International Journal of Gender, Science and Technology
http://genderandset.open.ac.uk

Selected papers presented at the 5th Network Gender & STEM Conference, 29–30 July 2021, in Sydney, Australia

In association with

NETWORK GENDER & STEM educational and occupational pathways and participation

"That Shocked Me": Physiological Arousal when Confronting Implicit Gender/STEM Biases

Matthew C. Graham[1], Katie Jacobson[2], Nicholette DeRosia[2], Katherine C. Cheng[3], & Madeleine Smith[4]
[1]Oregon Education Science Laboratory, University of Oregon; [2]Critical and Sociocultural Studies in Education, University of Oregon; [3]Educational Psychology, University of Arizona; [4]Education Studies, University of Oregon

ABSTRACT

Gendered differences in the "leaky STEM pipeline" start in the early grades. One explanation may be the implicit biases teachers hold regarding who does and does not belong in STEM disciplines. As a result, many teacher preparation programs provide pre-service teachers opportunities to discuss and confront their implicit biases. However, individuals do not always respond positively to these opportunities. The nature of this defensive response needs to be examined to support pre-service educators in confronting implicit gender biases so they can provide all students access to high quality STEM careers.

In this pilot study, we examined pre-service teachers' (*n* = 20) electrodermal activity, a physiological indicator of sympathetic nervous system arousal, during both the Gender/STEM implicit association test and a control activity. Our findings suggest that participants may have found it more cognitively demanding to associate "Female" with "Science" than to associate "Males" with "Science," that they might have experienced a greater physiological stress response while receiving feedback on their implicit biases than at baseline, and that the difference in both factors may relate to their scores on the IAT. These findings are a first step towards understanding and overcoming the barriers pre-service teachers face when confronting these biases.

KEYWORDS

Implicit Biases, Gender, STEM, Teacher Education, Electrodermal Activity

"That Shocked Me": Physiological Arousal when Confronting Implicit Gender/STEM Biases

INTRODUCTION

Gendered differences in the retention of students in STEM disciplines, or the "leaky STEM pipeline" (a metaphor for how individuals from historically underrepresented groups are pushed out of STEM career paths, Alper, 1993), start in the early grades (Wieselmann et al., 2020). One explanation may be the implicit biases teachers hold regarding who does and does not belong in STEM disciplines (Carlana, 2019). Implicit gender biases may influence how teachers interact with students and lead to differences in student persistence (Carlana, 2019; Tiedemann, 2000; Li, 1999; Hand et al., 2017). Prior research how found that implicit gender/STEM biases can impact teachers' grading and the attributional assumptions they make about female-identifying students (Thacker et al., 2022). Therefore, supporting educators in confronting these biases may be an important step towards addressing gendered differences in the leaky STEM pipeline. To achieve this end, many teacher preparation programs provide pre-service teachers opportunities to discuss and confront their implicit biases (e.g., Desai, 2019). However, individuals do not always respond positively to these opportunities (Howell et al., 2015, 2017). The nature of this defensive response needs to be examined to support pre-service educators in confronting implicit gender biases so they can provide all students access to high quality STEM careers.

In this pilot study, we examined pre-service teachers' (*n* = 20) experiences during a task designed to provide feedback about their implicit gender/STEM biases. Using a multimodal research design (Villanueva, 2019), we evaluated participants' electrodermal activity, a physiological indicator of sympathetic nervous system arousal, during both the Gender/STEM implicit association test (IAT; Greenwald & Krieger, 2006) and a control activity. Although prior research has examined individuals' emotional response around IATs using self-report survey methods (e.g., Howell et al., 2015, 1017), this study provides unique insight by using real-time, objective measures of participants' physiological response. Our findings suggest that participants may have found it more cognitively demanding to associate "Female" with "Science" than to associate "Male" with "Science," that they may have experienced a greater physiological stress response while receiving feedback on their implicit biases than at baseline, and that the difference in both factors may have related to their scores on the IAT. These findings are a first step towards understanding and overcoming the barriers pre-service teachers face when confronting these biases.

Theoretical Framework

This study is theoretically grounded in the Cognitive Reconstruction of Knowledge Model of conceptual change (CRKM; e.g., Sinatra & Pintrich, 2003). CRKM research emphasizes the mediating influence of an individual's affective state on the relation between self-concept and conceptual change (e.g., Trevors et al., 2016). During knowledge reconstruction, challenges to an individual's self-concept can lead to a defensive emotional response (Nyhan & Reifler, 2015). This emotional response can be a barrier to knowledge revision (Gregoire, 2003). When pre-service teachers are confronted with their implicit gender biases, they may perceive a threat to their identity and respond in a defensive manner that inhibits conceptual change (Larkin, 2012). Understanding pre-service teachers' experiences during a task designed to provide feedback about their implicit biases can provide insights that can be used to develop targeted interventions that can help support future educators in changing their understanding of implicit biases and confronting them.

Implicit Biases and the Leaky STEM Pipeline

In the United States, gender segregation in STEM disciplines persists in both higher-education (Bystydzienski, 2020) and labor markets (National Science Foundation [NSF], 2019; Schührer et al., 2015). As we move towards a "New Work Order" (Kelly, 2009) where opportunity, stability, and income are complexly connected to education (Ravn & Churchill, 2019), it is imperative that we ensure that pre-service teachers are prepared to provide all students to have equitable access to high quality career opportunities in STEM fields (Kurup et al., 2019).

Research has found that gendered differences in the "leaky STEM pipeline" (Alper, 1993) start early (Ambady et al., 2001; Beilock et al., 2010); by 6 years old, children may already associate STEM disciplines such as mathematics more with boys than girls (Cvencek et al., 2011). By middle grades, many women have experienced numerous social indicators that they do not belong in STEM disciplines and choose to pursue non-STEM pathways as they move into secondary school and beyond (Kim et al., 2018). To address this issue, researchers have increasingly sought to understand the underlying causes of this early divergence in career-trajectory. One explanation may be the implicit gender biases elementary teachers have regarding who does and does not belong in STEM disciplines (Carlana, 2019; Wang & Degol, 2017).

