They're comparing me to her: Social comparison perceptions reduce belonging and STEM engagement among women with token status

Veronica Derricks^{1,2} and Denise Sekaquaptewa²

¹Department of Psychology, Indiana University – Purdue University Indianapolis

²Department of Psychology, University of Michigan

Author Note

Correspondence concerning this article should be addressed to Veronica Derricks, Department of Psychology, Indiana University – Purdue University Indianapolis, 402 N Blackford St. Indianapolis, IN 46202, United States. Email: vderrick@iu.edu.

This is the author's manuscript of the work published in final edited form as:

Derricks, V., & Sekaquaptewa, D. (2021). They're Comparing Me to Her: Social Comparison Perceptions Reduce Belonging and STEM Engagement Among Women With Token Status. *Psychology of Women Quarterly*, 45(3), 325–350. <https://doi.org/10.1177/03616843211005447>

Abstract

Belonging and academic engagement are important predictors of women's retention in STEM. To better understand the processes influencing these outcomes, we investigate how numerical underrepresentation (i.e., token status) triggers *social comparison perceptions* – concerns that others are comparing oneself to another person – that can undermine women's STEM outcomes. Across four experiments, female college students recruited via subject pool (Study 1a) and MTurk (Studies 1b-3) read a hypothetical scenario in which another female (Studies 1a-3) or male student (Study 2) performed well or poorly in an engineering course. Findings showed that having token (versus non-token) status in the course increased social comparison perceptions (i.e., perceptions about being compared to an ingroup peer), which subsequently reduced course belonging (Studies 1a-1b). Study 2 found that (a) token status increased social comparison perceptions in response to the ingroup (versus outgroup) peer, and (b) social comparison perceptions decreased belonging through stereotype threat concerns, particularly when the peer performed poorly. Study 3 directly manipulated social comparison perceptions to further establish their causal role on negative outcomes and demonstrated that these perceived direct comparisons predicted additional consequences signaling STEM disengagement. Collectively, findings identify a novel process that can diminish belonging and academic engagement for women in STEM. *Additional online materials for this article are available on PWQ's website at PRODUCTION INSERT WEB ADDRESS.*

Keywords: belonging, identity threat, social comparison perceptions, social identity, stereotype threat concerns, tokenism

They're Comparing Me to Her: Social Comparison Perceptions Reduce Belonging and STEM Engagement Among Women with Token Status

Female students in male-dominated fields such as science and engineering may often find themselves to be one of few women in the classroom, a situation with demonstrated negative effects on women's academic outcomes including performance (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2002), participation (Dasgupta et al., 2015; Lewis et al., 2019), motivation (Leaper & Starr, 2019), and importantly, sense of belonging, a key psychological predictor of women's retention in the field (Dennehy & Dasgupta, 2017; Murphy, et al., 2007; Stout et al., 2011). In this paper, we examine one process by which being one of very few women in a male-dominated setting can diminish women's sense of belonging, defined as having a sense of stable, positive relationships with others in the setting, as well as their academic engagement (e.g., motivation to persist; Baumeister & Leary, 1995; Walton & Cohen, 2011). Previous work has identified several routes through which numerical underrepresentation can reduce women's academic outcomes, such as activation of identity-relevant stereotypes or greater apprehension about being the target of prejudice (Murphy & Taylor, 2012; Schuster & Martiny, 2017). In the current work, we focus specifically on the role of social comparison perceptions, which we define as perceptions that others are directly comparing oneself to a specific other person. Using a hypothetical scenario, we propose that when only one other woman is present in a setting (relative to many women), a woman may perceive that others are directly comparing her performance to that of the other woman, the consequences of which may depend on whether the other woman performs well or poorly.

Investigating the effect of social comparison perceptions on sense of belonging is particularly important because motivation to belong in one's environment often guides

subsequent attention and behavior (Baumeister & Leary, 1995; Walton et al., 2012). Within empirical literature, having a strong sense of belonging in one's field of study is a robust predictor of positive outcomes, such as psychological well-being, motivation and academic achievement, intentions to pursue a career in one's field, and the likelihood of recommending that field to others (Good et al., 2012; Lewis et al., 2017; Master et al., 2016; Nadal & Haynes, 2012; Smith et al., 2013; Walton & Cohen, 2007, 2011). Other indicators of academic engagement, such as persistence and participation, are also frequently associated with important outcomes (e.g., retention in one's field; Dasgupta & Stout, 2014; Hernandez et al., 2013).

Therefore, belonging and academic engagement have been studied extensively in the context of women in science, technology, engineering, and mathematics (STEM) fields given persistent disparities showing that women withdraw from STEM fields at significantly higher rates than their male counterparts (Ellis et al., 2016; Lewis & Sekaquaptewa, 2016; Master & Meltzoff, 2020; National Science Foundation, 2019; Pietri et al., 2019; Strenta et al., 1994). We focus our investigation on women at a relatively early stage in the academic pipeline—college students—given the influence of their sense of belonging on decisions to enter, withdraw, or persist in STEM majors. One reason why women may be especially likely to withdraw from male-dominated academic STEM fields is that their sense of belonging in these settings is undermined by being in the numerical minority.

When women are underrepresented in a field of study, they are likely to find themselves to be one of few women present in specific settings in that field, such as a classroom, which can make their gender identity more salient than when many other women are present. As such, we develop a theoretical model in which being one of very few women (specifically, one of two) present in a STEM setting can trigger social comparison perceptions that are identity-based, as

women feel that others are directly comparing them to the other woman in the setting. Identity-based social comparison perceptions, in turn, may activate concerns about being viewed through the lens of one's gender identity and result in worse outcomes (e.g., a lowered sense of belonging) for women.

The Consequences of Experiencing Token Status

Researchers have differentiated between minority status (being in the numerical minority; e.g., Taylor et al., 1978), token status (being in a situation in which one's social group is less than 15% of the larger group present; e.g., Kanter, 1977), duo status (when there are two and only two minority individuals present in a group of six or larger; Loyd et al., 2008), and solo status (being the only member of one's social category present; e.g., Thompson & Sekaquaptewa, 2002). Given that women comprise 20% or less of many STEM majors (National Science Foundation, 2019), it seems likely that college women studying STEM may often be in classrooms or lab sections in which two women comprise 15% or less of the total class enrollment. Therefore, we place our studies of social comparison perceptions between two women (who are in an otherwise all-male STEM classroom) within the context of token status. In this situation, the two women are also experiencing duo status. According to duo status theorizing, when there are two and only two women in the setting, they each experience peak intergroup pressure (expectations of being stereotyped by the outgroup), which can induce identity-based threat (Loyd et al., 2008). Therefore, this work also contributes to the study of duo status, which has heretofore been solely focused on theorized interpersonal dynamics between the duo pair, as opposed to their individual sense of belonging in STEM settings. However, our studies remain placed in the token status literature because we are primarily focused on the proportion of women present, rather than the absolute number of women present.

Some investigations have described token status in terms of attributions for inclusion, as token individuals are perceived by others as having been selected for inclusion in the setting by virtue of their social identification (e.g., to meet gender quotas; Craig & Feasel, 1998; Heilman & Herlihy, 1984; Kanter, 1977). However, even without such selection attributions, numerical token status often leads to subjective perceptions of gender inequity among tokens in the setting, which may often reflect the actual negative treatment of women (King et al., 2010). Therefore, it is likely that consequences of perceived social comparisons associated with token status are psychologically connected to the perceived threat of gender inequity or gender stereotyping, regardless of attributions for the gender imbalance in the setting.

Differential Outcomes Associated with the Presence of Ingroup Members

If token status is associated with more negative outcomes, then it seems likely that having more ingroup members present would be beneficial for the token. Social identity theory posits that identification with social groups enhances psychological attachment to other ingroup members, which can, subsequently, give people a greater sense of belonging in their social environments (Tajfel & Turner, 1986). Thus, the presence of ingroup members can yield psychological benefits because their representation is often perceived as an identity-safety cue (Avery et al., 2004; Hagedorn et al., 2007; Johnson et al., 2019). The presence of ingroup members may also be beneficial because individuals often derive their self-esteem from their group membership. For example, when people view an ingroup member who excels, their physiological responses mirror how they would respond if they had achieved the success themselves (Bernhardt et al., 1998). Given the ways in which ingroup members may signal safety or enhance self-esteem, people should exhibit psychological benefits in response to their presence.

Indeed, prior research on tokenism (Kanter, 1977) and solo status (Lord & Saenz, 1985; Sekaquaptewa & Thompson, 2002, 2003) have demonstrated the negative effects of not having ingroup members in one's environment. For example, people who find themselves in environments with few or no ingroup members generally show decreased participation in the setting, deficits in task performance, increased physiological and cognitive vigilance, stronger collective self-construals, and greater performance apprehension (Inzlicht & Ben-Zeev, 2000; Murphy et al., 2007; Saenz, 1994; Sekaquaptewa et al., 2007). Moreover, women who anticipated having token status in a group were more likely to prefer a different group, express an interest in changing the gender composition of the group, and expect to be the target of stereotypic treatment than non-token women (Cohen & Swim, 1995).

Although people often perceive the presence of ingroup members as an asset, particularly when their representation is otherwise limited, there are also contexts in which the presence of ingroup members may produce negative outcomes. In particular, adverse effects may emerge in settings marked by gender bias, which tend to promote more negative perceptions and interactions between women. Consistent with this idea, prior research has identified a "queen bee" phenomenon, which typically arises when women encounter hostile, sexist environments. In response to these settings, women may increase psychological distance from their gender identity and engage in gender stereotyping of subordinate women due to social identity threat (Derks et al., 2011). Although the mechanism underlying these negative behaviors has yet to be fully tested directly, psychologists theorize that queen bee behavior may be particularly likely to emerge when women in sexist environments are motivated to distance themselves from unfavorable gender representations (Derks et al., 2016). Because women experience both underrepresentation and gender stereotyping in STEM fields, it is possible that

token women's perceptions of each other are negatively affected by the identity threat aroused in this situation.

Identity threat, also referred to as stereotype threat, involves concerns about being evaluated on the basis of negative, group-based stereotypes (Steele et al., 2002). When individuals experience stereotype threat, they frequently exhibit concerns about confirming these stereotypes, which subsequently elicits negative arousal and worse outcomes (Ben-Zeev et al., 2005; Schmader et al., 2008). Token women may be particularly susceptible to stereotype threat given the salience of their gender identity; for instance, token women may be vigilant about monitoring their environment for cues to determine how they are being viewed by others (e.g., whether they are being negatively stereotyped). In male-dominated settings, one cue that may be especially informative is the presence of another woman who shares the distinctive gender identity. Given their shared group membership, women may perceive that the way in which one group member is evaluated will be used to make generalizations about the group as a whole (Cohen & Garcia, 2005). These comparisons, consequently, can remind women of the limited way in which they are viewed (e.g., being evaluated on the basis of their gender) and impede their academic outcomes (e.g., belonging). Therefore, we assert that token status in a STEM setting will facilitate social comparison perceptions made under threat, which can ultimately diminish women's STEM outcomes.

Social Comparison Perceptions Under Threat

Because women in male-dominated environments (such as many STEM classrooms) may experience identity threat aroused by token status, the heightened salience of their gender identity may lead them to perceive that others are not only viewing them in terms of their gender, but are also comparing them directly with the other person of their gender in that setting.

These social comparison perceptions—perceptions that one is being compared to a specific other person in the setting—may have an important influence on women’s sense of belonging in those settings, and may depend on whether the other woman performs well or poorly. Classic research on social comparison shows that people frequently make self-comparisons, comparing themselves to others, to inform their self-concepts (Festinger, 1954). General trends show that making downward social comparisons (to a lower performing individual) enhances self-esteem, whereas making upward social comparisons (to a higher performing individual) decreases self-esteem. Therefore, it could be predicted that within a token status situation when one woman performs poorly, the other woman would benefit from downward social comparisons, and when one woman performs well, the other woman would be threatened by upward social comparisons. However, comparisons perceived to be made by others (versus self-generated) about oneself may become particularly salient in contexts marked by threat, to the extent that concerns about being evaluated based on gender stereotypes are heightened in this situation. If token status highlights one’s salient social identity, this situation will likely increase concerns about whether others are attending to the behavior of people who share the salient identity and drawing comparisons between them. Thus, in the token situation, the influence of these perceived social comparisons may outweigh those of self-comparisons.

Previous work on social identity threat shows that threat is driven primarily by the perception that *others* endorse a stereotype about one’s group; one’s own endorsement or internalization of the stereotype is not required for social identity threat to affect one’s outcomes and experiences (Aronson et al., 1999; Schmader et al., 2004; Steele et al., 2002). Therefore, in the situation of social identity threat (e.g., having token status), identity-based social comparison perceptions (involving perceptions of comparisons made by others, as opposed to those made by

oneself) may be particularly relevant due to the heightened salience of one's gender and its associated stereotypes in that situation. In support of this idea, previous research shows that when members of marginalized groups have solo status, they are more likely to perceive that they are the focus of others' attention, particularly when their group identity is salient (Crosby et al., 2014; Pollak & Niemann, 1998). Consequently, when one other ingroup member is present, the salience of one's identity may increase concerns that external observers are attending to this identity and making direct comparisons with the other ingroup member (Marques et al., 1988; Pinto et al., 2010). When that salient identity is negatively stereotyped in the situation, the consequence may be that both upward *and* downward social comparison perceptions are detrimental to women's sense of belonging.

