Advances in Engineering Education

2024: VOLUME 12 ISSUE 4

Improving Judgements of Performance Among Black Engineering Students Highly Susceptible to Stereotype Threat: An Intervention

KATHERINE PICHO-KIROGA
Department of Human Development & Psychoeducational Studies
Howard University
Washington, D.C.

MARCELA VALENCIA SERRANO
Department of Psychology
Pontificia Universidad Javeriana Cali
Cali, Colombia

NOEL BOURNE
Department of Mathematics
Tennessee State University
Nasheville, TN

JERRIEL HALL
Department of Human Development & Psychoeducational Studies
Howard University
Washington, D.C.

ABSTRACT

Stereotype threat (ST) is implicated as a contributory factor to attrition in Science Technology Engineering and Math (STEM) fields. One of the mechanisms by which ST degrades performance is by impairing metacognitive monitoring (Schmader et al. 2008), which is positively related to learning and performance (Hadwin et al., 2017; Winne & Hadwin, 1998, 2008). Several interventions to improve metacognition exist, but to our knowledge, none have evaluated their effectiveness in improving performance outcomes among ST susceptible students who are more likely to suffer situational deficits in metacognitive processing under threat.

The present investigation fills this gap in the literature through a brief intervention to improve performance judgments. 25 Black Engineering students participated in a four-week intervention to improve judgments of learning. Results showed that the intervention improved student calibration among Black engineering students susceptible to race-related ST and attenuated the negative relationship between calibration bias and performance. Specifically, study participants underestimated their performance, which was negatively and significantly associated with performance at the onset of the study. However, calibration bias diminished significantly over the intervention and subsequently had no impact on performance at four and six weeks (two weeks post-intervention).

Key words: stereotype threat, engineering, intervention, metacognitive monitoring, self-regulated learning

INTRODUCTION

The United States' global position in Science, Technology, Engineering and Mathematics (STEM) has experienced a steady but marked decline in the last decade. The nation currently ranks 38 out of 71 countries in math and science (Pew Research 2017) and 30 out of 35 among members of the Office of Economic Co-operation and Development (OECD). The current challenges faced in producing and retaining STEM talent are partly because the field has not fully accessed the human talent pool: Women and people of color (POC) have been and still are under-represented in fields like in engineering and computer science; Black and Hispanic people constitute only 5% and 6% of the Science and Engineering workforce relative to their participation in the U.S. workforce (i.e., 15% and 16%), respectively (NSF 2017).

The under-representation of people of color in science and engineering stem from a span of complex structural, cultural and psycho-social factors, all inter linked. These disparities are perpetuated in part by (a) cultural beliefs that raw talent (brilliance) is critical to success in STEM, and stereotypes that associate this brilliance with men (e.g., Bian, Leslie, & Cimpian 2017; Elmore & Luna-Lucero 2017; Kirkcaldy, Noack, Furnham, & Siefen, 2007; Lecklider 2013), as well as (b) stereotypes that disparage the intellectual abilities of women and non-Asian minorities in these domains. Regarding the latter, individuals belonging to marginalized groups can sometimes experience concern and anxiety over confirming as self-characteristic, negative stereotypes associated with their social group (Steele et al. 2002). This phenomenon, called stereotype threat, is experienced in situations that signal subtly, that a person, by virtue of belonging to a stigmatized group, is socially devalued (Steele et al. 2002). Despite the gradual entry of non-Asian minorities and women in science and engineering, stereotypes questioning their competence in these domains persist, with serious implications for long-term their engagement and achievement. Large scale, longitudinal studies also implicate ST as a deterrent to further involvement of females in STEM (Beasley & Fisher 2012; Woodcock et. al 2012) and as a contributory factor to the increased attrition of women and people of color from these disciplines (Beasley & Fisher 2012).

A primary mechanism through which ST degrades performance is by impairing metacognitive planning, monitoring and attention regulating processes (Forbes et al. 2008; Forbes & Leitner 2014; Forbes et al. 2015). Although several interventions to improve metacognition exist, none have examined their effectiveness among stereotype threat susceptible students, who are more likely to suffer situational deficits in metacognitive processing under threat. Likewise, interventions to remedy metacognitive deficits evoked by ST are also lacking in the ST literature, despite evidence in the broader metacognition literature that metacognitive skills such as those implicated in the ST process, are not only trainable but can also reverse the negative trajectory of performance among poorly self-regulated learners in several domains (MacDaniel et al. 2021; Theobald, 2021), including engineering (Cervin-Ellqvis et al. 2021; Cunningham et al. 2016; Lawanto et al 2017; Saez et al. 2018; Saez et al. 2020).

The present investigation fills this lacuna in both literatures (metacognition and ST) by assessing whether skills-based training to enhance metacognitive monitoring can translate to positive gains in mathematics performance for Black engineering students identified as highly susceptible to ST. Black and Brown students susceptible to stereotype threat represent the most vulnerable within the sub-population of ethnically underrepresented groups in fields like Engineering. As the deleterious impacts of ST on performance are mediated in part by metacognitive monitoring, interventions to remedy such metacognitive deficits experienced under threat could possibly attenuate short and long-term effects of the phenomenon (under-performance, domain disidentification and attrition), for students belonging to this vulnerable group.

The study focused on engineering students because metacognition, which is critical to the self-evaluation of one's knowledge and abilities (Paris & Winograd 1990), is essential in both mathematics (Carr & Biddlecomb 1998; Schoenfeld, 1992), and engineering (Case, Gunston & Lewis 2001; Lawanto 2010; Newell et al. 2004). Also, studies show that that training engineering students in metacognitive skills has been effective in improving their use of effective learning strategies and related metacognitive judgments regarding the effectiveness of these strategies (Kupriyanov et al. 2021; Manganello & Falsetti 2019; Lawanto et al. 2014; Santangelo et al. 2021; Sedraz et al. 2018; Zheng et al. 2016). Further, metacognitive processes aid engineers in identifying, defining and mentally representing problems, planning solution procedures, evaluating solution progress and the final solution (Davidson, Deuser & Sternberg 1994). Metacognitive strategies used in problem-solving include establishing task demands, formulating and executing action plans, finding similarities between problems (task knowledge), recognizing inconsistencies, and identifying constraints, to name a few (Meijer, Veenman & Hout-Waters 2006). All these strategies support one's ability to navigate and persist through solving ill-structured problems, which is at the heart of innovation in engineering. Therefore, honing metacognitive skills is essential to success as an engineer, and subsequently, to increasing the competitive edge of future engineers on the global landscape.

Mathematics performance was selected as a performance metric in this study because it is the gateway subject to the physical sciences. engineering, in particular involves the application of mathematical and scientific principles to solve real world problems (Lawanto & Febrian 2017; Sánchez-Arevalo et al. 2018). Secondly, metacognitive skills are not only inseparable from math knowledge (Schoenfeld 1985), but they are also essential to mathematical problem solving (Borkowski 1992; Carr & Biddlecomb 1998; De Corte, Verschaffel, & Op't Eynde 2000), and account for 37–42% of the variance in mathematics performance (Desoete et al. 2001; Desoete & De Craene 2019; Muncer et al. 2022; Ozsoy et al. 2011).

THEORETICAL FRAMEWORK

Pintrich's Model of Self-Regulated Learning

Metacognition refers to knowledge about one's cognitive processes (Pintrich 2000). Self-regulated learning (SRL) refers to the deliberate control of one's thoughts, feelings, and actions before, during and after a task to achieve valued goals (Zimmerman 2008). Pintrich (2000) posits SRL as a cyclical process that occurs in four time-ordered sequential phases during the completion of a task. These are: planning and goal setting, monitoring, control, and reflection / evaluation. Learners use these phases to regulate their own cognition, behavior, and affect (motivation), as well aspects of the environment where learning occurs.

In the area of cognition (the primary focus of this study), planning, monitoring, control, and evaluation are key elements of the metacognitive component of the SRL process, for they allow one to adapt or change cognitive strategies to foster student learning and achieve the academic goals (Pintrich et al. 1991; Pintrich & Zusho 2002; Zimmerman 2000, 2002). Planning involves allocating cognitive resources to facilitate efficient and effective problem solving and select appropriate problem-solving strategies. Monitoring refers to one's awareness or forethought of one's understanding and performance; it involves self-monitoring and regulating behaviors during tasks (Pintrich 2000; Schraw & Moshman 1995), more specifically, actively monitoring task progress against established standards and making performance-related self-judgments (i.e., judgements of learning or performance) during and after a task, that provide internal feedback for the learner. Finally, evaluation refers to fine-tuning and continuous adjustment of one's cognitive activities (Pintrich et al. 1991). It involves reflection on one's personal weaknesses and strengths relative to established goals, and how strategies can be improved upon to optimize performance on future academic tasks (Greene & Azevedo 2007; Pintrich 2000).