Implicit biases (e.g., Greenwald & Banaji, 1995) are unconscious attributions individuals make based solely off the observed traits of another individual. Internationally, individuals hold a stronger association with males and STEM disciplines compared to their association between females and STEM disciplines, or a Male/STEM implicit bias (Nosek et al., 2007). Regardless of explicit beliefs, these implicit biases may influence how teachers interact with students (e.g., Tiedemann, 2000; Li, 1999; Hand et al., 2017). Ultimately, implicit teacher gender/STEM biases may lead to differentiated outcomes for male and female students in STEM disciplines (Carlana, 2019). As a result, there is interest in supporting pre-service teachers in confronting the gender biases before entering the classroom (e.g., Batchelor et al., 2019).

The Implicit Association Test

One way in which colleges and universities engage pre-service teachers, individuals training at the institution to become primary and secondary grade teachers, in conversations about implicit biases is through the use of Implicit Association Tests (IATs; Greenwald et al., 1998). Implicit Association Tests measure the strength of participants' automatic association between mental representations of different concepts (e.g., male/female gender identities and STEM/non-STEM disciplines) using observed response latencies through a series of computer-administered categorization tasks (Greenwald & Krieger, 2006). Theoretically, it should be less cognitively demanding and, therefore, faster for an individual to categorize paired concepts that are more closely related in their mind than paired concepts that are less closely related. For example, if an individual has a stronger association between males and STEM disciplines, they will be quicker to categorize those two concepts together compared to the speed with which they categorize females and STEM disciplines. After the test is completed, individuals receive feedback on the concepts that are more implicitly associated in their mind; of the international sample of 214,501 individuals who completed the Gender/STEM IAT from 2000 to 2006, 72% held a Male/STEM implicit bias (Nosek et al., 2007).

Despite the use of these tests as a pedagogical tool (e.g., Desai, 2019; Matias, 2016), individuals do not always respond positively to the feedback they receive through IATs (e.g., Batchelor, et al., 2019). Instead, individuals may degrade the feedback (e.g., Howell et al., 2015, 2017) or refuse to discuss implicit biases (e.g., Bauer & Clancy, 2018). Building off prior research on the role of emotions in moderating conceptual change (e.g., Trevors et al., 2016), one possible explanation for these responses could be that individuals perceive the concept of implicit biases and evidence of their own biases as a threat to their egalitarian self-concept, triggering an ego-protective defensive response (Howell et al., 2015, 2017; Morris & Ashburn-Nardo, 2009). This defensive response may be a barrier to conceptual change regarding the nature of implicit gender biases. There exists a need to understand pre-service teachers' experiences during IATs in order to support them in confronting these biases in order to promote the development of these cultural competencies.

Passive Measures of Student Experiences during the IAT

Understanding pre-service teachers' experiences during an IAT, particularly immediately after receiving feedback regarding their implicit biases, presents numerous methodological challenges. Prior research examining emotional response to IATs (e.g., Howell et al., 2015, 2017) has used self-report surveys of individual emotions immediately after participants have completed the test. Although this research has been useful in identifying psychological components of individuals' emotional response to the IAT, these findings have been constrained by the inherent limitations of self-report surveys of emotions (Pekrun, 2006), including susceptibility to social desirability biases, limitations to recall reliability, and inability to capture moment-to-moment changes in emotions. Fortunately, recent methodological and technological advances have provided new passive methods for collecting physiological data regarding individuals' experiences during a task (Roos et al., 2020).

One commonly used and minimally invasive methodology for collecting physiological responses during a specific task is electrodermal activity (EDA; Benedek & Kaernbach, 2010; Boucsein, 2012). During periods of arousal, an individual's sympathetic nervous system (SNS) is activated in order to support the body in responding to a stimulus (Schmidt & Thews, 1989). This SNS arousal causes changes in the secretion of sweat which, in turn, modulates the electrical properties of the skin (Boucsein, 2012). Variation in the electrical properties of the skin can then be measured by applying an electrical potential between two contact points in the skin and measuring changes in electrical current flow, or EDA. Fluctuations in the EDA can be connected to emotional regulation and cognitive processes, among other brain functions (Boucsein, 2012).

The totality of the EDA response is composed of both a background tonic signal and a fast-varying phasic signal. These signals can be separated using a variety of deconvolution methods that allow for separate analysis of each component of the total EDA response (Benedek & Kaernbach, 2010). Researchers have often focused on variabilizations of the faster-varying phasic signal in order to evaluate time- and event-dependent changes in SNS activity (Benedek & Kaernbach, 2010). Additionally, there are also multiple methodological approaches to variabilizing the phasic EDA signal component, such as latency, means, peaks, amplitudes, and magnitudes (Boucsein, 2012). The method of decomposing the phasic signal allows researchers to identify different aspects of SNS arousal (Setz et al., 2010). It is left for scholars to decide upon and interpret which characterization of the signal is best suited for their experimental design (Villanueva, 2019). Depending on the variabilization that are used, researchers can gain insight into different aspects of an individual's experience during a task.

The Present Study

In this study, we used two different variabilizations of the phasic EDA signal to understand pre-service teachers' experiences while taking two different IATs – the Gender/STEM IAT and a control task. First, we analyzed the average phasic EDA signal (mean EDA), which is an indicator of cognitive load (Villanueva et al., 2021). Second, we analyzed the frequency of changes in the phasic EDA signals (peak EDA), as an indicator of stress (Villanueva et al., 2021). Additionally, we examined differences in both peak and mean EDA during different tasks and examined the relation between these differences in EDA and performance on the IAT. Specifically, we answered the following research questions:

RQ1) Do participants experience greater cognitive load, as indicated by mean EDA, during the Gender/STEM IAT compared to their experience during the control task? Given the theorized underlying psychological processes driving how individuals respond to IATs (e.g., Greenwald & Krieger, 2006), we hypothesize that mean EDA will be higher during the Gender/STEM IAT than during the control activity.

RQ2) Do participants experience greater stress, as indicated by peak EDA, during the Gender/STEM IAT compared to their experience during the control task? Building on prior work using self-report survey methodologies (e.g., Howell et al., 2015, 2017), we hypothesize that peak EDA will be higher during the Gender/STEM IAT than during the control activity.