Mechanisms Through Which Social Comparison Perceptions Undermine Outcomes

Social comparison perceptions may lead to negative outcomes for women with token status due to different mechanisms, depending on whether the other woman they feel compared to is performing worse (downward social comparison perception) or better than oneself (upward social comparison perception). In the case of downward social comparison perceptions, token women may experience diminished belonging because the threat elicited by the situation reminds women of the limited way in which their group is seen (Fryberg et al., 2008). For instance, past research shows that when racial minority students saw an ingroup member perform poorly on a stereotype-relevant task (e.g., a standardized exam), they exhibited increased stereotype activation and greater physical distancing from the ingroup member due to concerns that the ingroup member's behavior would reflect negatively on their racial group as a whole (Cohen & Garcia, 2005). Therefore, perceiving that oneself is being compared to an ingroup member performing poorly on a stereotype-relevant task may increase perceptions of

being seen through the lens of one's negatively stereotyped identity (rather than as an individual). As such, we hypothesized that downward social comparison perceptions would be particularly likely to activate stereotype threat-related concerns among women with token status (relative to contexts where many women are present).

In the case of upward social comparison perceptions, women may also experience diminished outcomes but through a different mechanism, specifically, concerns about not being able to achieve the level of performance displayed by the higher-performing woman. Although perceptions that one is being compared to a high-performing ingroup member can be a source of pride in that one's group is being represented favorably (e.g., by challenging negative stereotypes about the group), research shows this event may also elicit threat if that success feels personally unreachable (Asgari et al., 2012; Betz & Sekaquaptewa, 2012; Brewer & Weber, 1994). This possibility is supported by research demonstrating that adolescent girls exhibited decreased STEM interest and self-rated ability when they were exposed to feminine STEM role models because the combination of femininity and STEM-relevant success felt particularly unattainable (Betz & Sekaquaptewa, 2012). Similar to downward social comparison perceptions, upward social comparison perceptions may also increase stereotype threat-related concerns. In particular, upward social comparison perceptions may strengthen token women's beliefs that they are greater risk of confirming negative stereotypes about the group (e.g., that women have low STEM ability), relative to another high-performing woman. Activating these beliefs, in turn, can evoke threat. Therefore, although past research focuses primarily on adverse outcomes that emerge in response to downward comparisons, upward comparisons to ingroup peers may also elicit stereotype threat outcomes among women who encounter novel token situations in STEM.

Overview of the Current Studies

Across four experiments, we focus on the specific situation of token status among women in academic STEM settings, building on current literature in social belonging, intragroup processes, and social comparison. We use the term “token status” to refer to the situation of being one of two women in a STEM classroom in which two is 15% or less of the class enrollment, without explicit attributions about inclusion in the class. Specifically, we test (a) one direct consequence of token status, social comparison perceptions (e.g., feeling compared to an ingroup member), that may undermine a sense of belonging in the setting (Studies 1a-1b), (b) why these comparisons might decrease belonging (Study 2), and (c) the extent to which this process is specific to ingroup peers (Study 2). Additionally, we experimentally manipulate social comparison perceptions to (1) further establish causal mediation, and (2) examine the effect of these perceived comparisons on other academic outcomes (e.g., STEM engagement; Study 3).

Using a hypothetical scenario, we theorized that women would be more likely to feel compared to another woman in engineering settings where they had token status (versus being in non-token, i.e., gender-balanced settings; Study 1a). Furthermore, we hypothesized that one route through which token status may reduce women’s sense of belonging is social comparison perceptions (their perceptions of being compared to the other female student: an ingroup peer; Study 1b). Additionally, we predicted that social comparison perceptions may decrease belonging through two different mechanisms (Study 2); when the ingroup peer performed well, women may feel concerned about being able to measure up to her performance. When the peer performed poorly, however, concerns about being evaluated on the basis of negative gender stereotypes may remind women of the limited way in which they are viewed and consequently, activate stereotype threat-related concerns. Moreover, we theorized that these processes would be particularly likely to emerge in response to an ingroup peer and would not generalize to an

outgroup peer (i.e., a male student; Study 2). Finally, we hypothesized that experimentally manipulating social comparison perceptions (e.g., by enhancing or minimizing concerns about being compared to an ingroup peer) would offer additional support for its role as a causal mechanism underlying negative outcomes (e.g., diminished belonging and STEM engagement; Study 3).

Study 1a

This pilot study sought to assess the feasibility of women perceiving ingroup comparisons when they have token status. As such, we utilized a hypothetical scenario about a college engineering course to examine how (a) the gender composition of the course, and (b) whether an imagined same-gender peer is performing well or poorly in the course, impacts women's perceptions of being compared to the peer.

Method

Participants

We recruited 94 college women (96.8% European American and 3.2% multiracial; $M_{\text{age}} = 18.81$; $SD_{\text{age}} = 0.89$; 42.6% perceived their major to be STEM-related) from a large, public Midwestern university. We aimed to recruit as many participants as possible across one academic semester, and data collection ceased when the semester ended. All participants were included in data analysis and received course credit in exchange for their participation.

Procedure

Participants were randomly assigned to one of four conditions in a 2 (gender composition: token status, non-token status) X 2 (academic event: success, setback) between-subjects design.

In this experiment, participants were told that the researchers were interested in

understanding college students' experiences. Participants read a vignette where they were asked to imagine that they were juniors in an engineering course. To manipulate the gender composition of the course, participants were told they were one of two (token condition) or thirteen (non-token condition) women and saw a course roster that contained a list of 29 student names (see Figure 1).

Following the gender composition manipulation, participants read that they were enrolled in a difficult engineering course where they have had a satisfactory performance thus far. Next, participants read about Jane M., another woman enrolled in the course. In the "academic success" condition, Jane performed well on an engineering presentation, receiving applause from her peers and being told that she did an incredible job on her presentation by the instructor. In the "academic setback" condition, Jane performed poorly, receiving confused looks and tepid applause from her peers. Moreover, the instructor stated that he would be speaking to some students about their project individually while looking at Jane (see the online supplement for the full vignettes).

After reading the vignette, participants completed items regarding their evaluations of the ingroup peer (Jane), attributions about the peer's performance, several individual difference measures (e.g., endorsement of feminism), and other assessments. For the sake of brevity, analyses using these items are reported in the online supplement. Results related to our primary outcome of interest, perceptions of being compared to the ingroup peer, are reported below.

Measure: Perceptions of Being Compared to an Ingroup Peer

Participants reported their perceptions of being compared to an ingroup peer on two items using a Likert-type scale ranging from 1 (*not at all*) to 9 (*very much so*); "How much do you think your class performance is being directly compared to Jane M's?"; "How much do you

think your ability in engineering is being directly compared to Jane M's?". These two items were highly correlated ($r = .85$), and as such, were collapsed into an index.

Analytic Strategy

Analyses were conducted in SPSS v.26.0 using univariate analysis of variance (ANOVA). To investigate the effect of gender composition and the ingroup peer's academic performance on participants' perceptions about being compared to the ingroup peer, we examined the main effects (gender composition and academic event). Although the analysis also included the Gender Composition x Academic Event interaction, our sample size was too small to draw any meaningful conclusions about this effect.

For the current and subsequent studies, all effect sizes are reported as Cohen's d (main effects; .2 = small effect, .5 = medium effect, .8 = large effect) or partial eta-squared (interactions; .01 = small effect, .06 = medium effect, .14 = large effect; Cohen, 1988). Statistical means and standard deviations that are not reported below are presented in the online supplement.

Results

Analyses revealed a significant main effect of gender composition, $F(1, 90) = 11.50$, $p < .001$, $d = .71$, showing that participants in the token condition ($M = 6.54$, $SD = 1.98$) reported stronger perceptions of being compared to the ingroup peer than participants in the non-token condition ($M = 5.16$, $SE = 2.06$). Neither the main effect of academic event, $F(1, 90) = 3.61$, $p = .061$, $d = 0.40$, nor the Gender Composition x Academic Event interaction, $F(1, 90) = 0.56$, $p = .456$, $\eta_p^2 = .006$, were significant.

Discussion

Findings from this pilot study showed that, in line with our predictions, college women

were more likely to feel compared to an ingroup peer when they had token status in the course (versus when the gender ratio was balanced). To our knowledge, this is the first empirical study to demonstrate that token status increases perceptions of being socially compared to an ingroup member. However, whether the ingroup member was performing well or poorly in the course did not significantly impact feelings of comparison.

Study 1b

Pilot data revealed that college women who imagined that they were one of two (versus many) women in an engineering course reported stronger perceptions of being compared to an ingroup peer. To replicate and extend Study 1a, Study 1b had two primary aims. First, Study 1b sought to determine the source of these comparisons (e.g., whether participants are personally drawing comparisons between themselves and an ingroup peer or whether they perceive that others – their instructor and peers – are drawing these comparisons). We theorized that having token status would be particularly likely to increase participants' perceptions that external observers are drawing comparisons between them and an ingroup peer (e.g., social comparison perceptions), but would not increase their own likelihood of making comparisons to her (self-comparisons), because the salience of their gender identity would increase women's perceptions that others are judging them in terms of the stereotype (according to stereotype threat theory; Steele & Aronson, 1995). Furthermore, we wanted to assess whether decreases in belonging were being driven by social comparison perceptions (rather than self-comparisons). Although one possibility is that self-comparisons might serve as an alternate mechanism that undermines belonging (independent from, or in conjunction with, social comparison perceptions), we expected that self-comparisons would be unrelated to belonging (or associated with increased

belonging) because people often make social comparisons as a means of restoring or protecting their self-esteem, particularly under threat (Suls & Wheeler, 2013).

Second, this study investigated the downstream consequences of these comparisons on belonging, an outcome that has important implications for retention in one's field. As such, Study 1b utilized a path model to (a) examine whether having token status undermines participants' sense of the belonging in the course, consistent with prior literature, and (b) test a mechanism underlying this relationship: social comparison perceptions. We hypothesized that having token status would increase participants' perceptions that others are comparing them to the ingroup peer, and these perceptions would subsequently reduce participants' sense of belonging in the course.

Method

Participants

Based on the effect size observed in Study 1a, we set out to recruit approximately 75 participants per cell and over-sampled given planned exclusions (e.g., participants who did not meet the demographic criteria). Post-hoc sensitivity analyses conducted in G*Power V. 3.1 (Faul et al., 2007) showed that we were able to detect an effect size of $f = 0.16$ for the 2 x 2 ANOVA (e.g., Gender Composition x Academic Event interaction).

We obtained a sample of 320 participants after directly recruiting college women on Turkprime. Towards the end of the study, we confirmed participants' eligibility by asking if they met our demographic criteria. Participants were told that they would be paid regardless of their answer. Before data analysis, we excluded 21 participants who reported that they did not meet the demographic criteria (identifying their gender as male ($n = 6$), identifying their gender as non-binary ($n = 1$), or reporting that they were not college students ($n = 14$))¹. After removing

these participants, we retained 299 college women (74.2% European American, 9.4% African American, 6.7% Asian American, 4.3% Latina, 5.4% other race/ethnicity or multiracial; $M_{\text{age}} = 22.43$; $SD_{\text{age}} = 1.59$; 44.1% perceived their major to be STEM-related; 78.3% were enrolled in a four-year college).

Procedure

The study design and procedure were identical to Study 1a. However, we included additional measures to test Study 1b's primary research aims.

Measures

After reading the vignette, participants responded to survey items measuring (a) whether they were directly comparing themselves to the ingroup peer, (b) whether they perceived that their instructor and peers were comparing them to the ingroup peer (i.e., social comparison perceptions), and (c) their sense of belonging in the engineering course. The complete wording for all survey items is reported in the online supplement.

Self-comparisons to the Ingroup Peer. We assessed whether participants were comparing their own engineering performance and ability to the ingroup peer with two items using a Likert-type scale ranging from 1 (*not at all*) to 9 (*very much so*); e.g., "How much do you think you would directly compare your class performance to Jane's" ($r = .72$).

Social Comparison Perceptions. We adapted the two items measuring self-comparisons to the ingroup peer to measure participants' social comparison perceptions (i.e., perceptions that the instructor and peers were comparing their engineering performance and ability to the ingroup peer). We measured these items using a Likert-type scale ranging from 1 (*not at all*) to 9 (*very much so*); e.g., "How much do you think your class performance is being directly compared to Jane's by the instructor and other students?". Because we theorized that having token status

would be particularly likely to increase social comparison perceptions, we increased statistical power of the index by measuring five additional items on a Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*); e.g., “I would worry that people in this class would draw conclusions about me based on Jane’s performance”. Because the items were measured on different scales², we computed z-scores for all seven items before averaging them into an index ($\alpha = .83$).

Sense of Belonging in the Course. We measured participants’ sense of belonging in the engineering course with eight items using a Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*); e.g., “I would probably feel like I belong in this engineering course”. These items were adapted from previous literature and showed high reliability ($\alpha = .92$; Good et al., 2012; Purdie-Vaughns et al., 2008; Walton & Cohen, 2007). As such, the items were aggregated into an index.