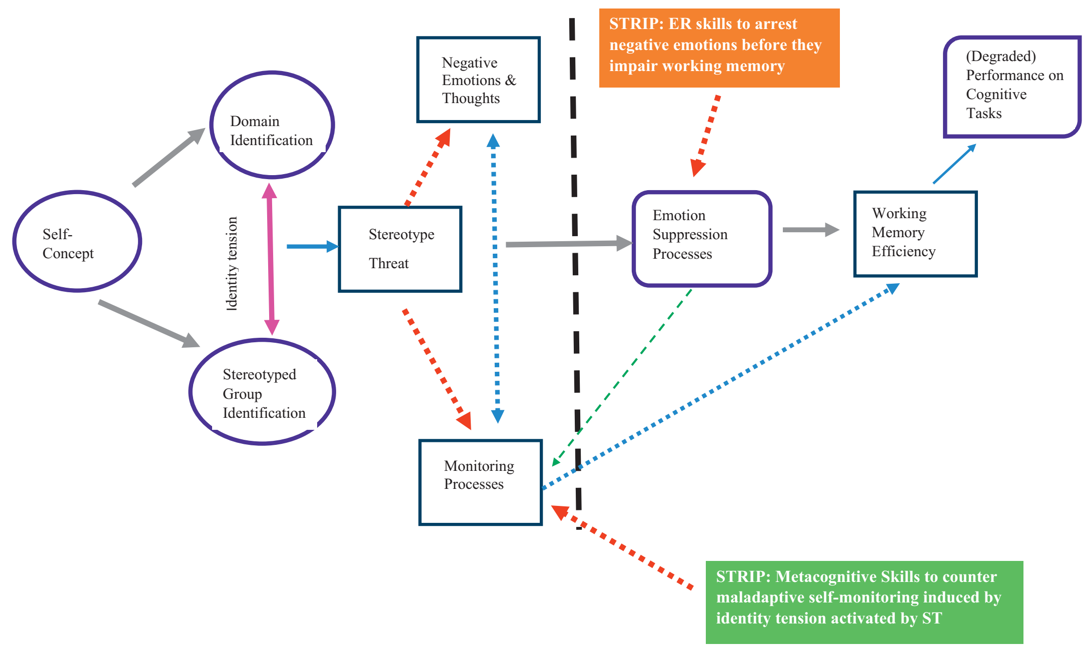

Figure 1. Stereotype Threat Reduction Intervention Program within the context of the Stereotype Threat Process Model.

Metacognitive Monitoring

Metacognitive monitoring allows learners to generate judgments that help them evaluate how well they are achieving the learning and performance goals that they set for themselves. These judgments become the starting point for learners to evaluate, change, or maintain the learning strategies that they use to solve problems. Thus, monitoring involves judgments of learning (JOLs) or monitoring comprehension (Flavell 1989; Nelson & Narrens 1990), which manifest when learners: (a) judge how well they are learning specific learning content; (b) monitor their understanding, by asking themselves questions as they perform activities like reading, writing or mathematical operations; (c) judge how well they can learn or remember after studying a specific material; and (d) assess how well they perform in particular academic task before and after performing it (Pintrich 2000).

Metacognitive Judgements of performance

Metacognitive judgments reference either judgments of learning or judgements of performance (Alexander 2013; Gutierrez de Blume 2022; Dunlonsky & Thiede 2013), which are distinguished as follows: JOL is refers to judgments made to evaluate how well specific content is understood or how well information can be recalled (Nelson & Narrens 1990; Kornell & Hausman 2017), whereas judgments of performance (JOP) refer to one's judgment about the possibility of doing well on a given task before or after performing it (Linchtestein et al. 1982; Rawson & Dunlonsky 2007; Schraw & Dennison 1994). Both types of judgments can be predictive (estimated prior to) or postdictive (estimated after) executing a task (Gutierrez De Blume et al. 2020; Nguyen et al. 2018).

Metacognitive judgments of performance reflect the gap between a learner's actual performance and his or her own judgment of the same (Schraw & Dennison 1994; Dunlonsly & Thiede 2013). These judgments offer insights on how well calibrated learners are, on a given subject. Calibration is typically assessed using calibration bias and calibration accuracy (Schraw 2009). Calibration bias occurs when there is a discrepancy between one's judgment of performance and one's actual performance. Thus, calibration bias refers to the degree to which individuals over or under-estimate their perceived judgments of performance (Dunlonsky & Rawson 2012; Kruger & Dunning 1999; Serra & Metcalfe 2009; Schraw et al 1993). Large discrepancies between students' estimated scores and their actual scores indicate high degrees of calibration bias and vice versa for small deviations. Calibration accuracy refers to the magnitude of judgment errors (Pajares & Graham 1999). It reflects the deviation of the absolute value of a bias score from 100 with higher values denoting smaller errors in judgement, and vice versa (Schraw 2009).

Studies show that calibration accuracy predicts student performance on exams (Gutierrez de Blume, 2022; Hacker et al., 2000; Koevoets-Beach et al. 2023; Nietfield et al. 2005). College students aren't always aware of their skills levels, however: Most tend to be poorly calibrated, and inclined to over or under-estimate their performance. This is especially so for low achieving students (Hawker et al. 2016; Fakcharoenphol et al. 2015; Morphew 2021; Tauber & Dunlosnky 2015), who tend to be poorly calibrated with respect to predictive and postdictive judgements of performance (Händel, et al. 2018). Multiple factors contribute to calibration accuracy and bias, among them, low self-efficacy and negative emotional reactions to the task (Niefield et al. 2005; Talsma et al. 2018), the level of difficulty of the task and teacher feedback (Kruger & Dunning 1999; Muis et al. 2016), learners' prior knowledge about (a) the task and (b) the most appropriate strategies to solve it, to name a few (Rawson & Dunlonsky 2007).

Stereotype Threat

Stereotype threat arises when one performs a challenging task and experiences anxiety over confirming negative stereotypes about the ability of their social group in a given domain (Steele 1997). Studies show that non-Asian ethnic minorities (Steele & Aronson 1995; Steele 1997; Armenta 2010) and women (Spencer, Steele & Quinn 2001) tend to experience stereotype threat (ST, hereafter) when reminded of negative stereotypes about their intellectual prowess, and quantitative ability, respectively. ST undermines performance in the short term (Steele 1997) and over time, chronic exposure to the phenomenon lowers sense of belonging to, and subsequently disengagement from the stereotype-relevant domain (Steele et al. 2002).

Most ST research has examined gender-related ST — the impact of negative gender stereotypes on the quantitative performance of women (Picho-Kiroga et al. 2021), although a reasonable amount of research on race-focused ST also exists. In typical race-ST experiments, Black or Hispanic students complete challenging tasks of verbal or quantitative reasoning under neutral or stereotype conditions where stereotypes alleging intellectual inferiority of these ethnic groups is made salient. A stereotype threat effect is present when members of these groups who are exposed to such stereotypes underperform compared to their White or Asian counterparts (Steele & Aronson 1995; Armenta 2010). Short-term, ST negatively affects performance, and over time, chronic exposure to ST leads to disidentification — a situation where the members of the stigmatized group disengage from and lose interest in the stereotyped domain (Aronson et al. 2001; Casad et al. 2018; Dennehy et al. 2018).

ST is not universal to an entire marginalized group; rather, within these groups, there is variability in level of susceptibility to ST based on essential conditions required for the phenomenon to occur: domain identification, group identification and stigma consciousness of the same, and lastly, negative emotions such as anxiety and worry, experienced in evaluative contexts (Steele 1997). According to theory, the sub-class of individuals who meet the essential conditions for ST are the most likely to be experience short and long-term effects of ST e.g., poor performance, and attrition from stereotype-relevant domains such as STEM. To our knowledge, this is the first study targeted towards students who meet the essential theoretical conditions for ST susceptibility (Steele 1997; Schmader et al. 2009).