RQ3) Is there a relation between participants' performance on the IAT and the difference in participants' cognitive load, as indicated by mean EDA, during the non-stereotypical tasks (i.e., "Female" with "Science" and "Birds" with "Large") and stereotypical tasks (i.e., "Male" with "Science" and "Mammals" with "Large"). Given the theorized underlying psychological processes driving how individuals respond to IATs (e.g., Greenwald & Krieger, 2006), we hypothesize that there will be a positive correlation between participants' performance on both IATs and the difference in their mean EDA during the non-stereotypical and stereotypical tasks.

RQ4) Is there a relation between participants' feedback on the Gender/STEM IAT and the difference in stress, as indicated by peak EDA, while they receive feedback on the Gender/STEM IAT compared to their baseline stress level. Given prior research using self-report survey methodologies (e.g., Howell et al., 2015, 2017), we hypothesize that there will be a positive correlation between participants' feedback (indicating a greater Male/STEM bias) on the Gender/STEM IAT and difference in peak EDA as they receive feedback compared to their baseline level.

METHOD

Participants

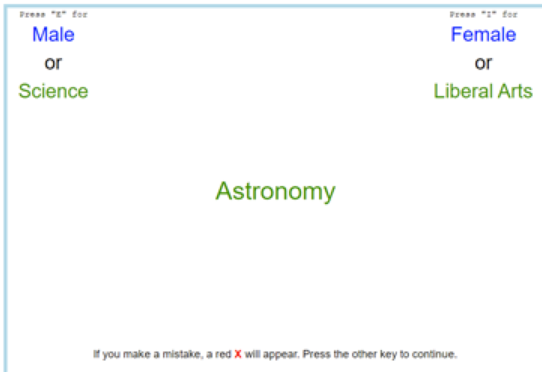
Pre-service teachers (n = 20) were recruited from an introductory Education course at a large research university in the Pacific Northwest of the United States during Spring (n = 12) and Fall (n = 8) Terms 2021. One participant identified as male and 19 identified as female. Using the U.S. federal guidelines for reporting race and ethnicity (e.g., Office of Management and Budget, 1997), of the 18 participants who provided demographic information 11 identified as White non-Hispanic, 5 as Hispanic, and 2 as some other race/ethnicity. Participants received a $40 gift card for participating in the study.

Procedure

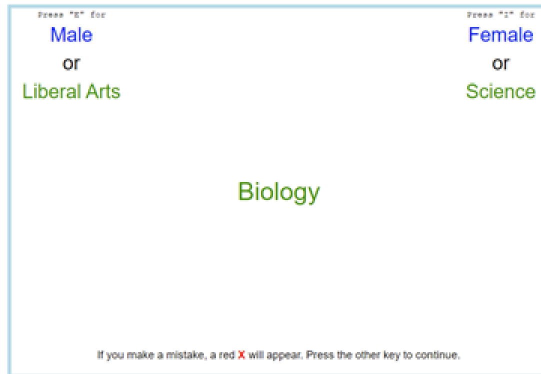
For this study, we used a crossover research design (Keppel & Wickens, 2004). The use of a crossover research design allowed us to evaluate within-subject differences in both mean and peak EDA while also accounting for period, or order, effects. This greatly increased our statistical power compared to a between-subjects comparison. All participants were asked to complete two different IATs over two successive data collection periods, the Gender/STEM implicit association test (IAT; Greenwald & Krieger, 2006) and the Mammal/Size IAT (Sriram & Greenwald, 2009). Using the Mammal/Size IAT as a control activity allowed us to analyze the degree to which changes in EDA were due to engagement with an IAT generally and what was specifically related to the sensitive content of the Gender/STEM IAT. Both IATs required participants to complete a series of sorting tasks where they were asked to classify a term populated in the center of the screen with categories either to the left or right side of the screen using the computer keyboard, including a series of practice tasks, a stereotypical sorting task (i.e., Male/STEM and for the Gender/STEM IAT and Mammals/Large for the Mammal/Size IAT) non-stereotypical sorting task (i.e., Female/STEM for the Gender/STEM IAT and Mammals/Small for the Mammal/Size IAT). We present examples from both IATs in Figure 1.

Figure 1

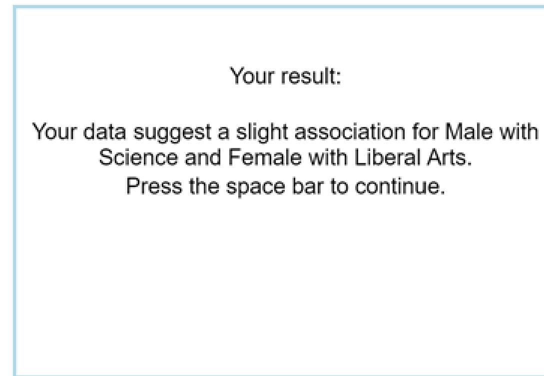
Example sorting tasks and feedback from the Gender/STEM (1a-1c) and Mammal/Size (1d-1f) IATs