Analytic Strategy

As in Study 1a, we used ANOVA to examine the effects of gender composition, academic event, and the Gender Composition x Academic Event interaction on the primary study outcomes. Significant interactions were further probed by examining the simple effects.

Results

Self-comparison to the Ingroup Peer

Analyses revealed a significant main effect of academic event, $F(1, 295) = 22.43, p < .001, d = .55$, showing that participants were more likely to compare themselves to the ingroup peer when she performed well ($M = 6.09, SD = 2.05$), versus poorly in the course ($M = 4.91, SD = 2.28$). Neither the main effect of gender composition, $F(1, 295) = 3.59, p = .059, d = -.22$, nor

the Gender Composition x Academic Event interaction, were significant, $F(1, 295) = 0.50, p = .480, \eta_p^2 = .002$.

Social Comparison Perceptions³

Analyses revealed a significant main effect of gender composition on social comparison perceptions, $F(1, 295) = 4.83, p = .029, d = .26$. Means showed that participants in the token condition ($M = 0.08, SD = 0.72$) were more likely to report social comparison perceptions (i.e., perceive that their instructor and peers were comparing them to the ingroup peer) than participants in the non-token condition ($M = -0.09, SD = 0.67$). Moreover, a significant main effect of academic event, $F(1, 295) = 6.47, p = .011, d = .30$, showed that participants were more likely to perceive social comparison perceptions when the ingroup peer performed well ($M = 0.10, SD = 0.65$), versus poorly ($M = -0.11, SD = 0.74$), in the course. The Gender Composition x Academic Event interaction was not significant, $F(1, 295) = 1.51, p = .220, \eta_p^2 = .005$.

Sense of Belonging in the Course

Analyses revealed a significant main effect of gender composition, $F(1, 295) = 4.38, p = .037, d = -.24$, showing that participants in the token condition ($M = 5.09, SD = 1.07$) reported a decreased sense of belonging in the course than participants in the non-token condition ($M = 5.33, SD = 1.06$). Moreover, a significant main effect of academic event, $F(1, 295) = 19.84, p < .001, d = .52$, showed that participants were more likely to perceive that they belonged in the course when the ingroup peer performed well ($M = 5.46, SD = 1.00$), versus poorly ($M = 4.93, SD = 1.08$), in the course. The Gender Composition x Academic Event interaction was not significant, $F(1, 295) = 1.51, p = .220, \eta_p^2 = .005$.

Testing the Proposed Mechanism: Social Comparison Perceptions

Our proposed model examined the mechanism underlying the relationship between the

gender composition of the course and sense of belonging: social comparison perceptions. Models examining the role of self-comparisons demonstrated that self-comparisons were not associated with decreased belonging. For the sake of brevity, these analyses are reported in the online supplement.

To assess the indirect effect, we generated models using AMOS v. 26 (see Figure 2).⁴ In the current and subsequent studies, we evaluated model fit using criteria outlined in prior research (e.g., nonsignificant chi-squared test, CFI > 0.90, TLI > 0.95, and RMSEA < 0.08; Hooper et al., 2008). Moreover, analyses utilized 5,000 bootstrap samples to produce 95% bias-corrected confidence intervals. For the current and subsequent studies, the paths of greatest theoretical interest are presented in the text, and all other model parameters and test statistics are reported in the tables or online supplement.

Because our proposed model was saturated (i.e., contained zero degrees of freedom), fit statistics could not be computed. Based on prior recommendations, we evaluated model fit by comparing the AIC of our proposed model to all alternative models (Kenny, 2018). Analyses revealed that our model had the second best fit (AIC = 12.00) relative to another model that eliminated the direct effect of gender composition on belonging (AIC = 11.95; $\chi^2(1) = 1.95$, $p = .163$, CFI = 0.97, TLI = 0.91, RMSEA = .06). However, given the minimal change in AIC and our theoretical interest in the direct effect, we conducted analyses for the saturated model.

As observed in the ANOVA, analyses revealed a significant main effect of gender composition on social comparison perceptions, $b = 0.17$, $SE = 0.08$, $p = .038$, 95% CI = [0.01, 0.33], showing that participants in the token condition were more likely to report social comparison perceptions than participants in the non-token condition. Stronger social comparison perceptions, in turn, predicted reductions in participants' sense of belonging in the course, $b =$

-0.45, $SE = 0.10$, $p < .001$, 95% CI = [-0.64, -0.27]. After accounting for the role of social comparison perceptions, the direct effect of gender composition on belonging was no longer significant, $b = -0.17$, $SE = 0.12$, $p = .194$, 95% CI = [-0.40, 0.09].

Although the Gender Composition x Academic Event interaction did not predict social comparison perceptions or belonging in the course, we were interested in testing whether social comparison perceptions predicted reductions in belonging regardless of whether the ingroup peer was performing well or poorly (see Figure 3). Therefore, we tested whether the relationship between social comparison perceptions and belonging was moderated by the ingroup peer's academic performance (e.g., academic event).

Analyses revealed that fit statistics for this model were somewhat inadequate, $X^2(3) = 7.64$, $p = .054$, CFI = .930, TLI = .767, RMSEA = .072. As observed in the previous model, analyses revealed a significant main effect of gender composition on participants' social comparison perceptions, $b = 0.16$, $SE = 0.08$, $p = .048$, 95% CI = [0.00, 0.32]. Although social comparison perceptions still predicted reductions in belonging, $b = -0.51$, $SE = 0.09$, $p < .001$, 95% CI = [-0.70, -0.34], this relationship was not moderated by academic event, $b = 0.12$, $SE = 0.18$, $p = .508$, 95% CI = [-0.23, 0.47]. Additionally, the direct effect of gender composition on belonging was not significant, $b = -0.15$, $SE = 0.12$, $p = .221$, 95% CI = [-0.37, 0.09].

Discussion

Replicating Study 1a, Study 1b revealed that women who had token status in an engineering course were more likely to perceive social comparison perceptions (i.e., that external observers were comparing them to the ingroup peer) than women in the gender-balanced course. Importantly, findings showed that having token status increased social comparison perceptions, but did not increase women's own likelihood of comparing themselves

to the peer.

Furthermore, mediation analyses identified social comparison perceptions as a mechanism underlying the relationship between the gender composition of the course and sense of belonging. Thus, modeling the indirect effect showed that having token status increased social comparison perceptions, and these perceptions subsequently predicted reductions in belonging.

Taken together, findings suggest that one reason why having token status may undermine women's belonging in male-dominated settings is that women perceive that others are comparing them to an ingroup peer. Therefore, under token status, decreased sense of belonging is driven by women's perceptions that external observers are making ingroup comparisons, rather than their own comparisons.

Study 2

Study 2 aimed to address two additional research questions. First, we sought to assess the generalizability of the model tested in Study 1b. Although the previous findings were examined in the context of an ingroup peer, Study 2 investigated whether the observed processes are specific to ingroup peers or if women show similar processes when they read about a male student in the class. We theorized that due to the salience of their gender identity, women would be particularly likely to perceive that external observers are comparing them to an ingroup (versus outgroup) peer, suggesting that social comparison perceptions are most likely to emerge in response to ingroup members.

Another aim of Study 2 was to examine why social comparison perceptions reduced belonging for women, regardless of the peer's performance in the course. We hypothesized that social comparison perceptions would diminish belonging through different routes. Specifically,

we expected that social comparison perceptions would mitigate women's sense of belonging because (a) when the peer is performing well, feeling compared to her may increase concerns about being able to measure up to her performance, and (b) when the peer is performing poorly, feeling compared to her may remind women of the limited way in which they are being evaluated (e.g., on the basis of their gender), which can activate stereotype threat-related concerns.

Method

Participants

In Study 2 we recruited as many participants as resources allowed, with a goal of at least 50 participants per cell. Using post-hoc sensitivity analyses in G*Power V. 3.1, we were able to detect an effect size of $f = 0.13$ for a 2 x 2 ANOVA (e.g., Gender Composition x Peer Gender interaction).

As in Study 1b, we directly recruited college women on Turkprime and obtained a sample of 520 participants. Before data analysis, we excluded 74 participants who reported that they did not meet our demographic criteria at the end of the study (identifying their gender as male ($n = 61$), identifying their gender as non-binary ($n = 4$), and reporting that they were not college students ($n = 9$)). After removing these participants, we retained 446 college women (60.5% European American, 18.2% African American, 5.6% Asian American, 7.2% Latina, 8.5% other race/ethnicity or multiracial; $M_{age} = 23.11$; $SD_{age} = 2.66$; 36.5% perceived their major to be STEM-related; 77.1% were enrolled in a four-year college).

Procedure

In addition to manipulating the gender composition of the course and the imagined peer's academic performance, Study 2 utilized another manipulation (peer gender) to account for

the role of the imagined peer's gender in the vignette. As such, participants were randomly assigned to one of eight conditions in a 2 (gender composition: token status, non-token status) X 2 (academic event: success, setback) X 2 (peer gender: female, male) between-subjects design.

To manipulate peer gender, participants read the vignettes utilized in the earlier studies that described the performance of an ingroup peer (Jane) or an outgroup peer (Eric). No additional changes were made to the study procedure.

Measures

Following the experimental manipulations, participants responded to the survey items used in Study 1b. We adapted these items to correspond with the imagined peer, Jane or Eric, that students read about in the vignette. Specifically, participants reported (a) the extent to which they were directly comparing themselves to the peer ($r = .78$), (b) their social comparison perceptions ($\alpha = .79$), and (c) their sense of belonging in the engineering course ($\alpha = .90$).

Additionally, participants answered items to test two possible mechanisms underlying the relationship between social comparison perceptions and decreased belonging: concerns about measuring up to the peer's performance and stereotype threat-related concerns. All items were measured on a Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Concerns About Measuring up to the Peer's Performance. We measured participants' concerns about measuring up to the peer's performance using four items (e.g., "I would be concerned that I wouldn't measure up to Jane/Eric's performance"). These items showed high reliability ($\alpha = .92$) and were aggregated into an index.

Stereotype Threat-related Concerns. We measured stereotype threat-related concerns using 12 items. Eight of the items were adapted from prior research (Ramsey et al., 2013) and measured concerns about the peer's performance reflecting on their gender group (e.g., "I would

worry that my instructor and peers would draw conclusions about members of my gender group based on Jane/Eric's performance"). Four additional items measured concerns about the peer's performance reflecting on the self (e.g., "I would be concerned that, because of Jane/Eric's performance, my instructor and peers would expect me to perform poorly"). All 12 items showed high reliability ($\alpha = .92$) and as such, were aggregated into an index. Analyses examining these subscales separately are reported in the online supplement.

Analytic Strategy

As in the previous studies, we used ANOVA to examine the main effects of gender composition, academic event, and peer gender, as well as the two- and three-way interactions, on the primary study outcomes. Significant interactions were further probed by examining the simple effects. Statistical means and standard deviations for significant main effects are presented below, and the means and standard deviations for significant interactions are reported in Table 1. All means, by condition, are reported in the online supplement.

Results

Self-comparison to the Peer

Analyses revealed a significant main effect of peer gender, $F(1, 438) = 7.67, p = .006$, $d = .26$, showing that participants were more likely to compare themselves to the female peer ($M = 5.53, SD = 2.21$) than the male peer ($M = 4.92, SD = 2.42$). Although the main effect of academic event was significant in Study 1b, neither this effect nor any of the remaining main effects, two-way interactions, or the three-way interaction were significant (all $ps > .147$).

Social Comparison Perceptions⁵

Analyses revealed a significant main effect of peer gender, $F(1, 437) = 15.98, p < .001$, $d = .38$, showing that participants reported stronger social comparison perceptions in response

to the female ($M = 0.12$, $SD = 0.63$), versus male, peer ($M = -0.12$, $SD = 0.67$). This main effect was qualified by a significant Gender Composition x Peer Gender interaction, $F(1,437) = 5.90$, $p = .016$, $\eta_p^2 = .013$. In line with our predictions, simple effects showed that participants were more likely to report social comparison perceptions in response to the female peer when they were in the token, versus non-token, condition, $F(1,437) = 8.54$, $p = .004$, $d = .28$. However, the gender composition of the course did not significantly impact their reported social comparison perceptions in response to the male peer, $F(1,437) = 0.27$, $p = .605$, $d = -.05$. Moreover, although participants' social comparison perceptions were unaffected by peer gender in the non-token condition, $F(1,437) = 1.23$, $p = .267$, $d = .11$, participants in the token condition reported stronger social comparison perceptions in response to the female, versus male, peer, $F(1, 437) = 20.59$, $p < .001$, $d = .43$.

However, in contrast to Study 1b, neither the main effect of academic event, $F(1, 438) = 2.26$, $p = .134$, $d = .14$, nor the main effect of gender composition, $F(1, 437) = 2.88$, $p = .090$, $d = .16$, were significant. Moreover, none of the remaining two-way interactions nor the three-way interaction were significant (all $ps > .112$).

