Supplemental Material

Two Brief Interventions to Mitigate a "Chilly Climate" Transform Women's Experience, Relationships, and Achievement in Engineering

By G. M. Walton et al., 2014, *Journal of Educational Psychology*

http://dx.doi.org/a0037461

Retention Rates

Retention rates were adequate and similar to past research (e.g., Walton & Cohen, 2011). At the end of the intervention session, students were asked to authorize the release of their university academic records. A total of 83.77% of students agreed to do so and could be matched to institutional records. This rate did not vary by gender or major type, $\chi^2(1, N = 228) < 2.25$, *p*s > .10; by condition, $\chi^2(2, N = 228) < 1$; or by condition for women, $\chi^2(2, N = 92) < 2.90$, $p > .20$, or for men, $\chi^2(2, N = 136) < 3.45$, $p > .15$. Students who authorized the release of their academic records did not differ on any preintervention measure from students who did not, *t*s < 1.35, *p*s > .15. Analyses of academic performance are thus based on 191 students (73 women, 118 men).

Among participating students, 91.23% completed at least one daily-diary survey, 80.26% completed at least three, and 64.04% completed all six. There was no difference by student gender, major, or condition in the mean number of daily-diary surveys completed, *F*s < 1.30, *p*s> .25. Exploratory analyses found no evidence that the number of daily-diary surveys completed moderated the daily-diary results.

About 4 months after the intervention, 67.54% of participating students responded to our communications and completed one or both second-semester surveys. Survey-completion rates did not vary by gender or major type, $\chi^2(1, N = 228) < 1.25$, *p*s >.25, or by condition, $\chi^2(2, N = 228) < 1.60$, *p*> .40. There was no difference between completers and noncompleters on the preintervention measures of students' evaluation of their current experience in engineering, their prospects for future success in engineering, or Percentage of Students' Friends of Each Gender × Major Category, *t*s < 1.40, *p*s >.15. However, completers had more positive implicit norms about female engineers in the preintervention survey, $t(221) = 2.73$, $p = 0.007$. There was no interaction between completion-status and either student gender, major type, or experimental condition, or any higher order interaction on this measure, $F\text{s} < 2.15$, $p\text{s} > .10$. Additionally, as noted, analyses of all second-semester measures, including implicit norms, controlled for relevant preintervention measures. Analyses of second-semester measures are thus based on 154 students (66 women, 88 men).

Classification of Engineering Majors

We categorized majors as gender-diverse or as male-dominated instead of treating the representation of women in each major as a continuous variable for three reasons.

First, rather than varying in a linear fashion, we expected that social marginalization and psychological threat would either arise in a setting or not in a manner consistent with the concept of critical mass (Etzkowitz, Kemelgor, Neuschatz, Uzzi, & Alonzo, 1994). That is, threat may not meaningfully differ between majors with 7% women and majors with 12% women. But majors with 10% women (no critical mass) may elicit a high level of threat while majors with 30% women (critical mass) may not. Thus, we anticipated that a dichotomous classification would index women's experiences more closely.

Second, as noted in the main text, this classification simultaneously tracks social stereotypes.

Third, majors tended to cluster below or above 20% women. As noted, across the 3 years women represented 32.57% of students enrolled in gender-diverse majors and 10.01% of students enrolled in male-dominated majors. Of the gender-diverse majors, all but two had at least 34% women across the 3 years (the exceptions, civil and systems-design engineering, did not seem strongly male-typed). Of the male-dominated interventions, all but one had fewer than 12% women across the 3 years (the exception, nanotechnology, did seem male-typed). Thus, the dichotomous classification was appropriate.

Subsample *N*s

The 228 participating students fell into the 12 Major × Gender × Cells as follows. Sample sizes for GPA analyses—students who authorized the release of their academic records who could be matched to institutional records—are in parentheses ($N = 191$, 83.77% of the sample).

	Gender-Diverse Majors			Male-Dominated Majors		
	Control	Social-Belonging	Affirmation-Training	Control	Social-Belonging	Affirmation-Training
Men	25 (23)	20 (19)	16 (13)	27 (19)	26 (24)	22 (20)
Women	21 (18)	21 (14)	18 (13)	8 (8)	12 (10)	12 (10)

As noted, the small sample size is a limitation of this study. It is important to test the replicability of the results in future research with larger (and more heterogeneous) samples; this would also support additional tests of moderation and mediation. With this limitation, it is also important to keep in mind strengths of the results, including (a) their statistical significance (in analyses that take into account the sample size); (b) the simplicity and robustness of the analyses, (e.g., all available participants were retained, statistical assumptions were met, there were no outliers, covariates were included on an *a priori* basis, alternative analyses yield similar results, and the results are consistent across diverse measures); and the facts that the results (c) were predicted *a priori* and (d) cohere with and contribute to an existing literature.

Measure of Implicit Normative Evaluations of Female Engineers

Implicit norms were measured using the Implicit Association Test (IAT; Greenwald, Nosek, & Banaji, 2003; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Nosek, Greenwald, & Banaji, 2007) modified to assess implicit norms (Peach, Yoshida, Spencer, Zanna, & Steele, 2011; Yoshida, Peach, Zanna, & Spencer, 2012). Participants were presented with category labels in the upper left and upper right of the computer screen. They were asked to categorize a series of words and images as quickly and as accurately as possible using keys on the left and right side of the keyboard to indicate to which category each word or image belonged. There were two practice blocks. In one, participants categorized words such as “most people like” or “most people don’t like” (e.g., “party,” “disease”), with *most people* defined as “most undergraduates at your university." In the other, participants categorized images as "female engineers" or "objects" (e.g., images of women building computers, of women doing math; images of desks, images of staplers).

During two subsequent critical blocks, participants used the combined categories of "most people like" and "female engineers" (vs. "most people don't like" and "objects") (Block 3) and, after a third practice block, "most people don't like" and "female engineers" (vs."most people like" and "objects"; Block 5). Order of critical blocks was not counterbalanced (Nosek et al., 2007). For individuals who hold negative associations with most people's evaluation of female engineers, the task should be more difficult when "most people like" and "female engineers" share a response key (Block 3) than when "most people don't like" and "female engineers" share a key (Block 5). They should thus be slower to respond in the former condition than the latter. IAT scores represent the difference between average response times in these critical blocks. Higher scores indicate more positive implicit normative evaluations of female engineers. We used the D600 algorithm to calculate IAT scores (Greenwald et al., 2003). The magnitude of the D-score is similar to an effect size for each individual participant.

Implicit norms were assessed in both the preintervention survey and in the second-semester surveys. On the preintervention assessment, three participants had high error rates (>30%; all others <20%); their scores were replaced with the Gender × Major mean. This has no effect on analyses. Additionally, because this measure was skewed, it was square-root-transformed prior to analysis.

To calculate implicit norms in the second semester, we averaged scores on the two second-semester assessments for participants who completed both assessments. For participants who had a high error rate (>20%) on one second-semester assessment but not the other, we used the score from the assessment with the lower error rate. For participants who completed only one second-semester assessment, we used the score from that assessment. Three participants had moderately high error rates (20%–33%) on either both second-semester assessments or the only second-semester assessment they completed. Primary analyses retain these participants' scores; dropping them yields similar results.

Intervention Session

Representative quotations from upper-year engineering students. For the complete quotations attributed to upper year engineering students in the social-belonging and affirmation-training conditions, see Supplemental Appendix S1.

Coding of students' "saying-is-believing" writings. To confirm that students were sensitive to the divergent content of the two interventions and the study-skills control condition, we coded the essays and letters students in the first cohort wrote. Two coders, blind to participants' condition, gender, and major, coded participants' written materials along six dimensions. Two dimensions assessed whether each participant's writings expressed each aspect of the key message conveyed in each condition:

1. *I/many students begin university with inadequate study skills*. (Study Skills Code #1)

2. *I/many students learn new study skills in university*. (Study Skills Code #2)

3. *I/many students worry at first about belonging in university*. (Social-Belonging Code #1)

4. *Worries about belonging dissipate with time*. (Social-Belonging Code #2)

5. *I/many students experience stress/feel overwhelmed/feel tunnel vision at first in university*. (Affirmation-Training Code #1)

6. *I/many students cope with stress/find a sense of balance by thinking about/engaging in activities outside direct coursework relevant to my/their personal values and identity and/or think about coursework in ways that are relevant to my broader values and identity*. (Affirmation-Training Code #2)

Each coder assigned each dimension a 2 if it represented a strong or explicit theme, a 1 if the theme was implied, and a 0 if the theme was absent. Interrater reliability was adequate, Cohen's $\kappa = 0.77$. Therefore, we averaged the two coders' ratings. We then averaged across the two items designed to pick up the key message in each condition. Analysis of these scores yielded a Coding-Dimension × Condition interaction, $F(4, 122) = 63.86$, $p < 0.001$. This interaction was not further moderated by participant gender or major, Fs < 1.35, ps > 0.25. The means are reported below. Means with a different superscript within column and within row differ significantly (ts > 5.75, ps < 0.0001):

	Study Skills Theme (Range: 0-2)	Social-Belonging Theme (Range: 0-2)	Affirmation-Training Theme (Range: 0-2)
Study Skills Control	1.44[a]	0.23[b]	0.24[b]
Social-Belonging	0.25[b]	1.28[a]	0.35[b]
Affirmation-Training	0.24[b]	0.15[b]	1.81[a]

The distribution of scores at the extremes of the range illustrates the same clear condition effect:

	Study Skills Items Mean		Social-Belonging Items Mean		Affirmation-Training Items Mean	
	Percent of Participants ≤0.50	Percent of Participants ≤1.50	Percent of Participants ≤0.50	Percent of Participants ≤1.50	Percent of Participants ≤0.50	Percent of Participants ≤1.50
Study Skills Control	13.64%	63.64%	86.36%	4.50%	77.27%	0.00%
Social-Belonging	81.82%	4.55%	18.18%	54.55%	72.73%	0.00%
Affirmation-Training	90.00%	0.00%	90.00%	0.00%	0.00%	95.00%

Analyses of Preintervention Measures

Check on random assignment. As reported in the main text, there was no difference by condition on any preintervention measure, Fs < 1 (see Table S3). We also tested separately for differences among men and for differences among women between each intervention condition and the control condition along all seven preintervention measures. Across 28 total comparisons, none were significant, ts < 1.70, ps > 0.095. There was one marginal pattern—among women, between affirmation training and control on the percentage of friends who were male engineers, $t(205) = 1.66$, $p = .098$—and one trend—among women, between social belonging and control on the same outcome, $t(205) = 1.56$, $p = .12$. Combining the intervention conditions, the effect was not significant, $t(207) = 1.86$, $p = .065$. All other comparisons were nonsignificant, ts < 1.15, ps > .25. As 28 comparisons were tested, 1.40 would be expected to be significant at $p < .05$ on the basis of chance alone. As none were, we conclude random assignment was successful.

