The Impact of Brief Mindfulness Training on Judgments of Confidence, Anxiety, and Difficulty While Answering Physics Questions

Avital Pelakh[1], Eric Kuo[2], Melanie Good[3], Michael J. Tumminia[1], Sara Jahanian[1], Brian Galla[1], and Timothy Nokes-Malach[1]

[1]Learning Research and Development Center, University of Pittsburgh

[2]Department of Physics, University of Illinois Urbana-Champaign

[3]Department of Physics and Astronomy, University of Pittsburgh

Author Note

Avital Pelakh https://orcid.org/0000-0003-1231-2442

Correspondence concerning this article should be addressed to Avital Pelakh, Learning Research and Development Center, University of Pittsburgh, 3420 Forbes Ave., Pittsburgh, PA 15217. Email: avp36@pitt.edu.

Deidentified data and analysis code are included in the supplementary materials and publicly available on GitHub: https://apelakh.github.io/pelakh-et-al-2025-supplement/.

Word Count: 11,379

Abstract

We tested the impact of a mindfulness training intervention to improve introductory physics students' experiences while answering physics questions. We expected the intervention to reduce physics threat and to increase students' confidence while lowering anxiety and judgments of difficulty. We also tested whether domain-level physics threat mediated the effects of the intervention on task judgments and whether the effects differed by gender. To test these hypotheses one hundred and forty-nine undergraduates were randomly assigned to receive either a 5-day mindfulness training intervention or no training (control). Both groups answered physics questions before and directly after the intervention and rated their confidence, anxiety, and difficulty for each question. Mindfulness training led to a greater increase in confidence and a reduction in anxiety among women and non-binary students, but not for men. The intervention also led to a reduction in judgments of difficulty for all students. The association between mindfulness training and self-reported anxiety among women and non-binary students was mediated by reductions in physics threat (measured mid-week using experience sampling). However, physics threat did not mediate any of the other mindfulness training outcomes for confidence or difficulty. The results are discussed in relation to a model of challenge and threat and mindfulness applications.

Public Significance Statement

Mindfulness training was found to reduce self-reported difficulty during a physics task among introductory physics students with pre-existing physics threat. Training also reduced self-reported anxiety and increased confidence judgments among women and non-binary students, but not for men. Physics threat was identified as a mediating variable under some conditions.

Keywords: Psychological Threat, Physics, Mindfulness, Problem Solving, Learning

The Impact of Brief Mindfulness Training on Judgements of Confidence, Anxiety, and Difficulty While Answering Physics Questions

Solving quantitative problems and answering conceptual questions are learning activities that are fundamental to many university-level STEM courses, especially physics. Scientific research on these kinds of activities was pioneered by cognitive psychologists investigating the mental processes involved in understanding problems and constructing their solutions from an information processing perspective (e.g., Chi et al., 1981; Larkin et al., 1980; Reif & Heller, 1982; for reviews see Kuo, 2023; Singh et al., 2023). However, this early work tended to focus on task performance and domain knowledge in isolation without considering how other individual and environmental factors like social identity, emotions, and psychological appraisals of stress impact physics problem solving in the classroom (Cwik & Singh, 2021b; Hazari et al., 2007; Jamieson et al., 2016, 2018, 2020; John-Henderson et al., 2015; Marshman, Kalender, Schunn, et al., 2018; Ramirez et al., 2018). These social and emotional factors have their roots in social, clinical, and health psychology. When these factors have been investigated in educational settings, they have been typically related to coarser, aggregated measures of learning and performance outcomes (e.g., exam scores, course grades). We hope to begin to bridge the gap between these approaches by investigating how social and emotional factors impact students experiences of problem solving and answering conceptual questions at the task level.

