Paper ID #37763

Investigation of a professor's feedback on student's divergent thinking performance: an electrodermal activity experiment

Elif Elcin Gunay

Elif Elçin Günay is an assistant professor in the Department of Industrial Engineering at Sakarya University, Turkey. She received her B.S.E., M.S.E., and Ph.D. degrees in Industrial Engineering from Sakarya University, Turkey, in 2007, 2009, and 2016 respectively. She worked in the Department of Industrial and Manufacturing Systems Engineering at Iowa State University as a post-doctoral research associate between 2017-2018. Her research interest includes optimization in manufacturing and service systems, stochastic processes, and engineering education. Her recent research interests focus on enhancing creativity in engineering classrooms.

Danielle S. Dickson

Danielle Dickson received her PhD from the University of Illinois at Urbana-Champaign in 2016 with a dissertation examining the memory system's representation of numerical information, using behavioral and electro-physiological (EEG, brainwaves) measures. She extended this work into comparisons of children and adults' arithmetic processing as a postdoctoral scholar at The University of Texas San Antonio. Her most recent research examines creative thinking processes as an area of postdoctoral research at The Pennsylvania State University.

Rafal Jonczyk (Dr)

Janet Van Hell (Professor of Psychology and Linguistics)

Zahed Siddique (Professor)

Zahed Siddique is a Professor in the School of Aerospace and Mechanical Engineering at the University of Oklahoma.

Gul E. Okudan Kremer (Wilkinson Professor and Senior Director)

Gül E. Kremer is Dean-elect of Engineering at University of Dayton. Kremer served as chair of the Department of Industrial and Manufacturing Systems Engineering (2016-2021) and Senior Director Presidential Projects (2021-2022), in addition to past leadership roles at Penn State. Dr. Kremer has degrees in industrial engineering from Yildiz Technical University, a masters in business from Istanbul University, and a PhD in Engineering Management from Missouri University of Science and Technology. She was a National Research Council-US AFRL Summer Faculty Fellow in the Human Effectiveness Directorate (2002-2004), a Fulbright Scholar (2010-2011), and Program Director in NSF's Division of Undergraduate Education (2013-2016). Dr. Kremer's research interests include applied decision analysis to improve complex products and systems, and engineering education. Her research has appeared in 3 books and over 360 refereed publications. She is a Fellow of the American Society for Mechanical Engineers and senior member of the Institute of Industrial Systems Engineers. In addition, she has significant contributions to research efforts that are directed toward improving engineering education.

Investigation of a professor's feedback on student's divergent thinking performance: an electrodermal activity experiment

Abstract

Throughout their education, college students receive feedback about their performance from professors, who are experts in their field. The way this feedback is expressed can influence students' future performances. One theory is that feedback from an authority figure invoking negative gender stereotypes, even unintentionally, might negatively affect performance. In this study, we aimed to investigate how an authority figure's feedback affects divergent thinking in male and female industrial engineering students. We targeted industrial engineers because of the relatively high gender balance in their student population. The divergent thinking abilities of the students were measured with a two-phase test consisting of the alternate uses task (AUT) and the utopian situations task (UST), with ideational fluency (number of ideas produced) as the critical output measure. Students were asked to complete both tasks while their electrodermal activity (EDA) was recorded, a biological measure that is thought to reflect engagement. The students' divergent thinking abilities and electrodermal levels were then compared before and after two forms of feedback: positive and negative (stereotype threat). Results showed that the number of ideas generated was significantly decreased after negative feedback. However, no significant change in ideation fluency occurred after the positive feedback delivery. There was no significant task-related EDA change under positive and negative feedback interventions. These results demonstrate that this type of research can contribute critical new information for educators on how to provide more effective feedback regarding student task performance.

1. Introduction

A stereotype is an overgeneralized and oversimplified common belief for a particular group of individuals that may or may not reflect the truth [1]. Any identity such as race, ethnic origin, color, or gender can be subject to stereotyping. Relating to stereotypical perspectives (e.g., Asians are good at math [2], African Americans have greater athletic ability [3]) can have a positive effect on an individual's behavior if they are encouraging and promising, but negative if they are unfavorable and unacceptable, e.g., Latinos perform poorly on academic performance tasks [4], women have weaker math ability [5]. These negative stereotypes can cause a threat. Spencer, Logel, and Davies describe stereotype threat as a situation that occurs when individuals are concerned about being evaluated or treated unfairly/negatively because of negative stereotypes related to their identity [6]. In this study, we consider gender-based stereotypes, including fixed beliefs about the general characteristics and behaviors expected from men and women. Many studies in the literature show lower performance by individuals in academic tests under sexist negative stereotypes conditions, e.g. [3], [7].