Stereotype Threat and Metacognition

Theoretical models of the ST process have implicated three primary pathways through which the phenomenon impairs performance: physiological stress responses that impair prefrontal processing, active metacognitive self-monitoring which affects regulation of attention, and efforts to suppress negative thoughts and emotions that arise during cognitive tasks (Schmader et al. 2008). To our knowledge these pathways have been little explored in a handful of studies (and hardly any interventions): Two experimental studies have examined emotion regulation and to our knowledge, none have looked at metacognitive monitoring. Results from the former provide evidence to support that using adaptive emotion regulation strategies like cognitive reappraisal can, in fact, reverse ST performance deficits observed when emotion suppression is deployed (Johns et al. 2008; Logel et al. 2009). ST disrupts performance by activating self-concepts relative to one's (a) social (stereotyped) group, and (b) ability in the stereotype-relevant domain; it also activates propositional links between these self-concepts, which underlie the ST experience. ST occurs in the presence of a negative

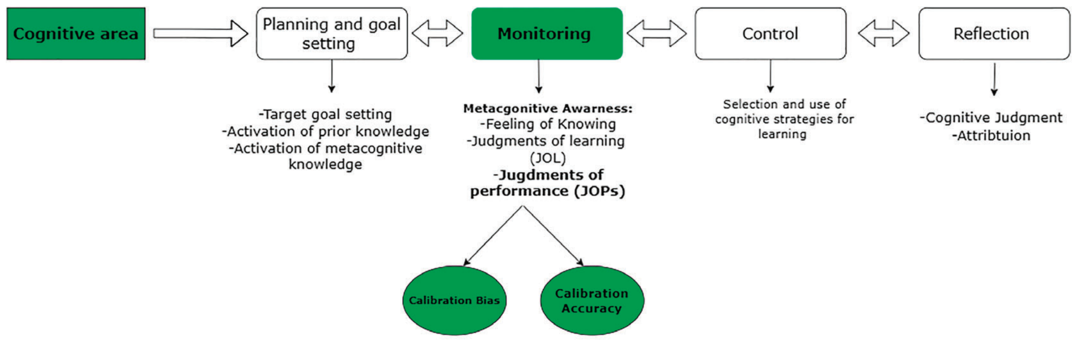

Figure 2. Pintrich's Model of Self-Regulated Learning. Adapted from Pintrich, R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds). Handbook of Self-Regulation (pp. 451-502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

propositional link where these aspects of self-concept are defined in opposition to another i.e., *my group does not have this ability, I am like my group, but I think I have this ability.* The negative link induces a state of cognitive imbalance that elicits (a) an acute physiological stress response (over confirming the stereotype), and (b) increased vigilance (self-monitoring) to internal and/ external cues that may help disambiguate this conflict. The latter increases the individuals' focus on his/herself and on their performance; for those susceptible to ST who are already concerned over confirming the group stereotype, and a motivated to disconfirm the same, which translates to a heightened vigilance to detect signs of failure (Schmader et al. 2008). This, however, taxes working memory and degrades the ability to regulate attention during complex tasks, subsequently impairing task performance under stereotype threat. Thus, under ST, metacognitive monitoring is misdirected to focus on alleviating stereotype-based concerns rather than outcome-based concerns (strategies to enhance successful problem solving), with negative consequences for performance.

LITERATURE REVIEW

Self-regulated Learning Intervention Research

Interventions to improve judgements of performance (or learning) have been conducted with freshmen, junior and sophomore students, mostly white and between ages of 18-21 years. The interventions also vary in design; some provide student feedback on their self-judgments of performance on exams regarding the magnitude of errors in their judgements (Callender et al. 2016; Neitefield et al. 2006; Urban & Urban 2019) and others focus on teaching the importance of calibration, the dangers of poor calibration bias, how to select appropriate study strategies and evaluate the usefulness of strategies used to understand the subject matter (Bruin et al. 2017; Osterhage et al. 2019; Morphew 2021).

There are also other interventions that offer practical applications to practice judgments of performance in class unit activities with opportunities to receive feedback on the accuracy of these judgments. Calibration bias – over-estimation specifically, declines significantly in interventions that combine teaching about judgements of performance (calibration bias and accuracy) with practice and feedback, compared to interventions where only practice and feedback are provided (Handel et al. 2020).

Studies show that over time, student practice with judgments of performance, coupled with feedback on the accuracy of these judgments substantially reduces calibration bias (Fakcharoenphol et al. 2015; Morphew et al. 2020; Tauber & Dunlosnky 2015), and consequently improves academic performance (Baars et al. 2014; Rawson & Dunlonsky 2007; Testa et al. 2023). These results have been consistent in ecological settings e.g., classrooms (Callender et al. 2016; Handel et al. 2020; Neitefield et al. 2006; Urban & Urban 2019) and among STEM students (Bruin et al. 2017; Hawker et al. 2016; Morphew 2021; Osterhage et al. 2019). However, despite the consensus that monitoring can be improved by enhancing self-judgments of learning and performance i.e., student calibration, (Mathabe et al. 2014; Foster et al. 2017; Nietfeld & Schraw 2002), the impact of calibration on performance appears to vary for high and low performing students. Interventions to improve student calibration tend to work better for high but not low-performing students (Hacker et al. 2000; Nietfeld et al. 2005; Foster et al. 2017; Gutiérrez & Price 2016), even in undergraduate disciplines like chemistry (Hawker et al. 2016; Mathabe et al. 2014; Pazicni & Baur 2014) and biology (Ziegler & Montplaisir 2014). These interventions appear particularly ineffective for low-performing students, who are inclined to overestimate their performance (Osterhage et al. 2019; Morphew 2021; Bruin et al. 2017). This is likely because low-performing students are less likely to use the results of previous performance on tests or exams to make new predictions on subsequent assessments. As such, their predictions about performance are likely to be influenced by their desired or aspired grades rather than by their actual performance (Saenz et al. 2017; Persky et al. 2020). Therefore, low performing students are likely to benefit from interventions that not only include practice on self-judgments, but also integrate instruction on selecting relevant cognitive study strategies, exercising awareness of when changes in study strategies are needed, evaluating strengths and weaknesses as learners, and lastly, increasing metacognitive awareness (awareness of progress towards learning goals) to improve understanding and performance (Emory & Luo 2022; Cogliano et al. 2021; Foster et al. 2017).

Stereotype Threat and Self-Regulation

Studies show that metacognitive monitoring is enhanced when individuals learn to gauge their performance on tasks against established standards more accurately, which in turn improves student performance (Bol & Hacker 2012; Crespo 2004; Gutierrez de Blume 2022; Hacker et al. 2000; Huff & Nietfeld 2009; Kim 2018; Morphew 2021; Nietfeld et al. 2005; Nietfeld et al. 2006; Thiede et al. 2003). However, empirical studies show that ST diminishes one's capacity to generate problem-solving strategies on difficult quantitative problems i.e. metacognitive planning (Quinn & Spencer 2001). It also compromises monitoring processes by redirecting their focus from evaluating one's performance relative to task goals, to (a) hypervigilance to error feedback (Forbes et al. 2008), discrimination (Kaiser et al. 2006) and (b) gauge whether one's behavior or performance is consistent with the negative group stereotype (Schmader et al. 2008). This impacts attentional regulation, and co-opts limited working memory, which subsequently impairs task performance. Electroencephalography (EEG) studies reveal robust neuro- responses of the brain to stereotype threat in the areas of metacognitive monitoring. Findings from these studies show that stereotype threat affects temporal regions of the brain that play a prominent role in performance monitoring processes (Forbes & Leitner 2014), and coactivates neural networks in the brain responsible for attention regulation (Forbes & Leitner 2014; Forbes et al. 2015) which impair working memory, and hence performance. ST and metacognition are thus interlinked, as the former disrupts performance by impairing metacognitive (self) monitoring, which plays a role in optimizing cognitive performance. Impediments to metacognitive processes can have a negative impact on performance, as demonstrated by the ST process model (Schmader et al. 2008). It follows then, that efforts to remedy deficits in metacognitive monitoring and to attenuate emotion suppression and stress induced during ST, might be effective in abating effects of the phenomenon altogether, particularly among those susceptible to the phenomenon.

To our knowledge, there exists one intervention to address metacognitive monitoring in relation to ST. The Stereotype Threat Risk Reduction Program (STRIP, hereafter) is an evidence- and skills-based intervention designed to address two of the three previously mentioned mediating mechanisms of ST: metacognitive self-monitoring and emotion suppression.

The STRIP Intervention

STRIP is a dual model for ST reduction based on Schmader et al.'s (2008) expanded ST process model. It is a 12-hour, skills-based intervention which consists of knowledge and skills building. The knowledge component of the intervention teaches students about ST, its processes and impact; it also describes metacognition and emotion regulation, making connections between these constructs and ST. Specifically, how they can be applied effectively to minimize ST. Knowledge, being a secondary focus of STRIP, accounts for a smaller proportion of the intervention i.e. 25% (or three of 12 hours) of the curriculum time, delivered didactically through lectures, and the use of video clips. The delivery format for skills training includes but is not limited to group discussions, scaffolding, and role play, to mention a few.

The skills building component is designed to counteract two of three previously mentioned mediating mechanisms through which ST impairs performance (self-monitoring and emotion suppression), by building metacognitive monitoring and adaptive emotion regulation skills (e.g., cognitive reappraisal), to minimize metacognitive deficits, and supplant and attenuate emotion suppression and its effect on working memory under threat. The program covers emotion regulation skills such as cognitive reappraisal and psychological distancing, which arrest negative emotions before they can be destructive. Findings from prior research show that cognitive reappraisal is effective in reversing the deleterious impact of emotion suppression on the performance of marginalized groups under conditions of stereotype threat (Johns et al. 2008; Logel et al. 2009).

Intervention Action Pertinent to Metacognitive Skills Training. The literature indicates that self-monitoring positively impacts performance. During ST however, self-monitoring is misdirected to focus on alleviating stereotype-based concerns rather than outcome-based concerns (i.e., strategies to enhance successful problem solving). This degrades the ability to regulate attention during complex tasks, leading to suboptimal performance. The metacognitive component of STRIP-- the focus of this investigation-- addresses this deficit by re-directing self-monitoring to focus on problem solving strategies which facilitate rather than impede efficient cognitive processing.

