Using a randomized, between-subjects experimental design, we tested hypotheses that self-critical perfectionism would moderate the effects of subtle stereotype threat (ST) for women and students in underrepresented racial/ethnic groups who are pursuing traditional degrees in science, technology, engineering, or math (STEM). A diverse sample of freshmen students \( N = 294 \) was recruited from 2 major universities. Students were blocked by gender and race/ethnicity and randomly assigned to experience subtle ST or no ST. Participants in the subtle ST condition were primed to consider their gender, race, and ethnicity prior to completing measures of science self-efficacy. Those in the control condition completed the measures without such priming. Controlling for prior academic performance and university context, ST priming significantly interacted (a) with self-critical perfectionism to predict coping self-efficacy scores and (b) with race/ethnicity to predict end-of-semester STEM grades. A 3-way interaction of ST priming, sex, and self-critical perfectionism also predicted students’ grades in courses wherein women and men were more proportionally represented. The Sex \( \times \) Self-Critical Perfectionism interaction was not significant for those in the ST group but was for those in the control group. Men in the control group had higher grade-point averages (GPAs) at low levels of self-critical perfectionism than they had at higher levels of perfectionism. In contrast, women had lower GPAs when self-critical perfectionism was low, but their GPAs were higher when self-critical perfectionism was high. The findings are discussed in terms of their implications for self-efficacy and performance in the pursuit of a STEM major.

**Keywords:** stereotype threat, perfectionism, science self-efficacy, STEM, academic performance

Considerable attention has been directed to channeling young people toward careers in science, technology, engineering, and mathematics (STEM) fields, and to understanding and possibly repairing disparities in interests, retention, and success in STEM careers. According to data compiled by the National Science Foundation, Division of Science Resources Statistics (2011), in 2009, 60% of the undergraduate degrees in the biological sciences went to women, as did 43% of the degrees in mathematics, and 42% of the degrees awarded in physical sciences. However, undergraduate degrees awarded to women in engineering (18%) and computer science (18%) indicated that women were still substantially underrepresented in those majors. Furthermore, some seemingly more reasonable percentages mask considerable within-field variability. For example, although women earned 42% of the physical science degrees, the bulk of those (82%) were awarded in chemistry. In physics, women accounted for 19% of the undergraduate degrees. Comparisons of racial/ethnic groups reveal additional disparities in terms of proportional representativeness. For example, in 2009, 67% of the students in undergraduate engineering majors were White, and 11% were of Asian descent. Only about 5% of those majors were African Americans, and 10% of the majors were Hispanic. In terms of overall enrollment at U.S. colleges and universities, African Americans make up 12.5% of the students, Hispanics make up 12.9% of the students, 58% are White, and just over 6% are of Asian descent (National Science Foundation, Division of Science Resources Statistics, 2011).

For many students, ongoing self-assessments of academic interests and performance capabilities can be considered a normative process likely to eventuate in reasonably accurate, performance-based career decisions over time. However, for students pursuing degrees in academic domains wherein (on the basis of their gender,
race, or both) they are proportionally underrepresented, these same self-assessments and decisional processes may be compromised by factors other than ability, such as contextual factors, interests, and psychological factors (Ceci, Williams, & Barnett, 2009). In this study, we examine stereotype threat and self-critical perfectionism as compromising factors.

Stereotype threat (ST) has been defined as a situational experience in which underrepresented peers feel pressurized by the possibility of being adversely judged by prevailing negative gender and/or racial stereotypes associated with a particular performance context (Smith, 2004; Steele & Aronson, 1995). Despite its wide appeal and integration within the broader STEM discourse, concerns have been raised as to veracity of ST (Sackett, Hardison, & Cullen, 2004), especially given the rather poor rates by which studies have been able to replicate ST in the lab; only 30–55% of studies attempting to replicate the original ST effects have been able to do so (Stoet & Geary, 2012). Thus, there are reasons to examine potential moderators of ST effects in predicting performance outcomes as well as the “real-world” generalizability of ST effects (Cullen, Waters, & Sackett, 2006). In the present study, we consider self-critical perfectionism as a performance-related individual differences factor that may exacerbate or mitigate potential threats to academic self-appraisals, and then influence the academic performance of underrepresented students in STEM fields.