Bringing these strands of research together is important because students with similar performance in STEM courses may experience learning and problem solving differently, and these differences can affect decision-making around academic persistence and career choices (Correll, 2001; Ellis et al., 2016). Furthermore, these differences are often reflected in systematic educational inequities, with women reporting greater anxiety and lower confidence, self-efficacy, and sense of belonging compared to. men (Goetz et al., 2013; Marshman, Kalender, Nokes-Malach, et al., 2018). This issue is especially relevant in college-level introductory physics, where gender-based stereotypes are deeply rooted in the discipline (Whitcomb et al., 2022). Understanding how social and emotional factors impact the experience of problem solving and learning activities at a fine-grained level, rather than in aggregate, can help inform teacher training and design academic interventions with greater precision.

In this study, we tested the impact of a brief mindfulness training intervention on introductory physics students' experiences while answering physics questions. This study used the biopsychosocial model of challenge and threat (Blascovich, 2008; Seery, 2013; Tomaka et al., 1997) as a holistic theoretical framework to better understand the interplay between psychological appraisals of stress at the domain level (physics threat) and judgements of specific experiences engaging in physics tasks among students reporting pre-existing physics-related stress. We examined the effects of mindfulness training on three dimensions of student experiences (judgments of confidence, anxiety, and difficulty), explored domain-level physics threat as a potential mechanism for effects of the intervention, and whether effects differ by gender.

Challenge and Threat in Introductory Physics

The biopsychosocial model of challenge and threat—hereinafter referred to as the challenge-threat model (Blascovich, 2008; Blascovich & Tomaka, 1996)—is a theoretical framework based on early theories of stress and coping (Lazarus & Folkman, 1984) that explains how individuals respond, physiologically, psychologically, and behaviorally, during high-stakes performance situations. The core theoretical contribution of the model is a distinction between two fundamental motivational states that can arise in response to stress: *challenge* or *threat*.

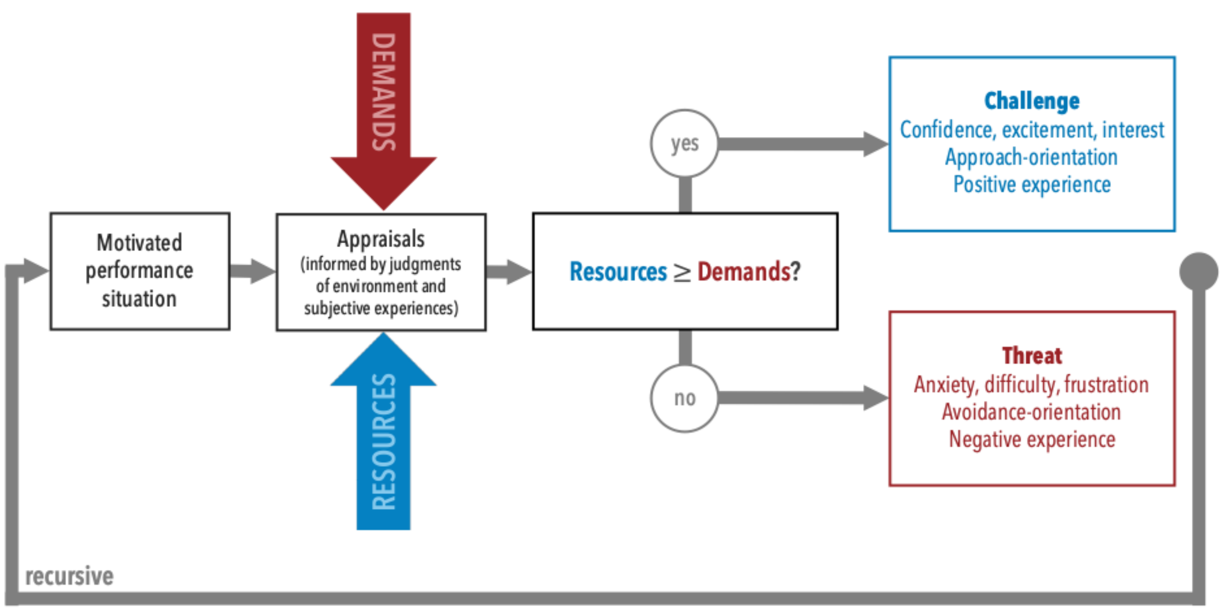
According to the challenge-threat model, whether people approach demanding situations as opportunities for growth (challenge) or as sources of potential harm (threat) is determined by their psychological appraisals of situationally relevant *resources* and *demands*. In the context of undergraduate physics, resources can include knowledge and skills, task familiarity, motivation to succeed, identity as a physics person, and expectations for success – which can be supported by one's own self-efficacy and availability of external support. Demands can include high workload, task difficulty, uncertainty about success, performance stakes, and judgments about effort required, but also physics anxiety, worry, or rumination. A challenge state occurs when perceived resources meet or exceed demands and is characterized by approach-oriented motivation, active engagement, and/or positive emotions. A threat state occurs when demands exceed resources and is marked by vigilance, avoidance, and/or negative emotions.

