**Evaluation Resilience: Women idea producers increase persistence and creativity under evaluation.**

Abstract

Evaluation systems in organizations produce trade-offs: they can increase employees’ innovative performance by enhancing accountability but also undermine it by creating anxiety. We consider how gender moderates the relationship between evaluation and creative performance, documenting a novel phenomenon that we call *Evaluation Resilience.* In contrast to the large body of research on evaluation apprehension showing that people facing evaluation suffer anxiety and diminished performance, we suggest that people, and women in particular, sometimes become more determined to prove themselves under evaluation. Three experiments reveal women idea producers under evaluation become particularly determined, leading to more creative idea production. Male and female innovators in organizations may experience distinct motivational responses to organizational design features, revealing new pathways to overcome typical trade-offs between evaluation, motivation, and creativity.

1. **INTRODUCTION**

A critical organizational design problem involves motivating employees to align their efforts with organizational goals (Barnard, 1938; Simon, 1947; March & Simon, 1958). To ensure effort, managers monitor, reward, and punish employees (Klapper and Reitzig 2018, Puranam 2018) Jensen & Meckling, 1976; Cruz, Gomez-Mejia, & Becerra, 2010). Although monitoring can improve performance by encouraging employee accountability, it can also have unintended consequences, if it makes employees feel anxious, thereby lowering their performance (Stinchcombe 2013, Ravid et al. 2020). Scholars have noted that when monitoring increases evaluation apprehension, defined as anxiety about being judged and negatively evaluated by others (Rosenberg, 1965; Bordia, et al., 2006), it might undermine employees’ ability to generate free flowing ideas necessary for creative work (Amabile, 1996; Deci & Ryan, 1985; Amabile & Gryskiewicz, 1989; Oldham & Cummings, 1996). As the economy shifts away from concrete tasks to knowledge-based creative work, organizations need to design their evaluation systems to manage this tradeoff between accountability and anxiety.

Thus far, however, the empirical evidence is mixed on how these trade-offs emerge in evaluation systems (e.g., Deci, Koestner, & Ryan, 1999; Gerhart & Fang, 2015). One reason why, according to scholars, is that individual differences in motivational orientation may influence employee responses to evaluation systems (Hennesey, 2019; Hennessey & Amabile, 2010; Naranjo-Gil et al., 2012). To determine how incentive systems, such as evaluations, align with organizational goals, it is essential to understand how employees perceive and react to these systems. Analyzing systematic differences across employee groups can reveal diverse responses to evaluations, potentially posing strategic tradeoffs.

Gender is a key individual difference (Eagly & Karau, 2002; Heilman, 2012), with women forming a significant portion of the workforce and critically affecting organizational outcomes. The strategic choices organizations make about their incentive systems can influence the motivation and commitment of men and women differently, and thereby affect firm performance and innovation (e.g., Dezsö & Ross, 2012; Lyngsie & Foss, 2017). Therefore, examining how gender moderates the relationship between evaluation processes and innovative performance is crucial.

This paper focuses on idea producers facing evaluation, examining gender differences in their performance and creativity. While we replicate anxiety-based evaluation apprehension among both genders, we suggest a distinctively gritty, determined, female response to certain kinds of evaluation that directly contrasts with traditional findings about the impact of evaluation apprehension on creativity (Cottrell, 1968; Cottrell, Wack, Sekerak, & Rittle, 1968). We call this response *Evaluation Resilience* and define it as increased persistence and thereby improved performance in the face of certain types of evaluation.

We conducted three controlled laboratory experiments to explore the interplay between gender, emotional state, and creative performance. We find that women's evaluation resilience, characterized by increased determination, only manifests under developmental evaluation, whereas competitive evaluation raises social anxiety in both men and women. By comparing male and female emotional responses to evaluation, we identify evaluation resilience as a potential counterpoint to evaluation apprehension and as a particularly female advantage in idea production. Since developmental evaluation appears to boost women’s innovative performance without harming men, we suggest organizations adopt this over competitive evaluation, such as innovation races or tournaments, to boost overall innovative performance in organizations.

1. **LITERATURE REVIEW**

Researchers have argued that the most crucial human resources system in organizations is performance evaluation (Judge & Ferris, 1993). Economic arguments assume a principal-agent problem, underscoring principals' difficulty in managing agents with divergent interests (Grossman & Hart, 1992). In this “Theory X” approach (MacGregor, 1960), evaluation ensures incentive alignment; increasing the intensity of monitoring addresses conflicts of interest between the employee and the organization (Reitzig & Maciejovsky, 2015). An extreme example is Taylor’s scientific management, which seeks to boost worker productivity by precise evaluation and measurement (Taylor, 1911). These ideas have profoundly influenced modern monitoring practices such as lean operations (Lund & Wright, 2001) and technological surveillance of employee attention and effort (Cascio & Montealegre, 2016; Zweig & Scott, 2007; Abraham et al., 2019). However, Taylorist monitoring practices also had unintended consequences. By making innovation a staff function, they often underutilized and undervalued the knowledge and creativity of line employees (Adler & Borys, 1996), and discouraged learning and experimentation (Bernstein, 2012). Thus, evaluation systems can create a tradeoff between efficiency and effort on the one hand and learning and innovation on the other.

Before considering how this trade-off might heterogeneously affect different employee groups, we review the roots of these opposing effects that arise from implementing formalized evaluation systems. Specifically, formalized evaluation systems support efficiency by providing clear performance metrics, reducing ambiguity in roles and expectations (Hall, 2008; Sims, Powell & Widgen, 2015). This clarity is essential for aligning individual efforts with organizational goals. Additionally, evaluation can improve employee motivation and effort (Condly, Clark, & Stolovitch, 2003; Jenkins et al., 1998). Firms use levers such as pay-for-performance and rewards to drive employees to work harder (Banker, Lee, Potter & Srinivasan 2000; Lazear, 2000) and increase their efforts.

Evaluation uses economic incentives to motivate goal achievement, and as such has the psychological effect of enhancing determination. Determination is defined as individual will that “manifests itself as the ability to set and achieve goals” (Jurgena & Cedere, 2015). By driving the persistence and commitment required to accomplish these goals, determination compels employees to exert the necessary effort, making it a key psychological predictor of behavioral effort. It has been captured in formulations such as resilience (Seligman, 1991; Dweck, 2006) or, more recently, grit (Duckworth, 2007). The determined individual is motivated to engage deeply with their work and resilient in persisting through challenges.

Well-designed evaluation systems can enhance positive emotional like determination by reducing ambiguity in what is expected of the employee and promoting a sense of fairness in their interaction with the organization. First, when evaluation systems create uniform performance standards, reduce ambiguity, and offer clear role expectations (Locke & Latham, 2002), employee focus and effort increase (Ashford & Cummings, 1983). This is particularly the case if people rely upon r**eflected self-esteem—i.e., focusing on** how others evaluate their performance—as compared to intrinsic self-esteem—i.e., self-esteem that derives from one’s innate qualities as compared to external achievements (Cooley, 1902; Leary & Baumeister, 2000; Zajonc, et al., 1970). A coherent evaluative structure reduces ambiguity by providing clear criteria for self-validation from others. Second, the evaluative structure can also reduce ambiguity around **fairness,** if evaluations apply consistent criteria across employees, promoting a sense of impartiality (Adams, 1965). This perceived fairness is especially critical for employees who may feel they have been subject to biases, and can enhance their determination to succeed (Greenberg, 1986).

However, in addition to reducing anxiety by providing a clear structure, monitoring and evaluation can create unintended forms of anxiety: i.e., evaluation apprehension, i.e. anxiety-toned concerns about winning positive evaluations from experimenters (Rosenberg, 1965). Further, evaluation leads to a focus on productivity or output generation as compared to innovative thinking (Amabile, 1996) and problem solving (Baer & Oldham, 2006), because it activates dominant responses rather than creative idea generation (Baas, De Dreu & Nijstad, 2008; Zajoncs et. Al., 1970). Bernstein (2012) found that curtaining off the assembly area for groups reduced their anxiety about being monitored, increased experimentation, and ultimately improved productivity. Whereas evaluation might temporarily boost productivity by focusing individual attention on specific performance criteria, these studies highlight the darker side of evaluation, whereby it can increase employee anxiety and undermine creativity and innovation (Shalley & Perry-Smith, 2001).

However, evidence for the claim that evaluation reduces innovation or creativity is mixed, suggesting this relationship is more complex than meets the eye. Some scholars find that evaluation reduces creativity (Baer & Oldham 2006; Byron & Khazanchi, 2012), whereas other found nonsignificant effects[[1]](#footnote-1) (Miller, 2007). One reason for these mixed results might be variations in how people perceive and experience evaluation, determining whether evaluation’s positive effects (increasing employee determination and effort) or negative effects (increasing anxiety) prevail. Individuals vary in what motivates them (Conti, et. Al., 1995), which might shift the balance between the positive and negative effects of evaluation.

Gender may shape construals of evaluation, if men and women differ in their preferences for workplace autonomy versus structure. Specifically, men typically score higher on measures of autonomy and control, which heightens their perceptions that evaluative contexts constraining (Niederle, 2016). On the other hand, women, who may face greater ambiguity in behavioral expectations due to conflicting gender and professional roles, could view structured evaluations as supportive frameworks that reduce uncertainty, providing external validation and guiding their efforts (Eagly & Karau, 2002).

In sum, gender could be a key boundary condition for evaluation’s effect on creative performance. While men may react negatively if they perceive evaluations as a threat to their autonomy and control, evaluations could enhance determination for women—by providing clear structure, motivating them to make a positive impression, and fostering a sense of fairness (Deci & Ryan, 1987; Hennessey & Amabile, 1988). We propose that these psychological responses are critical mediators between the organizational system as designed-in-theory and the ultimate performance effects as people respond to the system-in- practice (Trist & Bamforth, 1951).