One of the prevalent negative stereotypes for women is the ascription of creativity and innovation to male gender, but not, or less so, to female gender. For example, Kabat-Farr and Cortina [8] reported that women's innovative work behavior was less valued than their male colleagues. Proudfoot et al. [9] also showed that the same ideas were rated more highly when they came from a man than from a woman. In an educational setting, what effect might this negative stereotype, i.e., the ascription of creativity to men, have on female students, especially when delivered by a professor? This potentially causes a stereotype threat situation for female students in which they feel negative pressure, stress, to disconfirm the negative stereotype while responding to a professor. Professors serve as authority figures and role models for students; they set an example for students' professional life, promote passion for learning, innovating, and succeeding [10]. Moreover, their culpability in the spread of stereotypes should not be underestimated. Gunderson et al. [11] demonstrated how negative stereotypes about women's math capabilities are passed down to girls by parents and instructors. Being aware of the potential of their mistakes may lead to generalizations for all women. Consequently, women - students and professionals alike - may experience distracting pressure that may increase their cognitive load [6]. Such increased cognitive load limits people's engagement [12], i.e., their attentional and emotional involvement with the task [13]. This type of engagement is vital for students to innovate [14], [15] and to think creatively [16].

Muldner and Burleson [17] showed that students' engagement differs between low and high creative students. Likewise, Koch et al., [18] and Eldor and Harpaz [19] reported that workers employ higher cognitive flexibility, which leads to creative performance when they are engaged. Reid and Solomonides [20] found that engagement and creativity support each other for student learning in design. Taken together, stereotype threat might then pose a challenge to creativity performance through distraction and disengagement on the critical task. The present study, therefore, aims to examine the relationship between stereotype threat, engagement, and creativity. More specifically, our objective is to investigate whether stereotype threat delivery impacts student's engagement, and whether that is linked to differentiation in student's divergent thinking performance.

Our approach adopts the definition of creativity under a divergent thinking perspective, which focuses on the generation of original and diverse options and ideas. According to Guilford [21], divergent thinking, *a.k.a. divergent production*, provides a potential reliable assessment of creative thinking. This does not mean that divergent thinking necessarily captures or reflects all aspects of creative thinking, but divergent thinking leads to original thinking, and originality is at the center of creativity [22]. Consistent with this line of thinking, Plucker [23] found that high scores on divergent thinking tasks were linked to real-life future performance, such as granting patents, inventions, creating new businesses or organizations. Due to its dominant use [24, 25], and its link to real-life creative behavior, we used divergent thinking tasks to assess the creative thinking potential of the engineering students.

We measured students' electrodermal activity (EDA) as an indicator of their engagement with the divergent thinking tests. EDA is used to express the change in the electrical properties of the skin. It is associated with autonomic, emotional, and cognitive processing and provides instant physiological feedback by capturing the skin's variation in electrical properties. As such, EDA is closely related to emotion, arousal, and attention. Recent research has used EDA as an index of attentional processing, where salient stimuli led to increased EDA levels. Therefore, it has been used as an objective proxy measure of engagement [26].

In our experimentation, male and female industrial engineering students completed two divergent thinking tasks (AUT, UST). Their EDA levels during experimentation and their responses for each task were recorded. In order to have a better understanding of how stereotype threat impacts students' engagement with the task, halfway through the experiment, some students received positive feedback while others received negative feedback, which we hypothesize presents a stereotype threat for female students when coming from a male professor. Our intention was to observe if stereotype threat lowered student's engagement with the task (as measured by EDA) and creativity (as measured in ideational fluency) in the second half of the experiment. Likewise, we examined whether positive feedback might lead to higher engagement with the task and, consequently, higher ideational fluency. After completion of all the divergent thinking tasks, a survey was conducted to understand how students perceived the feedback. In sum, students' divergent thinking performance and engagement levels were compared pre-and post-feedback intervention with the aim to reveal how stereotype threat affects student engagement and creative performance.

2. Method

2.1 Participants

Five female and six male senior industrial engineering students consented to participate in the study, which was approved by Iowa State University's Institutional Review Board (Appendix 2). The study was conducted in English; the participants were highly proficient speakers of English. At the end of the study, participants were debriefed about the experiment and the feedback from a male professor. Critically, participants were informed that the feedback wasn't real. Prior to the experiment, all participants signed an informed consent form.

2.2 Materials

We used two divergent think tests: Alternate uses task (AUT) and Utopian situation task (UST). In AUT, first, we exemplified one common and one alternate use of an object, e.g., a shoe used for wearing is a common use; a shoe used as a plant pot is an alternate use. Then, we asked attendees to list as many possible alternate uses for the experimental items as they could. In the UST, participants were asked to put themselves in a hypothetical situation and to generate unusual solutions and ideas to the presented problems, e.g., "What would happen if an ice-age suddenly occurred?".