While there has been some work connecting the challenge-threat model to other stress-related theories of academic motivation, such as test anxiety (Beilock et al., 2017) and math anxiety (Ramirez et al., 2018) in academic contexts, less work has examined how domain-level threat (i.e., threat experienced while learning physics) manifests at a granular level while completing specific tasks and activities (i.e., answering physics questions or solving physics problems). We are especially interested in the association between the broader appraisals that can lead to challenge/threat and momentary judgments of specific experiences.

We use the terms *appraisal* and *judgment* to represent two distinct but related cognitive processes that operate at different levels of granularity. We use the term *judgment* to represent a quick subjective assessment of a task or an internal state at a specific moment in time; whereas the term *appraisal* is used to represent a broader evaluative process that is informed by cumulative judgments of one's environment and subjective experiences but may involve assigning meaning, significance, or emotional value to those experiences (see Figure 1). Connecting momentary judgments of specific experiences to broader beliefs and attitudes about physics (i.e., whether students view physics as threatening or challenging) can help us understand the nature of their association and identify the most effective levers for intervention.

Figure 1

Conceptual Diagram of the Challenge-Threat Model



Confidence, Anxiety, and Difficulty During Physics Tasks

Judgments of confidence, anxiety, and difficulty reflect interrelated facets of experience associated with student engagement during learning. We acknowledge that these terms have varying connotations in everyday life and across disciplines. Therefore, we define what we mean by each below.

We define *confidence* as a subjective, metacognitive judgment of one's performance on a specific task, absent explicit feedback. It is similar to, but distinct from self-efficacy (Bandura, 1997), which is a belief in one's capability to execute the behaviors necessary for future performance, and self-concept (Marsh & Martin, 2011), which is focused on more general self-judgments of ability, performance, and worth. Momentary confidence judgments are commonly used in the study of metacognition to calculate metacognitive accuracy, but they have also been shown to predict intention to pursue a STEM career (Bench et al., 2015).

Anxiety is a response to potential threat that has both cognitive (worry, rumination) and somatic (physiological arousal) components (Beilock et al., 2017). It has been well studied in educational contexts in terms of math anxiety and test anxiety. One explanation of how anxiety can affect task performance or learning in the moment is by coopting cognitive resources (limited by working memory) that could otherwise be used to engage in task-related behaviors (Beilock et al., 2017). Another explanation is that students' judgements of their learning experiences play a crucial role in the development of persistent domain-level (e.g., math or physics) anxiety, with more negatively judged experiences leading to worse self-appraisals and greater anxiety over time (Ramirez et al., 2018). This broader domain-level anxiety can then affect students' learning behaviors: for example, increasing task avoidance (Jenifer et al., 2023).

We conceptualize *difficulty* as a measure of immediate demands. Judgments are influenced by perceived task complexity, student knowledge and skills, and the resulting mental effort and cognitive load.

Figure 1 shows a conceptual map depicting the relationship between appraisals and judgments. As seen in the figure, our perspective is that association between domain-level appraisals and momentary judgments of task-specific experiences is bi-directional. On one hand, challenge and threat states can act as an interpretive lens for sensory information, influencing judgments. For example, a student in a threat state may be more sensitive to bodily sensations of arousal and quicker to interpret arousal negatively, as anxiety, frustration or worry. Further, they may be quick to overestimate the difficulty of a task to account for their discomfort or desire to disengage. Alternatively, momentary judgments can inform future appraisals. Experiencing lower confidence and greater anxiety and difficulty repeatedly while engaging in physics tasks informs resource and demand appraisals, strengthening physics threat over time.