Baseline differences by gender and major-type. To examine baseline differences by gender and major-type, we conducted an ANVOA involving these two factors on each preintervention measure.

Analysis of students' evaluation of their current experience in engineering yielded a main effect of gender, $F(1, 219) = 5.12$, $p = .025$, with no effect of or interaction with major type, Fs < 1.30, ps > .25. Women evaluated their experience in engineering ($M = 4.92$) more negatively than men ($M = 5.20$). Women in male-dominated majors ($M = 4.99$) did not differ from women in gender-diverse majors ($M = 4.88$), $t<1$.

Analysis of students' assessment of their prospects of succeeding in engineering yielded a main effect of gender, $F(1, 219) = 6.67$, $p = .010$, with no effect of or interaction with major-type, Fs < 2.60, ps > .10. Women evaluated their prospects in engineering ($M = 66.07$) more negatively than men ($M = 71.43$). Women in male-dominated majors ($M = 69.67$) were somewhat more confident about their prospects than women in gender-diverse majors ($M = 64.07$), $t(219) = 2.04$, $p = .043$.

Analysis of students' implicit norms about female engineers yielded a main effect of gender, $F(1, 206) = 8.75$, $p =.003$, with no effect of or interaction with major type, Fs < 1. Women's implicit norms ($M = 0.62$) were more positive than men's ($M = 0.48$). There was no effect of major type among women, $t < 1$.

Analysis of the representation of male engineers in students' friendship groups yielded a main effect of gender, $F(1, 207) = 41.59$, $p < .001$, a main effect of major type, $F(1, 207) = 7.91$, $p = .005$, and no interaction, $F < 1$. Unsurprisingly, men and students enrolled in male-dominated majors had more male-engineer friends than women and students enrolled in gender-diverse majors ($M_{\text{men}/\text{male-dominated}} = 70.04\%$; $M_{\text{men}/\text{gender-diverse}} = 59.94\%$; $M_{\text{women}/\text{male-dominated}} = 46.45\%$; $M_{\text{women}/\text{gender-diverse}} = 35.69\%$).

Analysis of the representation of female nonengineers in students' friendship groups yielded only a trend on the main effect of gender, $F(1, 209) = 2.19$, $p = .14$. Women tended to have more female nonengineer friends ($M = 12.53\%$) than men ($M = 8.87\%$).

Analysis of gender identification yielded no main or interaction effects of either factor, Fs < 1.

Dummy Variables in Multiple Regression Analyses

As noted in the main text, data were analyzed using multiple regression including dummy codes for student gender, major type (gender-diverse vs. male-dominated), experimental condition, and all two- and three-way interactions. Separate analyses tested the combined and separate effects of the two interventions.

Analyses testing the combined effects of the two interventions:	Analyses testing the separate effects of the two interventions (including two dummy variables for condition):
Gender (male vs. female)	Gender (male vs. female)
Major-type (gender-diverse vs. male-dominated)	Major-type (gender-diverse vs. male-dominated)
Condition (combined intervention vs. control)	Condition Dummy 1 (e.g., social-belonging vs. affirmation-training and control)
Gender × Major Type	Condition Dummy 2 (e.g., affirmation-training vs. social-belonging and control)
Gender × Condition	Gender Major-type
Major Type × Condition	Gender × Condition Dummy 1
Gender × Major Type × Condition	Gender × Condition Dummy 2
	Major Type × Condition Dummy 1
	Major Type × Condition Dummy 2
	Gender × Major Type × Condition Dummy 1
	Gender × Major Type × Condition Dummy 2

Supplemental Analyses

For each supplemental analysis, we conducted a multiple regression and focused on four key statistical tests:

(1) the Gender × Major × Condition interaction;

The three critical contrasts for women in male-dominated majors:

(2) the two interventions versus control;

(3) social-belonging versus control; and

(4) affirmation-training versus control.

First-year engineering GPA. We conducted preliminary analyses to ensure that primary analyses met statistical assumptions. For instance, as noted, there were no outliers. There was also no difference in variance by participant gender, major type, or condition, *F*s < 2.00, *p*s $>$ .15. There was a negative skew, $Z=-2.88$, $p=0.004$. After a square-root correction removed this skew, $Z = 1.53$, $p = .13$, analyses yielded an identical pattern of results as did analyses of the uncorrected outcome. Therefore, primary analyses examine the uncorrected variable. In analyses of the square-root-corrected outcome, (1) the Gender × Major × Condition interaction and (2–4) the three critical contrasts for women in male-dominated majors were all significant, $B = 1.63$, $t(182) = 2.26$, $p = .025$; $B = 1.26$, $t(182) = 2.71$, $p = .007$; $B = 1.27$, $t(178) = 2.40$, $p = .018$; and, $B = 1.24$, $t(178) = 2.34$, $p = .020$, respectively.

To further check on the robustness of intervention effects, we subjected first-year engineering GPA to two further analyses. First, we examined the outcome without the mean within-major GPA covariate. The results were similar: In this analysis, (1) the Gender × Major × Condition interaction was marginal and (2–4) the three critical contrasts for women in male-dominated majors were all significant, $B = 12.85$, $t(183) = 1.87$, $p = .063$; $B = 10.95$, $t(183) = 2.46$, $p = .015$; $B = 11.44$, $t(179) = 2.25$, $p = .025$; and $B = 10.47$, $t(179) = 2.06$, $p = .041$, respectively.

Second, we examined the difference between participants' first-year GPA and the mean GPA earned by students in their major. Again, the results were similar: In this analysis, (1) the Gender × Major × Condition interaction and (2–4) the three critical contrasts for women in male-dominated majors were all significant, $B = 16.21$, $t(183) = 2.45$, $p = .015$; $B = 11.61$, $t(183) = 2.71$, $p= .007$; $B = 11.77$, $t(179) = 2.41$, $p = .017$; and $B = 11.44$, $t(179) = 2.35$, $p = .020$, respectively.

Daily diaries: Perception of daily events. Primary analyses examined the sum-total perceived importance of negative events averaged across days, with the sum-total perceived importance of positive events averaged across days controlled. We conducted three supplementary analyses.

First, we examined the average daily perceived importance of negative events without the covariate. The results were similar: In this analysis, (1) the Gender × Major × Condition interaction and (2–4) the three critical contrasts for women in male-dominated majors were all significant, $B = -5.29$, $t(183) = -2.23$, $p = .027$; $B = -3.79$, $t(183) = 2.33$, $p = .021$; $B = -3.71$, $t(179) = -1.99$, $p = .048$; and $B = -3.86$, $t(179) = -2.18$, $p = .031$, respectively.

Second, we examined the difference score—the average daily perceived importance of positive events minus the average daily perceived importance of negative events. The results paralleled the covariate analysis: In this analysis, (1) the Gender × Major × Condition interaction was a trend, $B = 2.86$, $t(195) = 1.36$, $p = .17$, and (2–4) the three critical contrasts for women in male-dominated majors were all significant or marginal, $B = 3.31$ $t(195) = 2.32$, $p = .021$; $B = 2.87$ $t(179) = 1.77$, $p = .079$; and $B = 3.67$ $t(191) = -2.34$, $p = .020$, respectively.

Finally, we examined the rated valence measure: The daily average sum-total valence of negative events with the daily average sum-total valence of positive events controlled. The pattern was similar but weaker: (1) the Gender × Major × Condition interaction was a trend, $B = -2.57$, $t(193) = -1.44$, $p = .15$, and the three critical contrasts for women in male-dominated majors (2–4) were all marginal or trending, $B = -2.20$ $t(193) = -1.83$, $p = .069$; $B = -2.14$, $t(189) = -1.56$, $p = .12$; and $B = -2.25$, $t(189) = -1.70$, $p = .090$, respectively.

Daily diaries: Stress. Primary analyses were conducted to examine students' confidence in their ability to handle daily school stress. We conducted three supplementary analyses.

First, we examined primary appraisals of school stress—how much school stress students reported experiencing. There was no effect of condition on this outcome among women in male-dominated majors, either combining the two interventions or testing them separately, ts < 1.

Second and third, we calculated the average daily level of stress students reported experiencing with personal relationship partners (i.e., "family members," "close friends," "romantic partners," and "a person you are interested in dating but are not dating") and how confident students felt they could handle these nonschool sources of stress. Illustrating the specificity of the intervention effects, there was no effect of condition on either primary or secondary appraisals of nonacademic relational stress for women in male-dominated majors, both combining the two interventions and testing them separately, ts < 1.