2.3 Procedure

Our study included two phases (see Figure 1). In the first phase, participants underwent an eligibility test to determine whether fluctuations in their EDA levels were detectable in response to a range of stimuli. We challenged students with questions that included puzzles, rotation of 2-D shapes, and finding the differences between two similar shapes. The set of eligibility questions can be seen in Gunay et al. [27]. Task-related EDA changes exceeded 0.05 μS [26] in all participants; hence, they were invited to the second phase of the study.

In the second phase, participants were exposed to two sets of the AUT and two sets of the UST. Both AUT and UST sets consisted of four items, with two minutes of idea generation per item. After completing the AUT and UST tasks, students rested for 5 minutes. At the end of the break, participants were exposed to either positive or negative feedback. According to stereotype threat paradigms, victims of the stereotypes are individuals that belong to negatively stereotyped identities [6]. Therefore, in our study female students were subjected to stereotype threat. In order to magnify the impact of stereotype threat, a male professor was trained to express both types of feedbacks. Depending on the gender of the student and the type of intervention, different scripts were used by the same male professor (see Table 1).

Randomly selected three females received negative feedback (stereotype threat), and two females received positive feedback. Six male students served as control subjects; four male students received negative feedback and two males received positive feedback. Control participants received the same treatment.

Negative feedback that included stereotype threat was adapted from [28], while positive feedback was modeled after the work of [5], [28], [29]. After delivering feedback (negative or positive), participants completed the second half of the AUT and UST, each consisting of four trials. The AUT and UST trials before and after feedback intervention, represented as AUT_{pre}, UST_{pre}, AUT_{post}, UST_{post}, were included in Table A.1 in Appendix 1. The prompts were presented in a fixed order across all participants. Having finished the main experiment, participants also completed an adapted Stereotype Vulnerability Scale (SVS) questionnaire to assess the participant's pre-existing vulnerability to stereotype threat [5].

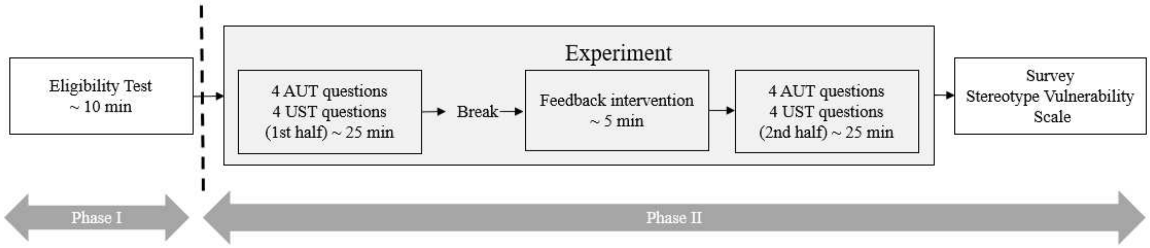

Figure 1. The flowchart of the experimental plan

Table 1. Negative and positive feedback scenarios

Gender of the student	Negative feedback scenario	Positive feedback scenario
Female	*We are looking at how you're doing. What we've been seeing so far is that **women** in particular are really struggling with this task, so please try to do the task to the best of your ability after the break.*	*We are looking at how you're doing. So, after the break, try to do the task to the best of your ability. What we've been seeing so far is that **women** in particular are really doing well with this task.*
Male	*We are looking at how you're doing. What we've been seeing so far is that **men** in particular are really struggling with this task, so please try to do the task to the best of your ability after the break.*	*We are looking at how you're doing. So, after the break, try to do the task to the best of your ability. What we've been seeing so far is that **men** in particular are really doing well with this task.*

2.4 Data collection procedure

We collected three types of data: (i) participants' EDA, (ii) participants' responses to the AUT and UST questions, and (iii) participants' responses to the stereotype vulnerability questionnaire. For collecting EDA, prior to both the eligibility test and the experiment, an unobtrusive, non-invasive wristband, Empatica E4, was worn on participants' left wrists for a 15- minute warm-up period. During the warm-up period, participants were asked to be relaxed, which enabled us to detect their baseline EDA. In order to avoid sudden EDA changes (noises), participants were asked to minimize their wrist movement. Empatica E4 continuously recorded participants' EDA during the eligibility test and the experiment to assess task-related changes. EDA includes two components: skin conductance level (SCL) and skin conductance response (SCR). The SCL, a.k.a., the tonic component, is a non-stimulus-specific, slow-changing portion of the signal. On the other hand, SCR refers to the sudden change in the conductivity of the skin in response to an arousing event. Therefore, task-related changes were measured by SCR component of EDA. Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA) was used to present stimuli and record participants' responses to AUT and UST questions. Participants were also instructed on how to fill out the survey questions. The data was collected through Qualtrics^{XM} online survey platform.