Using Mindfulness Training to Target Psychological Appraisals

Mindfulness training may be particularly well-suited to help students respond to stress with a state of challenge as opposed to threat because it targets the meaning-making processes that occur between judgment and appraisal (Luberto et al., 2020). Mindfulness is a practice of focusing attention to the present mental and physical experience with curiosity, openness, and acceptance (Bishop et al., 2004; Kabat-Zinn, 2003). Mindfulness training focuses on cultivating awareness of automatic patterns of thought; reframing sensations, thoughts, and emotions as transient mental events rather than literal and fixed truths; and allowing them to rise and fall without reactivity or identification. In this way, mindfulness creates a space between incidental thoughts and self-belief.

During a difficult experience with a learning activity, a thought such as, “I’m no good at this,” may arise along with a sensation of bodily discomfort. Without mindfulness, “I’m no good at this” may be taken to mean “I probably got this wrong” (low confidence); discomfort is labeled as anxiety, and the whole experience is perceived as more difficult due to the added discomfort and emotional cost. A student practicing mindfulness would be primed to recognize the thought as it appears and allow it to come and go without assigning it meaning or significance. Because they are able to let bodily sensations pass without additional interpretation, they perceive less anxiety and have the freedom to focus more attention and cognitive energy toward the task at hand; thus, feel more confident in their effort and perceive the experience as less costly.

Mindfulness-based interventions have been used to successfully treat a wide range of clinical and non-clinical psychological conditions related to stress. For example, they have been shown to reduce stress reactivity (Lindsay et al., 2018), support emotion regulation (Bai et al., 2020), and improve cognitive performance (Mrazek et al., 2013). Here, we explore whether mindfulness training improves judgments of physics task experiences for students who see physics as threatening, and if so, whether it does so through the reduction of domain-level physics threat or by impacting momentary judgments of those experiences directly.

Gender Differences in Self-Judgments and Beliefs

Many studies have reported gender differences in self-judgments and beliefs among college students with similar performance. For example, studies show women have lower self-efficacy than men with equivalent course grades and conceptual knowledge (Marshman, Kalender, Nokes-Malach, et al., 2018), tend to underestimate their performance relative to men (Bench et al., 2015), and report greater habitual math anxiety and lower perceived competence compared to men, despite having similar midterm grades (Goetz et al., 2013). These self-assessments can influence academic persistence (Ellis et al., 2016), and career choices (Bench et al., 2015; Correll, 2001). These findings suggest that men and women likely experience physics learning activities differently. However, it is unclear whether gender differences persist among students experiencing similar levels of threat. Put another way, do domain-level appraisals of resources and demands explain observed gender differences in how students experience learning activities? Or is there an aspect of gender identity that influences judgments above and beyond challenge and threat as defined by the challenge-threat model?

Further, it is unclear whether receptiveness to mindfulness training differs by gender. One study with college students showed that mindfulness training reduced negative affect and increased self-compassion for women more than men (Rojiani et al., 2017). However, a meta-analysis of the effects of mindfulness-based programs on psychological distress in non-clinical settings found no differences associated with gender or baseline distress levels among adults (Galante et al., 2023). Furthermore, to our knowledge, there is no prior work looking at the interaction of gender and mindfulness training in stressful academic environments. We investigate this interaction in our study by testing whether gender identity moderates the effects of a mindfulness intervention in introductory physics.

Current Study

We present results from a randomized controlled experiment testing the effects of mindfulness training on psychological appraisals of stress in physics (physics threat), as well as judgments of confidence, anxiety, and difficulty while answering physics questions. Students enrolled in undergraduate physics and who reported experiencing threat in their course were randomly assigned to a mindfulness training or control condition. Participants completed a set of physics tasks at baseline (day 1; before the intervention) and at posttest (day 5; directly after the intervention), during which item-level judgments of confidence, anxiety, and difficulty were measured. Physics threat was measured using self-report survey at baseline (before the first physics task), and twice per day on days two through four (between baseline and posttest) using experience sampling methods (EMA: ecological momentary assessment). These data were used to address the following research questions.