Daily diaries: Self-esteem. We conducted primary analyses to examine the level and stability of students' self-esteem across days and secondary analyses to examine the level (mean) and stability (reverse-scored standard deviation) of self-esteem separately. In analyses of self-esteem level, (1) the Gender × Major × Condition interaction and (2–4) the three critical contrasts for women in male-dominated majors were all trends, $B = 0.75$ $t(196) = 1.20$, $p = .23$; $B = 0.54$ $t(196) = 1.28$, $p = .20$; $B = 0.50$ $t(192) = 1.04$, $p = .30$; and $B = 0.58$ $t(192) = 1.23$, $p = .22$, respectively. In analyses of self-esteem stability, these effects (1–4) were all significant, $B = 0.93$, $t(186) = 3.61$, $p < .001$; $B = 0.39$ $t(186) = 2.23$, $p = .027$; $B = 0.42$ $t(182) = 2.07$, $p = .040$; and $B = 0.36$, $t(182) = 1.91$, $p = .058$, respectively.

Representation of female engineers in students' friendship groups. Notably, in male-dominated majors, women reported marginally *fewer* female engineer friends in the intervention conditions (combined) than in the control condition, $B = -0.16$, $t(136) = -1.80$, $p = .074$. Although laboratory and longitudinal field research show that exposure to successful women in STEM can buffer women against psychological threat and improve their outcomes (Marx & Roman, 2002; Stout, Dasgupta, Hunsinger, & McManus, 2011), less research has examined the role of peer relationships among women in STEM. Such relationships could play an important protective role; indeed, in gender-diverse majors, the higher representation of women in students' majors presumably forestalled feelings of social marginalization (Stout et al., 2011; Yoshida et al, 2012). But in settings where women are severely underrepresented, an emphasis on developing ingroup relationships may risk creating in women the feeling of occupying a social bunker (Akcinar, Carr, & Walton, 2011). In these contexts, it may be more helpful to either integrate with members of the dominant group or find ways to express broader aspects of self-identity (Walton & Carr, 2012).

Implicit normative evaluations of female engineers. Primary analyses were conducted to examine effects of the social-belonging intervention on implicit norms about female engineers in the second semester controlling for the preintervention assessment.

As noted, we also examined how implicit norms changed over time. We conducted a mixed-model ANOVA with time (preintervention vs. second semester) as a within-subject factor and gender, major type, and condition (social-belonging vs. affirmation-training/control) as between-subjects factors. This analysis yielded a main effect of gender (M_{women} = 0.61, M_{men} = 0.47), $F(1, 144) = 10.52$, $p = .001$, and a main effect of time, $F(1, 144) = 4.59$, $p = .034$. Replicating Yoshida and colleagues (2012), students' implicit norms about female engineers became more negative as they spent more time in engineering ($M_{preintervention}$ = 0.58; $M_{second\ semester}$ = 0.50). Finally, there was a marginal four-way interaction, $F(1, 144) = 3.10$, $p = .080$.

We broke down this interaction by condition. In the control and affirmation-training conditions (i.e., absent effective intervention to change implicit norms), only the main effects of gender and time were significant, $F(1, 144) = 4.51$, $p = .035$, and $F(1, 144) = 12.88$, $p < .001$, respectively. Women had more positive implicit norms ($M = 0.55$) than men ($M = 0.44$). and for both men, $F(1, 144) = 7.80$, $p = .006$, and women, $F(1, 144) = 5.45$, $p = .021$, implicit norms became more negative over time ($M_{preintervention}$ = 0.57; $M_{second\ semester}$ = 0.42). Both patterns replicate the findings of Yoshida and colleagues. The full pattern of means is as follows:

Implicit normative evaluations of female engineers: Control and affirmation-training conditions.

	Gender-Diverse Majors		Male-Dominated Majors	
	Men	Women	Men	Women
Preintervention	0.54 (0.06)	0.62 (0.06)	0.51 (0.06)	0.63 (0.08)
Second Semester	0.38 (0.06)	0.52 (0.06)	0.35 (0.06)	0.43 (0.08)
Effect of Time	$t(144) = -1.94$, p= .054	$t(144) = -1.31$, p = .19	$t(144)$=-2.01, p = .047	$t(144)$ = --1.94, p = .055

Note. Higher values represent "most people like"="female engineers." Results are based on the sample of students with data at both time points, n=102. Standard errors shown in parentheses.

In the social-belonging condition, the main effect of gender was again significant, $F(1, 144) = 6.14$, p =.014. However, the main effect of time was nonsignificant, $F < 1$. Instead, there was a marginal Time × Major × Gender interaction, $F(1, 144) = 2.94$, $p = .088$. As shown below, men in male-dominated majors continued to show more negative implicit norms about female engineers over time; women in gender-diverse majors showed the same trend though nonsignificantly. By contrast, women in male-dominated majors as well as men in gender-diverse majors showed nonsignificant reversals—more positive implicit norms about female engineers over time. Thus, the social-belonging intervention reversed the normative decline in implicit norms about female engineers typical over time.

Implicit normative evaluations of female engineers: Social-belonging condition.

	Gender-Diverse Majors	Male-Dominated Majors

	Men	Women	Men	Women
Preintervention	0.53 (0.09)	0.67 (0.09)	0.51 (0.07)	0.68 (0.12)
Second Semester	0.62 (0.09)	0.60 (0.09)	0.35 (0.08)	0.78 (0.12)
Effect of Time	$t < 1$	$t < 1$	$t(144) = -1.71, p = .090$	$t < 1$

Note. Higher values represent "most people like"="female engineers." Means at both time points are based on the sample of students with data at both time points, n=50. Standard errors shown in parentheses.

Reported frequency of jokes about female engineers. In the preintervention and second-semester surveys, students reported how often they had heard "jokes about female engineers," "sexist jokes about female engineers," "jokes about engineers that were based on ethnicity," and "jokes about engineers who are not originally from Canada" in their university (1 = *never*, 7 = *frequently*). The first two measures correlated at each time point (rs = .62 and .72, respectively, ps < .001) as did the final two measures (rs = .80 and .90, ps < .001), so they were combined to form respective scales. Because the reported frequency of jokes about female engineers was highly positively skewed at both time points (Zs > 4.65, ps < .001), primary analyses examined change scores, which were not skewed, Zs < 1.84, p = .066. In addition, because preliminary analyses revealed an overall main effect of condition (i.e., social-belonging vs. affirmation-training/control), we used analysis of covariance (ANCOVA) rather than multiple regression. To isolate effects on jokes about female engineers, we controlled for the reported frequency of jokes about engineers based on ethnicity and national origin, both the average across the two time points and the change score. Both covariates were predictive, $F(1, 142) = 3.50$, p =.063, and $F(1, 142) = 21.24$, $p < .001$, respectively.

The analysis yielded a marginal main effect of gender, $F(1, 142) = 3.40$, $p = .067$, d=0.36, a significant main effect of condition, $F(1, 142) = 7.11$, $p = .009$, $d = 0.39$, and a marginal Major × Condition interaction, $F(1, 142) = 2.93$, $p = .089$. As noted in the main text, in the theoretically relevant group, women in male-dominated majors, those in the control and affirmation-training conditions reported hearing more jokes about female engineers in the second-semester than preintervention survey ($M_{diffadj}$= 0.95) but those in the social-belonging condition showed no such change ($M_{diffadj}$= –0.04), a marginal condition difference, $t(142) = 1.77$, $p = .079$, $d = 0.71$. The control condition ($M_{diffadj}$ = 1.08) and the affirmation-training condition ($M_{diffadj}$= 0.85) did not differ, $t < 1$, and each tended to differ from the social-belonging condition, $t(138) = 1.71$, $p = .090$, $d = 0.80$, and $t(138) = 1.41$, $p = .16$, $d = 0.64$, respectively.

Interestingly, men in male-dominated majors also showed a difference between the social-belonging ($M_{diffadj}$=-0.40) and affirmation-training/control conditions ($M_{diffadj}$=0.63), $t(138) = 2.69$, $p = .008$, $d = 0.73$. However, unlike among women, among men it was the affirmation-training condition ($M_{diffadj} = 1.04$) that differed from the other two conditions, differing from the control condition ($M_{diffadj}$= 0.09), $t(138) = 2.03$, $p = .044$, and from the social-belonging condition ($M_{diffadj} = -0.40$), $t(138) = 3.35$, $p = .001$, $d = 1.03$; the control and social-belonging conditions did not differ, $t < 1.10$, $p > .25$.

In contrast to these results among students in male-dominated majors, in gender-diverse majors, the condition effect was nonsignificant for men, women, and combined, *t*s < 1.

We conducted two supplemental analyses. First, we tested the same analysis without the covariates. The results were similar. The main effect of condition remained significant, $F(1, 144) = 4.79$, $p = .030$, and the Major × Condition interaction became a trend, $F(1, 144) = 2.15$, $p = .14$.

Second, we conducted a repeated-measures ANCOVA including time of assessment (preintervention vs. second-semester) as a within-subject factor, gender, major, and condition as between-subjects factors, and the two covariates mentioned above. The Time × Condition interaction in this analysis is identical to the main effect of condition on the change score reported above. However, the repeated-measures analysis allowed us to formally examine change over time. Among women in male-dominated majors, those in the control and affirmation-training conditions reported hearing more jokes about female engineers in the second semester than before the intervention, combined: $t(142) = 2.88$, $p = .005$; control: $t(138) = 2.25$, $p = .026$; affirmation-training: $t(138) = 1.89$, $p = .060$. But women in the social-belonging condition showed no change over time, $t < 1$.