3. Results

The analysis focused on investigating how students are impacted by stereotype threat by comparing three types of data (number of ideas generated for AUT&UST trials, task-related EDA changes, survey questions) for pre-and post-feedback intervention. Due to the limited number of participants, non-parametric Wilcoxon signed-rank tests and Friedman tests were used in the analysis. Wilcoxon signed-rank tests were used to compare the two results of the same participant, such as the number of ideas generated before and after the feedback. Friedman tests can be considered an extended version of the Wilcoxon signed-rank test that is used when a comparison is needed for more than two data points. We used Friedman tests to reveal the differences in scores across multiple tests for the same individual, such as comparing the number of ideas generated in four divergent thinking tasks (AUT_{pre}, UST_{pre}, AUT_{post}, UST_{post}).

3.1 Responses to AUT and UST

We focused on ideational fluency (the number of ideas) as the dependent variable. The number of ideas generated for the AUT and the UST before and after feedback intervention are presented in Table 2.

Table 2. Number of ideas generated for AUT and UST before and after intervention

ID	Gender	Intervention	Ideas generated pre-intervention			Ideas generated post-intervention		
			AUT_{pre}	UST_{pre}	$Total_{pre}$	AUT_{post}	UST_{post}	$Total_{post}$
1	Male	NF	22	22	44	16	13	29
2	Male	NF	21	17	38	18	15	33
3	Female	NF	18	19	37	15	12	27
4	Female	NF	35	34	69	43	48	91
5	Female	PF	18	25	43	18	24	42
6	Female	PF	40	44	84	42	46	88
7	Male	PF	28	28	56	25	29	54
8	Male	PF	18	22	40	28	26	54
9	Male	NF	24	26	50	21	27	48
10	Female	NF	23	31	54	21	28	49
11	Male	NF	27	29	56	21	31	52

Note. NF = Negative feedback, PF = Positive feedback

Outliers were determined by the interquartile range (IQR) method. IQR is the difference between the third quartile (Q_3) and the first quartile (Q_1). Accordingly, if the student's

number of responses was below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$, the student was considered as an outlier. Only student ID4 was excluded in all the analyses since the number of total responses ($Total_{post}$) was more than $Q_3 + 1.5 \times IQR$. After this data pre-processing step, statistical analyses were conducted as reported next.
The total number of ideas generated in AUT and UST for pre-and post-feedback intervention were compared by Wilcoxon signed-rank test. The results indicated the number of responses generated significantly decreased after the delivery of negative feedback (Md_{post} = 40.50, n=6; Md_{pre}= 47, n=6), z=-2.21, p=.027, r=-.64. The total number of responses before and after positive feedback implementation was again compared by Wilcoxon signed-rank test. However, no significant difference was observed (Md_{pre} = 49.50 vs. Md_{post} = 54, z=-.73, p=.465, r=-.37).

Additionally, the Friedman test was conducted to evaluate whether there was a significant change in ideational fluency, taking into account the impact of negative feedback (before and after) and the task type (AUT and UST). The results revealed no significant effect of task on the number of generated ideas, χ^2(3, 24)=4.93, p=.177. The medians for the number of generated ideas for AUT and UST before and after negative feedback intervention were Md_{AUT_pre}=22.5, Md_{UST_pre}=24, Md_{AUT_post}=19.5, Md_{UST_post}=21. For the positive feedback, the Friedman test was also insignificant, χ^2(3, 16)=5.76, p=.124. The medians for number of generated ideas for positive feedback were Md_{AUT_pre}=23, Md_{UST_pre}=26.5, Md_{AUT_post}=26.5, Md_{UST_post}=27.5, respectively.

The stereotype threat included in negative feedback was further analyzed considering the gender effect. Our intention was to examine whether female students reacted differently under the stereotype threat intervention. Therefore, the number of generated ideas for both tasks before and after the stereotype threat intervention were compared for female and male students (control subjects). For female students, the Wilcoxon signed-rank test was insignificant (Md_{pre} = 45.5 vs. Md_{post} = 38, z=-1.34, p=.18, r = -.67). Similarly, Wilcoxon signed-rank test was also insignificant (Md_{pre}= 47 vs. Md_{post} = 40.50, z=-1.83, p=.068, r =.65) for control subjects (male students). However, numerically, fewer ideas were generated after negative feedback in both groups.

3.2 Survey questions

Student 4, as an outlier, was excluded from the survey analysis. Below we present the results of the stereotype threat vulnerability scale:

1) All female students who were exposed to negative stereotype intervention responded with “agree” and “somewhat agree” to the question “The experimenters expected me to do poorly on the test because of my gender.” Their number of responses to the AUT and UST questions decreased after negative stereotype implementation (ID 3: 37 responses to 27 responses; ID 10: 54 responses to 49 responses in Table 2).
2) 40% of the total participants (4 students) responded with “disagree” and “somewhat disagree” to the question “People of my gender rarely face unfair evaluations in engineering classes.” These four students were female, representing 80% of female participants.
3) 75% of female students (3 out of 4) responded with “disagree” and “somewhat disagree” to the question “People of my gender rarely face unfair evaluations in engineering classes.”