According to Slaney, Rice, and Ashby (2002), a core feature of self-critical perfectionism is discrepancy, or the perceived gap between expected levels of performance and self-appraisal of one’s adequacy in meeting those expected performance levels; larger perceived gaps correspond to higher levels of self-critical perfectionism. Although multiple dimensions of perfectionism have been identified (Stoeber & Otto, 2006), self-critical perfectionism in general, and the discrepancy dimension in particular, have consistently been linked to a host of ST-relevant variables, such as emotion dysregulation, cognitive test anxiety, avoidance-related achievement goal orientations, dissatisfaction with academic performance, and even low overall grade point average (e.g., Aldea & Rice, 2006; Elion, Wang, Slaney, & French, 2012; Eum & Rice, 2011; Grzegorek, Slaney, Franze, & Rice, 2004). Effects of other perfectionism dimensions, such as personal standards perfectionism, appear to be smaller or negligible in studies examining perfectionism and stress reactions (e.g., Dunkley, Blankstein, Halsall, Williams, & Winkworth, 2000).

Although not directly addressing self-critical perfectionism, Dasgupta (2011) described how the mere experience of underrepresentation in a negatively stereotyped performance context is likely to activate ST-related and often self-critical attributions (e.g., self-doubt regarding adequacy and belongingness). She and others (e.g., Sekaquaptewa, 2011) noted that these effects can occur regardless of whether the stereotypes are consciously recognized or internalized. Among other things Smith et al. (Smith, 2004, p. 194; Smith, Sansone, & White, 2007) proposed a “stereotyped task engagement process” model in which ST is presumed to have a direct impact on performance-avoidance goals, whereas individual characteristics have indirect or moderating impacts on the adoption of these persistence-threatening motivations. In short, the experience of underrepresentation and its attendant stereotype threats are presumed to heighten performance-related self-doubts and motivations to avoid failure, as well as to weaken the adoption of mastery goal orientations that have been empirically linked to sustained academic engagement (Dweck & Leggett, 1988; Elliot & Church, 1997).

In the present study, we were interested in the moderating effects of self-critical perfectionism on ST associations with proximal indicators of science self-efficacy and on the distal outcome of later performance in STEM courses. We expected science and math self-efficacy would be highly salient for students majoring in a STEM field, and therefore likely to be affected by ST and self-critical perfectionism. Nguyen and Ryan (2008) found stronger gender-related effect sizes in studies that primed ST by using indirect or subtle cues (e.g., requiring participants to indicate gender prior to test performance) than in studies wherein test performance was preceded by “blatant” or “moderately explicit” ST cues (e.g., presenting information that the impending test was diagnostic of gender differences). Because of this evidence, subtle text priming for ST by enhancing awareness of gender and race/ethnicity was manipulated in the current study. We examined ST, perfectionism, self-efficacy, and actual academic performance of college freshmen enrolled in a STEM-major at two different state universities. We also examined whether these ST and perfectionism effects were further moderated gender and race/ethnicity. To control for possibly confounding influences of prior academic abilities and contextual factors, we covaried effects of high school grade-point average (GPA) and university in all analyses.

We expected that self-critical perfectionism (discrepancy) would moderate the effects of subtle ST on STEM-related science self-efficacy and academic performance. Those effects were expected to be stronger for women and members of historically underrepresented racial/ethnic minority groups in STEM-related majors (African Americans and Latinos) compared with men and racial/ethnic groups more typically represented in STEM fields (Whites and Asians). Operationalizations of academic performance were end-of-semester GPAs in those STEM courses within fields where women and some minority groups are underrepresented (underrepresented GPA) and GPAs in courses where those groups are typically more represented (representative GPA).

Method

Participants

A random sample of freshmen students was recruited at two major state universities. University 1, located in a nonmetropolitan area in the southeastern United States, was a large, major research university with 35,000 undergraduates (54% women, 33% racial/ethnic minorities). University 2 was located in a large, urban, metropolitan area in the southwest United States with a student population of approximately 30,000 undergraduates (52% women, 57% racial/ethnic minorities). Participants from both sources were included only if they were 18 years of age or older and indicated an intention to major in one of the traditional STEM areas. The initial sampling plan called for both universities to provide 100 randomly selected women and 100 randomly selected men from each of four racial/ethnic groups: non-Hispanic White, Hispanic, non-Hispanic African American, and non-Hispanic Asian or Pacific Islander. Because of the low number of several minority and gender groups in STEM majors attending both universities, all of those students were invited to participate (total recruited N = 1,431). A total of 15 students started the study by entering the
PERFECTIONISM MODERATES STEREOTYPE THREAT