It is critical for organizational designers to address gender differences in their evaluative systems because such disparities can impact key strategic outcomes, innovation, and risk-taking behavior within firms. Gender has been studied as an important factor in strategic management, with research highlighting its influence on leadership, decision-making, and organizational outcomes (Lee & James, 2007; Guldiken et al., 2019). Evaluative systems that inadvertently penalize women may thus stifle innovation and reduce the firm’s competitive edge, making it essential to design efficient and fair evaluation processes that recognize differences.

1. **THEORY AND HYPOTHESES**

***Evaluation, Social Anxiety, and Creativity.***

While evaluation enables accountability in performance management, it can create counterproductive emotional responses such as anxiety that can negatively affect creative idea generation (Leary, Barnes, Griebel, Mason, & McCormack Jr, 1987; Levy & Williams, 2004). High levels of social anxiety can diminish intrinsic motivation, reducing the enjoyment of tasks (Ariely et al., 2009; Daker et al., 2020). This is particularly important for creative work, as the fear of judgment may lead individuals to conform to safe, conventional ideas rather than exploring novel and innovative solutions (Mueller, Melwani & Goncalo, 2012; Amabile, 1996). High levels of social anxiety can diminish intrinsic motivation, reducing the enjoyment of tasks critical to creative expression and risk-taking (Amabile, 1985; Ryan & Deci; 2000; Byron & Kazanchi, 2012). Because individuals fear making negative impression on others, they also filter out unconventional ideas and opt for safer, more conventional responses, ultimately stifling creativity (Lau & Buckland, 2001; Byron & Khazanchi, 2011). Therefore, in line with prior work on evaluation apprehension, our first hypothesis is that high levels of social anxiety negatively affect the idea-generation process stifling creativity or innovation.

*H1: Evaluation reduces creativity through increased social anxiety.*

***Evaluation Resilience: Gender Differences in Response to Evaluation***

While evidence of evaluation apprehension dominates prior research, evaluation’s psychological benefits, which potentially enhance innovative output beyond economic benefits such as accountability and reward, are less explored. Prior work emphasizes evaluation’s negative psychological effects on autonomy: it feels controlling and therefore demotivates employees (Deci & Ryan, 1985; Gagné & Deci, 2005; Harackiewicz & Sansone, 1991). Importantly, men typically score higher on measures of autonomy, independence, and self-direction (Niederle, 2016). Thus, evaluation could feel particularly controlling and demotivating to men.

As we argued earlier, the positive effects of evaluation on determination and thereby effort, have been somewhat overlooked in prior work. We hypothesize that women, may experience more psychological benefits than men when participating in a coherent evaluative structure with fair, clear feedback for two reasons. First. due to their relational orientation (Schwalbe & Staples, 1991), women may focus more on external evaluations (Nolen-Hoeksema, 1990), and exhibit more extrinsic motivation (Boggiano, Main, & Katz, 1991; Davis, Winsler, & Middleton, 2006; D'Lima, Winsler & Kitsantas, 2014). For example, adult approval had a stronger effect on female vs. male students (Boggiano et al., 1991), including parental and teacher rewards (D'Lima et al., 2014). This sensitivity to social feedback implies that clear external feedback plays a vital role in shaping women's determination. Second, since women experience gender bias in the workplace (Heilman, 2012; Eagly & Karau, 2002), having a unified evaluation that holds everyone to the same evaluation standards might give women a sense of fairness and hence increase their determination towards the task.

On the other hand, men facing evaluation might experience a decrease in determination, because evaluation can feel coercive, triggering a perceived loss of autonomy and intrinsic motivation, which men value more highly (Deci & Ryan, 2000). As a result, rather than enhancing effort, evaluations may undermine men's sense of agency, leading to reduced effort or disengagement from the task.

This highlights the potential tradeoff: when evaluation systems are perceived as a fair developmental structure, evaluation can create resilience and determination (more likely among women), whereas when they are perceived as controlling and monitoring, evaluation can produce anxiety and apprehension (more common among men). Thus, we hypothesize that:

*H2a: Women’s determination increases more than men’s under evaluation which, in turn, increases effort.*

High effort, driven by determination, significantly contributes to creative performance. Effort enables individuals to move beyond conventional thinking (Kahneman, 2011). Instead of fixating on top-of-mind solutions, it is highly effortful to develop multiple alternatives (Harvey & Berry, 2022), challenge ordinary assumptions, and generate novelty (Amabile, 1996; Kachelmeier, Wang & Williamson, 2019; Cardinaels & Feichter, 2021). Thus, following prior research (Dweck, 2006; Duckworth, et al., 2007) that directly links mental state (i.e determination) with behavioral manifestation (i.e. effort) and performance outcomes (i.e. creativity), we suggest:

*Hypothesis 2b: There is a serial mediation between evaluation and creativity moderated by gender, such that evaluation will increase determination, which will increase effort and, ultimately, creativity.*

While evaluation can often induce social anxiety, which typically harms creativity, our arguments above suggest that its psychological effects may differ between genders. If women exhibit increased determination and effort in response to evaluation (H2a), these positive effects could largely offset the negative impact of social anxiety. Consequently, the overall impact of evaluation on women’s creativity may be less detrimental compared to men’s, as the boost in their determination and effort helps counteract the negative effects of evaluation-induced anxiety.

Women’s increased determination and effort in response to evaluation may translate into heightened creativity, yet it is crucial to recognize the tradeoff posed by evaluation-induced social anxiety. Increased effort, particularly among women, can fuel creativity by encouraging them to engage more deeply with tasks and push their boundaries (Byron & Khazanchi, 2012). However, the anxiety caused by evaluation may still inhibit risk-taking and experimentation, key components of creative output (Amabile, 1996). This suggests a nuanced dynamic: although women's heightened determination may enhance their creativity, especially in structured tasks, the persistent anxiety associated with being evaluated could temper this effect. The positive impact of effort may, therefore, offset but not fully eliminate the creativity-limiting aspects of social anxiety, reflecting a balance between the motivating and inhibiting forces of evaluation.

*Hypothesis 3: Evaluation will decrease men’s creativity more than women’s.*

***Gendered response to different types of evaluation***

In the previous hypotheses, we have argued that men and women will respond to evaluations differently, with one key reason being how evaluations are perceived by employees. Evaluation systems in organizations can also vary significantly in their nature. The two most critical and distinct types of evaluations commonly used in organizations are competitive evaluations and developmental (constructive) evaluations (Shalley & Perry-Smith, 2001; Stephan & Dorfman, 1989). The type of evaluation can induce different psychological responses, again shifting the balance between social anxiety and determination, and may have heterogeneous impact on men and women. Therefore, it is important to consider how different kinds of evaluations affect employees in general, and potential gender differences.

We argue that competitive, rather than developmental, evaluation is more likely to heighten social anxiety among all employees. Competitive evaluations, which focus on comparison among employees, highlight areas of deficiency without providing constructive guidance. Employees subjected to such evaluations may feel scrutinized and judged harshly, leading to stress and anxiety about their performance. This type of evaluation can undermine an employee’s confidence and sense of job security, making them more preoccupied with avoiding errors than with engaging creatively in their work.

*H4a: Competitive evaluations will increase social anxiety among employees.*

In contrast, developmental evaluations, which offer constructive feedback aimed at personal and professional growth, are less likely to provoke anxiety. By emphasizing improvement and support rather than criticism and comparison, developmental evaluations can promote a growth mindset that can enhance motivation and effort by offering clarity and reducing uncertainty about performance expectations.

Below, we argue that competitive evaluation is more detrimental to women than men, whereas developmental orientation is particularly helpful for women compared to men. When evaluations prompt competition or social comparison, women, as well as men, might perceive them negatively, potentially leading to stereotype threat (Steele & Aronson,), performance anxiety, and decreased motivation. Economists find that women in particular exhibit reduced performance in competitive conditions, particularly against men (Gneezy & Rustichini, 2000). Competitive orientation, which involves striving to demonstrate competence relative to others, is linked with autonomy orientation, where individuals prioritize personal values and self-direction (Orosz et al., 2018; Ryckman et al., 1996). Conversely, given that women place greater value on external feedback and opinions, evaluations that focus on personal development and learning could motivate women. Given role incongruity (Eagly & Karau, 2002), constructive feedback clarifies role and performance expectations (Bullough, Guelich, Monolova & Schjoedt, 2022), creating more supportive and productive work environment. Therefore, we hypothesize:

*H4b: Developmental, but not competitive, evaluations will increase women’s determination and effort more than men’s.*

**4. METHODS**

**4.1 Experiment 1**

Experiment 1 explores how evaluation affects effort and creativity, along with gender’s moderating role. The experiment is guided by two hypotheses. First, replicating past research, we test the prediction that evaluation negatively affects creativity, an effect mediated by social anxiety (H1). Second, we test the novel H3, which proposes that evaluation’s detrimental effects of creativity are lower for women than men, driven by women’s higher effort.

**4.1.1 Participants and design**:

The design was a 2x2 between-subject design with evaluation (present or absent) and gender (male, female). Our sample consists of 136 undergraduate students at a large public university in the U.S. After removing observations from 4 participants who failed the attention check,1 we had 132 participants. These students were enrolled in an undergraduate business class (97% were business majors). They were an appropriate sample given that they planned to enter the business world and would be subjected to these types of evaluation (Frechette, 2015). The sample consisted of 43 females (32.58%), 89 males (67.42%), 80 Caucasians (60.61%), 40 Asians (30.30%), and 12 from other races (9.09%).