Sense of Belonging in the Course

Replicating Study 1b, analyses revealed a significant main effect of gender composition, $F(1, 438) = 11.72$, $p = .001$, $d = -.33$, such that participants in the token condition ($M = 5.04$, $SD = 1.07$) reported a decreased sense of belonging in the course compared to participants in the non-token condition ($M = 5.36$, $SD = 0.91$). None of the remaining main effects, two-way interactions, or the three-way interaction were significant (all $ps > .064$).

Concern About Measuring up to the Peer's Performance

Analyses revealed a significant main effect of academic event, $F(1, 438) = 34.90$,

$p < .001$, $d = .56$, showing that participants reported stronger concerns about measuring up to the peer's performance when they performed well ($M = 4.24$, $SD = 1.59$), versus poorly ($M = 3.32$, $SD = 1.70$), in the course. However, none of the remaining main effects, two-way interactions, or the three-way interaction were significant (all $ps > .273$).

Stereotype Threat-related Concerns

Analyses revealed a main effect of peer gender, $F(1, 438) = 3.85$, $p = .050$, $d = .19$, showing that participants who read about a female peer ($M = 3.80$, $SD = 1.36$) reported stronger stereotype threat-related concerns than participants who read about a male peer ($M = 3.55$, $SD = 1.39$). This effect was qualified by a significant Peer Gender x Academic Event interaction, $F(1, 438) = 10.74$, $p = .001$, $\eta_p^2 = .024$; means showed that although participants reported equal stereotype threat-related concerns when the female and male peer performed well, $F(1, 438) = 0.88$, $p = .349$, $d = -.09$, they reported greater stereotype threat-related concerns when the female (versus male) peer performed poorly, $F(1, 438) = 13.48$, $p < .001$, $d = .35$. Furthermore, although the male peer's performance did not significantly impact stereotype threat-related concerns, $F(1, 438) = 1.48$, $p = .224$, $d = .12$, participants reported stronger stereotype threat-related concerns when the female peer performed poorly, versus well, in the course, $F(1, 438) = 11.72$, $p = .001$, $d = -.33$. None of the remaining main effects, nor the remaining two- and three-way interactions, were significant (all $ps > .089$).

Testing the Proposed Mechanism: Social Comparison Perceptions

In an effort to replicate the model tested in Study 1b, we examined whether social comparison perceptions served as a mechanism underlying the relationship between gender composition and belonging in the course (see Figure 4). As such, we tested whether (a) the relationship between gender composition and social comparison perceptions was moderated by

peer gender, and (b) stronger social comparison perceptions reduced participants' sense of belonging in the course. Given the inclusion of peer gender in this study, we hypothesized that this process would be particularly likely to emerge in response to the ingroup (female) peer.

Fit statistics for this model were good, $X^2(3) = 0.36$, $p = .949$, CFI = 1.00, TLI = 1.21, RMSEA = .000. As observed in the ANOVA, analyses revealed a significant Gender Composition x Peer Gender interaction on social comparison perceptions, $b = 0.30$, $SE = 0.12$, $p = .013$, 95% CI = [0.07, 0.54], such that having token status increased social comparison perceptions in response to the ingroup (female) peer, but not the outgroup (male) peer. Stronger social comparison perceptions, in turn, predicted reductions in participants' sense of belonging in the course, $b = -0.29$, $SE = 0.07$, $p < .001$, 95% CI = [-0.42, -0.15]. Additionally, a significant direct effect of gender composition on belonging revealed that having token status predicted decreased belonging in the course, $b = -0.28$, $SE = 0.09$, $p = .004$, 95% CI = [-0.46, -0.09]. Thus, this analysis indicates that women who have token status feel lower belonging in the course because they perceive that others are comparing them to the other woman in the course.

Next, we sought to investigate why social comparison perceptions predicted reductions in belonging (see Figures 5-6). We expected that decreased belonging would emerge through two routes. When the peer performed well, we hypothesized that social comparison perceptions would increase participants' concerns about measuring up to the peer's performance. When the peer performed poorly, however, we expected that social comparison perceptions would activate stereotype threat-related concerns. As such, this model tested the extent to which (a) social comparison perceptions predicted these mechanisms, (b) this relationship was moderated by academic event, and (c) these mechanisms predicted reductions in belonging. The model accounted for (a) the effects of gender composition on each path, as well as (b) the direct effect

of social comparison perceptions on belonging.

Concerns About Measuring up to the Peer's Performance

Fit statistics for this model were good, $\chi^2(14) = 16.96$, $p = .259$, CFI = .989, TLI = .977, RMSEA = .022. As observed in ANOVA, analyses revealed that the Gender Composition x Peer Gender interaction predicted social comparison perceptions, $b = 0.30$, $SE = 0.12$, $p = .011$, 95% CI = [0.07, 0.55], such that having token status increased social comparison perceptions in response to the ingroup, but not outgroup, peer. Stronger social comparison perceptions predicted greater concerns about measuring up to the peer's performance, $b = 1.44$, $SE = 0.10$, $p < .001$, 95% CI = [1.24, 1.64], but this relationship was not moderated by academic event, $b = -0.23$, $SE = 0.20$, $p = .275$, 95% CI = [-0.61, 0.18]. Additionally, these concerns did not predict belonging in the course, $b = 0.00$, $SE = 0.04$, $p = .999$, 95% CI = [-0.09, 0.08].

Stereotype Threat-related Concerns

Fit statistics for this model were good, $\chi^2(14) = 12.28$, $p = .584$, CFI = 1.00, TLI = 1.01, RMSEA = .000. As observed in ANOVA, analyses revealed a significant Gender Composition x Peer Gender interaction on social comparison perceptions, $b = 0.30$, $SE = 0.12$, $p = .011$, 95% CI = [0.07, 0.55], such that having token status increased social comparison perceptions in response to the ingroup, but not outgroup, peer. Stronger social comparison perceptions predicted greater stereotype threat-related concerns, $b = 1.46$, $SE = 0.06$, $p < .001$, 95% CI = [1.34, 1.59], and this relationship was moderated by academic event, $b = -0.52$, $SE = 0.13$, $p < .001$, 95% CI = [-0.76, -0.27]. The pattern of means showed that social comparison perceptions increased stereotype threat-related concerns both when the peer performed well and poorly in the course, but this effect was stronger when the peer performed poorly. Stereotype threat-related concerns, in turn, predicted a decreased sense of belonging in the course, $b = -0.11$, $SE = .05$, $p = .044$, 95% CI =

$[-0.21, -0.00]$.⁶ Thus, this analysis indicated that women who have token status report stronger perceptions of being compared to the other woman (but not the man) in the course. These perceptions lead to lower belonging in the course because they elicit stereotype threat-related concerns, and these concerns are particularly strong when the peer performs poorly.

Discussion

Study 2 addressed two primary aims. First, Study 2 assessed whether the process observed in Study 1b was specific to an ingroup (female) peer, or whether this process also generalized to an outgroup (male) peer. Consistent with our hypotheses, we find that having token status increased social comparison perceptions in response to a female student (Jane), but did not increase social comparison perceptions in response to a male student (Eric). Replicating Study 1b, the gender composition of the course did not significantly impact women's own likelihood of making comparisons to the peer. Further replicating Study 1b and prior literature, results revealed that women who had token status reported a decreased sense of belonging in the course.

Replicating the indirect effect modeled in Study 1b, findings showed that having token status only increased social comparison perceptions in response to the ingroup peer, and increased social comparison perceptions subsequently predicted reductions in belonging. Given that this consequence was particularly likely to emerge in response to the ingroup (versus outgroup) peer, these findings offer some evidence that the observed process is specific to ingroup peers. Taken together, these findings are theoretically important and support our predictions regarding token women in male-dominated environments; having token status is particularly likely to increase women's perceived comparisons to ingroup, but not outgroup, members, and these comparisons have downstream consequences on belonging.

Study 2 also identified why social comparison perceptions reduce women's sense of belonging in an engineering course. Although concerns about measuring up to the peer's performance did not predict belonging, findings revealed that social comparison perceptions predicted a lower sense of belonging via increased stereotype threat-related concerns. Although these concerns emerged regardless of how the peer was performing, they were particularly strong when the peer was performing poorly. Although feeling compared to a low-performing peer may be particularly likely to activate token women's concerns about being negatively stereotyped, stereotype-threat related concerns may also be relevant for upward social comparison perceptions because token women may believe that they are at greater risk of confirming negative stereotypes (relative to the high-performing peer). Collectively, findings suggest that having token status may increase women's perceptions of being compared to an ingroup peer (Jane), and these perceptions may reduce their sense of belonging in the course by activating concerns about being evaluated on the basis of their gender identity, particularly when the peer performs poorly.

Study 3

To extend the previous studies, Study 3 had three primary goals. First, we sought to strengthen the causal link between our proposed mechanism, social comparison perceptions, and adverse outcomes. Although Studies 1b-2 offered initial evidence that social comparison perceptions served as a mechanism underlying the relationship between token status and decreased belonging in the engineering course, unmanipulated mediators can produce biases that overestimate mediation effects (Bullock et al., 2010). As such, we experimentally manipulated social comparison perceptions to gain additional support for its causal role in facilitating worse academic outcomes among women. We hypothesized that women would be more likely to

exhibit adverse outcomes when they perceived that they were being directly compared to an ingroup member (e.g., when their instructor and peers glanced in their direction following the other woman's public performance), relative to contexts where they did not perceive these comparisons (e.g., when class performance was anonymized). Consistent with the previous studies, we also expected that token women would exhibit worse outcomes (e.g., stronger social comparison perceptions and lower belonging) than non-token women. Furthermore, due to the heightened threat associated with token status, we theorized that token women might show greater sensitivity to the comparison manipulation than non-token women.

The previous studies examined the effect of social comparison perceptions specifically on belonging given that belonging is a particularly robust predictor of retention in one's field (Murphy et al., 2020; Rainey et al., 2018). However, there are several other psychological and behavioral outcomes that may be negatively impacted when women perceive that others are directly comparing them to an ingroup member. As such, the second aim of Study 3 was to identify additional consequences resulting from social comparison perceptions. In particular, Study 3 examined the effect of social comparison perceptions on three outcomes that have also been identified as important indicators of STEM engagement: (a) course participation, (b) motivation in the course, and (c) STEM interest (Dasgupta et al., 2015; Murphy et al., 2007).

Finally, Study 3 investigated the individuals for whom threat is most likely to emerge. According to stereotype threat theory (Steele et al., 2002), stereotype threat effects are strongest for individuals who are highly identified with the stereotyped domain. Therefore, we sought to examine whether the obtained results would be moderated by women's STEM identification (e.g., operationalized using their status as a STEM major)⁷. Because female STEM and non-STEM students may have different perceptions about, and experiences with, STEM settings, it is

possible that they will exhibit differential responses to the events depicted in the vignettes. In particular, STEM-identified women may be more susceptible to stereotype threat effects because (a) they are more likely to be threatened by pervasive stereotypes that women have low STEM ability, and (b) they may be more sensitive to cues about how they are being evaluated by others in a STEM context (Forbes et al., 2008). Therefore, we hypothesized that STEM- identified women would experience greater levels of threat (as indicated by their responses to the primary study outcomes).

Method

Participants

A priori power analyses in G*Power V. 3.1 indicated that we needed 245 participants for a 2 x 2 ANOVA when expecting a small-to-medium effect size ($f = 0.18$), with 80% power and $\alpha = 0.05$. To assess moderation by participants' major (STEM vs non-STEM), we set a goal of approximately 50 STEM majors per cell. Therefore, we over-recruited participants to ensure sufficient power, account for exclusions (e.g., participants who did not actually identify as female college students), and obtain an adequate number of STEM majors in our sample.

As in the previous studies, we directly recruited college women on Turkprime and obtained a sample of 585 participants. Before data analysis, we excluded 40 participants who reported that they did not meet our demographic criteria at the end of the study: identifying their gender as male ($n = 30$), identifying their gender as non-binary ($n = 7$), and reporting that they were not college students ($n = 3$). After removing these participants, we retained 545 college women (68.3% European American, 11.2% African American, 9.4% Asian American, 5.5% Latina, 5.7% other race/ethnicity or multiracial; $M_{age} = 23.02$; $SD_{age} = 3.24$; 34.3% perceived their major to be STEM-related; 77.8% were enrolled in a four-year college).

Procedure

In addition to manipulating the gender composition of the course, we included a new experimental factor (comparison) to manipulate our proposed mechanism (i.e., social comparison perceptions). Thus, participants were randomly assigned to one of four conditions in a 2 (gender composition: token status, non-token status) X 2 (comparison: high, low) between-subjects design. Because the previous studies revealed few differences for participants who saw an ingroup member perform well (versus poorly) in the course, all participants read about an ingroup member who performed poorly in Study 3.

To manipulate comparison, participants read versions of the vignettes utilized in the earlier studies with a few minor changes. In the “high comparison” condition, participants read that the ingroup member (Jane) struggled during her final project classroom presentation. Following Jane’s poor performance, the instructor and some of the other students in the course glanced directly at them (the participant). In the “low comparison” condition, participants read that Jane was messaging a friend about struggling with her final project; a few moments later, the instructor made an announcement to the class stating that in response to new university guidelines, students should include their ID numbers (instead of their name) on their final projects so that he could grade anonymously and other students in the class could not determine how their peers were performing. Following the vignette, participants responded to the survey measures.