Statistical tests of mediation. As noted, statistical powers limit the value of statistical tests of mediation in the present study (these analyses suffer from a lack of power even with reasonable sample sizes; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2004). However, we conducted a series of exploratory meditational analyses focusing on women in male-dominated majors. As described below, some of these analyses yielded intriguing patterns; others were less fruitful. Given the number of analyses conducted, significant results should be viewed tentatively. In addition, nonsignificant analyses (i.e., where a condition effect remained significant, controlling for the ostensible mediator and/or the mediator was nonsignificant) could reflect either a lack of power or a lack of mediation. As Cohen and colleagues (2009) wrote of a value-affirmation-intervention, “the intervention might have discrete effects on a host of education-relevant psychological and behavioral outcomes” (p. 402)

Analyses examined whether each measure that exhibited a condition effect among women in male-dominated majors mediated any downstream condition effect. First, we tested whether women’s reports of the quality of their current experience in engineering immediately after the intervention mediated any downstream effect (i.e., on daily-diary measures, second-semester measures, and GPA). Second, we assessed whether each daily-diary measure mediated any downstream effect (i.e., on second-semester measures and GPA). In analyses examining outcomes that both interventions affected, the two intervention conditions were combined. In analyses examining outcomes that only one intervention affected (e.g., friendships with male engineers), the control condition and the second intervention condition were combined. We also explored interactions between condition assignment and ostensible mediators (i.e., if a given outcome predicted a subsequent outcome more strongly in one condition than another). In each analysis, outcomes that were also assessed at baseline are residual scores with the baseline measurement controlled.

Two significant effects emerged. First, women’s reports of the quality of their current experience in engineering immediately after the intervention mediated the intervention effect on the degree to which women saw daily adversities as manageable. This analysis examined an outcome combining women’s perception of the “importance” of daily negative events relative to daily positive events (reverse-scored) and women’s reports of their confidence in their ability to handle daily school stress. We standardized and averaged the two measures. We combined these measures because they correlated, $r = .38, p < .001$, assessed the same critical construct, and showed similar meditational patterns. Among women in male-dominated majors, there was a significant effect of the interventions on this measure assessing the degree to which women saw daily adversities as manageable, ß = .52, $t(27) = 3.20, p = .004$ ($R^2 = .27$). Controlling for women’ reports of the quality of their experience in engineering immediately after the intervention rendered this condition effect nonsignificant, ß = 0.27, $t(26) = 1.53, p=0.14$;

simultaneously, the mediator was significant, ß = .45, $t(26) = 2.50$, $p = .019$ ($R^2 = .42$). The mediation was significant, asymmetric distribution of products test (ADPT) 95% confidence interval (CI) [0.30, 0.86], $p < .0.05$.

Second, women's reports of the quality of their current experience in engineering immediately after the intervention mediated the intervention effect on the same outcome in the second semester. Among women in male-dominated engineering majors, there was a significant effect of the interventions (combined) on this outcome, ß = .45, $t(19) = 2.20$, $p = .040$ ($R^2 = .20$). Controlling for women's reports of the quality of their experience in engineering immediately after the intervention eliminated the condition effect, ß = .13, $t < 1$; simultaneously, the mediator was significant, ß = .57, $t(18) = 2.62$, $p = .017$ ($R^2 = .42$). The mediation was significant, ADPT 95% CI [0.42, 1.14], $p < .05$.

Although these analyses are exploratory, they suggest that the interventions may have helped women view daily adversities as manageable and sustained the perception of a positive experience in engineering over time by inducing a more positive perspective immediately after the treatment.

Supplemental References

Akcinar, E. N., Carr, P. B., & Walton, G. M. (2011). Interactions with men and Whites matter too. *Psychological Inquiry, 22,* 247–251

Aronson, E. (1999). The power of self-persuasion. *American Psychologist*, *54*, 875–884.

Dal Cin, S. D., MacDonald, T. K., Fong, G. T., Zanna, M. P., & Elton-Marshall, T. E. (2006). Remembering the message: The use of a reminder cue to increase condom use following a safer sex intervention. *Health Psychology, 25,* 438–443.

Etzkowitz, H., Kemelgor, C., Neuschatz, M., Uzzi, B., & Alonzo, J. (1994). The paradox of critical mass for women in science, *Science, 266,* 51–54.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28,* 1183–1193.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), S*ocial psychology and the unconscious: The automaticity of higher mental processes (*pp. 265–292). New York, NY: Psychology Press.

Peach, J. M., Yoshida, E., Spencer, S. J., Zanna, M. P., & Steele, J. R. (2011). Recognizing discrimination explicitly while denying it implicitly: Implicit social identity protection. *Journal of Experimental Social Psychology*, *47*, 283–292.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality Social Psychology*, *100*, 255–270.

Walton, G. M., & Carr, P. B. (2012). Social belonging and the motivation and intellectual achievement of negatively stereotyped students. In M. Inzlicht & T. Schmader (Eds.). *Stereotype threat: Theory, processes, and application* (pp. 89–106). New York, NY: Oxford University Press.

Walton, G. M., Paunesku, D., & Dweck, C. S. (2012). Expandable selves. In M. R. Leary & J. P. Tangney (Eds.) *Handbook of self and identity* (pp. 141–154). New York, NY: Guilford Press.

Yeager, D. S., & Walton, G. M. (2011). Social–psychological interventions in education: They're not magic. *Review of Educational Research, 81*, 267–301.

Yoshida, E., Peach, J. M., Zanna, M. P., & Spencer, S. J. (2012). Not all automatic associations are created equal: How implicit normative evaluations are distinct from attitudes and uniquely predict meaningful behavior. *Journal of Experimental Social Psychology, 48,* 694-–706.

Table S1

Supplemental Analyses of Individual Attitudinal Measures Assessed in the Intervention Session and in the Second Semester

Construct (Composite)	Time of Assessment	Measure	Intervention Effect in Gender-Diverse Majors (>20% women)		Intervention Effect in Male-Dominated Majors (<20% women)		Gender × Major × Condition
			Men	Women	Men	Women	
Evaluation of Current Experience in Engineering (7-point scales)	Intervention Session	Sense of Belonging	$B = 0.27, t = 1.56, p = .12, d = 0.30$	$B = -0.11, t < 1$	$B = 0.02, t < 1$	$B = 0.59, t = 2.19, p = .030, d = 0.65$	$B = 0.94, t = 2.36$ $p = .019$
		Self-Efficacy	$B = -0.19, t < 1$	$B = -0.24, t < 1.15$	$B = -0.11, t < 1$	$B = 0.56, t = 1.80, p = .073, d = 0.48$	$B = 0.73, t = 1.57, p = .12$
		Enjoyment	$B = 0.04, t < 1$	$B = -0.18, t < 1$	$B = 0.02, t < 1$	$B = 0.75, t = 2.63, p = .009, d = 0.65$	$B = 0.96, t = 2.27$ $p = .024$
	Second Semester	Sense of Belonging	$B = 0.10, t < 1$	$B = -0.21, t < 1$	$B = 0.30, t = 1.33, p = .19$	$B = 0.56, t = 1.79, p = 0.075, d = 0.64$	$B = 0.57, t = 1.14$ $p = .25$
		Self-Efficacy	$B = 0.30, t < 1$	$B = -0.52, t = -1.66, p = .099, d = -0.44$	$B = 0.15, t < 1$	$B = 0.77, t = 1.70, p = 0.092, d = 0.66$	$B = 1.44, t = 2.01, p = .046$
		Enjoyment	$B = -0.18, t < 1$	$B = -0.13, t < 1$	$B = -0.09, t < 1$	$B = 0.71, t = 1.74, P = 0.084, d = 0.63$	$B = 0.76, t = 1.17, p = .24$
Perceived Prospects of Succeeding in Engineering (100-point scales)	Intervention Session	Possible Selves	$B = -0.83, t < 1$	$B = 2.29, t < 1$	$B = -1.63, t < 1$	$B = 3.81, t < 1.05$	$B = 2.32, t < 1$
		Self-Assessed Potential to Succeed	$B = 0.35, t < 1$	$B = -0.37, t < 1$	$B = -4.71, t = -1.87, p = .063, d = -0.28$	$B = -2.02, t < 1$	$B = 3.42, t < 1$
	Second Semester	Possible Selves	$B = -0.27, t < 1$	$B = -2.38, t < 1$	$B = 0.25, t < 1$	$B = 12.46, t = 2.35, p = .020, d = 0.74$	$B = 14.32, t = 1.71, p = .090$
		Self-Assessed Potential to Succeed	$B = 2.61, t < 1$	$B = -11.84, t = -2.41, p = .017, d = -0.61$	$B = -1.37, t < 1$	$B = 20.58, t = 2.42, p = .017, d = 0.88$	$B = 33.03, t = 2.95, p = .004$

Note. Primary analyses examine composite measures (see Figures 3 and S2, and Tables 2, S7, and S8). Supplemental multiple regression tested the effects of student gender, major type, condition (social-belonging and affirmation-training vs. control) and all higher order interaction terms on each measure with the same measure assessed in the preintervention survey controlled.