**4.1.2 Procedure:**

All participants were recruited through an email announcement from their instructor. Participants were informed that the experiment would take 30-45 minutes to complete. After reading the consent form and agreeing to participate, participants were randomly assigned to one of two experimental conditions in a between-subject design: 1) evaluation and 2) control condition with no expected evaluation.

Following prior literature, we created the evaluation manipulation based on expert evaluation (Collaros & Anderson, 1969; Diehl & Stroebe, 1987; Henchy & Glass, 1968; Zhou et al., 2019). Participants in the evaluation condition received the following prompt: “Before we begin, we shall tell you that two members of the faculty here at [large Midwestern University] will evaluate your performance based on the feasibility and originality of your proposals. They are specialists in management and business administration, and they'll be making systematic observations of your performance. You will also receive critical feedback about your performance as well. Please note that experts will receive your name, age, gender, and your [College ID] headshots along with your proposals. Please see next page for an example of the content that will be shared with experts."

[Insert figure 1 here]

The evaluation manipulation was the only additional information the treatment group received in addition to what the participants in the control group received. The participants received this prompt right before they received the writing task. Participants in the control group did not receive any manipulation in Study 1. All participants then completed manipulation checks and mediating variables.

Following previous literature on idea generation (Keum & See, 2017), participants were asked to write three proposals to solve common business problems. Participants were told that their proposals should be original and feasible and that they should write a minimum of 50 words. To avoid invoking stereotype threat (Steele & Aronson, 1995), the participants writing tasks involved three gender-neutral prompts – one each about marketing, innovation, and leadership – that business majors should feel comfortable providing opinions about and are unlikely to invoke gender-specific knowledge or make gender more salient in subjects' minds. We did not find significant order effects or gender-based differences in creativity between the three categories in an ex-post analysis. Each participant was given 5 minutes to write their proposal. At the end of 5 minutes, the student was directed to the next prompt. The exact prompts are shown in Appendix 1. After the study was complete, participants were debriefed on the purpose of the study (McNallie 2017).

**4.1.3 Measures:**

In this study, *Gender* and *Evaluation* where the independent variables, whereas *Creativity* was the dependent variable. *Social anxiety* and *Effort* were the mediating variables.

*Gender*. A 0 (male)/1 (female) dummy variable was assigned to participants based on the participant's self-reported sex. Although a non-binary option was provided, all participants identified as either male or female.

*Evaluation*. A 0 (control)/1 (treatment) was assigned to participants based on condition.

*Manipulation Check.* All participants were asked to report the extent to which they agreed with the following statement: “When I was writing the proposal, I was apprehensive about the quality of my ideas.” (1=strongly disagree, 7=strongly agree). We confirmed that participants in the evaluation condition (M=5.07, SE= 0.099) reported greater apprehension than those in the control group (M=4.49, SE=0.127, d= -0.58, t(394)= -3.63, p= <0.001).

*Social Anxiety*. We asked each participant 17 questions from the SPIN social anxiety survey (Connor et al., 2000). The scale reliability coefficient for social anxiety was 0.89.

*Creativity*. Two independent coders judged each proposal to assess the creativity of the submitted proposals. They rated the proposals on a Likert scale from 1 to 5 on the two components of creativity: originality and feasibility (Amabile, 1993). When the coders differed by more than two points on either attribute, we had them discuss the proposal together and come to a consensus. We averaged the originality and feasibility scores to compute the creativity variable. Interrater agreement was 77.19% for originality and 83.75% for feasibility, both of which are acceptable according to the interclass correlations (Cicchetti & Sparrow, 1981).

*Effort*. We measured effort by counting the total number of words in each proposal.

*Survey Duration.* We control for time taken by participants (in seconds), as outliers can indicate attention failure.

**4.1.4 Results:**

Descriptive statistics and correlations are displayed in Table 1. H1 predicted a mediated negative effect of evaluation on creativity through increased social anxiety. We examine this potential mediating effect of social anxiety between evaluation and creativity. First, we conducted a t-test which revealed that participants had considerably higher social anxiety in the evaluation condition (M=2.23, SD=0.05) than in the control condition (M=2.06, SD= 0.05; d=0.17, t(394)= 2.24, p=0.026). Then, we proceeded to test the mediation model with creativity as the dependent variable. The mediation analysis results in Table 2 indicate that evaluation has a positive effect on social anxiety (α1=0.14, SE=0.08, p=0.106), whereas social anxiety has a negative and significant effect on creativity (β1= –0.11, SE=0.05, p=0.017). We also find a significant indirect effect of social anxiety on creativity (α1β1= –0.01, SEa1β1=0.01, 95% CI= –0.04, 0.00). These results provide support for H1.

We next examine the interaction between gender and evaluation on effort, followed by effort's effect on creativity. For the first leg of our mediation, the positive interaction between gender and evaluation on effort emerges (2=15.46, SEa2=7.86, p=0.050, Table 2). We also see a significant positive effect of effort on creativity as well (β2=0.004, SEβ2=0.001, p<0.001). These effects are also evident in the index of moderated mediation, or the differences between conditional indirect effects, where we see significant results (α2β2=0.05, SEa2β2=0.03, 90% CI= 0.00, 0.11).

 Although determination was not measured in Experiment 1, we sought to explore how evaluation influences the efforts of men and women as a preliminary test and motivation for our next study. We expanded the mediation model which investigates how evaluation influences creativity through social anxiety, incorporating effort as a parallel mediator, moderated by gender. This analysis was conducted using SPSS PROCESS (Hayes, 2017, model 7) with Monte Carlo analyses using 5,000 bootstraps resamples to construct 90% confidence intervals (C.I.s) for indirect effects. We entered evaluation as the independent variable (coded as 1 = expected evaluation, 0= no expected evaluation), social anxiety and effort as parallel mediators, gender as the moderator between evaluation and effort, and creativity as the dependent variable. Table 2 presents the results of our analysis.

Finally, to test H3, which concerns gender's moderating role in the relationship between evaluation and creativity, we used ANOVA. Results revealed that there is a significant interaction effect between evaluation and gender on creativity (F (1, 392) = 4.49, p = 0.035). A split-sample analysis further indicates that for males, those in the evaluation condition have significantly lower creativity scores (M = 2.32) compared to the control group (M = 2.57), with a difference of 0.25 (t = 3.12, p = 0.001). In contrast, for females, there is no significant difference in creativity between the evaluation (M = 2.51) and control conditions (M = 2.46), with a difference of -0.05 (t = -0.42, p = 0.66). Our analysis suggests that while evaluation decreased creativity among men, it does not have the same impact on women. Figure 3 illustrates the mean creativity scores categorized by gender across two experimental conditions: control and evaluation.

**4.1.5. Discussion:**

In support of Hypothesis 1, our t-test and the mediation analysis showed that participants randomly assigned to the evaluation condition experienced higher social anxiety, which mediates a negative effect of evaluation on creativity. A follow-up mediation analysis revealed that evaluation increases effort for women but has no effect on men’s effort. Lastly, in support of H3, we found that gender moderates evaluation's effect on creativity. Evaluation increases social anxiety of the general sample, which has negative creativity consequences, but it increases the effort of only women, which has positive creativity consequences. Furthermore, we found that while evaluation negatively affects male creativity scores, it did not affect the women’s scores.

These findings suggest that evaluation can generate distinct psychological and behavioral forces, with contrasting effects upon creativity. Evaluation produces social anxiety (which negatively affects creativity) and also effort. We observed the resilience pattern among women, but not men, whose creativity scores were significantly lower in the evaluation condition than in the control condition.

This experiment supports our proposed model and captures key behavioral outcomes, including effort and creativity. However, to understand the psychological experiences of men and women facing evaluation in more detail, Experiment 2 involves more nuanced measures of affect. We were particularly interested in understanding the gender differences in psychological responses to evaluation that might have led to increased effort among women but not men, and when evaluation translated into affective responses such as anxiety, which could disrupt performance, (Beilock & Carr, 2005).

**4.2 Experiment 2**

Experiment 2 hypothesizes that a sense of determination precedes women’s increased effort under evaluation. The purpose was to test this proposed psychological mechanism in H2a and H2b, in addition to testing the robustness of our results from experiment 1, H1 and H3. This experiment was preregistered at https://osf.io/973uj. We followed our preregistered plan except for the following modifications. First, we do not present the analysis of evaluation’s effect on self-esteem due to challenges in measuring self-esteem as a state variable, as opposed to a trait, within the scope of this study. Secondly, we do not to report on idea sharing in this paper, and focus on the mechanisms of creativity during idea generation. Both of these analyses are available from the first author.

**4.2.1 Participants and design:**

The design was a 2x2 between-subject design with evaluation (present or absent) and gender (male, female). We recruited 260 undergraduate students at a large public university in the U.S. for an in-person lab experiment. After excluding 36 observations that failed attention checks2 and one participant who self-identified as non-binary, the final sample consisted of 223 observations. 100 women (44.8%) and 124 men (55.2%); 153 Caucasians (68.3%), 43 Asians (19.2%), and 28 participants from other races participated (12.5%).

**4.2.2 Procedure:**

All participants were recruited from a class via their professors’ announcement and an email. Upon arriving at the classroom, participants were informed that the experiment would take 15-30 minutes and completed the consent form. They were randomly assigned to one of two experimental conditions in a between-subject design: 1) treatment condition with expected evaluation and 2) control condition with no expected evaluation.