Measures

Following the experimental manipulations, participants responded to the survey items used in the previous studies. Because the gender composition of the course did not affect participants’ own comparisons to the ingroup peer in Studies 1b-2, we dropped the self-

comparison outcome in Study 3. As such, participants reported their (a) social comparison perceptions ($\alpha = .87$), and (b) sense of belonging in the engineering course ($\alpha = .91$). To examine additional outcomes that may be negatively impacted by the study manipulations and social comparison perceptions, participants responded to three new measures assessing STEM engagement: anticipated participation in the engineering course, motivation in the engineering course, and STEM interest.

Course Participation. We measured participants' anticipated participation in the engineering course using four items on a Likert-type scale ranging from 1 (*not at all likely*) to 7 (*very likely*); e.g., "How likely would you be to answer questions in this engineering class". These items were adapted from prior research and showed high reliability ($\alpha = .85$; Stout et al., 2011). As such, they were aggregated into an index.

Motivation in the Course. We assessed participants' motivation in the engineering course using three items on a Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*); e.g., "I would probably feel motivated to perform well in this engineering class". These items were also adapted from previous research (Canning et al., 2019). Although the three items showed inadequate reliability ($\alpha = .61$), two items measuring participants' motivation to perform well were highly correlated ($r = .80$). Because the reliability coefficient for these two items was larger than the reliability coefficient across all three items, we combined the two highly correlated items and assessed the third item ("I would probably feel little motivation to work hard in this engineering class" [reverse scored]) separately. Thus, the two-item composite reflected participants' "motivation to perform well" in the engineering course and the single-item measure reflected participants' "motivation to work hard" in the engineering course.

STEM Interest. We measured participants' interest in STEM using two items on a Likert-type scale ranging from 1 (*not at all interested*) to 7 (*very interested*); e.g., "How interested would you be in pursuing graduate studies in a STEM or STEM-related field?". These items were adapted from previous research and were highly correlated ($r = .90$; Stout et al., 2011). As such, they were collapsed into an index.

Analytic Strategy

As in the previous studies, we used ANOVA to examine the main effects of gender composition, comparison, and STEM status, as well as the two- and three-way interactions, on the primary study outcomes. Despite our hypothesis that STEM majors would be more likely to report social comparison perceptions and exhibit adverse outcomes (e.g., decreases in belonging and STEM engagement), analyses generally showed that STEM status did not moderate the reported findings. Thus, ANOVA analyses including STEM status as a moderator are presented in the online supplement.

Statistical means and standard deviations for significant main effects are reported below, and means (by condition) are presented in the online supplement.

Results

Social Comparison Perceptions

Analyses revealed a significant main effect of gender composition on social comparison perceptions, $F(1, 541) = 43.59, p < .001, d = .57$. As in the previous studies, the pattern of standardized means showed that participants in the token condition ($M = 0.20, SD = 0.76$) were more likely to perceive social comparison perceptions (i.e., concerns that others were comparing them to the ingroup member) than participants in the non-token condition ($M = -0.19, SD = 0.68$). Additionally, a main effect of comparison, $F(1, 541) = 30.82, p < .001, d = .48$, showed

that our manipulation was effective; participants in the high comparison condition ($M = 0.16$, $SD = 0.74$) were more likely to perceive social comparison perceptions than participants in the low comparison condition ($M = -0.16$, $SD = 0.71$). The Gender Composition x Comparison interaction was not significant, $F(1, 541) = 0.30$, $p = .582$, $\eta_p^2 = .001$.

Sense of Belonging in the Course

Replicating the previous studies, analyses revealed a significant main effect of gender composition, $F(1, 541) = 57.43$, $p < .001$, $d = -.65$, such that participants in the token condition ($M = 4.47$, $SD = 1.21$) reported a lower sense of belonging in the course than participants in the non-token condition ($M = 5.17$, $SD = 0.95$). Moreover, a significant main effect of comparison, $F(1, 541) = 4.81$, $p = .029$, $d = -.19$, showed that participants in the high comparison condition ($M = 4.73$, $SD = 1.16$) reported a lower sense of belonging in the course than participants in the low comparison condition ($M = 4.92$, $SD = 1.11$). The Gender Composition x Comparison interaction was not significant, $F(1, 541) = 0.08$, $p = .781$, $\eta_p^2 = .000$.

Course Participation

Analyses showed a significant main effect of gender composition, $F(1, 541) = 5.14$, $p = .024$, $d = -.19$. Study means indicated that participants in the token condition reported lower anticipated course participation ($M = 4.78$, $SD = 1.48$) than participants in the non-token condition ($M = 5.05$, $SD = 1.32$). However, neither the main effect of comparison, $F(1, 541) = 1.05$, $p = .307$, $d = -.09$, nor the Gender Composition x Comparison interaction, $F(1, 541) = 0.42$, $p = .517$, $\eta_p^2 = .001$, were significant.

Motivation in the Course

Analyses revealed a significant main effect of gender composition on participants'

motivation to perform well in the course, $F(1, 541) = 8.26, p = .004, d = -.25$. Study means showed that participants in the token condition ($M = 5.81, SD = 1.28$) reported lower motivation to perform well in the course than participants in the non-token condition ($M = 6.09, SD = 1.04$). However, the main effect of gender composition on participants' motivation to work hard in the course was not significant, $F(1, 541) = 0.70, p = .405, d = -.07$. Findings also revealed a significant main effect of comparison on participants' motivation to perform well, $F(1, 541) = 8.29, p = .004, d = -.25$, and work hard in the course, $F(1, 541) = 7.23, p = .007, d = -.23$. Means indicated that participants in the high comparison condition ($M_{perform\ well} = 5.82, SD_{perform\ well} = 1.26; M_{work\ hard} = 4.38, SD_{work\ hard} = 2.23$) reported lower motivation to perform well and work hard in the course than participants in the low comparison condition ($M_{perform\ well} = 6.09, SD_{perform\ well} = 1.07; M_{work\ hard} = 4.88, SD_{work\ hard} = 2.16$). The Gender Composition x Comparison interaction for participants' motivation to perform well, $F(1, 541) = 2.32, p = .129, \eta_p^2 = .004$, and work hard in the course, $F(1, 541) = 1.49, p = .223, \eta_p^2 = .003$, was not significant.

STEM Interest

A significant main effect of gender composition, $F(1, 541) = 4.84, p = .028, d = -.19$, showed that participants in the token condition ($M = 4.99, SD = 1.82$) reported lower STEM interest than participants in the non-token condition ($M = 5.31, SD = 1.71$). Additionally, a significant main effect of comparison, $F(1, 541) = 4.31, p = .038, d = -.18$, revealed that participants in the high comparison condition ($M = 5.00, SD = 1.84$) reported lower STEM interest than participants in the low comparison condition ($M = 5.30, SD = 1.70$). The Gender Composition x Comparison interaction was not significant, $F(1, 541) = 0.58, p = .447, \eta_p^2 = .001$.

Testing the Proposed Mechanism: Social Comparison Perceptions

Given Study 3's primary interest in assessing the causal link between social comparison perceptions and the primary study outcomes (e.g., belonging and STEM engagement), we sought to replicate the general model tested in Study 1b using the comparison manipulation in place of gender composition. Thus, we examined whether (a) perceiving strong direct comparisons to the ingroup peer predicted greater social comparison perceptions, and (b) stronger social comparison perceptions predicted reductions in the study outcomes (e.g., participants' sense of belonging in the course, course participation, motivation in the course, and STEM interest; see Figures 7a-e). Results for the models using gender composition in place of comparison are reported in the online supplement for interested readers.

As in Study 1b, fit statistics for the following models could not be computed because the models contained zero degrees of freedom. Examining the AIC for these saturated models revealed that they had (a) the best fit relative to alternative models, or (b) a slightly worse fit ($\Delta AIC < 2$) than an alternative model eliminating the direct effect.

Sense of Belonging in the Course. Analyses revealed a significant effect of comparison on social comparison perceptions showing that participants in the high comparison condition were more likely to report social comparison perceptions than participants in the low comparison condition, $b = 0.16$, $SE = 0.03$, $p < .001$, 95% CI = [0.10, 0.22]. Stronger social comparison perceptions, in turn, predicted reductions in participants' sense of belonging in the course, $b = -0.71$, $SE = 0.06$, $p < .001$, 95% CI = [-0.83, -0.59]. After accounting for this indirect effect, the direct effect of comparison on belonging was no longer significant, $b = 0.02$, $SE = 0.05$, $p = .658$, 95% CI = [-0.07, 0.11].

Course Participation. Analyses revealed a significant effect of comparison on social comparison perceptions showing that participants in the high comparison condition were more

likely to report social comparison perceptions than participants in the low comparison condition, $b = 0.16$, $SE = 0.03$, $p < .001$, 95% CI = [0.10, 0.22]. Stronger social comparison perceptions predicted reductions in participants' anticipated course participation, $b = -0.52$, $SE = 0.09$, $p < .001$, 95% CI = [-0.70, -0.34]. As observed in the ANOVA, the direct effect of comparison on course participation was not significant, $b = 0.03$, $SE = 0.06$, $p = .675$, 95% CI = [-0.09, 0.14].

Motivation to Perform Well in the Course. Analyses revealed a significant effect of comparison on social comparison perceptions showing that participants in the high comparison condition were more likely to report social comparison perceptions than participants in the low comparison condition, $b = 0.16$, $SE = 0.03$, $p < .001$, 95% CI = [0.10, 0.22]. Stronger social comparison perceptions predicted reductions in participants' motivation to perform well in the course, $b = -0.28$, $SE = 0.07$, $p < .001$, 95% CI = [-0.42, -0.13]. After accounting for the indirect effect, the direct effect of comparison on motivation to perform well in the course was no longer significant, $b = -0.09$, $SE = 0.05$, $p = .058$, 95% CI = [-0.19, 0.00].

Motivation to Work Hard in the Course. Analyses revealed a significant effect of comparison on social comparison perceptions showing that participants in the high comparison condition were more likely to report social comparison perceptions than participants in the low comparison condition, $b = 0.16$, $SE = 0.03$, $p < .001$, 95% CI = [0.10, 0.22]. However, social comparison perceptions did not predict participants' motivation to work hard in the course, $b = -0.17$, $SE = 0.14$, $p = .220$, 95% CI = [-0.45, 0.10]. As reported in ANOVA, the direct effect of comparison on participants' motivation to work hard in the course remained significant, $b = -0.22$, $SE = 0.10$, $p = .021$, 95% CI = [-0.42, -0.04].

STEM Interest. Analyses revealed a significant effect of comparison on social comparison perceptions showing that participants in the high comparison condition were more

likely to report social comparison perceptions than participants in the low comparison condition, $b = 0.16$, $SE = 0.03$, $p < .001$, 95% CI = [0.10, 0.22]. However, social comparison perceptions did not predict STEM interest, $b = -0.15$, $SE = 0.11$, $p = .198$, 95% CI = [-0.36, 0.07]. After accounting for the indirect effect, the direct effect of comparison on STEM interest was no longer significant, $b = -0.13$, $SE = 0.08$, $p = .103$, 95% CI = [-0.29, 0.03].

Discussion

Study 3 replicated and extended the previous studies in three ways. First, Study 3 experimentally manipulated social comparison perceptions to further establish its role as a causal mechanism that undermines women's STEM outcomes. Moreover, although the previous studies focused exclusively on belonging, Study 3 investigated whether social comparison perceptions impacted additional academic outcomes that serve as important indicators of STEM engagement (e.g., course participation, motivation, and STEM interest). Finally, Study 3 examined for whom threat was most likely to emerge (by assessing whether STEM-identified women were more susceptible to threat).

Findings demonstrated that manipulating the gender composition of the course (gender composition) and perceptions about being compared to an ingroup peer (comparison) independently impacted women's outcomes. Replicating the previous studies, women who had token (versus non-token) status in the engineering course reported stronger social comparison perceptions and a decreased sense of belonging in the course. Moreover, token (versus non-token) women reported reductions in course participation, motivation to perform well in the course (one of the two motivation measures used), and STEM interest.

Examining the effects of comparison provided supporting evidence for the causal role of social comparison perceptions on negative outcomes; women who learned that their course

performance was being directly compared to that of an ingroup member (relative to women who learned that their performance was anonymous) reported stronger social comparison perceptions, indicating that this experimental manipulation was effective. Moreover, women in the high (versus low) comparison condition exhibited decreased belonging in the course, lower motivation to perform well and work hard in the course, and reduced STEM interest. However, experimentally manipulating perceived comparisons to the ingroup member did not impact course participation. Despite our predictions that women with token status might be more sensitive to the comparison manipulation, the Gender Composition x Comparison interaction was not significant for any of the study outcomes.

Additionally, modeling the indirect effect revealed that the comparison manipulation increased social comparison perceptions, which subsequently predicted reductions in belonging and motivation to perform well in the course. Although social comparison perceptions also predicted decreased course participation, this outcome was unaffected by the comparison manipulation. However, social comparison perceptions did not predict motivation to work hard in the course or STEM interest. Taken together, study findings provide additional evidence for our hypothesis that social comparison perceptions can perpetuate negative effects on belonging and STEM engagement.