1a. Gender/STEM Stereotypical Task



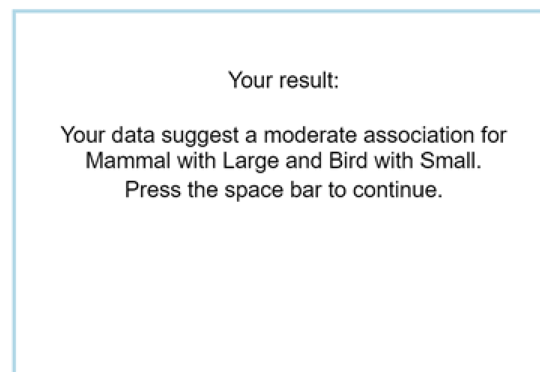
1b. Gender/STEM Non-stereotypical Task



1c. Gender/STEM Feedback

1d. Mammal/Size Stereotypical Task

1e. Mammal/Size Non-stereotypical Task

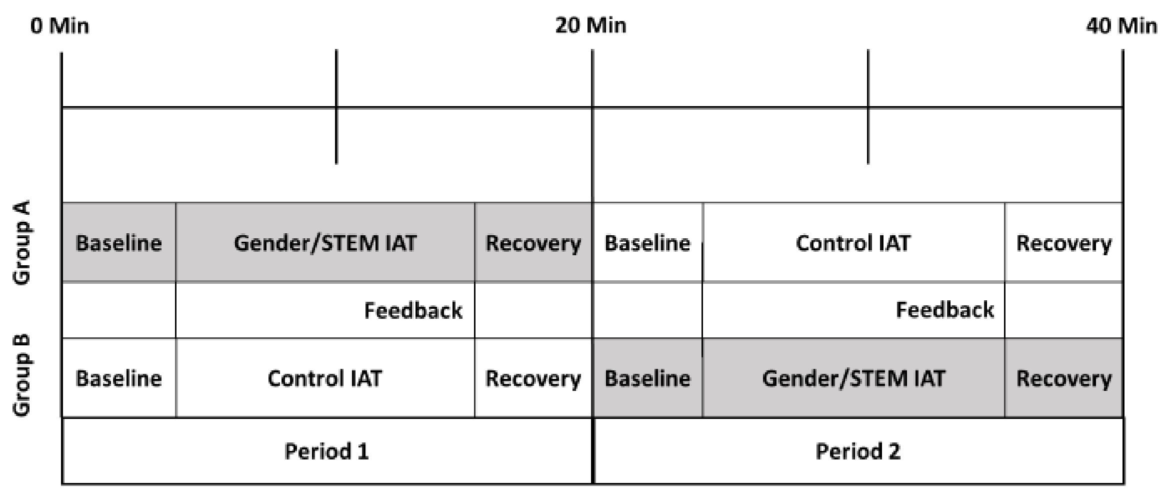


1f. Mammal/Size Feedback

Participants were randomly assigned to one of two groups where the order in which they completed the two IATs was randomly assigned using a random number generator to assign each participant a number from 1-100 and odd-numbered participants selected into the group completing the Gender/STEM IAT first followed by the Mammal/Size IAT and even-numbered participants completing the Mammal/Size IAT first followed by the Gender/STEM IAT.

Following procedures approved by the University of Oregon IRB, participants arrived at the research laboratory at 4:00pm on an assigned weekday, were provided a computer workstation, and given an overview of the study procedures. Research personnel then assisted participants in affixing the Empatica E4 wrist sensors (Empatica, Boston, MA) in order to collect EDA data. During the first period, participants were first asked to stare at a black circle on the screen for two minutes. After two minutes, participants then completed the first IAT. After completing the first IAT, participants received feedback on their implicit associations. Participants then had a two-minute recovery period staring at the black circle before the process started over again for the second period. Procedures for the second period were identical to the first, with participants taking the IAT they did not take during the first period. Total data collection time took approximately forty minutes. We present data collection procedures in Figure 2.

Figure 2
Diagram of Research Design



Measures
Measures used in the present study include IAT score and two different variabilizations of the phasic EDA signal.
IAT score. A D-score (Greenwald et al., 2003) was calculated for both IATs using the following formula:

$$D = \frac{\bar{\bar{x}}_b - \bar{\bar{x}}_a}{s_{ab}}$$

where $\bar{x}_b$ is the mean time it took an individual to sort terms into the non-stereotypical categorical pairings (e.g., "Female" and "Science" on one side and "Male" and "Liberal Arts" on the other), $\bar{x}_a$ is the mean time it took an individual to sort the stereotypical pair ("Male and "Science" on one side and "Female" and "Liberal Arts" on the other), and s_{ab} is the pooled standard deviation for both tasks. Therefore, IAT score can be interpreted as differences in latency expressed in standard deviations. Participants were provided feedback regarding their implicit biases using their D-score and pre-determined breakpoints (Greenwald et al., 2003). For example, on the Gender/STEM IAT participants were informed that they had a strong (-0.65), moderate (-0.35), or slight (-0.15) implicit association of "Female" with "Science" and "Male" with "Liberal Arts", no implicit association (-.15 to 0.15), and slight (0.15), moderate (0.35) or strong (0.65) implicit association of "Male" with "Science" and "Female" with "Liberal Arts." Scoring and feedback were identical for the Mammal/Size control IAT, with "Mammals" with "Large" and "Birds" with "Small" as the stereotypical pair and "Birds" with "Large" and "Mammals" with "Small" as the non-stereotypical pair.

Electrodermal activity. Electrodermal activity was collected from each participant at the recommended sample rate of 4 Hz (Boucsein, 2012) for the entirety of the study using Empatica E4 wrist sensor (Empatica, Boston, MA). After data collection, the fast-varying phasic EDA signal was decomposed from the tonic signal using the continuous deconvolution method in the Ledalab (Benedek & Kaernbach, 2010).

The resultant processed phasic EDA signal was aligned to timestamps for a specific event (i.e., receiving feedback on the Gender/STEM IAT) to extract the event-specific phasic EDA, and elicit how near-real-time stimuli was to a given event related to EDA (Benedek & Kaernbach, 2010). Timestamps were triangulated using 1) participant recorded timestamping using the E4 sensor event mark button, 2) timestamps extracted from screen recordings of the session, and 3) Qualtrics-generated timestamps. In the event that timestamp data were missing or did not align between sources, data were imputed using the preponderance of evidence and from the averaged time all participants took to complete each task. Premised on recent work (e.g., Villanueva et al., 2021), we variabilized two different aspects of the phasic EDA signal. First, we examined the number of peaks in the signal per second, reporting the highest points in the waveforms of a signal as an indication of stimuli-eliciting reaction associated with SNS reactivity (i.e., stress), which we refer to as "peak EDA.". Second, we examined the overall average of the waveforms of a signal during a given period as an indication of the cognitive load, which we refer to as "mean EDA" (Villanueva et al., 2019, 2021).

Statistical Analyses

To answer our first two research questions, we calculated the participants' total mean and peak EDA responses during both the Gender/STEM and Mammal/Size IAT. We then compared these responses using a specialized two-factor within-subject ANOVA specifically for the analysis of a within-subject research design with a counterbalanced factor (Keppel & Wickens, 2004, p. 387). A key attribute of the crossover research design is that the three main factors specified within this statistical model (IAT, period, and subject) are mutually orthogonal and can be disaggregated into separable portions of the observed variability in both mean and peak EDA. This allowed us to compare within-subject differences in participants' mean and peak EDA responses during the Gender/STEM and Mammal/Size IATs after accounting for the nuisance period factor, decreasing residual variance and increasing statistical power.