4) 50% of all participants (5 out of 10) responded with “agree” and “somewhat agree” to the question, “In engineering classes, people of my gender often face biased evaluations from others.” Among five students, four were female students, accounting for all the female participants (4 out of 4).
5) 50% of all participants (5 out of 10) responded with “disagree” and “somewhat disagree” to the question “My gender does not affect people’s perception of my problem-solving ability”. Among these participants, 60% were female (3 out of 5).

3.3 Task-related EDA results

First, the data were inspected visually to identify and exclude time windows in which the EDA signal connection was lost [30]. Second, the EDA signal was decomposed into tonic (skin conductance level – SCL) and phasic (skin conductance response – SCR) components. SCL is a slowly varying portion of the EDA, while the SCR differs in response to the external stimuli, i.e., task-related events [26]. We deployed continuous decomposition analysis to extract the SCR using the Ledalab plugin for MATLAB [31].

For the eligibility test, 0.05 µS SCR change in the 4s time window after stimulus presentation was considered a task-related change. For the experiment, the SCR above 0.01µS for [-1s, 1s] response time window (starts 1 s before ideation and continues until 1 second after ideation) was considered significant. The reason for using different thresholds was due to the characteristics of the tests. The eligibility test contains questions that challenge students from different angles, such as rotation of 2-D shapes, puzzle-like questions; thus, it resulted in higher task-related EDA changes. By contrast, the divergent thinking tasks were more monotonous. Hence, lower EDA changes were used as the threshold.

Integrated skin conductance response - ISCR [31] that calculates the area of SCR in response time window was used to represent the task-related SCR changes. In order to remove intra-individual ISCR difference, ISCR data per participant were normalized by range correction in which individual’s ISCR were re-rated considering the same individual’s minimum ISCR level as 0 and maximum ISCR level as 1 [32]. Table 3 presents the average ISCR level for AUT and UST before and after the intervention.

Table 3. Average ISCR level in AUT and UST for pre-and post-intervention

ID	Gender	Intervention	Average ISCR pre-intervention			Average ISCR post- intervention		
			AUT_{pre}	UST_{pre}	$Average_{pre}$	AUT_{post}	UST_{post}	$Average_{post}$
1	Male	NF	0.036	0.021	0.028	0.028	0.005	0.016
2	Male	NF	0.107	0.088	0.098	0.065	0.152	0.109
3	Female	NF	0.078	0.031	0.055	0.031	0.021	0.026
4*	Female	NF	0.150	0.110	0.130	0.120	0.107	0.113
5	Female	PF	0.073	0.141	0.107	0.136	0.125	0.131
6	Female	PF	0.148	0.043	0.095	0.040	0.051	0.045
7	Male	PF	0.207	0.205	0.206	0.214	0.269	0.241
8	Male	PF	0.185	0.124	0.154	0.128	0.144	0.136
9	Male	NF	0.000	0.000	0.000	0.036	0.000	0.018
10	Female	NF	0.022	0.010	0.016	0.078	0.056	0.067
11	Male	NF	0.045	0.035	0.040	0.075	0.021	0.048

Note. NF = Negative feedback, PF = Positive feedback. *Outlier, excluded in the analysis.

Friedman test was conducted to investigate whether the stereotype threat caused a differentiation in ISCR for female students. Friedman test showed no significant ISCR difference between pre- and post-feedback intervention for the female students who received negative feedback (χ^2(3, 8)=1.2, *p*=.753). Similarly, there was no significant ISCR change between pre- and post- intervention for male students, control subjects, (χ^2(3, 16)=2.70, *p*=.44). For positive feedback interventions, Friedman test showed no significant ISCR difference for the control subjects (male group) in pre- and post-implications (χ^2(3, 8)=4.20, *p*=.24). In the same vein, there was no significant ISCR change for female group before and after positive feedback intervention (χ^2(3, 8)=0.60, *p*=0.896).

3.4 Correlation analysis

Correlation analysis in Table 4 presents the relationship among the variables of interest. There was a significant positive correlation among average ISCR for both divergent thinking tests, both pre-intervention and post-intervention, which indicated that students' average ISCR were robust across different tasks. Students with higher average ISCR continue with higher ISCR in the subsequent tasks; similarly, students with lower ISCR showed lower ISCR for the consecutive tasks. There were also significant positive correlations between the number of responses generated in the divergent thinking tasks. These correlations indicated that students' performance among different tasks did not change.