Research Question 1

What is the effect of mindfulness training on item-level judgments of confidence, anxiety, and difficulty while answering physics questions?

H1. Compared to control, students assigned to a mindfulness training condition will show an increase in item-level judgments of confidence during physics tasks at posttest.

H2. Compared to control, students assigned to a mindfulness training condition will show a reduction in item-level judgments of anxiety during physics tasks at posttest.

H3. Compared to control, students assigned to a mindfulness training condition will show a reduction in item-level judgments of difficulty during physics tasks at posttest.

Research Question 2

Does domain-level physics threat mediate the association between mindfulness training and item-level judgments of confidence, anxiety, and difficulty?

H4. Associations in hypotheses 1-3 will be mediated by decreases in physics psychological threat, measured using ecological momentary assessment (EMA) between baseline and posttest.

Exploration of Gender Moderation

In addition to the stated questions and hypotheses, we tested whether or not the associations between mindfulness training and item-level judgments of confidence, anxiety, and difficulty were moderated by gender. Specifically, is the effect of mindfulness training on judgments stronger for women and non-binary students, after controlling for physics psychological threat at baseline?

Method

Transparency and Openness

This paper reports on one strand of a broader research project. This study was registered on clinicaltrials.gov (NCT04589377). Documents describing the study design, methods, research questions, and hypotheses of this project were preregistered on the Open Science Framework (OSF) after data were collected but before analysis began. The project overview document (osf.io/723rd) provides detailed descriptions of the study design, materials, and procedure. A comprehensive appendix of the entire study's survey items (osf.io/5368j) and physics task items (osf.io/fm8qx) are also available for reference. The current analyses focus on a subset of the survey items and physics task items. Multiple preregistration documents are posted on OSF, each one detailing hypotheses and planned analyses for a specific strand of the full study. The primary analyses described here follow Preregistration 3 – Physics Task Outcomes (osf.io/b57et).

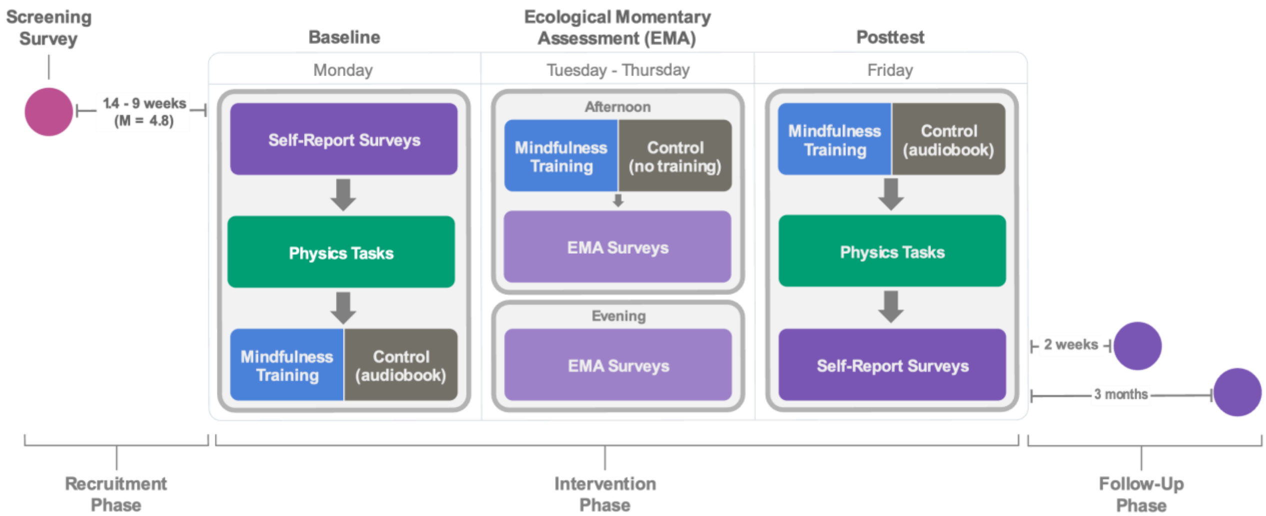
There were several minor deviations from the preregistration regarding additional covariates and other model specification details that did not change our pattern of results. There were also several analyses planned for task accuracy and learning outcomes that were preregistered and included only in the online supplement. We did not find any effects of mindfulness training on these outcomes. A more detailed description of changes from the preregistration document and additional results are provided in the supplemental materials. All data and analysis code have been made publicly available on GitHub and can be accessed at https://apelakh.github.io/pelakh-et-al-2025-supplement/.