To answer our third research question, we took the ratio of mean EDA during the non-stereotypical sorting task (i.e., "Female" with "Science" and "Birds" with "Large") over mean EDA during the stereotypical sorting task (i.e., "Male" with "Science" and "Mammal" with "Large"). This allowed us to evaluate the degree to which, in general, participants experienced a greater mean EDA response, an indicator of cognitive load, during the sorting task associated with the non-stereotypical sorting task. Then, using a bivariate Pearson correlation, we examined the relation between participants' performance on each IAT and the changes in mean EDA response during each sorting task.

To answer our fourth research question, we took the ratio of peak EDA as participants' received feedback on their implicit biases over their baseline peak EDA prior to starting each IAT. This allowed us to evaluate the degree to which participants' experienced a greater peak EDA response, an indicator of stress, when receiving feedback regarding their implicit associations. Then, using a bivariate Pearson correlation, we examined the relation between participants' performance on each IAT and the changes in peak EDA response when receiving feedback compared to baseline peak EDA.

RESULTS

We analyzed all data using SPSS 26.0 (IBM Corp., 2019). Overall, 9 participants were identified and received feedback that they had an implicit association between "Male" and "Science" and "Female" and "Liberal Arts", 4 no implicit association, and 7 had an implicit association between "Female" and "Science" and "Male" and "Liberal Arts." For the Mammal/Size control IAT, 10 had an implicit association between "Mammals" and "Large" and "Birds" and "Small," 5 had no implicit association, and 2 had an implicit association between "Birds" and "Large" and "Mammal" and "Small." We present descriptive statistics for mean and peak EDA by IAT (Gender/STEM v. Mammal/Size) and Period in Table 1.

Differences in Mean and Peak EDA by IAT

In order to answer our first two research questions, we conducted specialized two-way within-subject ANOVAs for crossover research design by extracted General Linear Model variance components in SPSS (IBM Corp.) and manually calculated residual variance, F-statistics, and partial η^2 (Keppel & Wickens, 2004, p. 388).

Table 1
Means & Standard Deviations (SD) for Mean and Peak EDA by Period and IAT (Gender/STEM v. Mammal/Size)

Mean (SD)	Period		
	Period 1	Period 2	Total
Mean EDA (μS)			
IAT			
Gender/STEM	0.23 (0.22)	0.03 (0.03)	0.13 (0.18)
Mammal/Size	0.06 (0.06)	0.12 (0.19)	0.09 (0.14)
Total	0.14 (0.18)	0.07 (0.14)	
Peak EDA (pps)			
IAT			
Gender/STEM	0.79 (0.28)	1.08 (0.13)	0.94 (0.26)
Mammal/Size	1.01 (0.17)	1.05 (0.27)	1.03 (0.22)
Total	0.90 (0.25)	1.07 (0.20)	

Note. IAT = implicit association test. μS = microsiemens. pps = peaks per second.

Differences in cognitive load by IAT. To answer our first research question, we evaluated differences in participants' mean EDA by IAT after accounting for period effects. Our dependent variable was participants' mean EDA, an indicator of cognitive load. Our independent variables were IAT type with two levels, Gender/STEM and Mammal/Size, and period with two levels, period 1 and period 2. We present results in Table 2. Using effect size, we found a moderate difference between the two IAT conditions, Gender/STEM (M = 0.13μS, SD = 0.18μS) and Mammal/Size (M = 0.09μS, SD = 0.14μS), η_p^2 = .16. However, these values were not statistically different, $F(1,15) = 2.91$, $p = .11$. These differences were overshadowed by the large (η_p^2 = .39) and significant differences in mean EDA by period, with period 1 (M = 0.14μS, SD =0.18μS) significantly higher than period 2 (M = 0.07μS, SD = 0.14μS), $F(1,15) = 9.60$, $p < .01$.

Table 2
Analysis of Within-subject Design with Counterbalance Factor for Mean EDA

Source	*SS*	*df*	*MS*	*F*	η_p^2
IAT (Gender/STEM v. Mammal/Size)	0.01	1	0.01	2.91	.16
Period	0.05	1	0.04	9.60**	.39
Subject	0.72	16	0.05		
Residual	0.07	15	<0.01		
Total	0.83	33			

Note. * $p < .05$, ** $p < .01$.

Differences in stress by IAT. To answer our second research question, we evaluated differences in participants' peak EDA by IAT after accounting for period effects. Our dependent variable was participants' peak EDA, an indicator of negative emotional response. Our independent variables were IAT type with two levels, Gender/STEM and Mammal/Size, and period with two levels, period 1 and period 2.

We present results in Table 3. We found that the difference between the two IAT conditions, Gender/STEM (*M* = 0.94pps, *SD* = 0.26pps) and Mammal/Size (*M* = 1.03pps, *SD* = 0.22pps), were modest in effect size (η_p^2 = .19), but not statistically different, *F*(1,15) = 3.50, *p* = .08. Similar to our findings for mean EDA, this finding was overshadowed by the large (η_p^2 = .46) and significant differences in mean EDA by period, with period 1 (*M* = 0.90pps, *SD* = 0.25pps) significantly lower than period 2 (*M* = 1.07pps, *SD* = 0.20pps), *F*(1,15) = 12.86, *p* < .01.

Table 3
Analysis of Within-subject Design with Counterbalance Factor for Peak EDA

Source	*SS*	*df*	*MS*	*F*	η_p^2
IAT (Gender/STEM v. Mammal/Size)	0.06	1	0.06	3.50	.19
Period	0.23	1	0.23	12.86**	.46
Subject	1.29	16	0.08		
Residual	0.27	15	0.02		
Total	1.85	33	0.06		

Note. * *p* < .05, ** *p* < .01.

Relation between IAT score and changes in Mean and Peak EDA during IATs

To answer our third and fourth research questions, we calculated the ratios between mean and peak EDA during different subtasks of the IAT and examined the relation between these ratios and participants' performance on the IATs. We present overall means, standard deviations, and correlations in Table 4 and separate results by IAT (Gender/STEM and Mammal/Size) in Table 5.

Table 4
Correlations (SE), Means, and Standard Deviations for IAT Score, Ratio of Mean EDA during Non-stereotypical/Stereotypical Tasks, and Ratio of Peak EDA during Feedback/baseline

	1	2	3
1. IAT Score	1.00		
2. Mean EDA Non-stereotypical/ Stereotypical Tasks	.26 (.17)	1.00	
3. Peak EDA Feedback/Baseline	.43** (.15)	-.08 (.17)	1.00
Mean	0.23	1.532	1.15
SD	0.50	1.950	0.66

Notes. ^{T} *p* < .10, * *p* < .05, ** *p* < .01.