Table 4. Correlation among variables

	1	2	3	4	5	6	7	8	9	10
1. Gender	-									
2. Scenario	.167	-								
3. Avg. ISCR for AUT_{pre}	.122	-.772**	-							
4. Avg. ISCR for UST_{pre}	.174	-.744*	.786**	-						
5. Avg. ISCR for AUT_{post}	.168	-.667*	.645*	.918**	-					
6. Avg. ISCR for UST_{post}	.210	-.626	.791**	.942**	.892**	-				
7. Responses for AUT_{pre}	-.109	-.270	.234	-.102	-.083	-.042	-			
8. Responses for UST_{pre}	-.389	-.389	.121	-.132	-.012	-.115	.886**	-		
9. Responses for AUT_{post}	-.163	-.626	.537	.116	.097	.148	.808**	.821**	-	
10. Responses for UST_{post}	-.204	-.523	.278	.054	.179	.067	.819**	.928**	.880**	-

**Correlation is significant at the .05 level (2-tailed).*
***Correlation is significant at the .01 level (2-tailed).*

4. Conclusions

This study, reporting preliminary findings, investigated how divergent thinking of male and female senior industrial engineering students was affected by positive feedback and negative feedback that expressed a stereotype threat for female students. In the experiment, the divergent thinking abilities of participants were measured by two tasks: AUT and UST. Three types of data were collected and used in the analyses: the number of ideas generated, responses to the stereotype vulnerability scale, and task-related EDA (ISCR).

Based on our first analyses, the following results were obtained: (i) The impact of negative feedback on ideational fluency was more pronounced than the impact of positive feedback, resulting in fewer ideas generated after negative rather than positive stereotype threat. However, the stereotype threat embedded in the negative feedback did not impact ideational fluency outcomes of women more than it impacted men's performance. Likewise, no difference between the two genders was observed in the number of generated ideas for the positive feedback scenario. (ii) All females exposed to stereotype threat thought that they had performed poorly on the second part of the experiment due to a male professor's negative feedback (stereotype threat). (iii) There was no significant effect on the EDA measure (average ISCR change) in AUT and UST under positive and negative feedback interventions. (iv) No significant impact of task type on the number of ideas generated and ISCR was observed. We will first address the implications of the behavioral outcomes (i-ii) and then address the differential findings between behavior and skin conductance measures (iii-iv).

Our study contributes to educators' increasing awareness of stereotypes and helps them create safe learning environments. Our results show that negative feedback reduces students' ideational fluency, regardless of gender. All female students exposed to stereotype threat state the reason for their poor performance was the feedback from faculty members. These preliminary results reveal the impact that feedback from faculty members can have on their students. Here, it reinforces the need to avoid negative feedback that includes gender-based comments in feedback to students. The negative aspects of stereotype threat can be alleviated by wise feedback [33], which provides constructive criticism that emphasizes professors' belief in the capacity of students to achieve. This way of delivering the feedback may be seen as an opportunity for students to improve themselves rather than an unfair evaluation of their current performance.

We did not find an association between average ISCR and number of ideas generated. This implies that the lower ideational fluency under stereotype threat was not connected to engagement levels, at least the type of engagement that can be captured by EDA. This suggests that, potentially, the hypothesis that stereotype threat leads to lower engagement and thus lower performance might not be right. It is also possible that there were differences in engagement caused by the feedback, but that EDA was not able to capture it. In future work, it may therefore be beneficial to use an additional engagement indicator, such as a self-reported survey or other behavioral tasks, whether to cross-check engagement level. Additionally, and most obviously, the experiment should be reproduced with a larger dataset.

These preliminary results naturally lead to additional remaining questions. It will be important in the future to verify these results in a larger student sample, which would also enable us to include additional experimental controls (including counterbalancing tasks), on the basis of which we can draw more solid conclusions and make broader generalizations. Because the limited number of participants prevented us from counterbalancing the tasks (AUT_{pre}, UST_{pre}, AUT_{post}, UST_{post}), it is hard to conduct full inter- and intra-block analyses across participants and items to determine how behavior and skin conductance changed over the course of the entire experiment. Second, the experimenter, who gave feedback, was a male faculty member, and it remains unknown whether the same outcomes would be reached if the experimenter was a female faculty member. Additionally, we focused on ideational fluency as a measure of creativity in the present study; however, future work should investigate how other dimensions of creativity, like flexibility and originality, might interact with the nature of feedback.

Acknowledgments

This study has been supported by NSF grants DUE 1561660 to Zahed Siddique and Janet van Hell; NSF DUE IUSE-1726811 to Janet van Hell, NSF DUE IUSE-1726358 to Zahed Siddique, and NSF DUE IUSE-1726884 to Gul Okudan-Kremer.