As in experiment 1, we informed participants in the evaluation condition that their ideas would be evaluated. Experiment 1, like several other prior manipulations of evaluation (Collaros & Anderson, 1969; Diehl & Stroebe, 1987; Henchy & Glass, 1968; Zhou et al., 2019), combined two distinct psychologies in their manipulations— identifiability (name, photo vs. anonymous responding) and evaluation. However, in this experiment, we removed the identifiability aspect to analyze evaluation’s effect on creativity in isolation from the threat that comes from being identifiable. This was particularly important in our context because both accountability and anxiety outcomes of evaluation could be driven by being identifiable and not necessarily by being evaluated. Additionally, in this experiment, we focused on a neutral form of evaluation by removing feedback statements and eliminating the expectation of feedback present in Experiment 1. This approach allows us to isolate the effects of evaluation without the influence of specific types of evaluation.

However, we anticipated that removing the identity disclosure requirement could weaken the manipulation: participants might not take the evaluation process seriously if they believed responses would not be linked to them personally. To address this, we made three adjustments to the evaluation process based on feedback from previous experiments. First, we emphasized the importance of the evaluation process, using language to highlight its significance. Specifically, participants in the treatment group were given the following prompt: “Before we begin, please note that we will **evaluate your performance** based on the feasibility and originality of your proposals and we will be making systematic observations of your performance. We will rank your creative performance on the following criteria: low/ below average/ average above average/ outstanding.”

Second, to emphasize the absence of evaluation in the control group and prevent participants from assuming that their performance would be assessed, we explicitly stated the lack of evaluation. Specifically, participants in the control condition received the following prompt: "Before we begin, please note that your suggestions today will **not** be evaluated in any way. We encourage you to write feasible and original solutions, though, as we very much look forward to reading them." Third, to enhance the effectiveness of the manipulations, participants in both the treatment and control groups were asked to reflect on a similar task they had encountered previously and to express their feelings about it in a couple of sentences. This additional step aimed to amplify the manipulations and ensure that participants were fully engaged in the task.

Next, all participants completed measurements of manipulation checks and mediating variables to assess the effectiveness of the manipulations and to understand potential underlying mechanisms. Finally, as in Experiment 1, participants wrote solutions to a general business problem to measure creativity behaviorally. We included only one creative writing task (the marketing task from Experiment 1) to reduce the potential for participant fatigue. Participants received the same directions as in Experiment 1. Following the completion of the study, participants were debriefed about the purpose of the study (McNallie, 2017).

**4.2.3 Measures:**

Evaluation, gender, effort, and survey duration were measured in the same way as in Study 1. We report the added or changed measures below.

*Manipulation Check.* We asked participants to report whether they expected to be evaluated (1 for yes and 0 for no). Participants in the treatment condition (M=0, SE=0) reported that they would be evaluated more than the participants in the control group (M=0.89, SE=0.29; t(222)= 30.43, p<0.000, d=0.894), passing the manipulation check. We also asked participants to report the degree to which they felt judged on a 5-point scale. The ANOVA results revealed that participants in the evaluation condition reported feeling significantly more judged than those in the control condition (F (1, 222) = 448.20, p < 0.001).

*Creativity.* As in experiment 1, we first calculated interrater agreement for originality (82.05%) and feasibility (82.68%). The average of these two measures comprised our creativity measure.

*Social Anxiety.* We used a shortened version of spin anxiety survey and included only the most relevant 5 items to this study. Items are listed in Appendix 2 and scale reliability coefficient across the items was 0.91.

*Determination.* To measure affective responses to evaluation, we use the PANAS scale which comprises 8 fundamental emotions: anxious, competitive, distressed, excited, determined, scared, nervous, and interested. Participants were instructed to rate their feelings on a scale from 1 to 5, indicating their emotional state after the manipulation and amplification procedures but before engaging in the writing task.

**4.2.4 Results:**

Descriptive statistics and correlations are reported in Table 3. Next, we tested H2a and H2b, where we analyzed our moderated mediation model, which proposed an affective pathway of determination, which drives women facing evaluation to exert more effort (H2a). We did not expect men to exhibit this affective pathway, resulting in more positive effects of evaluation on creativity for women than men (H2b). To test this model, we turned to PROCESS analysis to construct 95% CI for indirect effect. We entered evaluation as the independent variable (coded as 1 = expected evaluation, 0= no expected evaluation), determination and effort as serial mediators, gender as the moderator between evaluation and determination, and creativity as the dependent variable. In examining Table 4a, we confirm H2a and H2b. The first leg of our serial mediation model examines the interaction between evaluation and gender on determination. Here, we see a positive and significant effect (α1=0.77, SEa1=0.30, p=0.012). As expected, we also see a significant positive relationship in our second leg, which examines the effect of determination on effort (β1=4.01, SEβ1=2.29, p=0.081). Finally, we look at effort's effect on creativity and again see a significant relationship (γ1=0.003, SEγ1=0.001, p<0.001). Our serial mediation is additionally supported by the index of moderated mediation between all three legs (α1β1γ1=0.014, SEa1β1γ1=0.010, 90% CI= 0.00, 0.03). Therefore, we find support for H2a and H2b.

Second, replicating experiment 1 and using ANOVA, we find that evaluation has a less negative effect on female as compared to male creativity (F (1, 220) = 3.08, p = 0.081), supporting H3. While experiment 1 shows the effects of evaluation on social anxiety, the serial mediation analysis in experiment 2 shows that evaluation increases determination among women (Figure 5), which in turn increases their effort and creativity.

Finally, replicating Experiment 1 and supporting H1a, a t-test revealed that Experiment 2 participants in the evaluation condition also experienced higher social anxiety (M=2.59, SD=0.09) than the participants in the control group evaluation (M=2.32, SD= 0.09; t(222)= 2.09, p=0.038, d=0.26). However, using the PROCEESS analysis, we failed to find support for social anxiety's mediating role in evaluation and creativity (Table 4b). We believe this is due to the lack of identifiability in our experimental design for study 2.

**4.2.5 Post-hoc Analysis for Alternative Mechanisms:**

Using post-experiment survey measures, we examined two alternative mechanisms to explain why women anticipating evaluation increase their determination: (1) they expect the evaluation to be unfair and must prove themselves, and (2) they view evaluation as a learning opportunity and are more interested in receiving feedback.

*Perceived Fairness*: To measure fairness perceptions, participants were asked if they believed their ideas would be evaluated fairly. We employed a moderated mediation model with perceived fairness as the mediator, determination as the dependent variable, evaluation as the independent variable, and gender as the moderator. The results showed no significant relationship between perceived fairness and determination, nor any interaction between gender and evaluation on the expectation of fair treatment. These results suggest that women’s heightened determination was not due to an expectation that they would be unfairly evaluated.

*Learning from feedback*: To measure openness to feedback, participants were asked if they would be interested in sharing their ideas in exchange for more feedback. We found evidence that women were 74.6% more likely than men to share their ideas for the sake of feedback (p=0.09), and the preference for receiving feedback was strongly associated with increased determination overall.

Our findings suggest that perceived fairness does not play a significant role in driving women’s increased determination under evaluation conditions. Instead, openness to feedback and the view of evaluation as a learning opportunity appear to be the primary mechanisms, with women showing a stronger association between feedback preference and determination compared to men. This indicates that women’s motivation in these contexts may stem from a desire for development and growth rather than fairness concerns.

**4.2.6 Discussion:** The results of experiment 2 generally supported our predictions. First, a positive affective pathway emerged, particularly among women. In Experiment 1, gender moderated evaluation's effect on effort and, ultimately, creativity.   In Experiment 2, gender moderated this relationship through the affective response of determination. Women’s determination increased more than men’s when expecting evaluation, leading to a gender difference in effort and creative performance. Therefore, experiment 2 replicated Experiment 1’s finding of the female advantage in creativity under evaluation. We suggest that determination might be a unique emotional signature in female responses to evaluation, explaining the female advantage in creativity under evaluation. Additionally, consistent with Study 1 and the prior literature, evaluation increased social anxiety. However, unlike experiment 1, the experiment 2 did not indicate that social anxiety mediates creativity. We suspect that removing the identifiability from the manipulation might have reduced the negative effect of social anxiety on creativity.

This experiment isolated the effects of evaluation, eliminating confounding psychologies such as identifiability vs. anonymity that are present in extant manipulations of evaluation apprehension. Our post-hoc analyses suggested that women’s determination drives from a learning orientation, as compared to other motivations like proving oneself in a biased environment. As such, Experiment 3 directly tests this mechanism by directly manipulating different types of evaluation. We expect that developmental evaluation drives the determination and effort pathway for women.

**4.3 Experiment 3**

Evaluation might either be construed as a competitive process, leading to increased social anxiety, or as an opportunity for self-improvement, resulting in heightened determination. Experiment 3 aimed to directly test the mechanism, i.e. that developmental evaluation inspires determination and effort in women. We did so by directly manipulating these forms of evaluation, and compared male and female emotional and behavioral responses, including increased effort and creativity. This experiment was preregistered at https://osf.io/anjbf.   We deviated from our preregistered hypotheses to streamline the paper. We chose not to explore gender differences in idea sharing because our analysis revealed that creativity and idea sharing outcomes were closely intertwined, making separate study of idea sharing less feasible. All data/analyses are available from the first author. These changes were made to streamline our study and enhance its clarity.

**4.3.1 Sample and design:**

Experiment 3 recruited undergraduate business students in a large Southeastern public university. 338 participants were recruited from nine classes. After applying the same restriction criteria as previous experiments and excluding observations that did not follow the instructions, our final sample consisted of 320 students, with 113 women (35.3%) and 207 men (64.7%). 286 students (89.38%) reported their race as Caucasian, 10 (3.13%) Black and 24 (7.5%) reported other races.

**4.3.2 Procedure:**

Building on prior research (Shalley & Perry-Smith, 2001), we developed manipulations that differentiated between two distinct evaluation approaches: competitive and developmental. Participants randomly assigned to the competitive evaluation were informed that this evaluation aimed to compare their performance against that of other participants in the same class, ultimately ranking their creative potential: "Before we begin, please note that the feasibility and originality of your proposals will be evaluated by experts to compare you against all other participants in this class and rank your creative potential". We also included Figure 1 in the prompt. Following this, participants were asked to reflect on a past situation where their performance had been evaluated for comparative purposes and to describe their emotional response to this scenario.