Design

Participants were randomly assigned to one of two experimental conditions: mindfulness training or no-training audiobook control. Study activities, including self-report surveys, physics assessments, and audio training sessions, were completed over the course of five days (Monday through Friday) using a pre-posttest design with ecological momentary assessment (EMA) between baseline and posttest. All study activities were completed remotely. The full study design is depicted in Figure 2 below[1].

Figure 2

Study Procedure and Activity Flow



Participants

Participants were recruited from introductory calculus-based physics 1 courses across three semesters at a large, public mid-Atlantic university. The final study sample (*N* =149) was mainly composed of first-year (82.6%) and second-year (11.4%) college students, with only 6% reporting being in their 3rd year or greater. The sample had a mean age of 18.64 years (*SD* = 1.78) and 98.7% expressed interest in stress management at screening. The vast majority (98.6%) of the sample reported the intention to obtain a STEM major, with 71.4% on an engineering track[2]. A description of self-reported racial and gender identification of the sample is provided in Table 1.

[1] There were additional dispositional and engagement measures assessed during the self-report survey sections on Monday and Friday and using EMA, which are not depicted or described here (See the Project Overview on OSF for a full description of measures). Only the specific measures relevant to the research questions and hypotheses presented earlier are included in this manuscript.

[2] Percentages for intended major are calculated out of 147 because two students in the sample did not complete the demographics portion of the baseline survey and had missing data for these items.

Table 1

Racial and Gender Demographics of the Study Sample

	n	Percent
Racial and Ethnic Identification (participants could select more than one)		
American Indian/Native American/Alaskan Native	2	1.3%
Asian/Indian/Pacific Islander	36	24.2%
Black/African American	11	7.4%
Hispanic/Latine	7	4.7%
Middle Eastern/North African	5	3.4%
White	114	76.5%
Students Who Selected Multiple Identities[a]	24	16.1%
American Indian/Native American/Alaskan Native; Black/African American; Hispanic/Latine; White		
American Indian/Native American/Alaskan Native; White		
Asian/Indian/Pacific Islander; Black/African American		
Asian/Indian/Pacific Islander; White		
Black/African American; Hispanic/Latine		
Black/African American; White		
Hispanic/Latine; White		
Middle Eastern/North African; White		
Gender Identification		
Men	66	44.3%
Women	81	54.4%
Non-binary	2	1.3%

Note. Items taken from Fernandez et al., 2016
[a]Participants were permitted to select multiple racial identities.

Materials

Demographic Information

Demographic characteristics collected were age, year in school, residence outside of the US, physics course information, interest in learning stress reduction techniques, and intended major. Gender and race/ethnicity were collected following Fernandez et al., 2016.

Psychological Threat

Psychological threat was measured using the Adapted Stress Appraisal Questionnaire (Jamieson et al., 2016), which was developed to measure stress appraisals with regard to math anxiety about exams. The instrument was further adapted to the current study by replacing the term “math test” with “physics work” (see Table 2). “Physics work,” is defined in the survey instructions as, “preparing for exams, solving problems, reading the textbook, and all other course-related activities.” The measure is composed of two subscales: *demands* and *resources*, measuring perceived situational demands and perceived coping resources, respectively. A short-form (4-item) version of the instrument was used for the screening survey ($\alpha_{\text{resources}} = .67$, $\alpha_{\text{demands}} = .36$) and ecological momentary assessments ($\alpha_{\text{resources}} =$ $.87$-$.93$, $\alpha_{\text{demands}} = .78$-$.90$), and the full-length (11-item) version of the measure was used during the baseline ($\alpha_{\text{resources}} = .79$, $\alpha_{\text{demands}} = .74$) and posttest assessments ($\alpha_{\text{resources}} = .82$, $\alpha_{\text{demands}} = .75$).

