3.61): Can understand difficult English texts; Can understand difficult English class; Can do excellent job on English assignments; Can do excellent job on English tests; Can master English skills.

Work-family values: Following Xie and Shauman (2003:281), we used the items below to calculate respondents' scores (response scale: not important, somewhat important, very important). We first summed the scores in the work and family domains separately and then calculated the difference between them (W-F). Positive values indicate greater importance to work relative to family; negative values imply the opposite. Work values: Importance of being successful in line of work; Importance of having lots of money. Family values: Importance of marrying right person/having happy family; Importance of having children.

Appendix B: STEM disciplines in which students took courses during the first year of college

Took at least one course in:	STEM-aspiring students			Non-STEM-aspiring students		
	Men	Women	Gap (M-W)	Men	Women	Gap (M-W)
Math and statistics	0.80	0.74	0.06	0.65	0.56	0.09
Physical sciences	0.56	0.51	0.05	0.38	0.34	0.04
Biological sciences	0.30	0.54	-0.24	0.24	0.32	-0.08
Engineering & engineering technologies	0.36	0.05	0.31	0.09	0.01	0.08
Computer and information science	0.27	0.08	0.19	0.16	0.10	0.06
Natural resources/agriculture/multidisciplinary	0.06	0.08	-0.02	0.04	0.06	-0.02
Average number of lab courses	1.04	1.15	-0.11	0.83	0.54	0.29

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Disciplines are based on the general code of the CCM 2010 National Center for Education Statistics (NCES) code for courses. Percentages do not sum to 100 because students can enroll in more than one course across multiple disciplines.

Source: Educational Longitudinal Study of 2002 (ELS:2002)

Appendix C: Replication of Table 4 with the number of courses. Regression coefficients for the number of STEM courses (Poisson regression) and the proportion of STEM credits (fractional regression) in the first year of college by occupational plans in 12th grade

Occupational plans in 12th grade:	Number of courses (Poisson regression)				Proportion of STEM credits (fractional regression)			
	STEM-aspiring (N=750)		Non-STEM-aspiring (N=2,560)		STEM-aspiring (N=750)		Non-STEM-aspiring (N=2,560)	
Model specification:	M1 unadjusted	M2 adjusted	M1 unadjusted	M2 adjusted	M1 unadjusted	M2 adjusted	M1 unadjusted	M2 adjusted
<i>Panel A: Estimated coefficients</i>								
Women	-0.23*** (0.047)	-0.11* (0.048)	-0.19*** (0.040)	-0.07 (0.039)	-0.33*** (0.098)	-0.13 (0.105)	-0.24*** (0.064)	-0.07 (0.065)
Total credits					0.04** (0.007)	0.02** (0.007)	0.02*** (0.004)	0.01** (0.004)
<u>Adjustment factors:</u>								
Social and demographic factors		Yes		Yes		Yes		Yes
Academic preparation		Yes		Yes		Yes		Yes
Attitudes		Yes		Yes		Yes		Yes
College factors		Yes		Yes		Yes		Yes
<i>Panel B: Change in the predicted gender gap across models (women- men)</i>								
Predicted gender gap in the number of courses	-0.689	-0.336	-0.343	-0.132	-0.079	-0.028	-0.042	-0.013
% gap explained		51%		62%		64%		74%

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Robust standard errors are in parentheses. *p<0.05, **p<0.01, ***p<0.001. Adjustment factors are detailed in Table 1. For brevity, we do not present the coefficients for all adjustment factors. They are available upon request.

The Poisson regression models account for exposure to overall courses

Source: ELS:2002

Appendix D: Coefficients from multinomial logit models predicting the field of degree attained by 2012 by occupational plans in 12th grade and gender (predicted probabilities are presented in Figure 2)