Participants assigned to the developmental condition received instructions that stated: “Before we begin, please note that the feasibility and originality of your proposals will be evaluated by experts to provide you with developmental feedback and grow your creative potential,” followed by Figure 1. Like the competitive group, participants were then prompted to recall a past instance where their performance had been assessed for developmental feedback and to express their emotional reactions.

Finally, participants in the control group were told, "Before we begin, please note that this task is electronic, and all you need is the electronic device you are presently using to complete the task.". That is, they engaged in a writing task as did the other participants, but one that did not involve any statement that they would be evaluated in the upcoming tasks. However, they did receive Figure 1 and the same prompt outside of the conditions. Subsequently, all participants completed manipulation checks, mediators, and the proposal writing task as in the prior two experiments. Participants were debriefed at the end.

**4.3.3 Measures:**

Evaluation, gender, effort, survey duration, and PANAS were measured in the same way as in the previous studies. We report the added or changed measures below.

*Manipulation Check*. As we have three different manipulations in this study, we asked participants not only whether they thought they were going to be evaluated but also the nature of that evaluation. Therefore, we asked participants to choose yes or no as the answer to the following three statements. First, to compare the control group with both evaluation groups, we asked participants to answer, "According to the directions on the previous page, my performance on this task will be evaluated." T-test results showed that both treatment groups thought they were evaluated more than the control group with the following statistics between competition (M=0.981, SD= 0.013) and control groups (M=0.639, SD= 0.046; t(212)= 7.03, p<0.001, d=0.342), and between developmental evaluation (M=0.934, SD= 0.024) and control group (M=0.639, SD= 0.046; t(212)= 5.6, p<0.001, d=0.295). Then, specifically to test whether participants in the competitive evaluation thought they were going to have a competitive evaluation more than the other two groups, we asked "According to the directions on the previous page, my performance will be compared against all other participants." Differences in means between competitive (M= 0.972, SD= 0.017) and control groups (M= 0.509, SD= 0.048; t(212)= 9, p<0.001, d=0.462); and between competitive (M=0.971, SD= 0.061) and feedback (M= 0.604, SD= 0.048; t(210)= 7.29, p<0.001, d=0.368) were significant, showing that manipulation check was successful. Finally, to test the difference in response to the developmental manipulation, we asked participants "According to the directions on the previous page, I will receive developmental feedback on my performance."  Participants in the developmental evaluation condition (M= 0.887, SD= 0.031) reported that they will receive developmental evaluation more than the participants in the control group (M=0.296, SE=0.044; t(212)= 10.92, p<0.001, d= 0.59),  and more than the participants in competition group (M=0.481, SE=0.488; t(210) = 7.026 p<0.001).

*Creativity*. Following our two other experiments, we calculated the interrater agreement for originality and feasibility between the two independent raters’ scores, which were 83.75% and 77.19% respectively and since they were acceptable, we averaged them to develop our creativity measure.

*Social Anxiety*. We used the same 5 items in the shortened version of the SPIN survey as we did in Experiment 2. The scale reliability coefficient was 0.89.

**4.3.4 Results:**

Descriptive statistics and correlations are displayed in Table 6. In hypothesis 4a, we argued that competitive evaluation, but not developmental evaluation, would increase social anxiety. In support of this hypothesis, t-test results showed that participants in competition condition reported higher social anxiety (M=2.55, SE= 0.09) than those in control (M=2.3, SE= 0.088; t(212)= 2.009, p= 0.045, d= 0.253) and developmental (M=2.3, SE= 0.091; t(210)= 1.93, p=0.054, d= 0.247). Additionally, t-tests did not reveal any statistically significant differences in social anxiety between the developmental evaluation and control conditions (t(212)= 0.044, p=0.965, d = 0.006). There was no gender moderation for the effect of different evaluation types on social anxiety.

In examining developmental feedback (H4b), we find significant relationships throughout our model. We see a positive and significant interaction between evaluation and gender's impact on determination (α1=0.59, SEa1=0.31, p=0.060). We also see a similar effect of determination on effort, as we saw in H2c (β1= 4.11, SE 1=1.65, p=0.013). We confirm the above measures by examining the index of moderated mediation, which shows similar results (a1β1= 0.0206, SEa1β1=0.0134, 95% CI=0.00, 0.0522). Overall, these findings support H4b and also parallel our findings in Experiment 2 – that determination mediates the relationship between developmental evaluation, gender, and effort.

Now, we shift to studying how this relationship holds under competitive evaluation. Unlike our developmental feedback model, we do not see any significant relationship between evaluation and gender's effect on determination. In fact, we even see a flipped sign, albeit with no significance (α2=–0.07, SEa2=0.31, p=0.825). We also do not see a significant relationship between determination and effort in the competitive evaluation condition (β2=4.17, SEβ2=1.65, p=0.0119). The index of moderated mediation substantiates these findings, as the confidence interval straddles 0 (α2β2=–0.012, SEa2β2=0.012, 90% CI= –0.04, 0.01). Therefore, we do not find support for the idea that determination mediates the relationship between competitive evaluation, gender, and effort. These results combined with those for developmental evaluation support H4b.

**4.3.5 Post hoc analyses:**

In line with the approach taken in Experiment 2, we conducted post hoc analyses to examine the mechanisms by which women respond to evaluation with increased determination. Consistent with the findings from Experiment 2, women did not report heightened concerns about unfairness or bias in the evaluation process, suggesting that determination is not employed as a defensive response to perceived bias. Instead, women in the evaluation conditions demonstrated greater openness to feedback, which was subsequently associated with increased determination.

**4.3.6 Discussion**:

Experiment 3 indicates that the type of evaluation affects people’s psychological and behavioral responses. We found support for competitive evaluation’s effect on social anxiety, which does not occur when evaluation is developmental. It is developmental evaluation in particular that activates the emotional pathway of determination among women, and the consequent gender differences in effort and creativity.

**5. General Discussion**

Performance evaluation and monitoring is perhaps the most important human resource system in organizations (Judge & Ferris, 1993). Evaluation allows managers to focus employee efforts, measure productivity, and ensure their efforts align with organizational goals. Despite these potential performance enhancing effects, apprehension is an unintended psychological by-product of evaluation, whereby employees experience anxiety when being monitored and judged, which is detrimental to innovation performance. Because of these trade-offs the outcomes of evaluation, in both research and practice, are often inconsistent (Baas, et al., 2008; Baer & Oldham, 2006; Miller, 2007; Mueller, Melwani, & Goncalo 2012; Shalley & Perry-Smith, 2001), sometimes improving performance and at other times interfering with it.

We propose that these divergent outcomes can arise from the employee’s perceptions of evaluation, due to both individual differences and the nature of the evaluation process. In this paper, we utilize three laboratory experiments to study how gender moderates the relationship between evaluation and creativity. Experiment 1 replicates the classic evaluation apprehension findings, showing that evaluation leads to social anxiety among men and women. However, we find an intriguing effect among women, whereby they exhibit increased effort and therefore creativity following evaluation. Experiment 2 explores the emotion-to-behavior pathway whereby women experiencing evaluation respond with heightened determination, translating into more effort and creativity. Experiment 3 identifies a boundary condition, showing that this pathway emerges when women see evaluation as developmental--an opportunity to learn and get feedback--rather than competitive--a process that defines winners and losers. When evaluation is framed as competitive, both men and women exhibit social anxiety and reduce creativity. Taken together, these results indicate that people’s construal of evaluation shapes how they respond to it and its consequences for their performance.

Organizational design elements are not objective tools, they are subject to the interpretations of the employees who experience them. If managers do not appreciate the nuanced psychological mediation between the practice and its performance effects, they may deploy them with counterproductive results. A rich literature in organization theory and organization design has considered how and why policies misfire, and consider how they get enacted in organizations, and how they might have differing effects on different groups of employees. For example, Adler and Borys (1996) considered the differing facets of bureaucracy on line versus staff functions. Bernstein, (2012; 2017) discussed the effects of measures that induce greater transparency on the observer versus the observed. Barley (1986) considered the interactive effects of power and experience on developing routines after new technology adoption. As the workforce has become increasingly more diverse, and especially, female participation rates have dramatically increased, organization theory and strategy scholars have considered how gender diversity influences the outcomes to organizational policies. We follow this rich literature to consider how evaluation systems might impact specific groups of employees, in this case of different genders, by studying this question using a series of laboratory experiments.

**Empirical and Experimental Studies on Gender Differences in Competition**

Bonte and Piegeler (2013) conducted an empirical analysis across multiple countries, demonstrating that women are generally less competitively inclined than men. Their findings indicate that this reluctance significantly contributes to the gender gap in both latent and nascent entrepreneurship. The authors suggest that differences in risk-taking behavior and competitiveness are critical factors influencing women's entrepreneurial participation. Smilarly, Bonte and Jarosch (2011) explored the role of personality traits in entrepreneurship, identifying Individual Entrepreneurial Aptitude (IEA) as a crucial determinant. They found that women exhibit lower levels of IEA due to reduced, which contributes to the overall gender gap in entrepreneurship.