STRIP Knowledge Component. This focuses on honing metacognitive monitoring skills to improve student calibration. The knowledge component of STRIP teaches students to (a) accurately judge task performance, (b) identify gaps in their knowledge (c) develop strategies to close these gaps (d) assess their own monitoring skills during task performance. Here, calibration is taught by integrating feedback in a quantitative exercise as follows: First, students complete a calculus test in which they are asked to estimate their performance at the end of the exercise. Using test scores from the calculus test, students are taught to compute calibration bias and accuracy scores. They are then introduced to four different classifications of learners along the dimensions of (a) level of skill and (b) awareness of skill- levels shown in Table 1 below i.e., students who are skilled and aware, skilled and unaware, unskilled and aware, unskilled and unaware. Student characteristics associated with each classification are presented in relation to tendencies to calibrate accurately, or inaccurately. Further connections are made between the magnitude of calibration bias and student classifications in each quadrant. Lastly, the researcher provides detailed feedback to students on (i) which quadrant they belong, based on their actual performance of the task, and related student-computed calibration scores, (ii) classification (quadrant)-specific strategies to improve performance, and lastly, (iii) the researcher and student work together to set performance goals for the next task.

STRIP Skills Training Component. One way good and poor problem solvers are differentiated is by their ability to reflect on and regulate their problem-solving activities (Davidson & Sternberg 1998). Metacognitive processes aid engineers in identifying and defining problems, mentally representing problems, planning solution procedures, and evaluating solution progress and the final solution (Davidson, Deuser & Sternberg 1994). Other metacognitive strategies employed in problem-solving include establishing task demands, formulating and executing action plans, finding similarities between problems (task knowledge), recognizing inconsistencies and confusion, identifying constraints, switching from one representation to another, activating prior knowledge, and assessing problem difficulty (Meijer, Veenman & Hout-Waters 2006). These metacognitive strategies support one's ability to navigate and persist through solving ill-structured problems, which is central to engineering. The metacognitive skills component of STRIP emphasized training of these skills with the primary objective of assisting learners in checking and correcting their behavior as they proceeded on a task. This component of the intervention was grounded in an engineering design task, which required students to create a window device that could help the elderly open a window with ease. The design task was used to scaffold training and enhance student practice of the aforementioned metacognitive skills, embedded in the SRL cycle of planning, monitoring and evaluation.

Additionally, skills pertinent to consciously directing attention to focus on selective information to magnify the experience of select stimuli (attentional regulation), are taught. Here, students are trained to focus on the foregoing metacognitive skills, to facilitate redirecting focus from the threatening stereotype to the task at hand.

THE CURRENT INVESTIGATION

The present study examined whether an intervention to improve judgements of performance could impact the performance of Black engineering students identified as highly susceptible to race-focused ST. We sought specifically to address two primary questions: For engineering students highly susceptible to race-related ST, (1) Does an intervention to improve judgments of performance impact student calibration, and if so (1b) what is the relationship between calibration and performance? Thus far, no formal empirical investigations have been conducted to examine whether improving metacognitive control can also improve the performance of individuals identified as susceptible to ST using STRIP or otherwise. The present investigation focused solely using STRIP to investigate what impact if any, intervening on metacognitive monitoring could have, on the performance of ST susceptible engineering students.

METHODS

Sample

Participants were 25 first year college students majoring in engineering at a mid-size private historically black college and university in the Northeastern United States. The university has a student body of approximately 10,000 students, predominantly (99%) Black and undergraduate (84.2%). 10.8% of undergraduate students major in Physical Sciences, with a relatively high proportion of females (46.4%). 8.5% of faculty at the university faculty are in the Physical sciences, which is also predominantly male (83%) and Black (66%). 63% of the study sample was female. Participants were paid $40 for their participation in the study. All participants had been previously identified as highly susceptible to stereotype threat regarding race (i.e., stereotypes that Black people are intellectually inferior) using a latent profile analysis.

Measures

The Social Identities and Attitudes Scale, SIAS (Picho & Brown 2011). The SIAS is a 30-item psychometrically sound measure of ST susceptibility, independently validated with different populations i.e., high school, STEM and non-STEM college students (Picho et al. 2021; Cokley et al. 2015; Cromley et al. 2013; Chun et al. 2016; Smith et. al. 2016). The instrument assesses six key constructs identified by theory (Schmader et al. 2008) as critical to the ST process: negative emotions (affect), domain identification, group (race and gender) identification and stigma consciousness of the same. Items are continuous and measured on a seven-point Likert scale where 1 = strongly disagree, 2= disagree 3= somewhat disagree 4 = neither agree nor disagree, 5= somewhat agree 6= agree, and 7 = strongly agree such that higher mean scores on each subscale indicate higher levels of the latent construct. The focus on race-related ST informed the decision to use only group identification and stigma consciousness factors related to ethnicity and not gender; thus, only math and ethnic identification, ethnic stigma consciousness, and negative affect subscales of the SIAS were used to classify participants into ST susceptibility classes. Sample items from math and ethnic identification, as well as ethnic stigma consciousness, include: *My ethnicity is an important reflection of who I am, My ethnicity affects how my peers interact with me, My ethnicity affects how I interact with people of other ethnicities, Math is important to me, and my math abilities are important to my academic success.* Scale reliabilities were very strong (math identification $\alpha = .89$, ethnic stigma consciousness $\alpha = .87$, ethnic identification $\alpha = .85$ and negative affect □ = .92).

The memory updating task. Participants completed a working memory updating task (Schmiedek, et al. 2009), which consisted of ten arithmetic problems. The items exerted an even mix of medium and high cognitive demand on working memory. Each item required four responses for a total of

40 answers. Therefore, results from this task were scored as total number correct, with a minimum score of 0 and a maximum score of 40.

Total completion time for each task is 10–12 minutes.

Procedures

Students were recruited from an 'Introduction to Engineering Design' course which is a required course for engineering majors, as part of a larger study (N=160). Students who consented to participate completed the SIAS scale in Qualtrics. 136 Black students completed the SIAS, and their responses were subjected to a latent profile analysis (LPA), which was used to classify participants into different profiles of susceptibility to race-related stereotype threat.

Latent Profile Analysis

Latent profile analysis is a statistical technique that classifies individuals into qualitatively different subgroups within a population based on similarity in pattern responses to measures of latent constructs. Thus, LPA identifies unobserved group membership such that individuals within a subgroup tend to be similar to one another but differ qualitatively from those in a different sub-group (Geiser 2013). The process is called latent class analysis when membership is based on categorical data and latent profile analysis when classification into latent groups is based on responses to continuous measures of latent construct (Samuelson & Dayton 2010).

Stereotype threat susceptibility typologies. ST susceptibility was assessed using a stereotype threat susceptibility measure, the social identities and attitudes scale, SIAS, (Picho & Brown 2012), which measures latent constructs that align with ST susceptibility classes postulated by ST theory (Steele 1997) and process models (Schmader et al. 2008). These typologies are stratified by one's level of identification to the domain and stigmatized group (Steele 1997), corresponding chronic awareness of stigmatized status (Steele 1997; Schmader et al., 2008) and negative emotions induced during threat (Schmader et al. 2008) such that: High susceptibility to ST individuals rank highly on the above constructs, while those classified as low ST susceptibility rate highly on domain identification but low on all other constructs. The non-risk, disidentified group are a product of chronic ST exposure, who have disidentified from the stereotype relevant domain to preserve their self-esteem. The profile of this is similar to that of the high ST susceptible class but for low domain identification. Finally, the non-risk unidentified group constitutes those immune to ST because they neither identify with the stereotyped domain nor their social group, which renders group stereotypes self-irrelevant (Steele 1997).

SIAS data from the broader sample ($N = 160$) were subjected to LPA (details of LPA classification and validation are reported elsewhere, see Picho & Kearse 2024), and ST classification was reported for the entire sample. However, only data from participants who self-identified as Black ($n = 136$) were used to recruit participants for the current study. Here, 40 students classified as highly susceptible to ST were invited to participate in the study. 35 completed baseline measures and attended the first session but attrition due to school challenges brought on by the COVID-19 pandemic was such that of only 25 participants had more than baseline data recorded.

LPA results showed that individuals classified as highly susceptible to race-related ST also performed less well on the working memory task compared to their low ST susceptibility counterparts. Participants also completed a ten-item working memory updating task, which was scored and used to validate latent classification into ST susceptibility groups. Therefore, students classified as highly susceptible to ST ($n = 75$), whose classification was validated by performance on the working memory task, were invited to participate in the present study. 40 accepted the invitation to participate, but only 25 completed the baseline survey. Participants also completed a math task in Qualtrics one week prior to the intervention. Participants were also asked to estimate their performance on the task upon completion. These data were used to establish baseline scores on calibration bias and accuracy.