Relation between IAT score and change in Mean EDA. To answer our third research question, we calculated the ratio between mean EDA during the non-stereotypical sorting task (i.e., "Female" with "Science" and "Birds" with "Large") over mean EDA during the stereotypical sorting task (i.e., "Male" with "Science" and "Mammal" with "Large"). Overall, the relation between cognitive load, as indicated by higher mean EDA, was approximately 53% higher during the non-stereotypical sorting task than during the stereotypical sorting task (*M* = 1.53, *SD* = 1.95). This relation was also more pronounced during the Gender/STEM IAT. Mean EDA was approximately 87% higher when sorting “Female” and “Science” together compared to mean EDA while sorting “Male” and “Science,” whereas mean EDA was only 16% greater during while sorting “Birds” and “Large” together when compared to sorting “Mammals” and “Large”. The overall relation between IAT score and differences in mean EDA during the non-stereotypical and stereotypical sorting task was small and not significant; $r = .26$, $SE = .17$, $p = .13$. However, when we examined this relation by IAT, we found that it was small and non-significant for the Mammal/Size IAT ($r = -.10$, $SE = .26$, $p = .72$) but was much larger and trending towards significance during the Gender/STEM IAT; $r = .43$, $SE = .25$, $p = .07$.

Table 5

Correlations (SE), Means, and Standard Deviations for IAT score, ratio of Mean EDA during Non-stereotypical/Stereotypical Tasks, and ratio of peak EDA during feedback/baseline by IAT (Gender/STEM and Mammal/Size)

		1	2	3
1. IAT Score	Gender/STEM Mammal/Size	1.00		
2. Mean EDA Non-stereotypical/Stereotypical Tasks	Gender/STEM	$.43^{T}$ (.25)		
	Mammal/Size	-.10 (.26)	1.00	
3. Peak EDA Feedback/Baseline	Gender/STEM	.49* (.23)	-.23 (.25)	1.00
	Mammal/Size	.36 (.24)	.40 (.24)	
Mean	Gender/STEM	0.14	1.868	1.13
	Mammal/Size	0.34	1.156	1.18
SD	Gender/STEM	0.54	2.535	0.69
	Mammal/Size	0.43	0.896	0.66

Notes. ^{T} $p < .10$, * $p < .05$, ** $p < .01$.

Relation between IAT score and change in peak EDA. To answer our fourth research question, we calculated the ratio between peak EDA, an indicator of stress, as participants received feedback regarding their implicit associations and their baseline peak EDA level. Overall, participants appeared to have experienced a 15% increase in peak EDA when receiving feedback. This was slightly higher (approximately 18%) for the Mammal/Size IAT than the Gender/STEM IAT (approximately 13%). The overall relation between IAT score and differences in peak EDA while receiving feedback compared to baseline was significant, $r = .43$, $SE = .15$, $p < .01$. This relation was also slightly more pronounced during the Gender/STEM IAT ($r = .49$, $SE = .23$, $p = .03$) than during the Mammal/Size IAT ($r = .36$, $SE = .24$, $p = .16$).

DISCUSSION

In this study, we sought to use electrodermal activity (EDA) as a physiological indicator of pre-service teachers’ experience during the Gender/STEM implicit association test (IAT) and compare these results to their experience during a control task. We used two different variabilizations of the phasic EDA signal; 1) mean EDA as an indicator of cognitive load and 2) peak EDA as an indicator of stress (Villanueva et al., 2019, 2021).

To answer our first research questions, we evaluated differences in mean EDA during the Gender/STEM IAT and Mammal/Size IAT. Using a specialized within-subject ANOVA (Keppel & Wickens, 2004), we did not find significant differences in mean EDA between the Gender/STEM IAT and control task. Similarly, in answering our second research question, we did not observe differences in peak EDA, an indicator of stress, between the Gender/STEM IAT and control task. However, there were several factors that made interpreting these results more difficult. First, less than half our sample had an implicit Male/STEM bias, compared to the 72% observed in previous research (Nosek et al., 2007). Our results may be attenuated by this factor. Second, for both models we observed a significant period effect; mean EDA was significantly higher during period 1 compared to period 2 and peak EDA was significantly lower in period 1 compared to period 2. Additionally, we observed significant interindividual variability for both mean and peak EDA further muddying the results. Together, these challenges make evaluating within-subject change in either factor difficult. Additional research is needed with a larger sample in order to better understand differences in pre-service teachers' experiences during these two tests.

To answer our third research question, we examined the relation between mean EDA during the non-stereotypical task (e.g., categorizing "Female" and "Science" or "Birds" and "Large") and mean EDA during the stereotypical task (e.g., categorizing "Male" and "Science" or "Mammal" and "Large"). We observed that, overall, participants may have experienced a greater cognitive load, as indicated by greater mean EDA, during the non-stereotypical task. While this was true for both IATs, the relation may be more pronounced for the Gender/STEM IAT when compared to the Mammal/Size IAT. Additionally, we examined the relation between this difference and participants' score on the IAT and observed that the relation may be stronger for the Gender/STEM IAT than for the Mammal/Size task. These findings support the underlying assumption of the implicit association test that it should be easier to sort into paired categories when there is a stronger mental connection between them than if that connection is weaker (e.g., Nosek et al., 2005). This finding suggests that, in general, participants may have experienced greater cognitive load when "Female" and "Science" were paired than when "Male" and "Science" were paired, and that this difference was related to their performance on the IAT. If this is the case, then this would validate the IAT as a measure of these implicit biases and the tests' utility as a tool to support pre-service teachers in identifying and confronting their biases. However, this result was less pronounced for the control task, the Mammal/Size IAT. It may be that participants were less engaged during the control activity and, as a result, did not experience the same cognitive engagement. Additional research is needed both to confirm these findings and to explore participants' experience during the control IAT.