study website and providing consent but did not provide study data or only completed a few items on the first questionnaire before discontinuing their participation. Those participants were not included in the analyses. The final, total sample of 294 participants (181 from University 1 and 113 from University 2) included 175 women (59.5%) and represented 21% of those recruited. By design, the sample was relatively balanced in terms of race/ethnicity: 28.2% White, 27.6% Asian or Asian American, 26.9% Latino/a, and 17.3% African American. Participants ranged in age from 18 to 28 ($M = 18.77$, $SD = 1.12$); approximately 97% of the students were either 18 or 19 years old. The proportions of students in the major STEM areas were: mathematics/statistics (3.6%), biological sciences (47.0%), physics or astronomy (3.2%), chemistry (7.5%), engineering (28.5%), computer sciences (4.3%), agricultural sciences (2.5%), environmental sciences (1.4%), and other (2.1%).

Measures

A recently developed short version of original Almost Perfect Scale–Revised (APS-R; Slaney, Mobley, Trippi, Ashby, & Johnson, 1996; Slaney, Rice, Mobley, Trippi, & Ashby, 2001) was used in the present study (Short Almost Perfect Scale [SAPS], Rice, Tueller, & Richardson, 2012). The SAPS contains four items that measure the self-critical perfectionism dimension of discrepancy, the perception of a gap between one’s expectations and self-evaluation of meeting those standards. Items are responded to using a 7-point scale. Reliability and validity results for the SAPS are comparable to the APS-R. Confirmatory factor analyses supported the SAPS factor structure and also demonstrated item-level measurement invariance between women and men on the SAPS (Rice et al., 2012). Internal consistency for SAPS Discrepancy scores was .83 (.82 in the present study). The correlation between the short and long versions of discrepancy was .95, suggesting substantial correspondence. In numerous past studies, APS-R subscale scores correlated in expected directions with other scales measuring maladaptive perfectionism and maladjustment (e.g., Wei, Mallinckrodt, Russell, & Abraham, 2004).

Four items on the Self-Efficacy for Scientific/Technical Fields scale (STF; Lent et al., 2005; Lent, Singley, Sheu, Schmidt, & Schmidt, 2007) were used to assess science self-efficacy. Seven items on the STF measure self-efficacy for coping with problem situations associated with completing the degree. Participants rate items using a 10-point scale that ranges from 0 (no confidence) to 9 (complete confidence). The scores have had good reliability (internal consistency) estimates in several studies of college students, with Cronbach’s alpha ranging from .88 to .94 (Gwilliam & Betz, 2001; Lent, Brown, & Larkin, 1986; Lent et al., 2003). In the present study, alphas ranged from .86 to .88.

Objective end-of-semester grade data were obtained from institutional records and grades were organized into one of seven groups based on course descriptions and course listings: life sciences, physics, engineering, computer sciences, mathematics, social sciences, and nonscience. After converting letter grades to numerical values (i.e., A = 4.0, A− = 3.67, B+ = 3.33, etc.), grade point averages were calculated. Different GPAs were created for courses in which women have been underrepresented (physics, engineering, computer sciences, and mathematics courses) and for courses in which women are have been proportionally or overrepresented (life sciences and social sciences). High school grade point average (HSGPA) also was obtained from participants’ respective institutional records and both HSGPA and university were treated as covariates in the major analyses.

Procedure

A link to a web-based survey (in Survey Monkey) was sent to participants. At the survey site, participants first provided informed consent then were directed to the study site. A priori random assignment was made to either the subtle ST threat or no ST threat priming condition, with half of the women and men within each racial/ethnic group assigned to the treatment condition and the other half assigned to the control condition. The result of this assignment resulted in nearly completely balanced treatment ($N = 148$) and control ($N = 146$) groups. The experimental (i.e., ST group) distribution of race/ethnicity and gender (women [W], men [M]) was as follows: White (28W, 14M), African American (18W, 9M), Asian (20W, 18M), and Latino (18W, 21M). In the control group, the distribution was White (22W, 19M), African American (20W, 4M), Asian (22W, 21M), and Latino (27W, 13M). The distribution of gender across the treatment and control groups was balanced, $\chi^2(1, N = 294) = 0.48, p = .553$, as was the distribution of race/ethnicity, $\chi^2(3, N = 294) = 0.50, p = .920$.