Table S2
Elements of the Interventions

Element	Description	Purpose
(1) Cover Story	The study was represented as an opportunity for students to learn about students' experiences entering engineering and to share their experiences with future students to improve their transition.	This representation prevents students from viewing their participation in the study as stigmatizing or from thinking that they are seen as in need of help. Instead, it treats students as experts in the academic transition and empowers them to use this expertise to help future students.
(2) Survey of Upper Year Students	Students read summary statistics and quotations from senior engineering students describing their transition to engineering.	These materials provide students the key psychological information—a new, more adaptive way to think about common difficulties in the academic transition (e.g., that many students worry at first about their belonging but these concerns abate with time, social-belonging). Representing this information as normative accomplishes three objectives: (1) Rather than attempting to persuade students of the validity of the process described, it assumes that this process is valid in general and invites students to elaborate on it reflecting on their own experience. (2) It conveys that difficulties participating students have experienced are typical not unusual and not evidence of a lack of fit. (3) It represents a path of growth from early difficulties to later success and belonging.
(3) "Saying-is-Believing" Exercises	Students wrote a brief essay about "why people's experience in university develops in the way the senior students described" illustrating their essays "with examples from your own experience" and a personal letter to a future student describing "what you've experienced, and what you've learned."	These exercises give students the opportunity to describe the key intervention message in their own words, encourage students to view their own experience through the lens of the intervention message, and allow students to advocate for this message as a normative aspect of students' transition to a receptive audience (next year's incoming students). This is a powerful and noncontrolling persuasive technique (Aronson, 1999) that facilitates active learning and deep processing (Yeager & Walton, 2011).
(4) Key chain	Students received either a key chain depicting University of Waterloo insignia (social-belonging), one composed of opaque plastic containing a slip of paper on which students wrote a word or phrase to remind them of an important value (affirmation-training), or a key chain of their choice (control),	A physical reminder cue can help people remember an intervention message, especially in times of stress. In one study, researchers found that giving participants a reminder bracelet increased the effectiveness of a safe-sex message (Dal Cin et al., 2006). Moreover, the reminder bracelet was especially effective when participants reported having had sex after drinking. It did so, they theorized consistent with alcohol myopia theory, because when people are drinking their attentional field narrows, and they become more responsive to local cues in the situation. Insofar as people under stress also become more attuned to local cues (Walton et al., 2012), the key chain may be especially effective—for instance, in reminding students of personal values (affirmation-training)—in times of stress.

Table S3

Effectiveness of Random Assignment

Variable	Control		Social-Belonging		Affirmation-Training		Test Statistics	
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*
Mean Within-Major grade point average	71.28	4.56	70.72	6.13	71.42	4.88	0.39	0.68
Preintervention Evaluation of Current Experience in Engineering	5.09	0.84	5.12	0.70	4.03	0.88	0.20	0.82
Preintervention Perceived Prospects of Succeeding in Engineering	68.70	14.06	69.91	12.12	68.96	12.43	0.19	0.83
Preintervention Proportion Friends Male Engineers	0.53	.30	0.54	.31	0.57	0.26	0.32	0.73
Preintervention Implicit Normative Evaluations of Female Engineers	0.51	0.37	0.55	0.29	0.55	0.31	0.34	0.71
Preintervention Gender Identification	4.10	1.23	4.21	1.25	3.94	1.22	0.84	0.43
Preintervention Proportion Friends Female Nonengineers	0.11	0.18	0.12	0.19	0.09	0.15	0.46	0.63

Table S4

Raw Means (and Standard Errors) Among Women by Time Assessed, Major Type, and Experimental Condition

Variable	Assessment	Women in Gender-Diverse Majors			Women in Male-Dominated Majors		
		Control	Social-Belonging	Affirmation-Training	Control	Social-Belonging	Affirmation-Training
Academic Performance	First Semester Engineering GPA	76.34	71.41	68.06	68.81	77.75	79.09
		(2.46)	(3.12)	(3.26)	(4.39)	(3.40)	(3.81)
	Within-Major Mean GPA	*69.56*	*67.21*	*68.48*	*74.36*	*76.63*	*73.93*
	Second Semester Engineering GPA	76.84	75.25	67.82	64.44	78.37	75.09
		(2.29)	(1.94)	(3.48)	(3.04)	(3.40)	(3.42)
	Within-Major Mean GPA	*70.94*	*68.41*	*69.34*	*72.53*	*71.81*	*69.54*
Evaluation of Current Experience in Engineering	Preintervention	5.08	4.75	4.82	4.58	5.18	5.06
		(0.18)	(0.14)	(0.27)	(0.38)	(0.17)	(0.17)
	Intervention Session	5.21	4.69	4.93	4.13	5.30	5.06
		(0.21)	(0.14)	(0.29)	(0.36)	(0.16)	(0.14)
	Second Semester	5.28	4.57	5.06	4.02	5.09	5.14
		(0.21)	(0.24)	(0.23)	(0.36)	(0.23)	(0.30)
Perceived Prospects of Succeeding in Engineering	Preintervention	65.03	63.29	63.85	61.43	71.36	73.48
		(2.92)	(2.73)	(3.27)	(4.80)	(2.79)	(2.55)
	Intervention Session	63.97	63.74	63.87	63.81	72.92	73.65
		(2.78)	(2.41)	(3.21)	(5.08)	(2.84)	(2.45)
	Second Semester	68.72	59.15	63.67	46.05	63.76	71.47
		(2.46)	(3.66)	(4.43)	(6.38)	(4.69)	(4.57)
Percentage of Friends Male Engineers	Preintervention	29.05%	37.68%	41.11%	40.00%	50.00%	47.20%
		(0.05)	(0.06)	(0.05)	(0.14)	(0.11)	(0.06)
	Second Semester	26.76%	40.28%	34.17%	34.64%	78.75%	35.00%
		(0.05)	(0.07)	(0.07)	(0.08)	(0.10)	(0.12)
Implicit Normative Evaluations of Female Engineers	Preintervention	0.60	0.59	0.56	0.65	0.68	0.61
		(0.07)	(0.06)	(0.07)	(0.15)	(0.07)	(0.09)
	Second Semester	0.52	0.60	0.51	0.44	0.78	0.42
		(0.07)	(0.06)	(0.08)	(0.10)	(0.11)	(0.13)
Gender Identification	Preintervention	4.24	4.41	3.89	4.13	4.03	4.35
		(0.25)	(0.22)	(0.21)	(0.62)	(0.32)	(0.45)
	Second Semester	4.26	4.67	3.95	4.00	3.95	4.98
		(0.15)	(0.27)	(0.30)	(0.61)	(0.23)	(0.46)
Percentage of Friends Female Nonengineers	Preintervention	16.67%	10.69%	8.89%	10.00%	16.67%	11.51%
		(0.05)	(0.03)	(0.03)	(0.08)	(0.08)	(0.05)
	Second Semester	14.71%	8.33%	14.58%	9.29%	7.50%	23.75%

	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.11)

Table S5

Raw Means (and Standard Errors) Among Men by Time Assessed, Major Type, and Experimental Condition

Variable	Assessment	Men in Gender-Diverse Majors			Men in Male-Dominated Majors		
		Control	Social-Belonging	Affirmation-Training	Control	Social-Belonging	Affirmation-Training
Academic Performance	First Semester Engineering GPA	72.80	71.59	76.36	74.19	79.33	79.24
		(2.27)	(2.55)	(2.61)	(2.40)	(2.33)	(2.61)
	Within-Major Mean GPA	*69.90*	*66.95*	*69.25*	*72.39*	*75.15*	*75.57*
	Second Semester Engineering GPA	73.92	73.54	72.20	72.89	79.94	77.11
		(2.20)	(2.80)	(4.71)	(2.77)	(1.59)	(2.98)
	Within-Major Mean GPA	*72.79*	*69.35*	*71.59*	*70.51*	*71.64*	*72.07*
Evaluation of Current Experience in Engineering	Preintervention	5.14	5.24	4.92	5.22	5.29	5.29
		(0.14)	(0.14)	(0.14)	(0.17)	(0.15)	(0.19)
	Intervention Session	5.18	5.28	5.02	5.23	5.29	5.26
		(0.15)	(0.17)	(0.21)	(0.16)	(0.16)	(0.22)
	Second Semester	5.10	5.42	4.86	5.17	5.36	5.30
		(0.24)	(0.23)	(0.17)	(0.26)	(0.22)	(0.25)
Perceived Prospects of Succeeding in Engineering	Preintervention	71.30	73.72	68.91	71.40	71.78	70.94
		(1.90)	(2.13)	(2.58)	(3.27)	(2.47)	(2.76)
	Intervention Session	71.36	71.65	70.11	73.59	71.51	69.10
		(1.93)	(3.00)	(2.67)	(2.83)	(2.43)	(2.91)
	Second Semester	68.21	71.87	67.47	70.44	68.94	70.03
		(4.22)	(4.24)	(3.70)	(4.81)	(3.25)	(4.45)
Percentage of Friends Male Engineers	Preintervention	58.79%	60.34%	61.24%	71.85%	65.12%	73.64%
		(0.05)	(0.06)	(0.05)	(0.04)	(0.06)	(0.05)
	Second Semester	73.00%	65.91%	65.45%	68.33%	61.62%	65.21%
		(0.07)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)
Implicit Normative Evaluations of Female Engineers	Preintervention	0.43	0.49	0.52	0.45	0.50	0.51
		(0.07)	(0.08)	(0.07)	(0.08)	(0.05)	(0.08)
	Second Semester	0.47	0.62	0.27	0.33	0.35	0.37
		(0.08)	(0.11)	(0.11)	(0.08)	(0.10)	(0.06)
Gender Identification	Preintervention	4.04	4.11	3.80	4.03	4.19	3.84
		(0.24)	(0.34)	(0.37)	(0.24)	(0.28)	(0.25)
	Second Semester	3.78	4.29	4.13	4.04	3.96	3.97
		(0.28)	(0.38)	(0.12)	(0.20)	(0.15)	(0.17)
Percentage of Friends Female Nonengineers	Preintervention	9.18%	12.98%	6.84%	6.97%	8.78%	8.92%
		(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.04)
	Second Semester	8.17%	12.00%	6.36%	7.50%	9.41%	10.63%

	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)