To answer our fourth research question, we examined the relation between peak EDA, an indicator of stress, while participants received feedback on the IAT compared to their baseline peak EDA. We found that, overall, participants' stress response may have been greater while receiving feedback on the IAT compared to their baseline level. Additionally, this difference may be related to participants' IAT score and resultant feedback; participants who held greater stereotypical biases (e.g., “Men'' with “Science”) had a greater stress response. This finding would support the notion that participants may experience a physiological stress response when receiving feedback on the IAT. This response may be part of the reason why previous studies (e.g., Howell et al., 2015, 2017) have found participants degrade the test and the feedback they receive. Building on CRKM (e.g., Sinatra & Pintrich, 2003), this stress response may be a barrier to knowledge revision (Gregoire, 2003) and inhibit conceptual change (Trevors et al., 2016). If pre-service teachers do experience a physiological stress response as they receive feedback about their implicit biases, they may be less able to accommodate this information and adjust their behavior. Although this result may have been slightly more pronounced for the Gender/STEM IAT, there was little difference in either the difference in stress response while receiving feedback compared to baseline or the relation between this difference and IAT score. Additional research is needed in order to better understand this stress response both during the Gender/STEM IAT and during the control task.

Together, these findings have several implications. Our findings regarding differences in mean EDA during the stereotypical and non-stereotypical task and the relation between this difference and performance on the IAT provide additional validity for the IAT as a measure of implicit association. This finding suggests that the insights provided by these instruments may be a fair measure of differences in cognitive load between these tasks and can provide individuals useful insight. Participants may experience greater cognitive load when pairing “Female” and “Science” as compared to “Male” and “Science,” suggesting a weaker connection consistent with previous work on implicit biases (e.g., Greenwald & Krieger, 2006). These biases may, in part, be responsible for differences in how teachers differently interact with male and female-identifying students in their STEM classes (e.g., Carlana, 2019; Tiedemann, 2000; Li, 1999; Hand et al., 2017) and be partially responsible for gendered-differences in the “leaky STEM pipeline” (Alper, 1993). As a result, IATs may continue to be a useful tool to help pre-service teachers engage in identifying and confronting their implicit biases. However, our findings regarding differences in stress response while receiving feedback on the IAT and baseline stress level and the relation between this difference and performance on the IAT suggest that there is a need to support pre-service teachers during these tasks. If pre-service teachers experience significant stress, they may be less able to assimilate the findings of the IAT and, therefore, less about to confront their biases. Interventions such as meditative breathing (Cooper et al., 2015) and mindfulness meditation (e.g., Morais et al., 2022) may be useful in supporting pre-service teachers as they receive feedback regarding their implicit biases.

There were several limitations to this study. First, this was a small-n pilot study. As a result, our analyses were underpowered for the size effect we anticipated and observed. A post-hoc power analysis revealed that based on the directly observed mean differences, an *n* of approximately 60 (mean EDA) to 100 (peak EDA) would be needed to obtain statistical power at the recommended .80 level (Cohen, 1988). This limited the statistical analysis we could conduct and prohibited the mediational analysis of the impact of EDA reactivity on the relation between feedback and conceptual change proposed in our theoretical framework. Additional research is needed with a larger, more diverse sample in order to test the observed relations and allow for these more complex analyses. Second, the Gender/STEM IAT requires participants to sort disciplines into "STEM" and "Liberal Arts" categories; however, some frameworks include STEM as part of the liberal arts. Further research is needed to understand the impact this might have on participants' responses. Finally, although the population was racially and ethnically diverse, it was made of almost exclusively female-identifying pre-service teachers. Additional research is needed with a more diverse sample in order to evaluate the degree to which the observed patterns or relations are generalizable to a broader population.

CONCLUSION

In this study, we sought to understand pre-service teachers' experiences during the Gender/STEM IAT (Greenwald & Krieger, 2006) in order to identify barriers that may prevent them from being willing to confront their implicit biases (e.g., Howell et al., 2015, 2017). Grounded in CRKM (e.g., Sinatra & Pintrich, 2003), we used two different variabilizations of Electrodermal Activity (EDA; Boucsein, 2012) to evaluate participants' cognitive load and stress during the Gender/STEM IAT and a control task. Although we did not see significant differences between the two tasks, we did observe that, overall, participants experienced greater cognitive load when completing the non-stereotypical task compared to the stereotypical task and greater stress response when receiving feedback on the test compared to their baseline level. These finding suggest that the IAT may be a valid measure of the strength of connection between different concepts, but there may also be a need to support pre-service teachers as they receive feedback regarding their implicit bias or they may experience a stress response that may inhibit their willingness and ability to accept and confront these biases. Addressing this stress response may be an important step in helping pre-service teachers in plugging the leaky STEM pipeline and supporting all students in accessing high-quality STEM careers.

ACKNOWLEDGMENTS

The research reported in the article was made possible by a grant from the Spencer Foundation #202100256. The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation. Additionally, this work was supported in part by the National Science Foundation under grant numbers DUE-1661117 and EEC-1654140 and the Baney Fund through the University of Oregon College of Education. Special thanks to Drs. Idalis Villanueva (University of Florida) and Jenefer Husman (University of Oregon) for their time, energy, and support for this project.

REFERENCES

Alper, J (1993). The pipeline is leaking women all the way along. *Science*, *260*(5106), 409-411. https://doi.org/10.1126/science.260.5106.409

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385–390. https://doi.org/10.1111/1467-9280.00371

Batchelor, K. E., DeWater, K., & Thompson, K. (2019). Pre-service teachers' implicit bias: Impacts of confrontation, reflection, and discussion. *Journal of Educational Research and Innovation*, *7*(1), 3. https://digscholarship.unco.edu/jeri/vol7/iss1/3

Bauer, K., & Clancy, K. (2018). Teaching race and social justice at a predominantly white institution. *Journal of Political Science Education*, *14*(1), 72-85. https://doi.org/10.1080/15512169.2017.1358175

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 1860–1863. https://doi.org/10.1073/pnas.0910967107

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80-91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer Science + Business Media. https://doi.org/:10.1007/978-1-4614-1126-0

Bystydzienski, J. M. (2020). Gender and STEM in higher education in the United States. *Oxford Research Encyclopedia of Education.* https://doi.org/10.1093/acrefore/9780190264093.013.1298

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, *134*(3), 1163-1224. https://doi.org/:10.1093/qje/qjz008

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.