Metacognitive Intervention Action. Students participated in a five one hour-session intervention conducted once a week over five weeks, where they learned metacognitive skills previously mentioned. The intervention was facilitated by the first author. The intervention was two pronged and consisted of knowledge and skills training, detailed earlier. Students met for one hour each week; they also completed weekly online journals. Calculus and algebra quizzes were administered online via Qualtrics at baseline, and during weeks 3 and 4 of the intervention and two weeks after the intervention (week 6). At the end of each quiz, participants were asked to provide a single estimate of their performance on the task and rate the level of confidence in their rating.

De-identified data were presented for independent scoring to a mathematics doctoral student with no knowledge about the study aims. Performance estimates were used to calculate calibration bias and accuracy scores based on data collected at baseline, during and after the intervention. These calculations were computed at the end of the study.

Calibration bias and accuracy: Scoring and interpretation. For calibration bias scores, positive and negative values indicated over- and under estimation of performance, respectively. When actual and estimated performance match, no bias exists, and perfect calibration is achieved. Hence a score of 0 would indicate perfect calibration as it would mean a match between one's estimated and actual performance score. When performance estimates exceed actual performance then bias in judgments of performance reflect over estimation; similarly, under-estimation is reflected when performance estimates are lower than actual performance (Serra & Metcalfe 2009).

Regardless of the sign (positive or negative), the magnitude of bias determines how well or poorly calibrated one is in one's judgement of learning. Smaller bias scores imply better calibration in judgements of learning, whereas larger scores suggest the opposite. For example, bias scores of 3 and –3 are the same in terms of magnitude of bias. Both suggest strong judgments of learning (good calibration). However, considering bias scores of 3 vs. 40, whereas both represent over estimation in performance on a task, the magnitude of bias is significantly larger in the latter bias score, pointing to relatively poorer calibration of the student with that score.

Calibration accuracy was computed by subtracting the absolute value of the calibration bias score from 100%. As calibration accuracy scores range from 0–100%, scores closer to 100 would suggest small errors in judgement regarding one's performance (stronger calibration accuracy), and vice versa. Taken together, poor calibration in judgements in learning are denoted by large bias and small accuracy scores, and vice versa. It should be noted that low performing students can be well calibrated if they are able to gauge with relative precision- their performance on a task. As such, calibration scores do not indicate how strong a student is, but rather how well calibrated they are in judging their relative strengths and/ or weaknesses in learning or performance.

RESULTS

Data from 7 participants were removed from analyses because they only completed baseline tasks and there were no comparison data for them- leaving a final sample of 18 students. Descriptive statistics of student calibration scores over eight weeks are provided in table 1.

Descriptive Analysis

All analyses were conducted in Stata 15. Participants in this study had high calibration accuracy scores at the onset, which did not change much over the course of the study. Results presented in Table 1 below show very high calibration accuracy scores at baseline and over time, with little variability around these means, indicating small errors in judgements of performance. Calibration bias scores were highest and recorded the most variability in at baseline, prior to the intervention. Thus, on average, participants tended to underestimate their performance prior to the study (although

Table 1. Calibration bias and accuracy scores.

	Calibration Bias			Calibration Accuracy		
	M(SD)	Min	Max	M(SD)	Min	Max
Baseline	–.56 (2.77)	–5	6	98 (1.94)	94	100
3 - Weeks	–1.89 (2.42)	–6	1	97.89 (2.20)	94	100
4 -Weeks	1.39 (1.95)	–1.72	5	98.2 (1.53)	95	99.7
Posttest	–1.33 (2.42)	–4	2	98 (1.79)	96	100

these estimates were also very small), and much greater variability in underestimation existed among participants prior to the intervention. Trends in accuracy scores remained consistently high over time, whereas changes in calibration bias reflected improvements from baseline to posttest.

Did the intervention impact calibration accuracy and bias?

To examine whether calibration bias and accuracy were impacted by the intervention, calibration data, collected over four time points i.e., before, (twice) during and after the intervention, were subjected to the Skillings-Mack test, which is a non-parametric analog to repeated measures one-way Analysis of Variance (ANOVA) (Skillings & Mack 1981). It is considered to be a generalization of the Friedman test, used in cases where missing data exist. The Skillings-Mack test is robust to violations of normality and accounts for missing data (Chatfield & Mander 2009). This test was used because of the relatively small sample size ($n < 25$), and also because missing data existed for some participants, at various points of the intervention and after.

Intervention effectiveness was operationalized as a reduction in calibration bias and improvement in calibration accuracy scores from pre to posttest. Changes in calibration bias and accuracy over time were examined using the Skillings-Mack test. Here, time was specified as the independent variable; bias and accuracy were specified as dependent variables in two separate tests. Results revealed a statistically significant difference in calibration bias over time $\chi2(3) = 8.05$, $p = .02$; generally, the weighted sum of centered ranks decreased over time – indicating a reduction in calibration bias over time. There were no statistically significant differences in calibration accuracy over time $\chi2(3) = 1.642$, $p = .62$, however.

Examining relationships between calibration and performance across time

Correlations analyses were conducted to examine relationships between bias and accuracy scores and performance across time. Here, we examined correlations between bias and accuracy scores collected (a) at baseline (b) at two time points during the study, and (c) post intervention. The correlations were analyzed separately for each time point. at the onset, during and after the intervention. Results revealed no significant relationships between calibration accuracy and performance over time. However, calibration bias was negatively related to performance at baseline and in the third week of the intervention ($r = -.61$, $p = .01$ and $-.68$, $p = .04$, respectively), but not significantly so at the end of the intervention, and two weeks post intervention (all ps $>.05$).

DISCUSSION

The present study examined whether a brief intervention on judgments of performance could improve the performance of Black engineering students identified as highly susceptible to race-related

ST. Results showed very high calibration accuracy at baseline, which remained stable during and after the study, yielding no significant changes on this construct over time. That changes in calibration accuracy were not significant seem reasonable because of ceiling effects. Given average accuracy scores of 98% at the onset of the study left little room to improve. The stability in calibration accuracy is consistent with findings found in pioneering studies conducted with undergraduate psychology students in authentic classroom settings (Nietfield & Schraw 2002; Nietfield et al. 2005), which showed that, over the course of a semester-long intervention, there was no change in judgment accuracy.

Results also revealed large variability in calibration bias at baseline, primarily characterized by a tendency to under-estimate performance, which reduced significantly over time. That participants underestimated their performance at baseline does appear inconsistent with previous studies showing that students, especially low performers, tend to overestimate their performance (Fakcharoenphol et al. 2015; Gutierrez de Blume 2022; Nietfield et al. 2005; Tauber & Dunlosnky 2015). A plausible explanation for this deviant result is that participants constituted a specialized group different from what has been previously studied in the literature: students highly susceptible to race-related ST. Generally, students who underestimate their performance have trouble with developing awareness of what they know and what they are capable of (Talsma 2018; Kezia et al. 2021) probably because they present low self-efficacy and high levels of anxiety when performing tasks, which can lead to erroneous judgments about performance (Ehrlinger & Dunning 2003), negatively affecting other components cognitive areas of SRL such as planning (Blackmore et al., 2021; Zimmerman, 2002). Studies also show that students who underestimate their performance are so focused on the negative thoughts about their ability, which significantly curbs the amount of attention that they pay to the demands of the task, which subsequently impairs performance (Kezia et al. 2021).

ST elicits concern and anxiety over confirming as true, the negative stereotype of their social group (Steele 1997), which increases mind-wandering and subsequently a marked decrease in attention on the task (Mrazek et al. 2011). The phenomenon also elicits feelings of self-doubt, and increases the prevalence of negative thoughts (Cadinu et al. 2005), which usurp limited working memory resources required for efficient cognitive processing and problem-solving, subsequently resulting in diminished task performance (Schmader et. al. 2008). Taken together, it is possible that the anxiety and performance related worries could lead students who are highly susceptible to stereotype threat to under-estimate their performance on a task, and the reduced attention induced by mind-wandering on tasks, for this group of students likely impacted their metacognitive monitoring on tasks prior to the intervention.

Fortunately, calibration bias decreased over time, indicating that students' judgements of performance improved significantly from baseline to posttest. These findings suggest that the intervention positively influenced the calibration of engineering students identified as highly susceptible to ST. This result corresponds with previous research in which interventions conducted with college students improved calibration bias (Händel et al., 2020: Mathabe et al. 2014). It does, however, contradict other research where such interventions either had no impact on calibration bias (Emory 2022) or had the opposite effect i.e., increased over-estimation post intervention (Foster et al. 2017). It remains unclear why mixed findings on calibration bias exist, although a plausible explanation for this discrepancy could be that in the foregoing studies (i.e., Emory (2022) and Foster et al. (2017)), students reported a tendency to overestimate their performance at baseline (pre-test), whereas in the present investigation, the reverse was true. Thus, results from this study suggest that calibration bias could possibly be reduced more efficiently in interventions with students who tend to under-estimate rather than to over-estimate their performance.