In the subtle ST prime condition, students were prompted with larger font and bold text immediately before completing the STF (which was first in the sequence of measures) to report their gender, race, and ethnicity, but no overt message was conveyed regarding the link between those variables and performance on the questionnaires or in their major. In the control condition, participants reported gender, race, and ethnicity after completing all other study questionnaires, and font size was not enlarged and bold text was not used.

Results

Data Screening and Descriptive Statistics

Analyses were conducted with IBM SPSS 20 (2011) and Mplus (Version 7; Muthén & Muthén, 1998–2012). Based on screening for extreme scores (|z| > 3.0), five STF science Self-Efficacy scores were outliers (three from the treatment group and two from the control group). Those participants were deleted from further analyses. None of the SAPS Discrepancy scores was in the extreme range. Multivariate normality was examined for the three questionnaire scores, but based on Mahalanobis distance, no multivariate outliers were detected. Missing data were generally not of concern in analyses involving self-efficacy; covariance coverage in those analyses ranged from .910 to 1.0. However, as might be expected, there were more missing data when analyses involved the GPA indicators. Covariance coverage ranged from .637 to .682 for underrepresented GPA and from .689 to .734 for represented GPA. Full information maximum likelihood (FIML) in the Mplus program was used to generate unbiased parameter estimates. Given the relative amount of missing grade data, and as recommended by Enders (2010), we incorporated the self-efficacy scores as auxiliary variables into the FIML procedures when analyzing effects on underrepresented GPA and represented GPA.

The correlation matrix for the scores, differentiated by gender, appears in Table 1. One-tailed Fisher's r-to-z tests indicated a
significant difference in the correlations between Discrepancy and Science and Coping Self-Efficacy for women and men (cs = 2.70 and 2.79, p < .005, respectively). Science Self-Efficacy was more strongly associated with Underrepresented GPA for women (.42) than it was for men (.21), z = 1.95, p < .05. None of the other correlations were significantly different between women and men.

Perfectionism as Moderator of ST Priming

High school GPA and university (coded 0 and 1 for the two schools) were used as covariates in multiple regression analyses. We mean-centered the SAPS Discrepancy score and created two-way, three-way, and a four-way multiplicative interaction terms involving Discrepancy, the ST condition (coded 0 = control, 1 = ST prime), Sex (0 = female, 1 = male), and Race (0 = African American or Hispanic/Latino, 1 = White or Asian). Of most initial interest were the interaction effects. Effects of covariates, conditional main effects, and other interaction terms were controlled in each of the four separate analyses with Science Self-Efficacy, Coping Self-Efficacy, Underrepresented GPA, and Represented GPA as dependent variables. Results for the full models are displayed in Table 2.

Although the predictors accounted for 17% of the variance in Science Self-Efficacy, none of the main effects or interactions of interest was significant in the model. The predictors combined to account for 11% of the variance in Coping Self-Efficacy, with the ST × Discrepancy interaction making a significant contribution to the prediction. We used the parameter estimates from the full regression equation and calculated values for Coping Self-Efficacy scores for the participants in the control and ST prime conditions, as a function of low (−1 SD) and high (+1 SD) Discrepancy scores. In the control condition, low and high Discrepancy scores predicted nearly identical Coping Self-Efficacy scores (7.25 and 7.18, respectively). However, for the ST prime participants, higher levels of self-critical perfectionism were associated with lower Coping Self-Efficacy scores (6.78) compared with scores associated with low Discrepancy (8.04), a difference of a full standard deviation.

The predictors accounted for 19% of the variance in underrepresented GPA; the ST × Race interaction was significant in that

Table 1
Correlations Between Perfectionism, Self-Efficacy, and GPA

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ST Prime</td>
<td>1.00</td>
<td>−0.01</td>
<td>.04</td>
<td>.03</td>
<td>−0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>2. Discrepancy</td>
<td>.01</td>
<td>1.00</td>
<td>−0.23*</td>
<td>−.24*</td>
<td>−0.05</td>
<td>−.11</td>
</tr>
<tr>
<td>3. Science Self-Efficacy</td>
<td>.09</td>
<td>−0.09</td>
<td>1.00</td>
<td>.54*</td>
<td>.42*</td>
<td>.11</td>
</tr>
<tr>
<td>4. Self-Efficacy for Coping</td>
<td>−.07</td>
<td>−0.09</td>
<td>.41*</td>
<td>1.00</td>
<td>.14</td>
<td>−.01</td>
</tr>
<tr>
<td>5. Underrepresented GPA</td>
<td>.02</td>
<td>−.13</td>
<td>.21*</td>
<td>.04</td>
<td>1.00</td>
<td>.70*</td>
</tr>
<tr>
<td>6. Represented GPA</td>
<td>.08</td>
<td>.00</td>
<td>.02</td>
<td>−.12</td>
<td>.50*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. GPA = grade-point average; ST = stereotype threat. Correlations for women appear above the diagonal and for men appear below the diagonal. ST Prime coded 0 = control (no ST prime), 1 = treatment (ST prime).