Table S6
Each Covariate Included in Each Analysis

Category	Outcome	Covariate	Significance of Covariate	A Priori Basis for Inclusion
Academic Performance	1. First-Year Engineering GPA	Mean GPA earned in students' major	$t(182) = 4.43$, $p < .001$	Controls for the difficulty of the major students enrolled in.
Daily Diaries	2. Perceived Importance of Negative Daily Events	Perceived importance of daily positive events.	$t(182) = 10.63$, $p < .001$	Tests effects on the perceived importance of daily negative events *relative to* the perceived importance of daily positive events (alternative analyses yield similar results, Supplemental Material).
	3. Confidence Handling Daily School Stress	[None]	-	-
	4. Day-to-Day Level and Stability of Self-Esteem	[None]	-	-
Attitudes Toward Engineering (Intervention Session)	5. Evaluation of Current Experience in Engineering	Preintervention assessment of the outcome	$t(218) = 19.53$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
	6. Perceived Prospects of Future Success in Engineering	Preintervention assessment of the outcome	$t(218) = 20.17$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
Attitudes Toward Engineering (Second Semester)	7. Evaluation of Current Experience in Engineering	Preintervention assessment of the outcome	$t(144) = 9.15$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
	8. Perceived Prospects of Succeeding in Engineering	Preintervention assessment of the outcome	$t(144) = 10.45$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
Friendship Groups (2nd Semester)	9. % Male Engineers	Preintervention assessment of the outcome	$t(137) = 5.20$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
	10. % Female Non-Engineers	Preintervention assessment of the outcome	$t(136) = 4.48$, $p < .001$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).
Other 2nd Semester Measures	11. Gender Identification	Preintervention assessment of the outcome	$t(141) = 1.28$, $p = .20$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).

	12. Implicit Normative Evaluations of Female Engineers	Preintervention assessment of the outcome	$t(143)=1.75$, $p=0.083$	Tests effects relative to what would be expected on the basis of baseline measurements (similar to change scores).

Table S7

Intervention Effects on Primary Outcomes Among Men in Male-Dominated Majors

Variable	Combined Intervention (1) vs. Control (0)	Social Belonging (1) vs. Control (0)	Affirmation Training (1) vs. Control (0)
First-Year Engineering GPA	$B = 3.30$, $t(182) = 1.18$, $p = .24$	$B = 3.88$, $t(178) = 1.23$, $p = .22$	$t < 1$
Daily Diaries: Perceived Importance of Daily Neg. Events	$t < 1$	$t < 1$	$t < 1$
Daily Diaries: Confidence Handling Daily School Stress	$t < 1$	$t < 1$	$t < 1$
Daily Diaries: Day-to-Day Level and Stability of Self-Esteem	$t < 1$	$t < 1$	$t < 1$
Immediate Postintervention: Evaluation of Current Experience in Engineering	$t < 1$	$t < 1$	$t < 1$
Immediate Postintervention: Perceived Prospects of Succeeding in Engineering	$B = -3.18$, $t(218) = -1.83$, $p = .069$, $d = -0.25$	$B = -2.38$, $t(214) = -1.19$, $p = .24$	$B = -4.13$, $t(214) = -1.98$, $p = .049$, $d = -0.33$
Second Semester: Evaluation of Current Experience in Engineering	$t < 1$	$t < 1$	$t < 1$
Second Semester: Perceived Prospects of Succeeding in Engineering	$t < 1$	$t < 1$	$t < 1$

Note. These contrasts were not predicted to be significant. Contrasts were derived from multiple regression analyses. For intervention effects among women in male-dominated majors, see Table 2. For intervention effects among students in gender-diverse majors, see Table S8.

Table S8

Intervention Effects on Primary Outcomes Among Students in Gender-Diverse Majors

Variable	Contrasts Among Men in Gender-Diverse Majors			Contrasts Among Women in Gender-Diverse Majors			Gender × Condition in Gender-Diverse Majors
	Combined Intervention (1) vs. Control (0)	Social Belonging (1) vs. Control (0)	Affirmation Training (1) vs. Control (0)	Combined Intervention (1) vs. Control (0)	Social Belonging (1) vs. Control (0)	Affirmation-Training (1) vs. Control (0)	
First-Year Engineering Grade Point Average	$t < 1$	$t < 1$	$t < 1$	$B = -5.04$, $t(182) = -1.63$, $p = .10$, $d = -0.46$	$t < 1$	$B = -7.56$, $t(178) = -2.03$, $p = .043$, $d = -0.69$	$B = 7.04$, $t(182) = 1.70$, $p = .091$
Daily Diaries: Perceived Importance of Daily Negative Events	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$B = 1.00$, $t(178) = 1.13$, $p = .26$	$t < 1$	$t < 1$
Daily Diaries: Confidence Handling Daily School Stress	$B = 0.47$, $t(197) = 1.80$, $p = .074$, $d = 0.47$	$B = 0.48$, $t(193) = 1.59$, $p = .11$, $d = 0.48$	$B = 0.46$, $t(193) = 1.36$, $p = .17$	$B = -0.27$, $t(197) = -1.00$, $p = .32$	$t < 1$	$t < 1$	$B = 0.74$, $t(197) = 1.97$, $p = .050$
Daily Diaries: Day-to-Day Level and Stability of Self-Esteem	$B = 0.64$, $t(196) = 3.12$, $p = .002$, $d = 0.81$	$B = 0.55$, $t(192) = 2.34$, $p =.020$, $d = 0.69$	$B = 0.77$, $t(192) = 2.94$, $p = .004$, $d = 0.97$	$t < 1$	$B = -0.30$, $t(192) = 1.24$, $p = .21$	$t < 1$	$B = 0.81$, $t(196) = 2.76$, $p = .006$
Immediate Postintervention: Evaluation of Current Experience in Engineering	$t < 1$	$t < 1$	$t < 1$	$B = -0.15$, $t(218) = -1.07$, $p = .29$	$B = -0.23$, $t(214) = -1.45$, $p = .15$, $d = -0.27$	$t < 1$	$t < 1$
Immediate Postintervention: Perceived Prospects of Succeeding in Engineering	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$
Second Semester: Evaluation of Current Experience in Engineering	$t < 1$	$B = 0.27$, $t(140) = -1.01$, $p = .31$	$t < 1$	$B = -0.26$, $t(144) = -1.16$, $p = .25$	$B = -0.40$, $t(140) =-1.52$, $p = .13$, $d = -0.44$	$t < 1$	$B = 0.35$, $t(144) = 1.10$, $p = .27$

Second Semester: Perceived Prospects of Succeeding in Engineering	$t<1$	$t<1$	$t<1$	B=-5.34, $t(144)$=-1.52, p=0.13, d=-0.35	B=-6.05, $t(140)$=-1.45, p=0.15, d=-0.39	B=-4.60, $t(140)$=-1.08, p=0.28	B=6.26, $t(144)$=1.24, p = .22

Note. These comparisons were not predicted to be significant. Contrasts derived from multiple regression analyses. For intervention effects among women in male-dominated majors, see Table 2. For intervention effects among men in male-dominated majors, see Table S7.

Table S9

Gender Differences (Men = 0, Women = 1) on Primary Outcomes by Condition Within Gender-Diverse Majors

Variable	In Gender-Diverse Majors	
	Control Conditions	Intervention Conditions
First-Year Engineering Grade Point Average	$B = 4.35$, $t(182)=-1.37$, $p = .17$	$t< 1$
Daily Diaries: Perceived Importance of Daily Negative Events	$t < 1$	$t<1$
Daily Diaries: Confidence Handling Daily School Stress	$B = 0.36$, $t(197) = 1.25$, $p = .21$	$B = -0.38$, $t(197) = -1.58$, $p = .11$, $d = -0.38$
Daily Diaries: Day-to-Day Level and Stability of Self-Esteem	$B = 0.28$, $t(196) = 1.25$, $p = .21$	$B = -0.52$, $t(196)=-2.82$, $p = .005$, $d = -0.66$
Immediate Post-Intervention: Evaluation of Current Experience in Engineering	$t < 1$	$t < 1$
Immediate Post-Intervention: Perceived Prospects of Succeeding in Engineering	$B = -2.39$, $t(218) = -1.11$, $p = .27$	$t < 1$
2nd Semester: Evaluation of Current Experience in Engineering	$t < 1$	$t < 1$
2nd Semester: Perceived Prospects of Succeeding in Engineering	$B = 5.48$, $t(144) = 1.40$, $p = .16$	$t < 1$

Note. Contrasts were derived from multiple regression analyses. For gender differences within male-dominated majors, see Table 3.

Table S10

Effects of the Social-Belonging Intervention on Outcomes Predicted to Yield Effects Only for This Intervention Among Women in Male-Dominated Majors Among Students in Gender-Diverse Majors and Among Men in Male-Dominated Majors

	(1) Contrasts Among Men in Gender-Diverse Majors		(2) Contrasts Among Women in Gender-Diverse Majors		(3) Contrasts Among Men in Male-Dominated Majors	
Second semester	Social Belonging (1) vs. Affirmation-Training and Control (0)	Affirmation-Training (1) vs. Control (0)	Social Belonging (1) vs. Affirmation-Training and Control (0)	Affirmation-Training (1) vs. Control (0)	Social Belonging (1) vs. Affirmation-Training and Control (0)	Affirmation-Training (1) vs. Control (0)
Percentage of Male Engineers in Students' Friendship Groups	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t<1$
Implicit Normative Evaluations of Female Engineers	$B = 0.24$, $t(143) = 2.19$, $p = .030$, $d = 0.72$	$B = -0.20$, $t(139) = -1.58$, $p = .12$, $d = -0.59$	$t < 11$	$t < 1$	$t < 1$	$t<1$

Note. Contrasts were derived from multiple regression analyses. For contrasts for women in male-dominated majors, see Table 4.