The improvement in students' judgement of performance was also corroborated by the observed relationship between bias and performance over the duration of the study; the moderate, negative relationships between bias and performance observed at baseline and three weeks into the study disappeared at the end of the study and remained so even two weeks after the intervention. This result confirms findings from previous studies which found that when performance judgments are improved and more calibrated, students tend to perform well academically in both controlled (Baars et al. 2014; Rawson & Dunlonsky 2007) and authentic environments (Callender et al. 2016; Handel et al. 2020; Neitefield et al. 2006; Testa et al. 2023; Urban & Urban 2019).

Limitations and Recommendations for Future Research

The current investigation was based on Black engineering students identified as highly susceptible to race-related ST so study findings are specific to this demographic and might necessarily generalize to engineering students overall. Additionally, as this is the first study to investigate self-regulation among ST susceptible students, it is recommended that more, confirmatory studies with ST susceptible students are conducted to test and possibly validate these findings.

The present study lacked a control group due to the relatively small participant sample size. While the exploratory, qualitative nature of the present investigation justifies the sample size, it is recommended however, that future studies include control groups. This would allow comparisons in the pattern of results between intervention and non-intervention groups, strengthening the validity of study findings. Additionally, the integration of feedback in judgments of performance was not adequately covered in the current intervention because the sudden switch to online learning mid-semester due to COVID-19 presented logistical challenges to a successful implementation. Informative feedback on performance associated with specific calibration exercises was only provided for the first calibration exercise. This resulted in a missed opportunity for participants to reflect on the degree to which their judgements of performance were close or far from their actual performance and why at different points during the intervention. Although interventions without feedback have certainly been applied in previous studies, we believe that this study of a such a unique group of students (high risk for ST) could have benefitted even more from having feedback integrated into the curriculum. Curriculum-integrated feedback takes place during class activities, which implies that it is the teacher who discuss feedback with students on their performance, on how well they were calibrated in estimating it and on which strategies they can use to improve their learning and performance. This offers the opportunity for educational researchers to work collaboratively with teachers to strengthen their feedback processes in the classroom, so that they can support students in improving metacognitive monitoring, specifically, performance judgments. Generating reflections on performance does support the learner to transform elements such as self-efficacy and understanding of possible failures and successes, which help significantly improve calibration and performance (Händel et al. 2020; Perky & Dinsmore 2019). A performance feedback and reflection component could have complemented the instruction provided to strengthen the planning, monitoring, and evaluation processes during the intervention.

CONCLUSION

The present study was motivated by a desire to explore SR through the lens of ST and examine whether skills-based SR training could improve the performance of students highly susceptible to ST. Preliminary evidence appears to suggest that the negative relationship between calibration bias and performance can be mitigated through an intervention to improve judgements of performance among students highly susceptible to the phenomenon. Consequently, this opens an avenue to further explore how and whether the development of metacognitive skills can moderate, specifically minimize, ST effects among those vulnerable to the phenomenon. To that end, further experimental research in this area is encouraged to assess the relative strength or impact of such interventions in attenuating the performance of this sub-group of students under ST conditions.

Most interventions to improve self-regulated learning in engineering students have focused on metacognitive awareness and evaluation, and motivational aspects of achievement like self-efficacy and academic goal orientation (Sáez et al. 2018; Sáez et al. 2020). None, to our knowledge, have, specifically, addressed judgments of performance despite evidence to suggest a strong, positive impact of metacognitive monitoring on self-regulation (Manganello & Falsetti 2019; Schraw, Crippen, & Hartley 2006; Sedraz et al. 2018; Zheng et al. 2016) and performance (Capote et al. 2017; Gadella et al. 2020; Gaeta 2016; Ramírez et al. 2016; Lawanto et al. 2014; Zambrano 2016). Results from this study contribute to fill this knowledge gap in the literature; they also lay the foundation for further inquiry into how, through the strengthening of metacognitive processes such as monitoring, students at risk of stereotype threat can be supported to make more calibrated judgments about their performance, to facilitate improvements in self-efficacy, learning and performance in the field of engineering.

Lastly, it would be remiss not to point out a need for researchers and practitioners to address the systemic issues that contribute to ST in the first place. Whereas ST is a psychosocial phenomenon, it does not occur in a vacuum; it is a highly contextual phenomenon that relies on cues in the environment to make denigrating stereotypes about social groups salient, and subsequently trigger the phenomenon. Black and Latino students switch out of STEM majors before obtaining a degree at higher rates (40% and 37%, respectively) compared to their White counterparts (29%) (EAB 2019). Discrimination, bias, isolation and exclusion in STEM fields (occupational and academic) have been implicated as primary reasons for these high attrition rates (EAB 2019). Intellectually hostile academic and organizational climates that subtly or overtly question the ability and competence of women and non-Asian minorities, have been cited as key factors responsible for their disproportionate departure from STEM (Funk & Parker 2018). For example, women abandon engineering in college because they are negatively stereotyped by peers, professors, and in internships (Seron, Silbey, Cech & Rubineau 2016). In the engineering workforce women, Black and LatinX professionals report that discrimination, and prolonged stress from gendered (or non-technical) tasks, which devalue their contributions and impede career advancement (Seron et al. 2016; Funk & Parker 2018), as primary reasons for exiting the field. Thus, the burden to alleviate the impact of psychosocial-contextual phenomena like ST on the performance and long-term engagement of underrepresented groups in STEM should not rest solely on building ST resilience among susceptible individuals, but also involve efforts to address sociocultural factors such as bias and discrimination that contribute to the problem, which are often systemic, embedded within academic and work cultures, and sometimes inadvertently sustained by hidden curricula and/or inequitable organizational policies, respectively. To that end, significant resilience to ST is best facilitated and supported by fostering identity-safe, and intellectually non-threatening environments to cultivate a sense of belonging to fields like STEM where so few members of marginalized groups are represented.

REFERENCES

Adekitan, A., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon 5*, 3–21. 10.1016/j.heliyon.2019. e01250

Al-Sheeb, B., Magid, A., & Abdella, G. (2018). Investigating Determinants of Student Satisfaction in the First Year of College in a Public University in the State of Qatar. *Education Research International, 1*, 1–14. https://doi.org/10.1155/2018/7194106

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition, 39*, 171–184. https://doi.org/10.3758/s13421-010-0002-y.

Baars, M., Vink, S., Van Gog, T., De Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction, 33*, 92–107. https://doi.org/10.1016/j.learninstruc.2014.04.004

Blackmore, C., Vitali, J., Ainscough, L., Langfield, T., Colthorpe, K. (2021). A Review of Self-Regulated Learning and Self-Efficacy: The Key to Tertiary Transition in Science, Technology, Engineering and Mathematics (STEM). I*nternational Journal of Higher Education, 10*(3), 169–177.

Bol, L., & Hacker, D. (2012). Calibration research: where do we go from here? *Frontiers in Educational Psychology, 3*, 1–6. https://doi.org/10.3389/fpsyg.2012.00229

Bruin, A., Kok, E., Lobbestael, J., De Grip, A. (2017). The impact of an online tool for monitoring and regulating learning at university: overconfidence, learning strategy, and personality. *Metacognition and Learning, 12*, 21–43. 10.1007/s11409-016-9159-5

Callender, A., Franco-Watkins, A., & Roberts, A. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning, 11*, 215–235. https://doi.org/10.1007/s11409-015-9142-6.

Capote, G., Rizo, N., & Bravo, G. (2017). La autorregulación del aprendizaje en estudiantes de la carrera ingeniería industrial. *Universidad y Sociedad, 9*(2), 44–52.

Cervin-Ellqvist, M., Larsson, D., Adawi, T., Stöhr, C., & Negretti, R. (2021). Metacognitive illusion or self-regulated learning? Assessing engineering students' learning strategies against the backdrop of recent advances in cognitive science. *Higher Education, 82*, 477–498 https://doi.org/10.1007/s10734-020-00635-x

Chatfield, M.. & Mander, A. (2009). The Skillings-Mack test (Friedman test when there are missing data). *The Stata journal*, *9*(2), 299–305

Cogliano, M., Bernacki, M. L., & Kardash, C. M. (2021). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. *Journal of Educational Psychology, 113*(7), 1421–1440. https://doi.org/10.1037/edu0000624

Crespo, N. (2004). La metacognición: las diferentes vertientes de una teoría. *Revista Signos, 22*(48), 91-115. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-0934200000480000

Cunninghan, P., Hulman, R., Matusovich, H., Morelock; J., Nicole, D. (Junio, 2016). Beginning to Understand and Promote Engineering Students' Metacognitive Development. *Anual Conference & Exposition.* New Orleans, LA

Cunninghan, P., Matusovich, H., Venters, C., Williams, S., Bhaduri, S. (Junio, 2018). The impact of metacognitive instruction on students' conceptions of learning and their self-monitoring behaviors. ASEE Anual Conference & Exposition

Davidson, J.E., Sternberg, R. J. (1998). Smart Problem Solving: How Metacognition Helps. In Metacognition in Educational Theory and Practice, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds., Hillsdale, NJ: Lawrence Erlbaum Associates, 1998.

Davidson, J. E., Deuser, R., & Sternberg, R. J. (1994). The Role of Metacognition in Problem Solving, in Metacognition: Knowing about Knowing, J. Metcalfe and A. P. Shimamura, Eds., Cambridge, MA: MIT Press.