*p < .05; two-tailed tests.

Table 2
Regression Analyses for the Effects of Stereotype Threat, Self-Critical Perfectionism, Sex, and Race Predicting Self-Efficacy and GPA

| Predictor | Science Self-Efficacy | | | Coping Self-Efficacy | | | Underrepresented GPA | | | Represented GPA | |
|-----------|------------------------|--|--|-----------------------|--|--|------------------------|--|--|------------------------|--|--|
|            | B (SE B) | β | B (SE B) | β | B (SE B) | β | B (SE B) | β | B (SE B) | β |
| HSGPA      | 0.42 (0.21) | .12 | −0.07 (0.23) | −.02 | 0.89 (0.20) | .24* | 0.76 (0.16) | .21* |
| University | −0.46 (0.15) | −.17* | −0.41 (0.17) | −.15* | −0.05 (0.17) | −.02 | 0.30 (0.15) | .09* |
| ST Prime   | −0.14 (0.26) | −.05 | 0.20 (0.29) | .08 | −0.56 (0.28) | −.22* | −0.14 (0.21) | −.05 |
| Sex        | 0.57 (0.34) | .22 | 0.66 (0.38) | .26 | 0.03 (0.33) | .01 | 0.00 (0.26) | .00 |
| Race       | −0.06 (0.25) | −.02 | 0.13 (0.27) | .05 | −0.39 (0.26) | −.15 | −0.24 (0.19) | −.09 |
| Discrepancy| −0.10 (0.13) | −.12 | −0.02 (0.14) | −.03 | −0.14 (0.14) | −.15 | −0.11 (0.11) | −.12 |
| ST × Sex   | −0.22 (0.45) | −.07 | −0.65 (0.50) | −.21 | 0.60 (0.44) | .21 | 0.31 (0.35) | .10 |
| ST × Race  | 0.31 (0.36) | .11 | −0.24 (0.39) | −.08 | 1.03 (0.37) | .37* | 0.39 (0.28) | .14 |
| ST × DIS   | −0.04 (0.18) | −.03 | −0.42 (0.20) | −.33* | 0.33 (0.20) | .26 | 0.19 (0.15) | .16 |
| Sex × Race | −0.24 (0.43) | −.08 | −0.74 (0.47) | −.25 | 0.11 (0.42) | .04 | 0.23 (0.33) | .08 |
| Sex × DIS  | 0.08 (0.25) | .06 | −0.02 (0.27) | −.01 | 0.09 (0.23) | .07 | 0.34 (0.19) | .25 |
| Race × DIS | −0.26 (0.18) | −.22 | −0.26 (0.20) | −.22 | −0.10 (0.19) | −.08 | −0.01 (0.14) | −.01 |
| ST × Sex × Race | 0.69 (0.58) | .17 | 0.73 (0.64) | .18 | −0.90 (0.57) | −.24 | −0.60 (0.45) | −.15 |
| Sex × Race × DIS | 0.13 (0.30) | .07 | 0.10 (0.33) | .06 | −0.05 (0.29) | −.03 | −0.04 (0.24) | −.02 |
| ST × Race × DIS | 0.31 (0.25) | .17 | 0.52 (0.28) | .29 | −0.09 (0.28) | −.05 | −0.14 (0.21) | −.08 |
| ST × Sex × DIS | −0.23 (0.31) | −.12 | 0.10 (0.34) | .05 | −0.12 (0.30) | −.07 | −0.51 (0.24) | −.29* |
| ST × Sex × Race × DIS | 0.28 (0.41) | .11 | 0.20 (0.45) | .08 | 0.06 (0.40) | .03 | 0.27 (0.32) | .11 |

Note. GPA = grade-point average; HSGPA = high school GPA; ST = stereotype threat; DIS = discrepancy. ST prime was coded 0 = control group (no ST prime) and 1 = treatment group (ST prime), sex was coded 0 = female and 1 = male, and race was coded 0 = African American or Hispanic/Latino and 1 = White or Asian. Discrepancy scores were mean-centered prior to creating the multiplicative interaction terms.