Horbach and Jacob (2018) provide insights into how female leaders often avoid competitive environments, preferring collaborative approaches that foster innovation. Their research highlights that complex innovation activities rely more on collaborative behavior, suggesting that women's empathetic and transformational leadership styles can enhance team dynamics and creativity. Eagly and Johannesen-Schmidt (2001) also emphasize that women’s leadership styles are often characterized by a greater focus on collaboration and inclusivity. Baer et al. (2014) (lab and field experiment evidence) found that the sex composition of teams profoundly affects creativity in competitive contexts. Their study indicates that while competition can stimulate creativity in male-dominated groups, it can be detrimental to female-dominated groups. They also emphasize the importance of collaboration, willingness to listen to others (can be related to openness to feedback) among women, which often leads to higher creative outputs compared to their male counterparts.

Croson and Gneezy (2009) reviewed the literature on gender differences in competitive preferences, concluding that women generally prefer non-competitive settings. Niederle and Vesterlund (2007) further illustrated this by demonstrating that only 35% of female participants opted for competitive scenarios, in stark contrast to 73% of male participants. The experimental work of Hogarth et al. (2012) highlighted that women are more likely to withdraw from competitive situation. Shurchkov (2012) showed that women’s likelihood of engaging in competition decreases under conditions of task stereotypes and time constraints, emphasizing the role of social context in shaping competitive behavior. Sutter and Rützler (2010) found similar results in their studies with children, showing that gender differences in competitiveness emerge early in life.

**Gender Differences in Grit**

Grit has been consistently linked to enhanced performance outcomes across various domains. Duckworth et al. (2007) defined grit as the combination of passion and perseverance toward long-term goals, which is crucial for achieving success. Numerous studies have found that grit positively correlates with academic achievement, engagement, and motivation (Bowman et al., 2015; Galla et al., in press). This relationship underscores the significance of grit in both academic and personal contexts.

Research on gender differences in grit presents mixed findings, although many studies indicate that women tend to exhibit higher levels of grit than men. For instance, Christensen and Knezek (2014) found a significant difference favoring females in grit scores among high school students, which aligns with findings from Rojas et al. (2012) and Batres (2011) that also report gender disparities in favor of women. However, not all studies agree; Hodge, Wright, and Bennett (2018) and Carpenter et al. (2018) reported no significant differences in grit scores between genders. Overall, while there are variations in findings, the trend of women displaying higher grit levels is noteworthy and deserves further exploration.

Moreover, research has suggested that mindset interacts with grit in meaningful ways, particularly among women. Sigmundsson et al. (2021) highlight that while passion and grit are significant predictors of achievement, their interplay may differ by gender. Specifically, they found a strong relationship among passion, grit, and mindset for women, suggesting that these attributes collectively contribute to their success. In contrast, for men, the correlation between passion and grit was more pronounced, indicating that passion may play a more critical role in their achievements while growth mindset may be more important for women (Sigmundsson et al., 2020b). Notably, studies indicate that for women, both mindset and grit are strongly associated, enhancing their academic performance, whereas this relationship is less pronounced for men. This observation suggests that a growth mindset, which emphasizes self-improvement and the value of feedback, may be particularly beneficial for women who also exhibit higher grit scores. This synergy between a growth mindset and grit aligns with our findings, indicating that supportive evaluation and constructive feedback may foster greater potential and performance among women compared to their male peers.

One key limitation of this study is its external validity, as the experimental setting may not fully capture the complexities of real-world evaluation scenarios. We acknowledge this limitation but focused on tightly controlled experimental conditions to isolate the effects of evaluation on creativity. We also included tasks that reflect real-world business problems to increase realism. Future research could explore these effects in more naturalistic settings to enhance generalizability. Additionally, we were unable to test idea sharing as a separate outcome due to its overlap with creative performance, potentially complicating the interpretation of results. We concentrated on creativity as a primary outcome and suggest that future studies examine idea sharing independently to provide a clearer picture of its relationship with evaluation. Finally, we have focused on women’s openness to developmental forms of evaluation as one mechanism. We suspect that there are multi-dimensional causal processes at play. Future studies might highlight threat, particularly stereotype threat in women, and this could elicit different psychological dynamics. For instance, women facing these threats could exhibit more anxiety-based performance losses or could exhibit forms of determination derived from reactance—proving that the negative stereotype is wrong. **Theoretical implications**

We consider three implications of this research, and avenues for further research. First, we describe Evaluation Resilience as a novel and understudied counterpoint to classic findings of Evaluation Apprehension, well-established over decades. Second, we consider Evaluation Resilience as a uniquely female form of responsiveness, that contrasts with prevailing findings that women underperform in competitive and other evaluative contexts. Finally, we consider implications for creativity research, suggesting that certain forms of evaluation might facilitate idea generation, and creativity.

*Evaluation as a motivational tool, not simply an instrument of control*

For decades, researchers have observed the negative psychological by-products of evaluation, while its positive psychological by-products have remained obscured. One reason why this might be is because of cultural values of independence. In a culture that values independence, evaluation is intrusive and controlling, undermining one’s freedom.

However, people are motivated by values other than just freedom. Performance, especially in creative work, is ambiguous, and people suffer without clarity of expectations and a clear path to progress and improvement. These motivations may be stronger among women and minorities, who are particularly likely to find themselves subjected to ambiguous and moving standards of evaluation.

While apprehension is the well-established dark side of evaluation, evaluation resilience is its potential bright side, allowing people to gain a sense of clarity, purpose, and determination. This research opens ways to see evaluation as a motivational tool, not simply as a tool to control, judge, and monitor employees. Rather than assuming that evaluation is inherently threatening, managers can design evaluation practices that generate positive psychological forces that motivate employees, reframing a practice that judges work into one that builds and develops people. When evaluation provides clarity, feedback, and milestones that capture progress, there may be greater alignment between managerial goals and employee responses. While there is much managerial concern about evaluation as intrusive surveillance, we see much potential for future research to capture the motivating potential of evaluation.

*Gender and a novel emotional pathway to resilience*

Second, extant research is laden with the assumption that women crumble in the face of evaluation. For instance, stereotype threat research finds that when women have to highlight gender identity prior to taking a math test, their performance suffers, due to anxiety about stereotypes that women perform worse at math (Shih, Pittinsky & Ambady, 1999; Specer, Steele & Quinn, 1999). Other research shows that women perform worse in competitive environments (Gneezy, 2003; Baer, 2014). The damning implication is that women may not be equipped to survive the evaluative pressure cooker. By contrast, our research shows a clear process by which women can thrive in evaluation, and indeed outperform men. By reframing the purposes and processes by which managers design evaluation, a very different image of female resilience emerges. While our research has focused on gender as an individual difference, the implication is that organizational designers might identify distinctive motivational pathways as they engage with diverse employees.

While these findings offer a sanguine view about the creative potential of women at work, we note a more sobering interpretation as well. A reason why women may find clear evaluation more motivating is because they are typically subjected to regimes with unclear, unfair, or limited feedback. These problems are also pertinent for minority employees (Hoyt, et al., 2006). People who anticipate bias may feel most motivated by structured processes that counteract it. Additionally, determination may also derive from bias, and the desire to prove one’s worth in the face of negative assumptions.

*Evaluation and idea generation*

Finally, the creative process is often separated into idea generation and idea selection phases. Evaluation dominates the idea selection phase, but it is typically segregated from the idea generation phase, where it could constrain and limit the pool of ideas people generate. Our research suggests that evaluation could, if carefully deployed, produce a positive, clarifying force that can focus idea generation. We hope that future research considers how evaluation is not simply a force that undermines creative idea generation, but potentially part of a system that could facilitates innovative thinking.

*Conclusion*

Technology has allowed managers to become ever more adept at forms of evaluation that measure, control and monitor employees. The question is whether employee reactions will prevent these anticipated performance gains from being realized. While a direct line between organizational design and performance is often assumed, this paper disambiguates the dual mediating psychological processes. By recognizing both of these emotional pathways, evaluation becomes, not just an anxiety-provoking, performance inhibiting threat, but a potential motivator as people pursue creative work. Our hope is that even a fraction of the amount invested into these performance-inhibiting forms of evaluation will be channeled into designing evaluation processes that unleash employee creativity.

**Appendix 1:** Questions participants answered in study 1.

**1) Leadership Prompt**

You are the CEO of a car company. Oil prices are increasing, and your firm has been under regulatory pressure to become cleaner. Electric car firms like Tesla have been taking up market share, while alternative forms of transportation have become more popular.

How would you approach the situation? What would your plan be in this scenario?

**2) Marketing Prompt:**

You are a marketing manager for a large fast food restaurant. Recently, food delivery services such as GrubHub and DoorDash have started gaining more Millennial and Generation Z customers at your expense.

How would you approach the situation? What would your marketing plan be in this scenario?

**3) Innovation Prompt:**

You are an innovation manager for a phone manufacturer start-up. Innovation has slowed down, and most phones have very similar features. You are under competitive pressure from larger manufacturers and most profits are going to the software and app designers.

How would you approach the situation? What would your innovation plan be in this scenario?

**Figure 1:** example of the content that will be shared with experts



**Figure 2:** Differences in effort between evaluated (1) and control (0) groups, based on gender (Experiment 1).



**Figure 3:** Differences in creativity between evaluated and control groups, based on gender (Experiment 1).



**Figure 4:** Differences in determination between evaluated and control groups, based on gender (Experiment 2).



**Appendix 2:** Social anxiety scale used in Study 2 and 3.

Please indicate the extent that you are feeling this way **at this moment.**

I am worried about what other people think of me.

I am afraid other people will notice my shortcomings.

I am afraid that others won't approve of me.

I am worried that I say or do the wrong things.

I am worried about looking foolish.