Desoete, A., & De Craene, B. (2019). Metacognition and Mathematics education: an overview. ZDM. Mathematic Education, 51, 565-575. https://link.springer.com/article/10.1007/s11858-019-01060-w

Dunlonsky, J., & Thiede, K. W. (2013). Metamemory. In D. Reisberg (Ed.), *The Oxford Handbook of Psychology* (pp. 283–298). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0019.

Education Advisory Board [EAB]. (2019, October 7th). *A third of minority students leave STEM majors. Here's why*. https://eab.com/insights/daily-briefing/student-success/a-third-of-minority-students-leave-stem-majors-heres-why/

Emory, B., & Luo, T. (2022) Metacognitive Training and Online Community College Students' Learning Calibration and Performance, Community College. *Journal of Research and Practice, 46*(4), 240–256. https://doi.org/10.1080/10668926.2020.1841042

Fakcharoenphol, W., Morphew, J., & Mestre, P. (2015). Judgments of physics problem difficulty among experts and novices. *Psychics Education Research, 11*, 1-14. https://journals.aps.org/prper/pdf/10.1103/PhysRevSTPER.11.020128

Fazirah, N., Mohd-Yusof, K., & Phang, F. (Noviembre, 2017). Metacognitive development in engineering students through cooperative problem-based learning (CPBL). 7th World Engineering Education Forum (WEEF), Kuala Lumpur, Malaysia. 10.1109/WEEF.2017.8467177

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of psychological inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*(1), 1–19. doi:10.1007/s11409-016-9158-

Funk, C., & Parker, K. (2018). *Women a men in STEM often at odds over workplace equity.* Pew Research Center. https://www.pewresearch.org/social-trends/2018/01/09/women-and-men-in-stem-often-at-odds-over-workplace-equity/

Gadella, L., Estévez, I., Arias, J., & Delgado, C. (2020, del 27 al 30 de abril). Application of learning analytics to study the accuracy of self-reported working patterns in self-regulated learning questionnaires. *IEEE Global Engineering Education Conference (EDUCON)*. 10.1109/EDUCON45650.2020.9125405

Gaeta, M. (2015). Aspectos personales que favorecen la autorregulación del aprendizaje en la comprensión de textos académicos en estudiantes universitarios. *Revista de Docencia Universitaria, 13*(2), 17-35.

Greene, J., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin´s model of self-regulated learning. *Review of Educational Research, 77*(3), 334–372. https://journals.sagepub.com/doi/abs/10.3102/003465430303953

Gutiérrez de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology, 114*(4), 681–700. https://doi.org/10.1037/edu0000674

Gutierrez, A., & Price, A. (2016). Calibration between Undergraduate Students' Prediction of and Actual Performance: The Role of Gender and Performance Attributions. *The Journal of Experimental Education, 85*(3), 486–500. http://dx.doi.org/10.1080/00220973.2016.1180278

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *Journal of Experimental Education, 83*(3), 386-404. https://doi.org/10.1080/00220973.2014.907230

Hacker, D., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170. https://doi.org/10.1037/0022-0663.92.1.160.

Hadwin, A., Järevelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. P. Alexander, Schunk, D., & Greene, J. (Eds), *Handbook of self-regulation of learning and performance* (pp.83-106). Routledge

Handel, M., Harder, B., & Dressel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction, 65*, 1–9. https://doi.org/10.1016/j.learninstruc.2019.101245

Hawker, M., Dysleski, L., Rickey, D. (2016). Investigating General Chemistry Students' Metacognitive Monitoring of Their Exam Performance by Measuring Postdiction Accuracies over Time. *Journal of Chemical Education, 93*(5), 832–840. https://doi.org/10.1021/acs.jchemed.5b00705

Huff, J., & Nietfeld, J. (2009). Using strategy instruction and confidence judgements to improve metacognitive monitoring. *Metacognition and Learning, 4,* 161–176. https://doi.org/10.1007/s11409-009-9042-8

Kim, J. (2018). The effect of metacognitive monitoring feedback on performance in a computer-based training simulation. *Applied Ergonomics, 61*, 193–202. https://doi.org/10.1016/j.apergo.2017.10.006

Koevoets-Beach, C., Julian, K., & Balabanoff, M. (2023). "I guess it was more than just my general knowledge of chemistry": exploring student´s confidence judgments in two-tiered assessments. *Chem. Educ. Res. Pract, 1-14*- 10.1039/D3RP00127J

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lawanto, O., & Febrian, A. (2017, del 12 al 15 de octubre). Student's self-regulation in senior capstone design proyects: A case study of two project teams. *Frontiers in Education Conference (FIE)*. Erie, USA. 10.1109/FIE.2016.7757481

Lawanto, O., Santoso, H., Lawanto, K., & Goodridge, W. (2014). Self-Regulated Learning Skills and Online Activities Between Higher and Lower Performers on a Web-Intensive Undergraduate Engineering Course. *Journal of Educators Online*, 11(3), 1–32. https://files.eric.ed.gov/fulltext/EJ1033324.pdf

Luo, T., Mui, W., Hong, Z., & Li, W. (2021). STEM stereotypes predict student´s STEM career interest via self-efficacy and outcome expectations. *International Journal of STEM Education, 8*(36), 1–13. https://doi.org/10.1186/s40594-021-00295-y

Manganello, F., Falsetti, C., & Leo, T. (2019). Self-Regulated Learning for Web-Enhanced Control Engineering Education. *Journal of Educational Technology & Society, 22*(1), 44–58. https://eric.ed.gov/?id=EJ1202779

Mathabate, K., Potgieter, M. (2014). Metacognitive monitoring and learning gain in foundation chemistry. *Chemestry Education Research and Practice, 1*, 1–12 https://doi.org/10.1039/C3RP00119A

Mayorga, C., Colomo, E., & Palmero, J. (2020). Causas de la deserción escolar en Ingeniería en Electrónica y Computación del Centro Universitario de los Valles de la Universidad de Guadalajara (México). *Revista Espacios, 41*(6), 1–15.

McDaniel, M., Einstein, G. & Een, E. (2021). Training college students to use learning strategies: A framework and pilot course. Psychology *Learning and Teaching, 20(*3), 364–382. https://doi.org/10.1177/1475725721989489

Morphew, J. (2021). Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacognition and Learning, 16*, 89–111. https://doi.org/10.1007/s11409-020-09239-3

Morphew, J., Gladding, G., & Mestre, J. (2020). Effect of presentation style and problem-solving attempts on metacognition and learning from solution videos. *Physical Review Physics Education Research, 16*(1), 010104. https://doi.org/10.1103/PhysRevPhysEducRes.16.010104.

Muncer, G., Highman, P., Gosling, C., Cortese, S., Wood-Downie, H., & Hadwin, J. (2022). A Meta-Analysis Investigating the Association Between Metacognition and Math Performance in Adolescence. *Educational Psychology Review, 34*, 301–334 https://link.springer.com/article/10.1007/s10648-021-09620-x

Muis, K., Winne, P., & Ranelucci, J. (2016). The role of calibration bias and performance feedback in achievement goal regulation. *International Journal of Educational Research, 4*(1), 14–36. 10.12735/ier.v4i1p14

Nelson, T. O., & Narens, L. (1994). Why investigates metacognition? En J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (p. 1–25). The MIT Press.

Nguyen, T., Abed, E., & Pezdek, K. (2018). Postdictive confidence (but no predictive confidence) predicts eyewitness memory accuracy. *Cognitive research: principles and applications, 3*(32), 1–13 https://doi.org/10.1186/s41235-018-0125-4

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research, 95*(3), 131–142. https://doi.org/10.1080/00220670209596583

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7–28.

Nietfeld, J., Cao, L., & Osborne, J (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition Learning 1*(2), 159–179 https://doi.org/10.1007/s10409-006-9595-6

Osterhage, J., Usher, E., Douin, T., Balley, W. (2019). Opportunities for Self-Evaluation Increase Student Calibration in an Introductory Biology Course. *CBE-Life Science Education, 18*(2), 1–10. https://doi.org/10.1187/cbe.18-10-0202

Pazicni, S., & Bauer, C. F. (2014). Characterizing illusions of competence in introductory chemistry students. *Chemestry Education Research and Practice, 15*(1), 24–34. 10.1039/c3rp00106g

Persky, A., & Dinsmore, D. (2019). Metacognitive changes and sources of confidence judgements in health professions classroom learning. *Current in Pharmacy Teaching and Learning, 11*, 338–345. https://doi.org/10.1016/j.cptl.2019.01.005

Persky, A., Lee, E., & Schlesslman, L. (2020). Perception of Learning Versus Performance as Outcome Measures of Educational Research. *American Journal of Pharmaceutical Education, 84*(7), 993–1000. https://doi.org/10.5688/ajpe7782

Picho-Kiroga, K., Turnbull*, A. & Rodriguez-Leahy, A. (2021). The misspecification of stereotype threat theory in empirical research: Consequences and remedies. Journal of Advanced Academics, 32(2), 231–264 doi: 10.1177/1932202X20986161

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. En M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). New York: Academic Press.