*p < .05.
equation. The predicted underrepresented GPA scores revealed relatively higher and comparable GPAs for the African American and Latino students in the control group (3.25) and the White and Asian students in the ST prime group (3.33). Relatively lower underrepresented GPAs were predicted for the White and Asian control group participants (2.87), with the lowest GPAs predicted for the African American and Latino students in the ST prime group (2.69).

Finally, the predictors accounted for 19% of the variance in represented GPA. Among the predictors of interest, the three-way ST × Sex × Discrepancy interaction was statistically significant. The Sex × Discrepancy effect was significant for the control group ($B = 0.36, SE = 0.18, p = .046$) but not the ST prime group ($B = -0.16, SE = 0.16, p = .311$). In the control group and at low levels of Discrepancy, men had higher represented GPAs (3.47) compared with women (2.91). At high levels of Discrepancy, the gap in GPA between women and men narrowed (3.15 vs. 3.25, respectively). Within gender, women with high levels of Discrepancy had higher GPAs compared with the GPAs for women at low levels of Discrepancy. The opposite was true for men, who had higher GPAs when Discrepancy was low compared to GPAs when Discrepancy was high. These results suggest some surprising academic value for self-critical perfectionism for women, provided they are in STEM classes with generally representative numbers of women and not primed to experience ST. Alternatively, self-critical perfectionism seemed to be a disadvantage for men in those more gender-representative STEM courses.

Discussion

Correlations revealed that Discrepancy scores were more strongly related to both Science Self-Efficacy and Coping scores for women than for men. In addition, Science Self-Efficacy scores were more robustly correlated with GPAs in those courses wherein women were underrepresented than in those courses wherein men were underrepresented. These findings are consistent with prior findings (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa, 2011) that women’s experience of underrepresentation in traditional STEM-related performance contexts may heighten their sensitivity to social-evaluative stress regarding their capabilities and thus increase their risk of experiencing ST.

Our regression analyses revealed that ST uniquely predicted underrepresented GPAs within our sample (with students in the control condition demonstrating higher grades). A significant ST × Discrepancy interaction in the prediction of Coping Self-Efficacy was generally aligned with our expectations, with elevated Discrepancy scores significantly predicting lower coping confidence among participants in the ST condition relative to those in the control condition. Similarly, we observed a significant ST × Race interaction in the prediction of underrepresented GPAs: In the control condition, African American and Latino students evidenced end-of-semester grades that were generally high and comparable to those of their White and Asian peers; however, the lowest underrepresented GPAs occurred for African American and Latino students in the ST condition. In their meta-analysis, Nguyen and Ryan (2008) reported that ST effects were somewhat more prominent among studies using a race/ethnicity-based stereotype than among those using a gender-based stereotype.

We unexpectedly found no significant interactions between sex and other variables in the prediction of performance in courses where women have been underrepresented. We did find a significant Sex × Discrepancy interaction among control group participants in the prediction of GPA in STEM courses wherein women have been proportionally represented. In the control group, self-critically perfectionistic women had higher GPAs than women who were not self-critically perfectionistic. In contrast, self-critically perfectionistic men had lower GPAs compared with men who were not self-critically perfectionistic. A general effect of maladaptive perfectionism being associated with academic performance has been inconsistently observed in other studies (e.g., Grzegorek et al., 2004; Rice & Slaney, 2002). In light of the present findings, prior inconsistencies might have been the result of not exploring gender, its interaction with perfectionism, or differentiating performance outcomes with respect to course- or field-related gender representation.

It is difficult to explain unexpected nonsignificant effects with regard to the interaction of sex, ST, and perfectionism in predicting underrepresented GPA. Nevertheless, several plausible explanations raised elsewhere could be discussed and addressed in future work. For example, an additional, subtle ST prime may simply be ineffectual in affecting GPAs among a potential subgroup of women who already are primed to be sensitive to performance-relevant primes by virtue of their presence in STEM classes where women are underrepresented. Perhaps a strong, compensatory re-actance effect (“I’ll show you that women can do science”) was already initiated that resulted in those women being unaffected by yet another subtle ST experience. Block, Koch, Liberman, Merriweather, and Roberson (2011, pp. 575–576), drawing on Klinger’s (1977) earlier work, referred to this possibility as “fending off the stereotype” through responses such as “invigoration” (overcompensating and working harder to meet goals) or making more productive than negative internal attributions. Such ideas are probably consistent with the effect observed by Forbes and Schmader (2010), wherein women who underwent an attitude and stereotype retraining exercise experienced increased motivation in math, a field in which women are typically underrepresented. In addition, Martens, Johns, Greenberg, and Schimel (2006) observed that women who made a self-affirmation statement (a statement of a valued aspect of the self) prior to a stereotype threat situation performed better than those who did not affirm a valued attribute.