Outcome: Field of BA (ref=non-STEM)	Occupational plans: STEM-aspiring students				Non-STEM-aspiring students			
	Men		Women		Men		Women	
	Model #: (1)	(2)	(3)	(4)	(3)	(4)	(4)	(4)
	No degree	STEM	No degree	STEM	No degree	STEM	No degree	STEM
Percentage of STEM courses in first year	0.03*** (0.009)	0.06*** (0.011)	0.01 (0.008)	0.03** (0.010)	0.01 (0.006)	0.05*** (0.009)	0.01* (0.005)	0.05*** (0.009)
<i>Adjustment factors:</i>								
Social and demographic factors								
Academic preparation								
Attitudes								
College factors								
Total number of courses in first year	-0.22** (0.068)	-0.12 (0.087)	0.00 (0.069)	-0.02 (0.081)	-0.18*** (0.047)	-0.13 (0.104)	-0.05 (0.030)	-0.06 (0.071)
Constant	1.26 (1.788)	-6.47** (2.140)	3.86 (1.985)	-1.37 (2.629)	0.95 (1.122)	-4.85** (1.739)	3.33*** (0.807)	-6.57** (2.187)
Observations	420		330		990		1570	
Pseudo R-squared	0.246		0.203		0.198		0.158	
AIC	181428.6		127162.1		402529.5		575577.7	
BIC	181687.3		127405.2		402842.8		575920.6	

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Adjustment factors are detailed in Table 1. The base outcome is a non-STEM degree. For brevity, we do not present the coefficients for the adjustment factors. They are available upon request.