Finally, given the importance of domain identification for stereotype threat outcomes, Study 3 examined whether experienced threat was stronger for women who were highly identified with STEM. Contrary to our hypotheses, analyses offered little evidence that women's STEM identification (operationalized by their status as a STEM major) moderated the results. Therefore, findings suggest that the threat of perceived ingroup comparisons generalizes across women who are STEM and non-STEM identified.

Collectively, Study 3 offers greater support for social comparison perceptions as a causal mechanism underlying negative STEM outcomes for women. In particular, Study 3 replicates and extends previous study findings by documenting the negative impact of (a) token status, and (b) perceived comparisons to an ingroup member, on social comparison perceptions, course belonging, and STEM engagement. Moreover, modeling the indirect effect showed that the consequences resulting from social comparison perceptions extend beyond belonging. Specifically, social comparison perceptions also predicted reductions in anticipated course participation and motivation to perform well in the course.

General Discussion

Many studies have identified features of male-dominated settings that can threaten women's sense of belonging and engagement in STEM, including token status (Inzlicht & Good, 2006; Murphy et al., 2007). In addition to identifying these factors, it is also imperative to understand the mechanisms through which deleterious outcomes emerge. To extend our knowledge of these processes, four experiments tested (a) whether the gender composition of a course (e.g., having token status, operationalized as being one of two women in the course) decreases women's sense of belonging and engagement in that course, and (b) the role of social comparison perceptions within these relationships. Investigating this process offers a more comprehensive understanding of psychological mechanisms that may predict women's likelihood of withdrawing from STEM fields.

Study 1a offered initial evidence that having token status (versus being in a gender-balanced course) increased women's perceptions of being compared to an ingroup peer (another woman) in the course. Study 1b replicated and extended these findings by identifying these perceived comparisons, labeled social comparison perceptions, as a mechanism underlying the

relationship between token status and decreased belonging in the course. Importantly, analyses also showed that token status increased social comparison perceptions (women's perceptions that external observers — the instructor and peers — were comparing them to the other woman in the course), but did not increase women's own likelihood of comparing themselves to her. As such, findings suggest that negative effects on belonging were being driven by perceived comparisons from external observers (rather than participants' own comparisons).

Extending the prior studies, Study 2 revealed that although token status increased social comparison perceptions in response to the ingroup peer, it did not increase social comparison perceptions in response to an outgroup peer (e.g., a man in the course). Further replicating the previous studies, modeling the indirect effect showed that token status diminished belonging through increased social comparison perceptions that emerged in response to an ingroup peer. Moreover, Study 2 tested two routes through which social comparison perceptions may decrease belonging, depending on the nature of the peer's performance: concerns about measuring up to the peer's performance and stereotype threat-related concerns. We predicted that when the peer performed well, social comparison perceptions would increase participants' concerns about measuring up to his/her success; however, this prediction was not supported. We did find evidence that social comparison perceptions activated stereotype threat-related concerns regardless of how the peer performed, but these concerns were particularly strong when the peer performed poorly. Stereotype threat-related concerns, in turn, predicted reductions in belonging. Taken together, Study 2 findings demonstrated that the processes observed across these studies are more likely to emerge in response to an ingroup peer (another woman in the course), relative to an outgroup peer (a man in the course).

Study 3 replicated previous study findings regarding the consequences of token status

and offered additional evidence for social comparison perceptions as the causal mechanism underlying negative outcomes; directly manipulating whether the instructor and peers were comparing the participant to the ingroup member revealed that perceiving these comparisons increased social comparison perceptions, decreased belonging, and undermined additional outcomes that serve as indicators of STEM engagement (e.g., motivation to perform well in the course). Although we predicted that the level of threat resulting from these perceived comparisons might be stronger for token (versus non-token) women due to increased sensitivity to the comparison manipulation, this hypothesis was not supported. Instead, the gender composition and comparison manipulations each had independent effects on outcomes, indicating that these two factors produced additive, but not multiplicative, effects. It is possible that participants in the token (versus non-token) condition did not display differential sensitivity to the comparison manipulation because all participants were placed in a STEM context. As such, our comparison manipulation may have activated gender-based stereotypes (and consequently, threat) in similar ways across conditions. Another reason underlying additive effects is that the comparison manipulation was blatant. As such, participants in the token condition may have shown greater sensitivity to the comparison manipulation if the comparison cues were more ambiguous (consistent with research demonstrating that members of minoritized groups show greater sensitivity to ambiguous stigma cues, such as gender microaggressions, relative to majority groups; Basford et al., 2014).

Furthermore, findings provided little evidence for our prediction that women who were highly identified with STEM (e.g., were STEM majors) would exhibit more negative outcomes in response to threat. Although these findings are inconsistent with stereotype threat theory (which would predict stronger threat effects among women with high domain identification),

extant literature regarding the effects of domain identification is complex. For instance, prior research suggests that individuals with moderate domain identification may be more strongly affected by stereotype threat than individuals with strong domain identification (Nguyen & Ryan, 2008). Other research indicates that the effects of domain identification may depend on an individual's social identities (e.g., whether they are a Black man or woman; Nadler & Komarraju, 2016). Therefore, it is possible that the events depicted in the hypothetical scenario reflect a situation where the role of domain identification is more complex. Alternatively, there may have been issues with our operationalization of domain identification; for example, examining STEM identification broadly may be a weak predictor of threat given the vignettes' specific focus on engineering.

Theoretical Contributions

The present work has important implications for understanding the processes through which having minority status undermines women's belonging and engagement in STEM settings. Research-to-date shows that being numerically underrepresented in a setting elicits harmful consequences for minority groups through several mechanisms, such as activation of identity-relevant stereotypes, greater feelings of anxiety or apprehension about being the target of prejudice, decreased performance expectations, and worse fit in a setting (Good et al., 2012; Inzlicht & Ben-Zeev, 2000; LaCosse et al., 2016; Master et al., 2016; Sekaquaptewa & Thompson, 2003; Thompson & Sekaquaptewa, 2002; Shapiro & Williams, 2012). The current studies contribute to extant literature by identifying a novel mechanism, social comparison perceptions, that may also be triggered by having minority status in a setting. In fact, this set of studies is the first to show that having token status increases perceptions of being directly compared to an ingroup member. This finding is particularly important given its implications for

understanding the consequences that can result when individuals find themselves in token status situations. In addition to being an independent predictor of belonging and academic engagement, findings show that social comparison perceptions also prompt identity threat processes (e.g., stereotype threat concerns) that are known to impede academic outcomes. Therefore, to gain a more complete understanding of the mechanisms that can reduce academic outcomes for minority groups, future work should continue investigating additional routes through which environmental cues may elicit threat.

This work also contributes to research on intragroup relations. First, these studies extend prior work focusing specifically on duo status; although previous research theorizes about the dynamics of the interactions between two ingroup members (e.g., their willingness to give and receive social support), this work demonstrates how having duo status in a setting can influence perceptions of the environment (e.g., beliefs that external observers are making direct comparisons to the other ingroup member) and produce downstream consequences for the self (e.g., beliefs about whether one belongs in a setting; Loyd et al., 2008).

Moreover, a large body of literature suggests that the presence of other ingroup members can alleviate threat, particularly in contexts where the group is underrepresented (Sekaquaptewa & Thompson, 2002, 2003; Shaffer et al., 2013; Stout et al., 2011; Tajfel & Turner, 1979). For instance, extensive research finds that the presence of ingroup role models can be especially effective for retention in a setting (Drury et al., 2011; Herrmann et al., 2016). In particular, the stereotype inoculation model posits that high-performing ingroup members can improve recruitment and retention by increasing social belonging and “inoculating” group members from threat (Dasgupta, 2011). However, the current findings demonstrate that responses to ingroup members, particularly when the group is underrepresented, may be more complex than

previously considered.

Although women may show more active participation when they are in female-majority groups (relative to male-majority or gender parity groups), these findings are often tested in the context of small work groups (Dasgupta et al., 2015). Even though these small groups may contain very few women, women may not perceive ingroup comparisons (or experience identity threat) because they are adequately represented in the gender ratio of the group (e.g., two women on a five-person team). This suggests the importance of the proportion of women to men in the group, as opposed to the actual number of women. In our studies, we operationalized token status as two women in a class size of 29, which represents both a duo status and token status situation. This choice not only reflects the likely gender composition of many actual classrooms in extremely male-dominated STEM fields, but also maximized the possibility that a female participant would feel compared to the only other woman in the scenario, allowing us a clear test of social comparison perceptions. But given our focus on proportions, we suggest that the social comparison perception processes revealed in this research would also emerge if there were more than just two women in the scenario, as long as the women still felt that they were in the minority and that their gender identity stood out to others in the situation. Therefore, we theorize that these conditions are most likely to promote social comparison perceptions.

The present work examined negative outcomes that emerge when the ratio of women to men is very low (e.g., two women in a 29-person course) because we expected that these comparison processes would be especially likely to emerge in settings where women are more severely underrepresented. As a result, this work contributes to extant literature on intragroup relations by demonstrating that the presence of ingroup members can produce negative effects when contexts activate women's concerns that external observers are comparing them to other

ingroup members. Therefore, although much of the current literature on women in STEM focuses on negative consequences that result from intergroup interactions, future work should develop a deeper understanding of (a) the conditions under which the presence of an ingroup member is consequential, and (b) additional consequences that may arise when an ingroup member represents the group unfavorably.

This work also informs our understanding of conditions that may evoke identity threat. Although extant research identifies several contexts that activate identity threat outcomes (e.g., performance apprehension in a stereotype-relevant domain), these studies are consistent with literature suggesting that identity threat outcomes can arise even when individuals are not directly performing (Cohen & Garcia, 2005; Lewis & Sekaquaptewa, 2016; Sekaquaptewa et al., 2007; Steele & Aronson, 1995; Steele et al., 2002). Instead, individuals with minority status may exhibit identity threat outcomes, particularly when an ingroup member is performing poorly, due to beliefs that the ingroup member's behavior will reflect directly on their own performance (Cohen & Garcia, 2005).

These studies also have implications for research on social comparison processes. Extant research in this area focuses primarily on the motives that shape an individual's likelihood of making social comparisons, as well as the psychological and behavioral consequences that result from making these comparisons (Suls & Wheeler, 2013). However, this literature often fails to account for another dimension of social comparisons: perceiving comparisons made by external observers (e.g., social comparison perceptions). Similar to individuals' own social comparisons, perceptions that one is being compared to others can also be influenced by personal motives and contexts. For example, these studies identify a specific context change, threat raised by having token status, that may be particularly likely to activate concerns about social comparison

perceptions. Moreover, this work highlights the importance of social contexts for understanding the consequences that result from different types of comparisons. For example, although past research shows that downward social comparisons are often beneficial (Gibbons & Gerrard, 1991), the current studies offer a novel contribution by showing that downward social comparison perceptions can be detrimental, specifically in stereotype threat situations.

Therefore, the current research demonstrates that social comparison perceptions are both theoretically and practically important because these perceptions may be just as consequential as self-generated comparisons, particularly in contexts marked by threat. Future work should test the various ways in which perceived comparisons (by external observers) may diverge from self-generated comparisons (e.g., the resulting consequences and/or conditions under which people make these comparisons).

Additionally, although we did not find that upward social comparison perceptions to an ingroup member were more harmful than downward social comparison perceptions, further investigation of these effects revealed that concerns about measuring up to the peer's performance predicted significant reductions in belonging when we did not account for the influence of other variables (e.g., gender composition and social comparison perceptions) on belonging. As such, these findings offer some evidence that identity threat processes (e.g., feeling compared to an ingroup member) may be stronger predictors of decreased belonging than feelings of personal threat or competition (e.g., not being able to measure up to a peer's performance). Moreover, although we expected that social comparison perceptions would be particularly likely to increase concerns about measuring up to the peer's performance when they performed well, findings showed that this relationship emerged regardless of how the peer performed. Follow-up analyses revealed that this lack of moderation was driven by women's

response to the male peer. This pattern of findings is consistent with prior research showing that women may be particularly concerned about measuring up to a male peer's performance, regardless of whether he does well or poorly, because men are positively stereotyped as having high engineering ability. As a result of these stereotypes, men are often viewed as a marker of success in STEM settings. Consequently, when women feel compared to a male peer, their concerns about measuring up to his performance may be less affected by the nature of his performance (relative to a female peer's performance).

Practice Implications

This work also has important implications for belonging interventions. Within prior research, belonging interventions often attenuate threat by affirming people's values, introducing ingroup exemplars, encouraging challenge and growth mindsets, and mitigating stereotype threat concerns (Dweck et al., 2014; Shnabel et al., 2013; Stout et al., 2011; Walton & Cohen, 2011). However, the current findings offer another point of intervention that can improve belonging: mitigating women's concerns that they will be directly compared to other women in male-dominated settings. Women's concerns about these comparisons are valid; prior research suggests that people are, in fact, more likely to make comparisons between group members who have distinctive identities in a setting (Lord & Saenz, 1985; Taylor et al., 1978). Therefore, efforts to retain women in academic STEM settings should instruct educators in these settings to (a) refrain from making comparisons between female students and (b) avoid behavior that may signal these comparisons (e.g., scheduling the token female students' class presentations back-to-back). Additionally, institutional efforts should emphasize the importance of focusing on individual (rather than group) characteristics in classroom settings.