References

[1] D. A. McFarlane, “A positive theory of stereotyping and stereotypes: is stereotyping useful?,” *Journal of Studies in Social Sciences*, vol. 8, no. 1, 2014.

[2] J. Aronson, M.J. Lustina, C. Good, K. Keough, C. M. Steele, and J. Brown, “When white men can’t do math: Necessary and sufficient factors in stereotype threat,” *Journal of Experimental Social Psychology*, vol. 35*,* no. 1, pp. 29-46, 1999.

[3] C. M. Steele, S. J. Spencer, and J. Aronson. “Contending with group image: The psychology of stereotype and social identity threat, ” in *Advances in Experimental Social Psychology*, vol. 34, pp. 379-440, 2002.

[4] T. Schmader, and J. Michael, “Converging evidence that stereotype threat reduces working memory capacity,” *Journal of Personality and Social Psychology,* vol. 85, no. 3, pp.440-452, 2003.

[5] S. J. Spencer, C. M. Steele, C. M. and D. M. Quinn, “Stereotype threat and women's math performance,” *Journal of Experimental Social Psychology*, vol. 35, pp. 4-28, 1999.

[6] S. J. Spencer, C. Logel, and P.G. Davies, “Stereotype threat,” *Annual review of Psychology*, vol. *67*, pp. 415-437, 2016.

[7] E. A. Eschenbach, M. Virnoche, E. M. Cashman, S. M. Lord, & M. M. Camacho, “Proven practices that can reduce stereotype threat in engineering education: A literature review” in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1-9).

[8] D. Kabat-Farr, and L M. Cortina, “Selective incivility: Gender, race, and the discriminatory workplace,” in Gender and the Dysfunctional Workplace, S. Fox, T. R. Lituch, Ed. Edward Elgar Publishing, 2012, pp. 107-119.

[9] D. Proudfoot, A. C. Kay, and C. Z. Koval, “A gender bias in the attribution of creativity: Archival and experimental evidence for the perceived association between masculinity and creative thinking,” *Psychological Science*, vol. 26, pp. 1751–1761, 2015.

[10] H. Zacher, and E. Johnson, “Leadership and creativity in higher education,” *Studies in Higher Education*, vol. 40, no. 7, pp. 1210-1225, 2015.

[11] E. A. Gunderson, G. Ramirez, S. C. Levine, and S. L. Beilock, “The role of parents and teachers in the development of gender-related math attitudes,” *Sex Roles*, vol. 66, no. 3, pp. 153-166, 2012.

[12] T. Schmader, J. Michael, and C. Forbes, “An integrated process model of stereotype threat effects on performance,” *Psychological Review*, vol. 115, no.2, pp. 336-356, 2008.

[13] C. Peters, G. Castellano, and S. De Freitas, “An exploration of user engagement in HCI,” in *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, 2009.

[14] F. Benyahia, and M. Khraishehla, “Innovation and creativity in engineering design: The role of student engagement, team spirit and industry support,” in *Fourth Interdisciplinary Engineering Design Education Conference*, IEEE, 2014. pp. 52-57.

[15] C. K. Prahalad, and V. Ramaswamy, “The new frontier of experience innovation,” *MIT Sloan Management Review,* vol. 44, no. 4, pp. 12-18, 2003.

[16] X. Oriol, A. Amutio, M. Mendoza, S. Da Costa, and R. Miranda, “Emotional creativity as predictor of intrinsic motivation and academic engagement in university students: the mediating role of positive emotions,”. *Frontiers in Psychology*, vol. 7, pp. 1-9, 2009.

[17] K. Muldner, W. Burleson, “Utilizing sensor data to model students’ creativity in a digital environment,” *Computers in Human Behavior*, vol. 42, pp. 127-137, 2015.

[18] A.R., Koch, C. Binnewies, C. Dormann, “Motivating innovation in schools: School principals’ work engagement as a motivator for schools’ innovation,” *European Journal of Work and Organizational Psychology*, vol. 24, pp. 505– 517, 2015.

[19] L. Eldor, I. Harpaz, “A process model of employee engagement: The learning climate and its relationship with extra-role performance behaviors,” *Journal of Organizational Behavior*, vol. 37, pp. 213– 235, 2016.

[20] A. Reid, I. Solomonides, “Design students' experience of engagement and creativity,” *Art, Design & Communication in Higher Education*, vol. 6, no.1, pp. 27-39, 2007.

[21] J. P. Guilford, “Creativity,” *American Psychologist*, vol. 5, pp. 444–454,1950.

[22] M. A. Runco, S. Acar, “Divergent thinking as an indicator of creative potential,” *Creativity Research Journal*, vol. 24, no.1, pp. 66-75, 2012.

[23] J. A. Plucker, “Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data,” *Creativity Research Journal*, vol. 12, no. 2, pp. 103-114, 1999.