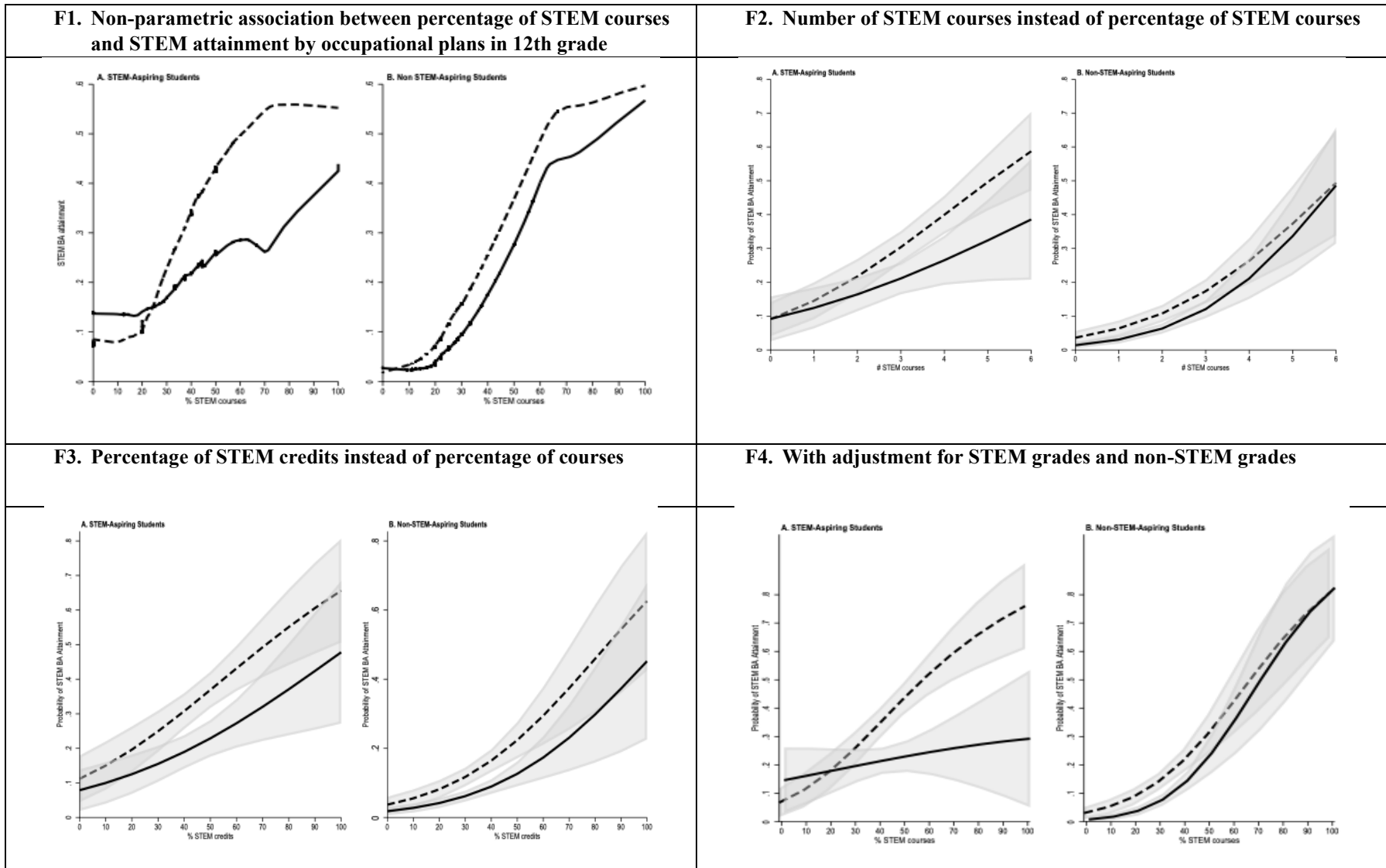
Source: ELS:2002

Appendix E: Coefficients from multinomial logit models predicting field of degree attained by 2012 in Table 5

E1: STEM-aspiring students (Panel A, N=750)									
Model #	(1)		(2)		(3)		(4)		
Outcome: Field of BA (ref=Non-STEM)	No degree	STEM	No degree	STEM	No degree	STEM	No degree	STEM	
Gender (women=1):	-0.70*** (0.209)	-1.15*** (0.228)	-0.53* (0.259)	-0.98*** (0.286)	-0.51 (0.270)	-0.90** (0.297)	0.23 (0.449)	-0.15 (0.662)	
Percentage of STEM courses in first year					0.02** (0.006)	0.04*** (0.007)	0.03*** (0.008)	0.05*** (0.010)	
Total number of courses in first year					-0.09* (0.044)	-0.05 (0.057)	-0.09* (0.044)	-0.06 (0.058)	
Female # Percentage of STEM courses in first							-0.02* (0.011)	-0.02 (0.014)	
Adjustment factors:									
Social and demographic factors				Yes		Yes		Yes	
Academic preparation				Yes		Yes		Yes	
Attitudes				Yes		Yes		Yes	
College factors				Yes		Yes		Yes	
Constant	0.49** (0.151)	0.39* (0.156)	3.33** (1.178)	-3.02* (1.376)	3.37** (1.224)	-3.60* (1.410)	2.96* (1.246)	-3.95** (1.421)	
Pseudo R-squared	0.02		0.16		0.20		0.20		
AIC	400015		344913		328738		327480		
BIC	400034		345200		329043		327794		
E2: Non-STEM-aspiring students (Panel B, N=2560)									
Gender (women=1):	-0.05 (0.111)	-0.89*** (0.182)	0.22 (0.125)	-0.62** (0.206)	0.24 (0.125)	-0.53* (0.213)	0.37 (0.215)	-0.53 (0.386)	
Percentage of STEM courses in first year					0.01** (0.004)	0.05*** (0.006)	0.02* (0.006)	0.05*** (0.009)	
Total number of courses in first year					-0.09*** (0.026)	-0.09 (0.058)	-0.09*** (0.026)	-0.09 (0.058)	
Female # Percentage of STEM courses in first							-0.01 (0.008)	-0.00 (0.010)	
Adjustment factors:									
Social and demographic factors				Yes		Yes		Yes	
Academic preparation				Yes		Yes		Yes	
Attitudes				Yes		Yes		Yes	
College factors				Yes		Yes		Yes	
Constant	-0.21* (0.093)	-1.19*** (0.127)	1.53* (0.631)	-4.90*** (1.232)	2.06** (0.658)	-4.77*** (1.320)	2.00** (0.662)	-4.78*** (1.329)	
Pseudo R-squared	0.01		0.12		0.16		0.16		
AIC	1185366		1049317		1005836		1005520		
BIC	1185389		1049679		1006222		1005917		

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Adjustment factors are detailed in Table 1. The base outcome is a non-STEM degree. Source: ELS:2002

Appendix F: Replication of Figure 2 with alternative model and variable specifications



Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Source: ELS:2002

Appendix G: Results for students with undecided plans

G1. Replication of Tables 2 and 3 for students with undecided occupational plans. High-school sophomores in 2002 who attended four-year colleges

	Men (N=690)		Women (N=730)		Gender difference (M-W)	Sig
	Mean/percent	SD	Mean/percent	SD		
<i>Panel A: Field of degree in 2012 (similar to Table 2)</i>						
STEM/Biomed-MD	18.2		7.9			
Other fields	42.2		53.6			
No degree	39.5		38.5			
<i>Panel B: STEM momentum (similar to Table 3)</i>						
Has STEM course	0.86	0.35	0.79		0.07	*
# STEM courses	2.15	1.66	1.57	1.28	0.58	*
Prop. STEM courses	0.27	0.21	0.19	0.17	0.08	*
# STEM credits	6.80	6.75	5.27	5.78	1.53	*
Prop. STEM credits	0.28	0.23	0.21	0.19	0.07	*
Grades:						
STEM average	2.7	0.84	2.8	0.88	-0.10	
Non-STEM average	2.82	0.71	3.06	0.6	-0.24	*
STEM vs. non-STEM difference in grades	-0.19	0.73	-0.28	0.7	0.09	
<i>Panel C: Took at least one course in (similar to Appendix B)</i>						
Math and statistics	0.70		0.57		0.13	
Physical sciences	0.38		0.31		0.07	
Biological sciences	0.22		0.28		-0.06	
Engineering/ engineering technologies	0.13		0.02		0.11	
Computer and information science	0.18		0.09		0.09	
Natural resources/ agriculture/ multidisciplinary	0.04		0.06		-0.02	
Average number of lab courses	0.71		0.51		0.20	

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges.

Source: ELS:2002

G2. Replication of Table 4 for students with undecided plans. High-school sophomores in 2002 who attended four-year colleges.

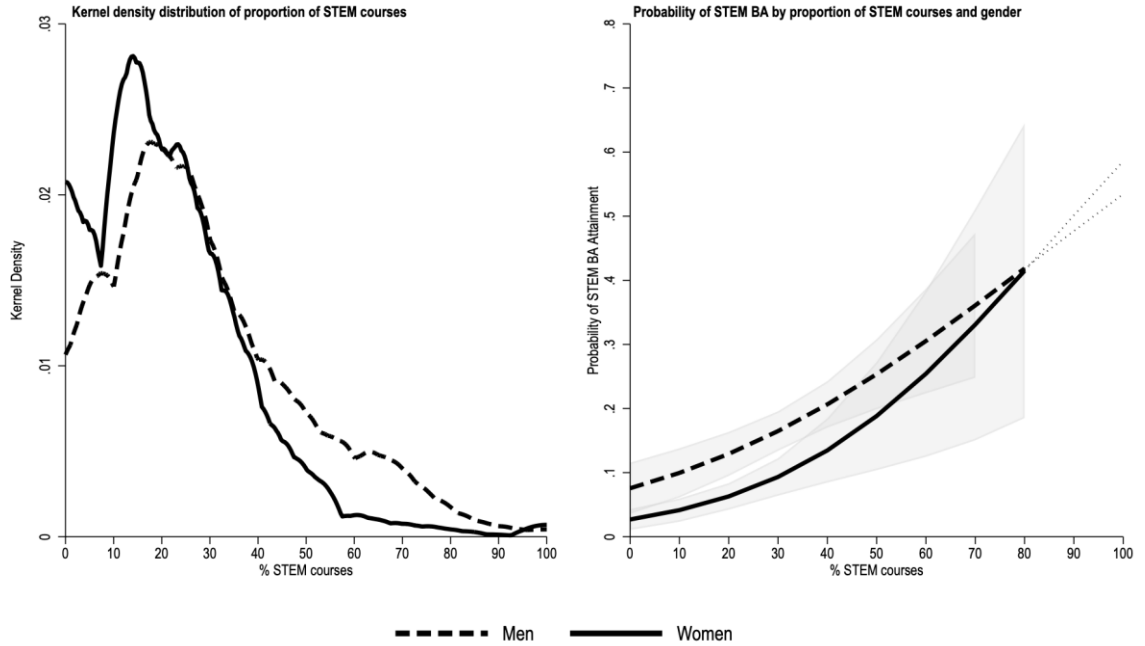
Selected coefficients from fractional logit regression predicting proportion of STEM courses