Future Directions

It is important to understand the various ways in which environmental factors may deter women from persisting in male-dominated environments. Although the current studies show that women's perceptions of being compared to another woman impedes their sense of belonging and academic engagement, particularly in male-dominated settings, one limitation of this work is that participants were presented with a hypothetical vignette. Although this methodology allowed us to test our research questions in a more controlled environment that minimized the effect of extraneous variables, future research should (a) test this scenario within laboratory and classroom settings, and (b) identify boundary conditions by examining other features of the environment that may moderate the observed effects. For instance, future work should manipulate aspects of the environment, such as the level of hostility and/or sexism directed towards women in the course, to determine their role on the observed pattern of findings, particularly given work showing that sexist environments promote queen bee behaviors (Derks et al., 2011). Moreover, although these studies focused specifically on participants who had token status, it is important to identify whether these processes change depending on the gender ratio of the course (e.g., whether women make up 15 or 30% of the course). Although past work would suggest that these comparison processes are likely to emerge any time one has minority status, there is some evidence that making up at least 35-40% of a setting is sufficient to minimize identity threat outcomes (Kanter, 1977; Tolbert et al., 1995). As such, future research should seek out the "tipping point" to better understand when the observed processes are most likely to arise. Furthermore, one remaining question from this work is whether beliefs about being compared to an ingroup member diminishes belonging, causing women to leave these settings, and/or if concerns about these comparisons reduce women's motivation to enter male-dominated fields in the first place. Future work should investigate these possibilities.

Another limitation of this work is that we focused on social comparison perceptions among women in male-dominated settings. Although similar variables have been examined among with racial/ethnic minorities in prior research (Cohen & Garcia, 2005), future work should examine whether the processes observed in this set of studies generalize to other social groups, such as racial/ethnic minorities. Previous work (Sekaquaptewa et al., 2007) shows that Black students who completed an academic task with solo status (i.e., in a group of Whites) perceived that their performance would be generalized to other members of their race, suggesting that they may experience social comparison perceptions if there were another Black student in that setting. Moreover, future research should investigate these processes in the context of intersectionality. For example, future work should test whether Black women with token status are more likely to perceive that they will be compared to a Black man (due to their shared racial identity) versus a White woman (due to their shared gender identity). The nature of such social comparison perceptions may depend on which aspects of identity are salient in the specific situation (Shih et al., 1999) or on long-standing differences in social privilege and disadvantage that favor Whites (Knowles et al., 2014; Unzueta & Lowery, 2008). Finally, one limitation of this work is that we did not observe responses for individuals who self-identify as gender non-binary. Because we focused specifically on recruiting a sample who self-identified as female, future research should examine how individuals across the gender spectrum respond to the hypothetical scenario. Considering diversity in gender expression is particularly important given the consequences that result from labels ascribed by external observers. For example, someone who has a traditionally “feminine” gender presentation may be more likely to perceive that they are being compared to another woman in the course than someone who has a “masculine” or more ambiguous gender presentation.

Finally, future research should consider additional consequences that result from making social comparison perceptions in response to an ingroup member. For instance, one unexamined consequence of these perceived direct comparisons, particularly when the ingroup member behaves in a negative or stereotype-consistent manner, is that people may exclude this individual from the ingroup to reduce threat (Eidelman et al., 2006). This response can have implications for the subsequent treatment of the ingroup member (e.g., reduced likelihood of conferring support and other benefits that would typically result from ingroup favoritism). Future research should explore the full range of consequences arising from social comparison perceptions to determine how they impact outcomes that extend beyond academic engagement and belonging.

Conclusion

Collectively, these experiments offer a greater understanding of the mechanisms that may undermine belonging and contribute to the leaky pipeline for women in STEM. Although extant literature focuses on the benefits of adding another woman into a setting for solo women, these studies suggest that the psychological processes activated by this event are complex and require further study. By identifying a mechanism, social comparison perceptions, that can have adverse effects on belonging and academic engagement for token women, the present work can inform the future development of belonging interventions to increase gender parity in STEM.

Endnotes

- 1 In the current and subsequent study, all reported findings hold when retaining female participants who were not currently college students.
- 2 Although we intended for all of the items to be reported on a 7-point scale, two of the items were measured on a 9-point scale due to a manual error. We caught this error after data collection, but retained these scales in the subsequent studies for consistency.
- 3 We also examined results using the 2-item composite, which was based on the two self-comparison items. The pattern of findings replicated the findings observed using the larger composite.
- 4 In the current and subsequent studies, the models were also tested using Hayes PROCESS macro v.3.5. Analyses are reported in the online supplement for interested readers.
- 5 As in Study 1b, the pattern of findings observed for the 7-item social comparison perception measure replicated with the 2-item version.
- 6 We also conducted multigroup analysis to test whether the model held for participants who read about the ingroup (versus outgroup) peer. Findings showed that all of the theoretically-relevant paths were significant for participants who read about the ingroup peer (with exception of a marginal effect of stereotype threat-related concerns on belonging). Analyses for the outgroup peer showed that token status did not increase perceptions of being compared to the outgroup peer, and social comparison perceptions made in response to the outgroup peer did not predict decreased belonging. Statistical parameters for these analyses are reported in the online supplement.
- 7 Although the samples in Study 1a-2 contained fewer STEM majors, we also assessed moderation by STEM status for these studies. Results were generally nonsignificant, particularly for the key outcomes of interest. We report test statistics for these analyses in the online supplement.

References

- Asgari, S., Dasgupta, N., & Stout, J. G. (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin*, 38(3), 370-383. <https://doi.org/10.1177/0146167211431968>
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1), 29-46. <https://doi.org/10.1006/jesp.1998.1371>
- Avery, D. R., Hernandez, M., & Hebl, M. R. (2004). Who's watching the race? Racial salience in recruitment advertising. *Journal of Applied Social Psychology*, 34(1), 146-161. <https://doi.org/10.1111/j.1559-1816.2004.tb02541.x>
- Basford, T. E., Offermann, L. R., & Behrend, T. S. (2014). Do you see what I see? Perceptions of gender microaggressions in the workplace. *Psychology of Women Quarterly*, 38(3), 340-349. <https://doi.org/10.1177/0361684313511420>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497-529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, 41(2), 174-181. <https://doi.org/10.1016/j.jesp.2003.11.007>
- Bernhardt, P. C., Dabbs, J. M., Jr., Fielden, J. A., & Lutter, C. D. (1998). Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. *Physiology & Behavior*, 65(1), 59-62. [https://doi.org/10.1016/S0031-9384\(98\)00147-4](https://doi.org/10.1016/S0031-9384(98)00147-4)
- Betz, D. E., & Sekaquaptewa, D. (2012). My fair physicist? Feminine math and science role

- models demotivate young girls. *Social Psychological and Personality Science*, 3(6), 738-746. <https://doi.org/10.1177/1948550612440735>
- Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, 66(2), 268–275. <https://doi.org/10.1037/0022-3514.66.2.268>
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, 98(4), 550-558. <https://doi.org/10.1037/a0018933>
- Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, 5(2), eaau4734. <https://doi.org/10.1126/sciadv.aau4734>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology*, 89(4), 566–582. <https://doi.org/10.1037/0022-3514.89.4.566>
- Cohen, L. L., & Swim, J. K. (1995). The differential impact of gender ratios on women and men: Tokenism, self-confidence, and expectations. *Personality and Social Psychology Bulletin*, 21(9), 876–884. <https://doi.org/10.1177/0146167295219001>
- Craig, K. M., & Feasel, K. E. (1998). Do solo arrangements lead to attributions of tokenism? Perceptions of selection criteria and task assignments to race and gender solos. *Journal of Applied Social Psychology*, 28(19), 1810-1836. <https://doi.org/10.1111/j.1559-1816.1998.tb01347.x>
- Crosby, J. R., King, M., & Savitsky, K. (2014). The minority spotlight effect. *Social*

- Psychological and Personality Science*, 5(7), 743-750.
<https://doi.org/10.1177/1948550614527625>
- Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry*, 22(4), 231–246.
<https://doi.org/10.1080/1047840X.2011.607313>
- Dasgupta, N., Scirclo, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 112(16), 4988–4993. <https://doi.org/10.1073/pnas.1422822112>
- Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 21-29.
<https://doi.org/10.1177/2372732214549471>
- Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceeding of the National Academy of Sciences*, 114(23), 5964–5969. <https://doi.org/10.1073/pnas.1613117114>
- Derks, B., Ellemers, N., Van Laar, C., & De Groot, K. (2011). Do sexist organizational cultures create the Queen Bee? *British Journal of Social Psychology*, 50(3), 519-535.
<https://doi.org/10.1348/014466610X525280>
- Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly*, 27(3), 456–469. <https://doi.org/10.1016/j.leaqua.2015.12.007>
- Drury, B. J., Siy, J. O., & Cheryan, S. (2011). When do female role models benefit women? The

- importance of differentiating recruitment from retention in STEM. *Psychological Inquiry*, 22(4), 265–269. <https://doi.org/10.1080/1047840X.2011.620935>
- Dweck, C. S., Walton, G. M., & Cohen, G. L. (2014). *Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning*. <http://k12education.gatesfoundation.org/resource/academic-tenacity-mindsets-and-skills-that-promote-long-term-learning/>
- Eidelman, S., Silvia, P. J., & Biernat, M. (2006). Responding to deviance: Target exclusion and differential devaluation. *Personality and Social Psychology Bulletin*, 32(9), 1153–1164. <https://doi.org/10.1177/0146167206288720>
- Ellis, J., Fosdick B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLOS ONE*, 11(7), e0157447. <https://doi.org/10.1371/journal.pone.0157447>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Forbes, C. E., Schmader, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience*, 3(3), 253–261. <https://doi.org/10.1093/scan/nsn012>
- Fryberg, S. A., Markus, H. R., Oyserman, D., & Stone, J. M. (2008). Of warrior chiefs and Indian princesses: The psychological consequences of American Indian mascots. *Basic and Applied Social Psychology*, 30(3), 208–218. <https://doi.org/10.1080/01973530802375003>

- Gibbons, F. X., & Gerrard, M. (1991). Downward comparison and coping with threat. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 317–345). Lawrence Erlbaum Associates Publishers.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–717. <https://doi.org/10.1037/a0026659>
- Hagedorn, L. S., Chi, W. Y., Cepeda, R. M., & McLain, M. (2007). An investigation of critical mass: The role of Latino representation in the success of urban community college students. *Research in Higher Education*, *48*(1), 73–91. <https://doi.org/10.1007/s11162-006-9024-5>
- Heilman, M. E., & Herlihy, J. M. (1984). Affirmative action, negative reaction? Some moderating conditions. *Organizational Behavior and Human Performance*, *33*(2), 204–213. [https://doi.org/10.1016/0030-5073\(84\)90021-7](https://doi.org/10.1016/0030-5073(84)90021-7)
- Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology*, *105*(1), 89–107. <https://doi.org/10.1037/a0029691>
- Herrmann, S. D., Adelman, R. M., Bodford, J. E., Graudejus, O., Okun, M. A., & Kwan, V. S. Y. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology*, *38*(5), 258–268. <https://doi.org/10.1080/01973533.2016.1209757>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1),

53-60. <https://doi.org/10.21427/D7CF7R>

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males.

Psychological Science, 11(5), 365–371. <https://doi.org/10.1111/1467-9280.00272>

Inzlicht, M., & Good, C. (2006). How environments can threaten academic performance, self-knowledge, and sense of belonging. In S. Levin & C. van Laar (Eds.) *Stigma and group inequality: Social psychological perspectives* (pp. 129–150). Lawrence Erlbaum Associates Publishers.

Johnson, I. R., Pietri, E. S., Fullilove, F., & Mowrer, S. (2019). Exploring identity-safety cues and allyship among Black women students in STEM environments. *Psychology of Women Quarterly*, 43(2), 131–150. <https://doi.org/10.1177/0361684319830926>

Kanter, R.M. (1977). *Men and women of the corporation*. Basic Books.