**Table 1:** Summary Statistics and Correlation Matrix – Study 1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|   Variables  |  Mean  |  Std. Dev.  |   (1)  |   (2)  |   (3)  |   (4)  |   (5)  |
|  (1) Social Anxiety  | 2.149  | .745  | 1.000  |   |   |   |   |
|  (2) Manipulation Check  | 4.788  | 1.621  | 0.106  | 1.000  |   |   |   |
|  (3) Effort  | 106.21  | 36.30  | -0.010  | 0.101  | 1.000  |   |   |
|  (4) Creativity  | 2.458  | .684  | -0.134  | -0.113  | 0.181  | 1.000  |   |
|  (5) Survey Duration  | 476.174  | 141.66  | 0.006  | -0.062  | -0.168  | 0.040  | 1.000  |
|   |   |   |

**Table 2:** PROCESS analysis results of Study 1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Social Anxiety |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$1.60^{\*\*\*}$$ | 0.54 |  | $$182.59^{\*\*\*}$$ | 27.31 |  | $$2.51^{\*\*\*}$$ | 0.52 |
| *Evaluation* | 0.14 | 0.08 |  | $$8.45^{\*}$$ | 4.54 |  | $$-0.16^{\*\*}$$ | 0.08 |
| *Social Anxiety* |  |  |  |  |  |  | $$-0.11^{\*\*}$$ | 0.05 |
| *Effort* |  |  |  |  |  |  | $$0.004^{\*\*\*}$$ | 0.001 |
| *Moderator - Female* |  |  |  | -2.52 | 5.57 |  |  |  |
| *Int. – Female X Eval.* |  |  |  | $$15.46^{\*\*}$$ | 7.86 |  |  |  |
| *Survey Duration* | 0.00 | 0.00 |  | $$-0.06^{\*\*\*}$$ | 0.02 |  | -0.00 | 0.00 |
| Direct & Indirect effects |  |  |  |  |  |  | $$β$$ | 95% CI |
| Direct effect - Evaluation |  |  |  |  |  |  | -0.16 (0.08) | [-0.31, -0.01] |
| Indirect - Social Anxiety |  |  |  |  |  |  | -0.01 (0.01) | [-0.04, 0.00] |
| Indirect - Effort |  |  |  |  |  |  | 0.05 (0.02) | [0.01, 0.09] |
| Index of Mod. Mediation |  |  |  |  |  |  | $$β$$ | 95% CI |
| Effort through Gender |  |  |  |  |  |  | 0.06 (0.03) | [0.00, 0.12] |

N.B. N = 396. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

**Table 3:** Summary Statistics and Correlation Matrix – Study 2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   Variables  |  Mean  |  Std. Dev.  |   (1)  |   (2)  |   (3)  |   (4)  |   (5)  |   (6)  |   (7)  |   (8)  |   (9)  |   (10)  |   (11)  |   (12)  |   (13)  |   |
|  (1) Social Anxiety  | 2.455  | .944  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |   |   |
|  (2) Manipulation Check  | 1.451  | .499  | -0.107  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |   |
|  (3) Effort  | 98.08  | 40.294  | -0.006  | -0.012  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |
|  (4) Creativity  | 2.515  | .587  | 0.019  | -0.042  | 0.295  | 1.000  |   |   |   |   |   |   |   |   |   |   |
|  (5) Duration  | 658.643  | 107.54  | 0.030  | 0.053  | -0.258  | -0.123  | 1.000  |   |   |   |   |   |   |   |   |   |
|  (6) PANAS Anxiety  | 2.183  | 1.147  | 0.474  | -0.027  | 0.058  | 0.039  | 0.065  | 1.000  |   |   |   |   |   |   |   |   |
|  (7) PANAS Competitive  | 2.326  | 1.237  | 0.102  | -0.130  | 0.090  | 0.057  | -0.005  | 0.293  | 1.000  |   |   |   |   |   |   |   |
|  (8) PANAS Distressed  | 1.848  | 1.013  | 0.418  | -0.095  | 0.118  | -0.006  | 0.028  | 0.614  | 0.355  | 1.000  |   |   |   |   |   |   |
|  (9) PANAS Excited  | 2.304  | 1.17  | -0.062  | -0.013  | 0.074  | -0.034  | 0.030  | 0.049  | 0.377  | 0.156  | 1.000  |   |   |   |   |   |
|  (10) PANAS Determined  | 2.683  | 1.141  | -0.001  | -0.008  | 0.117  | 0.042  | -0.017  | 0.116  | 0.544  | 0.179  | 0.660  | 1.000  |   |   |   |   |
|  (11) PANAS Scared  | 1.491  | .878  | 0.409  | -0.098  | -0.071  | -0.029  | 0.104  | 0.578  | 0.302  | 0.538  | 0.208  | 0.187  | 1.000  |   |   |   |
|  (12) PANAS Nervous  | 1.951  | 1.133  | 0.517  | -0.088  | -0.067  | 0.028  | 0.054  | 0.748  | 0.331  | 0.607  | 0.099  | 0.144  | 0.714  | 1.000  |   |   |
|  (13) PANAS Interested  | 2.598  | 1.011  | 0.098  | -0.102  | -0.026  | -0.034  | 0.187  | 0.064  | 0.335  | 0.010  | 0.490  | 0.538  | 0.153  | 0.081  | 1.000  |   |
|   |   |   |

**Table 4a:** PROCESS Analysis Results for Study 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Determination |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$3.17^{\*\*\*}$$ | 0.49 |  | $$149.48^{\*\*\*}$$ | 17.77 |  | $$2.98^{\*\*\*}$$ | 0.29 |
| *Evaluation* | $$-0.43^{\*\*}$$ | 0.20 |  | 1.50 | 5.21 |  | 0.00 | 0.07 |
| *Determination* |  |  |  | $$4.01^{\*}$$ | 2.29 |  | 0.01 | 0.03 |
| *Effort* |  |  |  |  |  |  | $$0.003^{\*\*\*}$$ | 0.001 |
| *Moderator – Female* | $$-0.68^{\*\*\*}$$ | 0.21 |  |  |  |  |  |  |
| *Int. – Female X Eval.* | $$0.77^{\*\*}$$ | 0.31 |  |  |  |  |  |  |
| *Survey Duration* | -0.00 | 0.00 |  | $$-0.10^{\*\*\*}$$ | 0.02 |  | -0.00 | 0.00 |
| Index of Mod. Mediation |  |  |  |  |  |  | $$β$$ | $$SE\_{β}$$ |
| Through Gender |  |  |  |  |  |  | 0.011 (0.008) | [0.000, 0.030] |

N.B. N = 224. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

\* p <0.10, \*\* p < 0.05, \*\*\* p < 0.01

**Table 4b–** Replication using the Study 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Social Anxiety |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$2.11^{\*\*\*}$$ | 0.39 |  | $$160.33^{\*\*\*}$$ | 17.29 |  | $$2.99^{\*\*\*}$$ | 0.29 |
| *Evaluation* | $$0.27^{\*\*}$$ | 0.13 |  | 2.54 | 7.06 |  | -0.00 | 0.08 |
| *Social Anxiety* |  |  |  |  |  |  | 0.00 | 0.04 |
| *Effort* |  |  |  |  |  |  | $$0.003^{\*\*\*}$$ | 0.001 |
| *Moderator - Female* |  |  |  | 1.97 | 7.43 |  |  |  |
| *Int. – Female X Eval.* |  |  |  | -3.10 | 10.61 |  |  |  |
| *Survey Duration* | 0.00 | 0.00 |  | $$-0.10^{\*\*\*}$$ | 0.02 |  | -0.00 | 0.00 |
| Direct & Indirect effects |  |  |  |  |  |  | $$β$$ | 90% CI |
| Direct effect - Evaluation |  |  |  |  |  |  | -0.00 (0.08) | [-0.15, 0.15] |
| Indirect - Social Anxiety |  |  |  |  |  |  | 0.00 (0.01) | [-0.02, 0.03] |
| Indirect - Effort |  |  |  |  |  |  | 0.00 (0.02) | [-0.04, 0.04] |
| Index of Mod. Mediation |  |  |  | $$β$$ | 95% CI |  |  |  |
| Through Gender |  |  |  | 0.00 (0.01) | [-0.02, 0.02] |  |  |  |

N.B. N = 224. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

\* p <0.10, \*\* p < 0.05, \*\*\* p < 0.01

**Appendix 3:** Manipulations and Amplifiers in Study 3

Participants in competitive condition received the following manipulation, followed by the picture we used in our manipulation in Study 1 (Figure 1) and the follow up question to emphasize the manipulation:

“Before we begin, please note that the feasibility and originality of your proposals will be **evaluated by experts to compare you against all other participants** in this class and **rank your creative potential**.”

followed by

“Think about a situation where your performance was evaluated for the purpose of comparing you against others. What comes to mind? Consider how it made you feel. Write a couple of sentences about it.”

Whereas participants in developmental condition received the following prompt:

“Before we begin, please note that the feasibility and originality of your proposals will be **evaluated by experts to provide you with developmental feedback** and **grow your creative potential.”**

followed by

“Think about a situation where your performance was evaluated for the purpose of providing you with developmental feedback. What comes to mind? Consider how it made you feel. Write a couple of sentences about it.”

Lastly, the control group received the following prompt:

“Before we begin, please note that this task is electronic and **all you need is the electronic device** you are presently using to complete the task.”

followed by

“Think about your typical routine in the morning. What comes to mind? Consider how it makes you feel. Write a couple of sentences about it.”