	M1:	M2:M1+	M3:M2+	M4:M3+	M5: M4+
Model specification:	Gender only	Social and demographic factors	High-school achievements and courses	Attitudes	College characteristics
Gender	-0.44*** (0.072)	-0.43*** (0.073)	-0.36*** (0.071)	-0.35*** (0.073)	-0.32*** (0.068)
Predicted gender gap (women-men)	-0.077	-0.076	-0.062	-0.060	-0.053
% of the gap explained		2%	20%	23%	31%

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. N=1420. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The models account for sociodemographic factors, academic preparation, attitudes, and college factors (see Table 1 in the main text). The models also account for the number of courses students took during their first year.

Source: ELS:2002

G3. Replication of Figures 1 (left) and 2 (right) for students with undecided plans



G4. Replication of Table 5 for students with undecided plans.

Predicted STEM degree attainment probability, estimated from multinomial logit models

Model specification	Adjusted probability of STEM degree attainment			Predictive scenarios	
	M1: Unadjusted	M2: M1+ Social background, academic preparation, attitudes, and college factors	M3: M2+ STEM momentum	Scenario 1: If men and women had men's distribution	Scenario 2: If both men and women had men's association between courses and attainment
Men	0.182	0.175	0.160	0.175	0.175
(se)	(0.020)	(0.017)	(0.016)	(0.017)	(0.017)
Women	0.079	0.083	0.094	0.114	0.141
(se)	(0.012)	(0.013)	(0.015)	(0.022)	(0.020)
Predicted gender gap (M-W)	0.103	0.092	0.066	0.061	0.034
% explained		11%	36%	41%	67%

Note: Data are weighted. High-school sophomores in 2002 who attended four-year colleges. Standard errors of the predictions are in parentheses. *Source:* ELS:2002.

Appendix H: Assessment of fit for the models in Table 5

Model fit indicators from cross-validation analyses for models comparable to Models 2-4 in Table 5 (STEM predictions only) with 80%-20% division of the samples. Model 2 contains gender, social and academic background, attitudes, and college factors. Model 3 adds the proportion of STEM courses (STEM momentum). Model 4 adds gender*STEM momentum interactions.

Model #	M2	M3	M4
Panel A: STEM-aspiring students			
Positive predictive value (precision): TP/(TP+FP)	0.50	0.44	0.61
Sensitivity (true positive rate; recall): TP/(TP/ FN)	0.30	0.34	0.48
Specificity (true negative rate; selectivity): TN/(TN+FP)	0.87	0.83	0.88
Balanced accuracy: average between the recall and specificity	0.59	0.59	0.68
F1 statistics: (2*precision*recall)/(precision + recall)	0.38	0.38	0.53
Mean absolute percentage error (MAPE)	0.21	0.18	0.15
Symmetric mean absolute percentage error (SMAPE)	0.59	0.61	0.48
Panel B: Non-STEM-aspiring students			
Positive predictive value (precision): TP/(TP+FP)	0.64	0.35	0.50
Sensitivity (true positive rate; recall): TP/(TP/ FN)	0.02	0.03	0.03
Specificity (true negative rate; selectivity): TN/(TN+FP)	0.99	0.97	0.97
Balanced accuracy: average between the recall and specificity	0.56	0.58	0.63
F1 statistics: (2*precision*recall)/(precision + recall)	0.21	0.25	0.37
Mean absolute percentage error (MAPE)	0.09	0.06	0.06
Symmetric mean absolute percentage error (SMAPE)	0.20	0.17	0.16

Notes: TP=true positive, TN=true negative, FP=false positive, and FN=false negative. Data are weighted. High-school sophomores in 2002 who attended four-year colleges. To incorporate complex weights and clustered standard errors in these models, we collapsed the non-STEM categories and estimated the predictions of STEM BA vs. other outcomes using logit models rather than multinomial.

Source: ELS:2002