Table S11

Effects of Affirmation-Training on Outcomes Predicted to Yield Effects Only for This Intervention Among Women in Male-Dominated Majors Among Students in Gender-Diverse Majors and Among Men in Male-Dominated Majors

	(1) Contrasts Among Men in Gender-Diverse Majors		(2) Contrasts Among Women in Gender-Diverse Majors		(3) Contrasts Among Men in Male-Dominated Majors	
Second semester	Affirmation-Training (1) vs. Social Belonging and Control (0)	Social Belonging (1) vs. Control (0)	Affirmation-Training (1) vs. Social Belonging and Control (0)	Social Belonging (1) vs. Control (0)	Affirmation-Training (1) vs. Social Belonging and Control (0)	Social Belonging (1) vs. Control (0)
Percentage of Female Nonengineers in Students' Friendship Groups[a]	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$	$t < 1$
Gender Identification	$t < 1$	$t < 1$	$B = -0.44$, $t(141) = -1.43$, $p = .16$	$B = 0.40$, $t(137) = 1.18$, $p = .24$	$t < 1$	$t < 1$

Note. Contrasts were derived from multiple regression analyses. For contrasts for women in male-dominated majors, see Table 5.

[a]Either retaining or dropping participants with missing preintervention data.

Figures

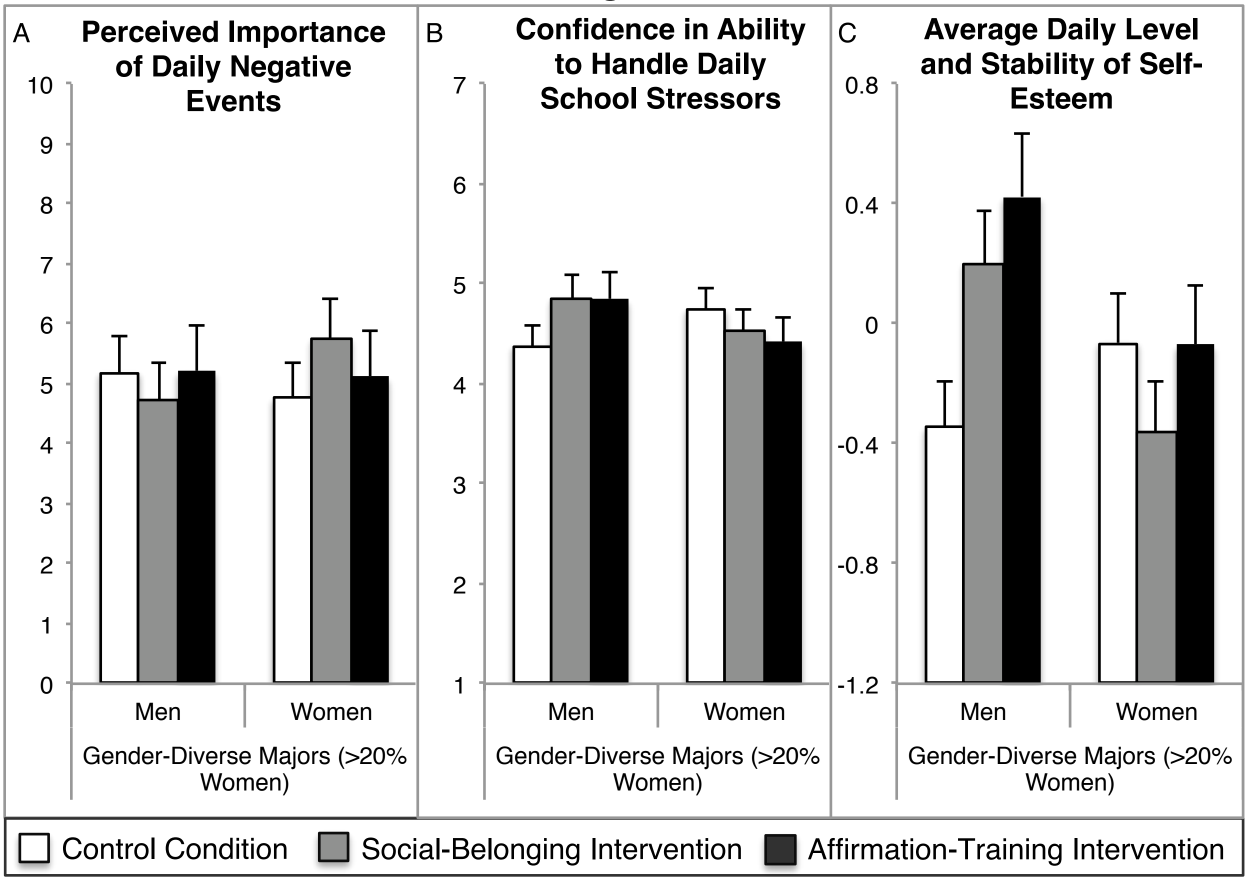

Figure S1. Daily functioning over 12days after the intervention among students in gender-diverse majors (>20% women). Error bars represent +1 standard error. (**A**) Perceived importance of negative events each day (adjusted for the perceived importance of positive events). (**B**) Confidence in ability to handle daily school stressors. (**C**) Level and stability (reverse-scored standard deviation) of self-esteem. The two measures were standardized and then averaged. The *y* axes in Panels (**A**) and (**C**) represent approximately 2.50 standard deviations. The *y* axis in Panel (**B**) represents the full range of the scale. Sample sizes (including students in male-dominated majors, Figure 2): N_{men} = 110–121; N_{women} = 81–84.

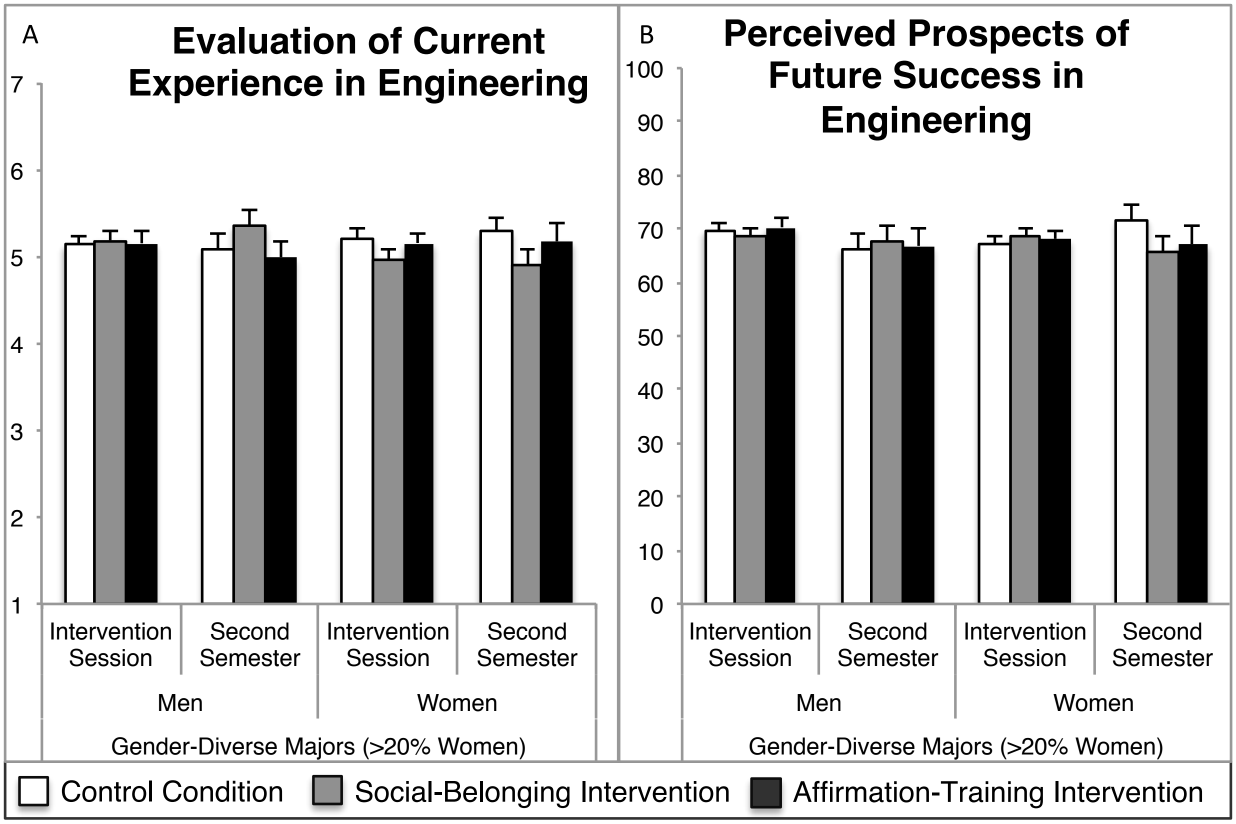

Figure S2. Attitudes toward engineering in the intervention session and in the second semester among students in gender-diverse majors (>20% women). Means are adjusted for preintervention reports. The *y* axes represent the full range of each scale. Error bars represent +1 standard error. (**A**) Students' evaluation of their current experience in engineering. (**B**) Students' perceived prospects of succeeding in engineering. Sample sizes (including students in male-dominated majors, see Figure 3): Men: $N_{\text{intervention session}} = 135$, $N_{\text{second semester}} = 88$; Women: $N_{\text{intervention session}} = 92$, $N_{\text{second semester}} = 65$).

.