**Table 5:** Manipulation Checks – Study 3

|  |  |  |
| --- | --- | --- |
|   | Sample   | Difference in means  |
| Variable   | Full  | Control   | Competitive Evaluation  | Developmental Evaluation   | Competitive vs. Control   | Developmental vs. Control   | Developmental vs. Competitive  |
| General evaluation manipulation check  | 0.85  | 0.64  | 0.98  | 0.93  | 0.342\*\*\*  | 0.295\*\*\*  | -0.047†  |
| (0.358)  | (0.483)  | (0.137)  | (0.249)  | (0.049)  | (0.052)  | (0.028)  |
| Competitive manipulation check  | 0.69  | 0.5  | 0.97  | 0.6  | 0.462\*\*\*  | 0.095  | -0.368\*\*\*  |
| (0.462)  | (0.502)  | (0.167)  | (491)  | (0.051)  | (0.068)  | (0.05)  |
| Developmental manipulation check   | 0.55  | 0.3  | 0.48  | 0.88  | 0.185\*\*  | 0.59\*\*\*  | 0.406\*\*\*  |
| (0.498)  | (0.458)  | (0.502)  | (0.318)  | (0.066)  | (0.054)  | (0.058)  |

**Table 6:** Summary Statistics and Correlation Matrix – Study 3

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   Variables  |  Mean  |  Std. Dev.  |   (1)  |   (2)  |   (3)  |   (4)  |   (5)  |   (6)  |   (7)  |   (8)  |   (9)  |   (10)  |   (11)  |   (12)  |   (13)  |   (14)  |   (15)  |
|  (1) Social Anxiety  | 2.384  | .93  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|  (2) Mani. Check 1  | .85  | .358  | 0.025  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |   |   |
|  (3) Mani. Check 2  | .694  | .462  | 0.004  | 0.499  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |   |
|  (4) Mani. Check 3  | .553  | .498  | -0.09  | 0.362  | 0.153  | 1.000  |   |   |   |   |   |   |   |   |   |   |   |
|  (5) PANAS Anxious  | 1.994  | 1.059  | 0.485  | 0.039  | 0.099  | -0.047  | 1.000  |   |   |   |   |   |   |   |   |   |   |
|  (6) PANAS Competitive  | 2.375  | 1.28  | 0.055  | 0.116  | 0.211  | 0.121  | 0.166  | 1.000  |   |   |   |   |   |   |   |   |   |
|  (7) PANAS Distressed  | 1.694  | .923  | 0.390  | 0.031  | 0.096  | 0.042  | 0.556  | 0.235  | 1.000  |   |   |   |   |   |   |   |   |
|  (8) PANAS Excited  | 2.266  | 1.051  | -0.085  | 0.040  | 0.104  | 0.048  | 0.078  | 0.441  | 0.110  | 1.000  |   |   |   |   |   |   |   |
|  (9) PANAS Determined  | 2.947  | 1.12  | -0.049  | 0.074  | 0.053  | 0.115  | 0.108  | 0.517  | 0.142  | 0.590  | 1.000  |   |   |   |   |   |   |
|  (10) PANAS Scared  | 1.45  | .821  | 0.280  | 0.006  | 0.026  | -0.036  | 0.443  | 0.238  | 0.558  | 0.213  | 0.200  | 1.000  |   |   |   |   |   |
|  (11) PANAS Nervous  | 1.85  | 1.012  | 0.426  | -0.019  | 0.015  | -0.040  | 0.625  | 0.189  | 0.565  | 0.161  | 0.151  | 0.628  | 1.000  |   |   |   |   |
|  (12) PANAS Interested  | 2.994  | 1.023  | -0.041  | 0.066  | -0.004  | 0.038  | 0.084  | 0.289  | 0.054  | 0.503  | 0.448  | 0.112  | 0.111  | 1.000  |   |   |   |
|  (13) Effort  | 92.35  | 33.223  | 0.254  | 0.021  | -0.011  | -0.025  | 0.150  | 0.101  | 0.181  | 0.027  | 0.135  | 0.110  | 0.162  | 0.088  | 1.000  |   |   |
|  (14) Creativity  | 2.433  | .749  | 0.183  | -0.006  | -0.008  | -0.011  | 0.204  | 0.097  | 0.195  | 0.063  | 0.023  | 0.127  | 0.169  | 0.074  | 0.346  | 1.000  |   |
|  (15) Duration  | 806.747  | 131.022  | 0.045  | 0.064  | 0.018  | 0.148  | 0.057  | 0.036  | 0.050  | 0.005  | 0.064  | -0.037  | 0.031  | 0.117  | -0.081  | 0.084  | 1.000  |
|   |   |   |

 **Experiment 3–** Competition vs All --> Social Anxiety --> Creativity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mediator |  |  | Outcome |
|  | Social Anxiety |  |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$1.77^{\*\*\*}$$ | 0.39 |  |  | $$1.64^{\*\*\*}$$ | 0.34 |
| *Competitive Evaluation* | $$0.22^{\*}$$ | 0.13 |  |  | 0.02 | 0.11 |
| *Social Anxiety* |  |  |  |  | $$0.12^{\*\*}$$ | 0.06 |
| *Survey Duration* | 0.001 | 0.001 |  |  | 0.001 | 0.000 |
| Direct & Indirect effects |  |  |  |  | $$β$$ | 95% CI |
| Direct effect - Evaluation |  |  |  |  | 0.02 (0.11) | [-0.19, 0.23] |
| Indirect - Social Anxiety |  |  |  |  | 0.03 (0.02) | [-0.01, 0.09] |

N.B. N = 214. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

**Experiment 3–** Feedback vs Control: Serial Mediation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Determination |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$3.10^{\*\*\*}$$ | 0.48 |  | $$91.91^{\*\*\*}$$ | 14.46 |  | $$1.46^{\*\*\*}$$ | 0.34 |
| *Feedback Evaluation* | 0.02 | 0.19 |  | 5.34 | 4.48 |  | -0.07 | 0.10 |
| *Determination* |  |  |  | $$5.14^{\*\*\*}$$ | 1.97 |  | -0.04 | 0.04 |
| *Effort* |  |  |  |  |  |  | $$0.007^{\*\*\*}$$ | 0.001 |
| *Moderator – Female* | $$-0.52^{\*\*}$$ | 0.22 |  |  |  |  |  |  |
| *Int. – Female X Eval.* | $$0.60^{\*}$$ | 0.32 |  |  |  |  |  |  |
| *Survey Duration* | -0.00 | 0.00 |  | -0.02 | 0.02 |  | 0.00 | 0.00 |
| Index of Mod. Mediation |  |  |  |  |  |  | $$β$$ | $$SE\_{β}$$ |
| Through Gender |  |  |  |  |  |  | 0.022 (0.017) | [-0.001, 0.064] |

N.B. N = 214. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

\* p <0.10, \*\* p < 0.05, \*\*\* p < 0.01

**Experiment 3–** Feedback vs Competition: Serial Mediation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Determination |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$2.67^{\*\*\*}$$ | 0.51 |  | $$111.21^{\*\*\*}$$ | 15.68 |  | $$1.12^{\*\*\*}$$ | 0.36 |
| *Feedback Evaluation* | -0.13 | 0.19 |  | -0.27 | 4.63 |  | -0.09 | 0.10 |
| *Determination* |  |  |  | $$3.58^{\*}$$ | 2.04 |  | -0.01 | 0.04 |
| *Effort* |  |  |  |  |  |  | $$0.009^{\*\*\*}$$ | 0.001 |
| *Moderator – Female* | $$-0.57^{\*\*}$$ | 0.23 |  |  |  |  |  |  |
| *Int. – Female X Eval.* | $$0.64^{\*}$$ | 0.33 |  |  |  |  |  |  |
| *Survey Duration* | 0.00 | 0.00 |  | -0.03 | 0.02 |  | 0.00 | 0.00 |
| Index of Mod. Mediation |  |  |  |  |  |  | $$β$$ | $$SE\_{β}$$ |
| Through Gender |  |  |  |  |  |  | 0.021 (0.018) | [-0.001, 0.064] |

N.B. N = 212. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

\* p <0.10, \*\* p < 0.05, \*\*\* p < 0.01

**Experiment 3–** Feedback vs All: Serial Mediation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mediators |  | Outcome |
|  | Determination |  | Effort |  | Creativity |
|  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |  | $$β$$ | $$SE\_{β}$$ |
| Intercept | $$2.83^{\*\*\*}$$ | 0.40 |  | $$98.67^{\*\*\*}$$ | 12.22 |  | $$1.20^{\*\*\*}$$ | 0.29 |
| *Feedback Evaluation* | -0.06 | 0.16 |  | 2.62 | 3.95 |  | -0.09 | 0.08 |
| *Determination* |  |  |  | $$4.11^{\*\*}$$ | 1.65 |  | -0.02 | 0.04 |
| *Effort* |  |  |  |  |  |  | $$0.008^{\*\*\*}$$ | 0.001 |
| *Moderator – Female* | $$-0.54^{\*\*\*}$$ | 0.16 |  |  |  |  |  |  |
| *Int. – Female X Eval.* | $$0.61^{\*\*}$$ | 0.27 |  |  |  |  |  |  |
| *Survey Duration* | 0.00 | 0.00 |  | $$-0.02^{\*}$$ | 0.01 |  | 0.00 | 0.00 |
| Index of Mod. Mediation |  |  |  |  |  |  | $$β$$ | $$SE\_{β}$$ |
| Through Gender |  |  |  |  |  |  | 0.021 (0.013) | [0.000, 0.052] |

N.B. N = 320. Gender was coded as 1 = Female, 0 = Male. CI, Confidence interval. Coefficients are presented as unstandardized estimates. Standard errors (SEs) are in parentheses.

\* p <0.10, \*\* p < 0.05, \*\*\* p < 0.01

1. Creativity involves both idea generation and idea selection, so the role of evaluation may differ across these stages (Keum & Sea, 2017). While evaluation disrupts divergent thinking by inducing self-censorship in idea generation phase (Amabile, 1983), it is an essential part of the idea selection, to identify the most viable ideas. Our focus here is on carefully managing the idea generation phase, and the intrusion of these detrimental evaluative processes on creativity. [↑](#footnote-ref-1)