Cooper, T., Hedrick, J., Murphy, B., Reil, L., Sayaovang, K., & Thomas, T. (2015). Meditative breathing yields inconclusive results in stabilizing physiological variables following fear-induced acute stress in college students. *Journal of Advanced Student Sciences, 2015.* http://digital.library.wisc.edu/1793/80162

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766-779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Desai, S. R. (2019). Humanizing Trayvon Martin: Racial profiling, implicit biases, and teacher education. *Urban Education, 54*(8), 1031-1057. https://doi.org/10.1177/0042085916646609

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4. https://doi.org/10.1037/0033-295x.102.1.4

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945-967. https://doi.org/10.2307/20439056

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of*

Personality and Social Psychology, *74*(6), 1464-1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216. https://doi.org/10.1037/0022-3514.85.2.197

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review, 15*(2), 147–179. https://doi.org/10.1023/A:1023477131081

Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, *20*(4), 929-945. https://doi.org/10.1007/s11218-017-9408-8

Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science*, *6*(4), 373-381. https://doi.org/10.1177/1948550614561127

Howell, J. L., Redford, L., Pogge, G., & Ratliff, K. A. (2017). Defensive responding to IAT feedback. *Social Cognition*, *35*(5), 520-562. https://doi.org/10.1521/soco.2017.35.5.520

IBM Corp. (2019). IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.

Kelly, P. J. (2009). Generation Y, flexible capitalism and new work ethics. In A. Furlong (Ed.), *Handbook of Youth and young Adulthood: New perspectives and agendas* (1 ed., pp. 399 - 405). Routledge.

Keppel, G., & Wickens, T. (2004). *Design and Analysis: A Researcher's Handbook (4th ed.).* Pearson.

Kim, A. Y., Sinatra, G. M., & Seyranian, V. (2018). Developing a STEM identity among young women: A social identity perspective. *Review of Educational Research*, *88*(4), 589-625. https://doi.org/10.3102/0034654318779957

Kurup, P. M., Li, X., Powell, G., & Brown, M. (2019). Building future primary teachers' capacity in STEM: Based on a platform of beliefs, understandings and intentions. *International Journal of STEM Education*, *6*(1), 1-14. https://doi.org/10.1186/s40594-019-0164-5

Larkin, D. (2012). Using the conceptual change model of learning as an analytic tool in researching teacher preparation for student diversity. *Teachers College Record, 114*(8), 1-35. https://doi.org/10.1177/016146811211400808

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, *41*(1), 63-76. https://doi.org/10.1177/016146811211400808

Matias, C. E. (2016). 'Why do you make us hate ourself?': Re-teaching Whiteness, abuse, and love in urban teacher education. *Teaching Education*, *27*(2), 194-211. https://doi.org/10.1080/10476210.2015.1068749

Morais, P.; Pinheiro, A.; Fonseca, M. and Quintão, C. (2022). The mindfulness meditation effect on states of anxiety, depression, stress and quality of life. In *Proceedings of the 15th International Joint Conference on Biomedical*

Engineering Systems and Technologies, 573-582. https://doi.org/10.5220/0010854900003123

Morris, K. A., & Ashburn-Nardo, L. (2009). The Implicit Association Test as a class assignment: Student affective and attitudinal reactions. *Teaching of Psychology*, *37*(1), 63-68. https://doi.org/10.1080/00986280903426019

National Science Foundation. (2019). *Women, minorities, and persons with disabilities in science and engineering: 2019*. (Special Report NSF 19-304). Arlington, VA: National Center for Science and Engineering Statistics.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*(2), 166-180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265–292). Psychology Press.

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine, 33*(3), 459–464. https://doi.org/10.1016/j.vaccine.2014.11.017

Office of Management and Budget. (1997). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*, 62(210) Fed. Reg. 58782-58790.

Pekrun, R. (2006). The Control-Value Theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. http://doi.org/10.1007/s10648-006-9029-9

Ravn, S., & Churchill, B. (2019). Making pathways? A mixed methods analysis of young women who have left school early in 'the new work order'. *Journal of Sociology*, *55*(4), 637-653. https://doi.org/10.1177/1440783319869831

Roos, A. L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2020). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 33*(2), 579–618. https://doi.org/10.1007/s10648-020-09543-z

Schmidt, A, & Thews, G (1989). Autonomic nervous system. In Janig, W. *Human Physiology 2 ed.* (pp. 333–370). New York, NY: Springer-Verlag.

Schührer, S., Bills, D. B., & Weiss, F. (2015). Gender differences in labor market entry and their long-Term consequences in the United States. In *Gender, Education and Employment*. Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781784715038.00014

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (200910). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410-417. https://doi.org/10.1109/titb.2009.2036164

Sinatra, G. & Pintrich, P. R. (Eds.) (2003). *Intentional conceptual change*. Erlbaum.

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology 65*(4): 283-294. https://doi.org/10.1027/1618-3169.56.4.283

Thacker, I., Copur-Gencturk, Y., & Cimpian, J. R. (2022). *Teacher bias: A discussion with special emphasis on gender and STEM learning.* Routledge. https://doi.org/10.4324/9781138609877-REE185-1

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*, 144–151. https://doi.org/10.1007/s11199-011-9996-2

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes, 53*(5-6), 339–370. https://doi.org/10.1080/0163853x.2015.1136507

Villanueva, I., Husman, J., Christensen, D., Youmans, K., Khan, M. T., Vicioso, P., ... & Graham, M. C. (2019). A cross-disciplinary and multi-modal experimental design for studying near-real-time authentic examination experiences. *Journal of Visualized Experiments*, (151), https://doi.org/10.3791/60037

Villanueva, I., Zorrilla, E. M., Husman, J., & Graham, M. (2021, July). Human-technology frontier: Measuring student performance-related responses to authentic engineering education activities via physiological sensing. In *International Conference on Applied Human Factors and Ergonomics* (pp. 338-345). Springer.

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational psychology review*, *29*(1), 119-140. https://doi.org/10.1007/s10648-015-9355-x

Wieselmann, J. R., Roehrig, G. H., & Kim, J. N. (2020). Who succeeds in STEM? Elementary girls' attitudes and beliefs about self and STEM. *School Science and Mathematics*, *120*(5), 297-308. https://doi.org/10.1111/ssm.12407