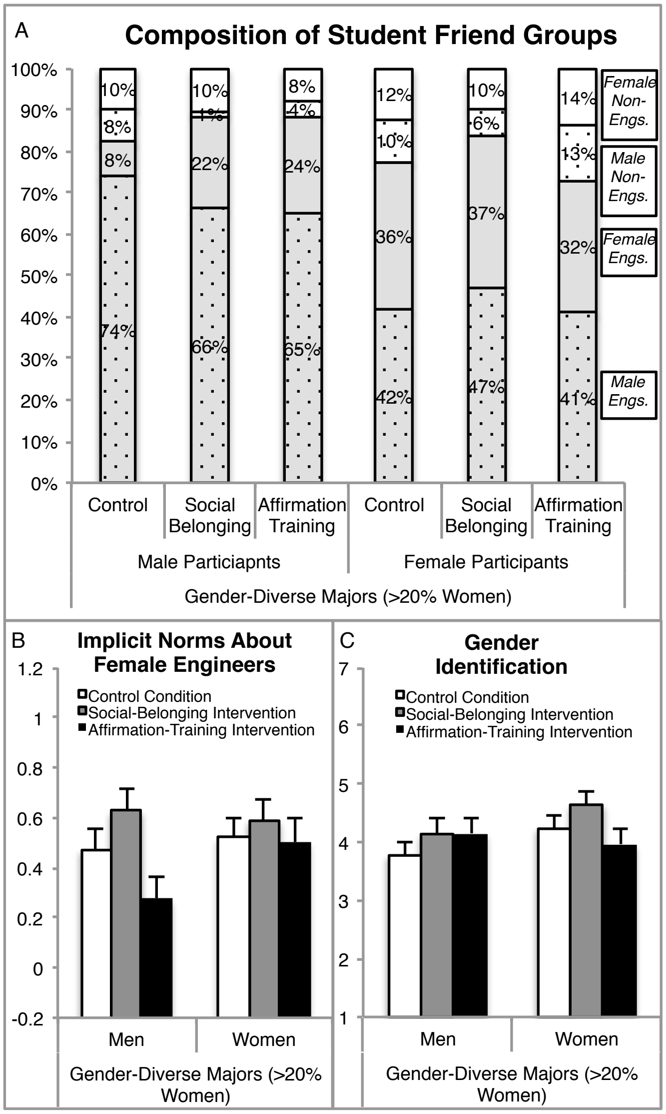

Figure S3. Friendship groups, implicit norms, and gender identification in the second semester among students in gender-diverse majors (>20% women). Means are adjusted for preintervention reports. Error bars represent +1 standard error. (**A**) Representation in students' friendship groups of male and female engineers and nonengineers. (**B**) Implicit normative evaluations of female engineers. Higher values represent more positive implicit norms about female engineers (i.e., "most people like"="female engineers"). The *y* axis represents approximately 3.75 standard deviations. (**C**) Self-reported gender identification. The *y* axis represents the full range of the scale. Sample sizes (including students in male-dominated majors, see Figure 4): N_{men} = 81–87; N_{women} = 64–65.

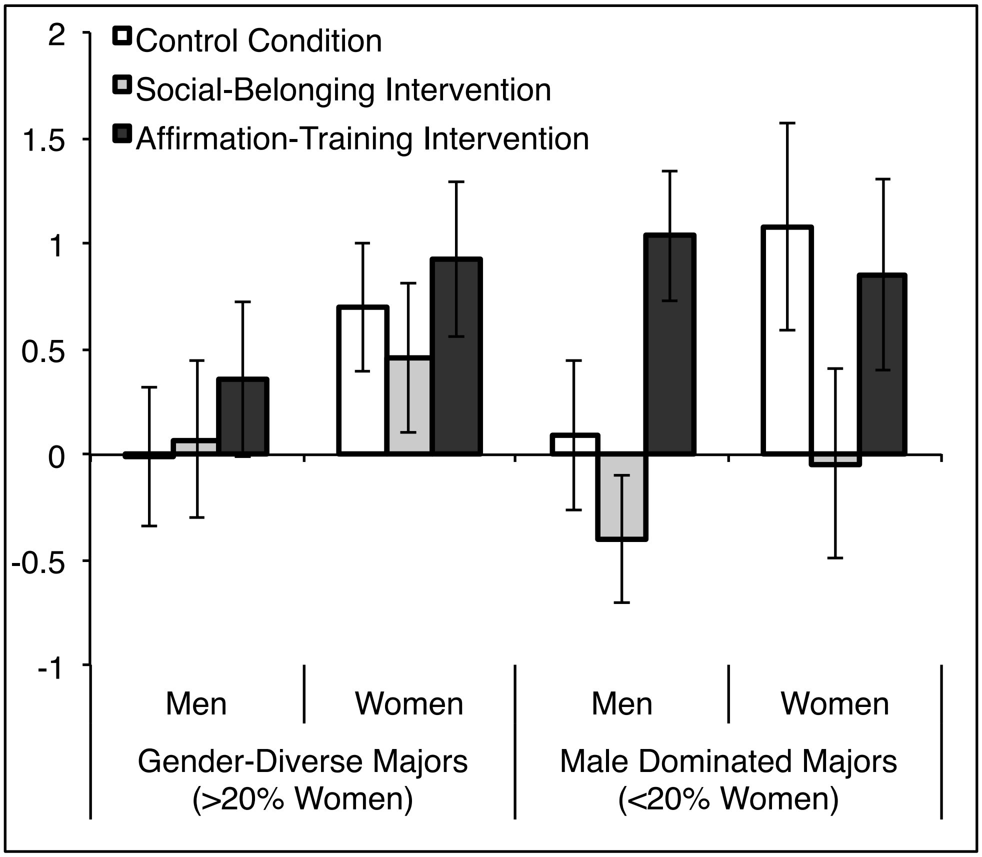

Figure S4. Change in reported frequency of sexist jokes about female engineers from before the intervention to the second semester. Positive values represent increases over time. Means are adjusted for the perceived frequency of ethnic jokes about engineers and jokes about non-Canadian engineers. Error bars represent ±1 standard error. The *y* axis represents approximately 2.15 standard deviations. $N_{\text{men}} = 87$; $N_{\text{women}} = 5$.

Supplemental Appendix S1
Example Quotations From Upper Year Engineering Students

Social-Belonging Condition

When I first got to Waterloo, I worried that I was different from the other students. Everyone else seemed so certain it was the right place for them and were so happy to be here. But I wasn't sure I fit in—if I would make friends, if people would respect me. Sometime after my first year, I came to realize that almost everyone comes to Waterloo and feels uncertain at first about whether they fit in. It's something everyone goes through. Now it seems ironic—everybody feels different first year, when really we're all going through the same things.

—"Karen," 4A Electrical

I didn't go to a very good high school, and I worried that my high school courses had not prepared me well for university. Honestly, when I got here, I thought professors were scary. I thought they were critical and hard in their grading, and I worried about whether other students would respect me. I was nervous about speaking in class, and I didn't want to ask people for help with assignments. After some time, I began to feel more comfortable—I made some close friends, and I started enjoying my classes more. I also became more comfortable asking for help when I had trouble with an assignment. And I saw that even when professors are critical or their grading is harsh, it didn't mean they looked down on me. It was just their way of pushing us. Since I realized that, I have been quite happy at Waterloo. It took time, but now I really feel like I belong in the intellectual community here. And to be honest, I'm glad I have been challenged. It's made me a better engineer.

— "Tom," 3B Chemical

Initially my transition to university wasn't bad. I enjoyed most of my classes. But it took a while to get to know my classmates. I remember once in my first term having lunch with some other civil engineers. They spent 90% of the time talking about hockey, about which I know next to nothing. I felt like I didn't belong. It was discouraging. But over time I got to know my classmates better, individually and as a group. Once I remember talking about the TV show *Monster Machines*, which I have to admit I love. We had a great time sharing stories about the different episodes. Even though I don't share their love of hockey, I realized that we do have a lot in common—an interest in how things work— and that's why we're all engineers. My major has turned out to be a lot of fun. I have made good friends with a number of my classmates, and I feel like I really belong here at UW.

—"Fatima," 4A Civil

Affirmation-Training Condition

When I first got to Waterloo, I worried that I was different from the other engineers. Everyone else seemed so excited and happy to be here but I just felt stressed and overwhelmed. There were so many new people; my classes were harder; it was a totally new environment. Sometime after my first year, I realized that almost everyone feels overwhelmed at times in the transition to university. It's just a process that everyone goes through. It takes time to find your own way of keeping things in balance in a new place. Now it seems ironic— everyone feels different first year, when really we're all experiencing the same things.

— "Karen," 4A Electrical

My first year was tough. I didn't know many people, and my classes were a ton of work. There was one particular stretch—I had a bunch of midterms and some nasty assignments, all at the same time. I was stressed. One night, I remember, I was trying to finish up an assignment, and I had to study for a test later. It was going to be a long night. But I took a break and called home. I talked to my mom. It was just a 5-min phone call, but when we hung up and I went back to studying, I felt so much better. I understand now the value of taking a time-out. Sometimes when I'm about to take a test, I take a mental break—and think about getting together with friends later or talking to my parents. There is so much going on, sometimes you have to take time to chill out.

—"Mike," 4A Mechatronics

In first year, I sometimes felt like I had tunnel vision—that I was just so completely caught up with life at Waterloo—with classes, with people I was meeting, the whole thing really—and I hardly thought of anything else and, it was hard at first, and it was stressful. But then I realized that, well, there are things outside of engineering that I do care about. I remembered that I had done volunteering in high school, and so I decided to get involved with an environmental group here on campus. And even though, objectively, I had less time, with volunteering on top of schoolwork, I found I felt really refreshed, and I could concentrate a lot better. I also met a lot of people while I was volunteering, and most of them shared similar interests as me, and we all became really good friends. I find that the longer I spend in Waterloo, the more I find things to do that are just broadening my life away from school work, and it's really good. It took me time to find those activities, but they've made a really big difference in my experience. And, I guess the one thing I had to learn was that it isn't the best thing for me to just study non-stop.

— "Mahesh," 3B Environmental