Significant effects for men occurred with grades in courses in which they are typically represented as women (represented GPA). Without the prime, perfectionistic self-criticism seemed beneficial, but with gender subtly placed into awareness, self-critical perfectionism became problematic. That these effects did not occur with underrepresented GPA suggests that there is something about the relative balance of women and men in the courses that plays a role in whether perfectionism will enhance or harm performance outcomes. Tagler (2012) found that self-conscious men underperformed when primed with gender-relevant instructions, as if the pressure to perform or avoid failure results in “choking.” Perhaps the prime introduced a competitive element, or a pressure to perform, that was irrelevant in courses where men outnumbered women.

Our findings offer some qualifications to conclusions from others that lab-based ST effects may not be generalizable to “real world” performance situations (e.g., Cullen et al., 2006). Consis-
tent with those views, we did not find across the board main effects for subtle ST on our dependent variables, although we did find some moderators of ST effects. With regard to the main effects, it is possible that our use of a subtle form of ST priming (i.e., calling attention to participants’ gender and race/ethnicity prior to having them complete the research measures) was insufficient to activate ST among students that would then alter their science self-efficacy perceptions. However, and unlike many of the studies reviewed in Nguyen and Ryan’s (2008) meta-analysis that incorporated a real or anticipated “task performance” that was presented immediately after ST priming, we relied on a more “distal,” albeit highly relevant, measure of objective task performance (i.e., end-of-semester course grades) and found significantly lower underrepresented GPA for those students in the ST priming condition. Future work in this area could be strengthened by expanding the coverage of more naturalistic proximal performance effects as well as distal outcomes that might be linked to subtle ST (e.g., test grades, persistence in major, post baccalaureate career and education choices and outcomes).

Although the present study represents a randomized field experiment and relatively few studies in the counseling literature represent bona fide experimental designs, there were potential limitations that could affect both the internal and external validity of the study. For example, subtle ST is, by definition, a subtle manipulation that precluded the use of a more direct manipulation check. Other than through random assignment to condition we did not incorporate a way to determine whether the independent variable was in fact manipulated as intended. Likewise, the direct link with the more distal outcome of semester performance in STEM-related courses should be considered tentative without knowing more about the stability of the ST effects. Nevertheless, the external validity of the study could be considered strengthened by this type of manipulation. Although it seemed reasonable to expect that the subtle priming condition was comparable to what women and minority students in underrepresented STEM majors might experience in their natural environments, after the experiment, both treatment and control groups remained part of those environments and were equally likely to experience ST in those contexts.

External validity could also be limited in that only one semester of data was collected, all participants were freshmen, and students at one of the universities had significantly higher scores on three out of the four criterion variables. Because Type I error rate was already likely inflated by the large number of coefficients tested, we did not create additional interaction terms involving university but future studies should consider university context as a potential moderator. Results were obtained from a relatively modest sample of men and African American men were in very short supply (as might be expected given their representation in STEM majors). Whether these effects, or the absence of effects, would hold up over time, with larger samples, or in other universities, remains to be seen. Another related concern is that students in their first semester may have had only a few STEM courses so future work should be done to increase the number of data points for performance-related data.

To summarize, the potential intersection between an individual tendency to be self-critically perfectionistic and the interpersonal context within which STEM-related performance emerges over time should be examined in future research, perhaps through gender-balanced and gender-imbalanced teams tasked to solve science- or math-relevant problems. Interventions addressing the pipeline to STEM careers may be improved by noting that some form of social comparison reactance may be triggered when their dispositional, evaluative-sensitive characteristic comes in contact with a primed context that has performance-related implications.

References

PERFECTIONISM MODERATES STEREOTYPE THREAT


Rice, K. G., Tueller, S., & Richardson, C. M. E. (2012). The short-form of the revised Almost Perfect Scale. Unpublished manuscript, Georgia State University, Atlanta, GA.