Ramírez, J., García, A., & Olarte, F. (2016). Adaptation and Validation of the Motivated Strategies for Learning Questionnaire -MSLQ- in Engineering Students in Colombia. *International Journal of Engineering Education, 32*(4), 1774–1787

Rawson, K., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology, 19*(4/5), 599–579. https://doi.org/10.1080/09541440701326022

Rebello, N. S. (2012). How accurately can students estimate their performance on an exam and how does this relate to their actual performance on the exam? In N. S. Rebello, P. V. Engelhardt, & C. Singh (Eds.), AIP Conference Proceedings (pp. 315–318). AIP. https://doi.org/10.1063/1.3680058

Sáez, F. M., Díaz, A. E., Panadero, E., & Bruna, D. V. (2018). Revisión Sistemática sobre Competencias de Autorregulación del Aprendizaje en Estudiantes Universitarios y Programas intracurriculares para su promoción. *Formación universitaria, 11*(6), 83–98. http://dx.doi.org/10.4067/S0718-50062018000600083.

Saez, F., Mella, J., Loyer S., Zambrano, C., & Zañartu, N. Self-regulated learning in engineering students: A systematic Review. *Revista Espacios*, 41(2), 1–7. https://www.revistaespacios.com/a20v41n02/20410207.html

Saenz, G., Geraci, L., & Tirso, R. (2019). Improving metacognition: a comparison of interventions. *Applied Cognitive Psychology, 33*(5), 918–929. https://doi.org/10.1002/acp.3556

Saenz, G., Geraci, L., Miller, T., & Tirso, R. (2017). Metacognition in the classroom: the association between student´s exam predictionand their desired grades. *Consciusness and Cognition, 51*, 125–139. https://doi.org/10.1016/j.concog.2017.03.002

Sánchez-Arévalo, M., Cruz-Hueso, L., & Ferro-Escobar, R. (2018). Behavior Approach Model of Voluntary University Desertion in Undergraduate Programs in Engineering Period 2015–2018. *Ingeniería Solidaria*, *14*(26), 19-27. https://doi.org/10.16925/in.v14i26.2433

Santangelo, J., Cadieux, M., & Zapata, S. (2021). Developing student metacognitive skills using active learning with embedded metacognition instruction. *Journal of Stem Education: Innovation and Researc*h, 22 (2), 51–63. https://www.jstem.org/jstem/index.php/JSTEM/article/view/2475

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). Lawrence Erlbaum. https://doi.org/10.4324/9780203876428

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness*. Contemporary Educational Psychology, 19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Sedraz, C., Zambom, E., Rodrigues, R., Ramos, J., & De Souza, F. (2018). Effects of learning analytics on students' self-regulated learning in flipped classroom. *International Journal of Information and Communication Technology Education, 14*(3), 91–107. https://doi.org/10.4018/IJICTE.2018070108

Seron, C., Silbey, S. S., Cech, E., & Rubineau, B. (2016). Persistence Is Cultural: Professional Socialization and the Reproduction of Sex Segregation. Work and Occupations, 43(2), 178–214. https://doi.org/10.1177/0730888415618728

Serra, M. J., & Metcalfe, J. (2009). Effective Implementation of Metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 278–298). Lawrence Erlbaum Associates, Inc.

Skillings J. H, Mack, G. A. On the use of a Friedman-type statistic in balanced and unbalanced block designs. *Technometrics.* 1981;23:171–177

Stolk, J., Martello, R., Ollin, F., Kohler, K., Chen, K., Herter, R. (2014, de 22 al 25 de octubre). Well that didn't work. A trouble attempts to quantitatively measure engineering student's lifelong learning development over two years college. *Frontiers in Education Conference (FIE).* Madrid, España. 10.13140/2.1.1450.9765

Reda, H., & Mulugeta, G. (2018). Investigating the Causes of Students' Less Academic Performance in Engineering College of Debre Berhan University. *American Journal of Theoretical and Applied Statistics, 7*(3), 126–131. https://www.sciencepublishinggroup.com/journal/paperinfo?journalid=146&doi=10.11648/j.ajtas.20180703.15

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe therefore I achieve (and vice versa): A meta-analytic Cross lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences,* 61, 136–150.

Tauber, S. K., & Dunlosky, J. (2015). Monitoring of learning at the category level when learning a natural concept: Will task experience improve its resolution? *Acta Psychological*, 155, 8–18. https://doi.org/10.1016/j.actpsy.2014.11.011

Testa, I., Galano, S., Tarallo, O. (2023). The relationships between freshmen's accuracy of self-evaluation and the likelihood of succeeding in chemistry and physics exams in two STEM undergraduate courses. *International Journal of Science Education, 45*(5), 358–382. 10.1080/09500693.2022.2162833

Theobald, M. (2021). Self-regulated learning training programs enhance university student´s academic performance, self-regulated learning strategies, and motivation: A metanalysis. *Contemporary Educational Psychology 66*, 1–19 https://doi.org/10.1016/j.cedpsych.2021.101976

Thiede, K., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning text. *Journal of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Urban, K., & Urban, M. (March, 2019). Improving the accuracy of the self-evaluation during on-screen self-regulated learning through calibration feedback. In INTED proceedings. https://library.iated.org/view/URBAN2019IMP

Zambrano, C. (2016). Autoeficacia, prácticas de aprendizaje autorregulado y docencia para fomentar el aprendizaje autorregulado en un curso de ingeniería de software. *Formación Universitaria, 9*(3), 51–60. http://dx.doi.org/10.4067/S0718-50062016000300007

Zheng, W., Yin, J., Shi, H., & Skelton, G. (2016). Prompted Self-Regulated Learning Assessment and Its Effect for Achieving ASCE Vision 2025. *Journal of Professional Issues in Engineering Education and Practice, 143*(2), 1–10. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000308.

Ziegler B., & Montplaisir L. (2014). Student Perceived and Determined Knowledge of Biology Concepts in an Upper-Level Biology Course. *CBE Life Sci Educ, 13*(2), 322–30. https://www.ncbi.nlm.nih.govnne/pmc/articles/PMC4041508/

AUTHORS

Dr. Picho-Kiroga is an assistant professor of Educational Psychology in the department of Human Development and Psychoeducational Studies at Howard University. She is also the director of Research In STEM Education (RISE) lab at Howard. Her research focuses on examining the role of social learning contexts in elevating stereotype threat among women and people of color in Science Technology Engineering and Mathematics (STEM). Specifically, she seeks to understand key contextual factors that exacerbate ST in K-12 school settings and how these contexts interact with psychological factors to sabotage the achievement of women and people of color in STEM. Within this line of research, Dr. Picho-Kiroga has conducted cross-cultural research examining stereotype threat in under-studied populations such as adolescents and students in non-Western cultural contexts. She has also developed interventions to attenuate the impact of this phenomenon on students belonging to underrepresented groups in STEM.

Marcela Valencia Serrano is a PhD Candidate at Pontificia Universidad Javeriana Cali in the Psychology Doctoral Program. She has a master's degree in education from ICESI University in Cali. She is also a teacher in the Social Science Department at Javeriana University where she teaches developmental psychology, research methods in psychology and motivation processes in education. Marcela's research interest is focused on understanding self-regulated learning in college students and its relation to teacher feedback processes. She is also interested and has worked in research about metacognition in reading and writing; self-regulation of motivation in engineering students and teacher-students interaction in highs school students. About this, she is certified in the Classroom Assessment Scoring System (CLASS, Pianta, Hamre & Mintz, 2008), which allowed her to observe and analyze teacher-student interaction in four domains: Emotional support, classroom organization, pedagogical support, and student behavioral engagement.

Dr. Noel Bourne is Guyanese American and is currently an Assistant Professor of Mathematics at Tennessee State University. He completed his bachelor's degree in mathematics at Xavier University of Louisiana, his master's degree in mathematics at North Carolina Central University, and his Ph.D. In Mathematics at Howard University, where he worked under Dr. Xingting Wang on Hopf Algebras. He then went to Carnegie Mellon University as a postdoctoral fellow where he worked with Dr. Michael Young on his NSF-Supported Project: Studying Successful Doctoral Students in Mathematics from Underrepresented Groups. He has recently been honored by the Network of Minorities in Mathematical Sciences and listed on their Mathematically Gifted and Black website.

Jerriel Hall is a doctoral student in the educational psychology program at Howard University. He has a Master of Education in Curriculum & Instruction from the University of Southern California.Jerriel has 15 years of experience teaching locally and abroad. In 2018, Hall was awarded a Fulbright Distinguished Award in Teaching grant in Ghana to research improving STEM teacher effectiveness and self-efficacy through differentiated professional development. Jerriel's primary research interest is reducing the psychological and emotional effects of stereotype threat for secondary school-aged African-American males within STEM inclusion classrooms with an identified specific learning disability.