[24] L. Zeng, R. W. Proctor, G. Salvendy, “Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?,” *Creativity Research Journal*, vol. 23, no.1, pp. 24-37, 2011.

[25] R. Reiter-Palmon, B. Forthmann, B. Barbot, “Scoring divergent thinking tests: A review and systematic framework,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 13, no. 2, pp. 144-152, 2019.

[26] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, “A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments,” *Psychophysiology*, vol. 49, no.1, pp. 1017-1034, 2013.

[27] E. E. Gunay, C. Y. Chu, L. Delgado Tapia, O. Elmenoufy, A. Yam, S. Raje, ... and G. E. Okudan Kremer, “The investigation of the relationship between emotional engagement and creativity,” in *ASEE North Midwest Section Annual Conference 2020 Publications*.

[28] R. Jonczyk, Y. Liu, D. S. Dickson, G. E. O. Kremer, Z. Siddique and J. van Hell, “Does stereotype threat affect creative thinking in female engineering students? A behavioral and neurocognitive study,” in *2020 ASEE Virtual Annual Conference Content Access*, 2020.

[29] H. J. Johnson, L. Barnard-Brak, T. F. Saxon, and M. K. Johnson, “An experimental study of the effects of stereotype threat and stereotype lift on men's and women's performance in mathematics,” *The Journal of Experimental Education*, vol. 80, pp. 137-149, 2012.

[30] E. Di Lascio, S. Gashi, and S. Santini, “Unobtrusive assessment of students’ emotional engagement during lectures using electrodermal activity sensors,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp.1-21, 2018.

[31] M. Benedek, and C. Kaernbach, “ A continuous measure of phasic electrodermal activity,” *Journal of Neuroscience Methods*, vol. 190, pp. 80-91, 2010.

[32] W. Boucsein, *Electrodermal activity*, 2nd ed. Springer Science & Business Media, 2012.

[33] D. S. Yeager, V. Purdie-Vaughns, J. Garcia, N. Apfel, P. Brzustoski, A. Master, W. T. Hessert, M. E. Williams, and G. L. Cohen, “Breaking the cycle of mistrust: wise ınterventions to provide critical feedback across the racial divide,” *Journal of Experimental Psychology*, vol. 143, pp. 804-824, 2014.

Appendix 1. AUT and UST questions

Table A.1. AUT and UST questions

	Question 1	Question 2	Question 3	Question 4
AUT 1	key	hanger	pipe	foil
AUT 2	pencil	brick	magnet	helmet
UST 1	Imagine, there is a creeping plant rising up to the sky. What would you expect at the top of the plant?	What would happen, or might be changed, if there were no longer door locks and all doors were unlocked?	What would happen, or might be changed, if there was no gravity in the world?	What would happen, or might be changed, if scientists discovered a material strong as steel and light as silk?
UST 2	What would be the consequences, what would happen if nobody could speak anymore?	What would be the consequences, what would happen, if a single pill contained sufficient food for the whole day?	What would be the consequences, what would happen, if buildings were made with organic materials?	What would be the consequences, what would happen, if energy was unlimited?

Appendix 2. IRB Approval Letter

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202

Date: 02/21/2020

To: Elif Gunay Gul Okudan-Kremer

From: Office for Responsible Research

Title: **NEUROCOGNITIVE EXPERIMENTATION TO ENHANCE STEM EDUCATION: STUDIES ON DIVERGENT THINKING IN FEMALE AND MALE ENGINEERING STUDENTS**

IRB ID: **19-558**

Submission Type: Initial Submission **Review Type**: Expedited

Approval Date: 02/21/2020 **Approval Expiration Date: N/A**

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, including the **recruitment materials and informed consent documents that have the IRB approval stamp**.
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required**.**
- **Obtain IRB approval prior to implementing any changes** to the study or study materials.
- **Promptly inform the IRB of any addition of or change in federal funding for this study.** Approval of the protocol referenced above applies only to funding sources that are specifically identified in the corresponding IRB application.
- **Inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project** with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an eligible PI to remain open.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g., student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

- Your research study may be subject to **post-approval monitoring by Iowa State University's Office for Responsible Research**. In some cases, it may also be subject to formal audit or inspection by federal agencies and study sponsors.
- Upon completion of the project, transfer of IRB oversight to another IRB, or departure of the PI and/or Supervising Investigator, please initiate a Project Closure to officially close the project. For information on instances when a study may be closed, please refer to the IRB Study Closure Policy.

If your study requires continuing review, indicated by a specific Approval Expiration Date above, you should:

- **Stop all human subjects research activity if IRB approval lapses**, unless continuation is necessary to prevent harm to research participants. Human subjects research activity can resume once IRB approval is re-established.
- **Submit an application for Continuing Review** at least three to four weeks prior to the **Approval Expiration Date** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.