Kenny, D.A. (2018, September 25). *Mediation*. <http://davidakenny.net/cm/mediate.htm>

King, E. B., Hebl, M. R., George, J. M., & Matusik, S. F. (2010). Understanding tokenism: Antecedents and consequences of a psychological climate of gender inequity. *Journal of Management*, 36(2), 482-510. <https://doi.org/10.1177/0149206308328508>

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How White Americans manage a privileged identity. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 9(6), 594–609. <https://doi.org/10.1177/1745691614554658>

LaCosse, J., Sekaquaptewa, D., & Bennett, J. (2016). STEM stereotypic attribution bias among women in an unwelcoming science setting. *Psychology of Women Quarterly*, 40(3), 378–397. <https://doi.org/10.1177/0361684316630965>

- Leaper C., & Starr C. R. (2019). Helping and hindering undergraduate women's STEM motivation: Experiences with STEM encouragement, STEM-related gender bias, and sexual harassment. *Psychology of Women Quarterly*, 43(2), 165–183.
<https://doi.org/10.1177/0361684318806302>
- Lewis, N. A., Jr., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology*, 11, 40-43.
<https://doi.org/10.1016/j.copsyc.2016.05.002>
- Lewis, N. A., Jr., Sekaquaptewa, D., & Meadows, L. A. (2019). Modeling gender counter-stereotypic group behavior: A brief video intervention reduces participation gender gaps on STEM teams. *Social Psychology of Education: An International Journal*, 22(3), 557-577. <https://doi.org/10.1007/s11218-019-09489-3>
- Lewis, K., Stout, J.G., Finkelstein, N., Pollock, S., Miyake, A., Cohen, G., & Ito, T. (2017). Fitting in to move forward. Belonging, gender, and persistence in the physical Sciences, Technology, Engineering, and Mathematics (pSTEM). *Psychology of Women Quarterly*, 41, 420-436. <https://doi.org/10.1177/0361684317720186>
- Loyd, D. L., White, J. B. and Kern, M. (2008). Duo status: Disentangling the complex Interactions within a minority of two. In K. W. Phillips (Ed.), *Diversity and groups (research on managing groups and teams, vol. 11)* (pp.75-92). Emerald Group Publishing Limited. [https://doi.org/10.1016/S1534-0856\(08\)11004-0](https://doi.org/10.1016/S1534-0856(08)11004-0)
- Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observers. *Journal of Personality and Social Psychology*, 49(4), 918–926. <https://doi.org/10.1037/0022-3514.49.4.918>
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J. (1988). The "black sheep effect": Extremity of

- judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, 18(1), 1–16. <https://doi.org/10.1002/ejsp.2420180102>
- Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, 108(3), 424–437. <https://doi.org/10.1037/edu0000061>
- Master, A., & Meltzoff, A. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology*, 12(1), 152–198.
- Murphy, M. C., Gopalan, M., Carter, E. R., Emerson, K. T. U., Bottoms, B. L., & Walton, G. M. (2020). A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university. *Science Advances*, 6(29), eaba4677. <https://doi.org/10.1126/sciadv.aba4677>
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10), 879–885. <https://doi.org/10.1111/j.1467-9280.2007.01995.x>
- Murphy, M. C., & Taylor, V. J. (2012). The role of situational cues in signaling and maintaining stereotype threat. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 17–33). Oxford University Press.
- Nadal, K. L., & Haynes, K. (2012). The effects of sexism, gender microaggressions, and other forms of discrimination on women's mental health and development. In P. K. Lundberg-Love, K. L. Nadal, & M. A. Paludi (Eds.), *Women's psychology. Women and mental disorders* (pp. 87–101). Praeger/ABC-CLIO.
- Nadler, D. R., & Komarraju, M. (2016). Negating stereotype threat: Autonomy support and

- academic identification boost performance of African American college students. *Journal of College Student Development*, 57(6), 667-679. <https://doi.org/10.1353/csd.2016.0039>
- National Science Foundation, National Center for Science and Engineering Statistics. (2019). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019*. Special Report NSF 19-304. Alexandria, VA. <https://www.nsf.gov/statistics/wmpd>
- Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334. <https://doi.org/10.1037/a0012702>
- Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Bailey, A. H., Moss-Racusin, C. A., & Handelsman, J. (2019). Addressing unintended consequences of gender diversity interventions on women's sense of belonging in STEM. *Sex Roles*, 80, 527–547. <https://doi.org/10.1007/s11199-018-0952-2>
- Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of Personality and Social Psychology*, 99(1), 107–119. <https://doi.org/10.1037/a0018187>
- Pollak, K. I., & Niemann, Y. F. (1998). Black and White tokens in academia: A difference of chronic versus acute distinctiveness. *Journal of Applied Social Psychology*, 28(11), 954–972. <https://doi.org/10.1111/j.1559-1816.1998.tb01662.x>
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, 94(4), 615–630. <https://doi.org/10.1037/0022-3514.94.4.615>
- Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender

differences in how sense of belonging influences decisions to major in STEM.

International Journal of STEM Education, 5(10).

<https://doi.org/10.1186/s40594-018-0115-6>

Ramsey, L. R., Betz, D. E., & Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Social Psychology of Education: An International Journal*, 16(3), 377–397. <https://doi.org/10.1007/s11218-013-9218-6>

Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition*, 12(1), 61–74. <https://doi.org/10.1521/soco.1994.12.1.61>

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50, 835–850. <https://doi.org/10.1023/B:SERS.0000029101.74557.a0>

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. <https://doi.org/10.1037/0033-295X.115.2.336>

Schuster, C., & Martiny, S. E. (2017). Not feeling good in STEM: Effects of stereotype activation and anticipated affect on women's career aspirations. *Sex Roles*, 76, 40–55. <https://doi.org/10.1007/s11199-016-0665-3>

Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high- and low-status groups. *Personality and Social Psychology Bulletin*, 28(5), 694–707. <https://doi.org/10.1177/0146167202288013>

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance

- expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39(1), 68–74. [https://doi.org/10.1016/S0022-1031\(02\)00508-5](https://doi.org/10.1016/S0022-1031(02)00508-5)
- Sekaquaptewa, D., Waldman, A., & Thompson, M. (2007). Solo status and self-construal: Being distinctive influences racial self-construal and performance apprehension in African American women. *Cultural Diversity and Ethnic Minority Psychology*, 13(4), 321–327. <https://doi.org/10.1037/1099-9809.13.4.321>
- Shaffer, E. S., Marx, D. M., & Prislín, R. (2013). Mind the gap: Framing of women's success and representation in STEM affects women's math performance under threat. *Sex Roles*, 68, 454–463. <https://doi.org/10.1007/s11199-012-0252-1>
- Shih, M., Pittinsky, T. L., & Ambady N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83. <https://doi.org/10.1111/1467-9280.00111>
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66, 175-183. <https://doi.org/10.1007/s11199-011-0051-0>
- Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying values-affirmation interventions: Writing about social belonging is a key to buffering against identity threat. *Personality and Social Psychology Bulletin*, 39(5), 663–676. <https://doi.org/10.1177/0146167213480816>
- Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality & Social Psychology Bulletin*, 39(2), 131–143. <https://doi.org/10.1177/0146167212468332>

- Strenta, A. C., Elliott, R., Adair, R., Matier, M., & Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in higher education*, 35(5), 513-547.
<https://doi.org/10.1007/BF02497086>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.
<https://doi.org/10.1037/0022-3514.69.5.797>
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, Vol. 34 (pp. 379-440). Academic Press.
[https://doi.org/10.1016/S0065-2601\(02\)80009-0](https://doi.org/10.1016/S0065-2601(02)80009-0)
- 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 and Social Psychology*, 100(2), 255-270. <https://doi.org/10.1037/a0021385>
- Suls, J., & Wheeler, L. (Eds.). (2013). *Handbook of social comparison: Theory and research*. Springer Science & Business Media.
- Tajfel, H. & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp.33-47). Brooks/Cole.
- Tajfel, H., & Turner, J. C. (1986). Social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Nelson-Hall.
- Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*,

- 36(7), 778-793. <https://doi.org/10.1037/0022-3514.36.7.778>
- Thompson, M., & Sekaquaptewa, D. (2002). When being different is detrimental: Solo status and the performance of women and racial minorities. *Analyses of Social Issues and Public Policy*, 2(1), 183–203. <https://doi.org/10.1111/j.1530-2415.2002.00037.x>
- Tolbert, P. S., Simons, T., Andrews, A., & Rhee, J. (1995). The effects of gender composition in academic departments on faculty turnover. *Industrial and Labor Relations Review*, 48(3), 562-579. <https://doi.org/10.2307/2524782>
- Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on White Americans' conceptions of racism. *Journal of Experimental Social Psychology*, 44(6), 1491–1497. <https://doi.org/10.1016/j.jesp.2008.07.011>
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96. <https://doi.org/10.1037/0022-3514.92.1.82>
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447–1451. <https://doi.org/10.1126/science.1198364>
- Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology*, 102(3), 513–532. <https://doi.org/10.1037/a0025731>

Tables

Table 1

1a. *Gender Composition x Peer Gender Interaction on Social Comparison Perceptions (Study 2).*

Measure	Token Condition, Female Peer		Token Condition, Male Peer		Non-token Condition, Female Peer		Non-token Condition, Male Peer	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Social comparison perceptions	0.25	0.61	-0.15	0.66	-0.01	0.63	-0.10	0.68

Note. Means and standard deviations reflect standardized *z*-scores

1b. *Peer Gender x Academic Event Interaction on Stereotype Threat-related Concerns (Study 2).*

Measure	Female peer, Success		Female peer, Setback		Male peer, Success		Male peer, Setback	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Stereotype threat- related concerns	3.49	1.24	4.10	1.41	3.66	1.38	3.43	1.40

Table 2

Statistical Parameters for the Model Examining Concerns About Measuring Up to the Peer's Performance (Study 2).

Measure	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>
Social comparison perceptions				
Gender composition	.10	.06	.097	-0.02, 0.22
Peer gender	.24	.06	<.001	0.13, 0.37
Gender composition x Peer gender	.30	.12	.011	0.07, 0.55
Concerns about measuring up to peer's performance				
Social comparison perceptions	1.44	.10	<.001	1.24, 1.64
Academic event	.79	.13	<.001	0.54, 1.06
Academic event x Social comparison perceptions	-.23	.20	.275	-0.61, 0.18
Gender composition	-.20	.13	.101	-0.45, 0.04
Belonging in course				
Concerns about measuring up to peer's performance	.00	.04	.999	-0.09, 0.08
Social comparison perceptions	-.29	.10	.005	-0.47, -0.09
Gender composition	-.28	.09	.004	-0.46, -0.10

Table 3*Statistical Parameters for the Model Examining Stereotype Threat Concerns (Study 2).*

Measure	<i>b</i>	<i>SE</i>	<i>p</i>	95% <i>CI</i>
Social comparison perceptions				
Gender composition	.10	.06	.097	-0.02, 0.22
Peer gender	.24	.06	<.001	0.13, 0.37
Gender composition x Peer gender	.30	.12	.011	0.07, 0.55
Stereotype threat-related concerns				
Social comparison perceptions	1.46	.06	<.001	1.34, 1.59
Academic event	-.34	.09	<.001	-0.52, -0.17
Academic event x Social comparison perceptions	-.52	.13	<.001	-0.76, -0.27
Gender composition	.05	.09	.606	-0.14, 0.23
Belonging in course				
Stereotype threat-related concerns	-.11	.05	.044	-0.21, -0.00
Social comparison perceptions	-.13	.10	.196	-0.32, 0.06
Gender composition	-.27	.09	.005	-0.46, -0.09

Figures

Figure 1

Course Roster (Studies 1a-3)

<u>Engineering Section Roster</u>	
Blue: Male students Pink: Female students	
Eric A.	Erica A.
Adam B.	Adam B.
Steven B.	Sarah B.
Michael C.	Michael C.
Chaoxiang C.	Chaoxiang C.
Joseph C.	Joseph C.
Andrew C.	Andrea C.
Trevor E.	Trevor E.
Eric F.	Eric F.
Peter H.	Patricia H.
Aaron L.	Aaron L.
Nathan L.	Nathan L.
Ryan M.	Rosie M.
Jane M.	Jane M.
Taylor M.	Taylor M.
Hong M.	Hong M.
Nicholas O.	Nicole O.
Christopher P.	Christopher P.
Jiang P.	Jiang P.
Paul R.	Paula R.
Garrett S.	Garrett S.
Justin S.	Justine S.
Logan T.	Logan T.
Ross T.	Rachel T.
Abram T.	Abram T.
Hunter V.	Hunter V.
Igor Y.	Isla Y.
Brandon Z.	Brandon Z.
You	You

Note. Participants in the token condition saw the roster on the left (7% female), and those in the non-token condition saw the roster on the right (45% female).

Figure 2

Testing the Proposed Mechanism: The Role of Social Comparison Perceptions (Study 1b).

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 3

Testing the Proposed Mechanism: The Role of Social Comparison Perceptions, Moderated by Academic Event (Study 1b).

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 4

Testing the Proposed Mechanism: The Role of Social Comparison Perceptions, Moderated by Peer Gender (Study 2).

Note. Coefficients are unstandardized. $^{\dagger} p < .10$, $* p < .05$, $** p < .01$, $*** p < .001$

Figure 5

Testing the Proposed Mechanism: The Role of Concerns About Measuring Up to the Peer's Performance (Study 2).

Note. Coefficients are unstandardized. $^{\dagger} p < .10$, $* p < .05$, $** p < .01$, $*** p < .001$

Figure 6

Testing the Proposed Mechanism: The Role of Stereotype Threat-related Concerns (Study 2).

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 7

a. *Testing the Manipulated Mechanism: The Role of Comparison on Belonging (Study 3).*

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

b. *Testing the Manipulated Mechanism: The Role of Comparison on Course Participation (Study 3).*

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

c. *Testing the Manipulated Mechanism: The Role of Comparison on Motivation to Perform Well (Study 3).*

Note. Coefficients are unstandardized. † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

d. *Testing the Manipulated Mechanism: The Role of Comparison on Motivation to Work Hard (Study 3).*

Note. Coefficients are unstandardized. † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

e. Testing the Manipulated Mechanism: The Role of Comparison on STEM Interest (Study 3).

Note. Coefficients are unstandardized. [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$