Table 2

Psychological Threat Self-Report Measure

Subscale	Item	Response Scale
Demands	[a]*Working on physics is very demanding.*	1 (strongly disagree) - 6 (strongly agree)
	I am uncertain about how I will perform in physics.	
	My physics work will take a lot of effort to complete.	
	[a]*Working on physics is very stressful.*	
	Poor performance on physics work would be very distressing for me.	
	I think physics work represents a threat to me.	
Resources	[a]*I feel that I have the abilities to succeed on my physics work.*	1 (strongly disagree) - 6 (strongly agree)
	It is very important to me that I perform well on physics work.	
	[a]*I’m the kind of person that does well on physics work.*	
	I expect to perform well on physics work.	
	I view physics work as a positive challenge.	

Note. This measure was adapted from the Adapted Stress Appraisal Questionnaire (Jamieson et al., 2016).
[a]Items used on the short-form for the screening survey and ecological momentary assessments (EMA)

Physics Tasks

Parts 1-3 of the physics tasks included a combination of multiple-choice and free-response items and were designed to measure three types of skills and knowledge related to physics knowledge and problem solving: Part 1: Quantitative problem solving (1 item), Part 2: Problem categorization (5 items), and Part 3: Qualitative (conceptual) problem solving (4 items). There were two equivalent versions of the physics tasks (A and B), which were counterbalanced between baseline and posttest. Testing order was distributed evenly between experimental groups. There was an additional task measuring preparation for future learning (Belenky & Nokes-Malach, 2012; Schwartz & Martin, 2004) given at posttest only. This additional task included two physics questions (each with a multiple-choice and free-response component) and a learning resource for students to study. Detailed descriptions of the tasks and sample items are provided in the supplementary materials.

Momentary Item-Level Task Judgments

Three judgments per physics question were each measured by one item: confidence ("I am confident in my answer."), anxiety ("I felt anxious working on this item."), and difficulty ("This item was difficult"). Judgments were rated on a Likert scale of 1 (*strongly disagree*) to 6 (*strongly agree*).

Mindfulness Training

The mindfulness training was structured in 5 daily, 20-minute audio-guided lessons. The training was built around a mindfulness practice called "RAIN," originally developed by Michele McDonald (https://learn.tricycle.org/p/rain). RAIN is an acronym that stands for Recognize, Accept, Investigate, and Non-Identify (Brach, 2013; Brewer, 2017). The practice was adapted to the context of learning physics by addressing common challenges and emotions that students reported in focus group interviews (Pelakh et al., 2025). Material for the program was developed in collaboration with a full-time mindfulness teacher and drew from prior written and recorded materials (e.g., Brach, 2019; Kornfield, 2008; Rumi, 2004). Attention check questions were included after each audio-guided lesson. These questions were in multiple-choice format at baseline and posttest (lessons 1 and 5), and short answer format during the week (lessons 2-4).

No-Training Control.

Students assigned to the no-training audiobook control condition listened to a 20-minute short story on day 1 ("The Stolen Dream" by Richard Le Gallienne, 1912) and day 5 ("The Bet" by Anton Chekhov, 1889). The stories were followed by a multiple-choice attention check question. The control condition received no audiobooks during days 2-4 of the study.

Procedure

Self-report surveys, physics tasks, and audio lessons were completed online using the Qualtrics survey platform. Study enrollment and informed consent were conducted online over Zoom. Study communications were conducted using phone calls, text messages, and email. An overview of the study design and procedure is provided in Figure 2.

Participant Recruitment and Screening Survey

Study recruitment occurred across three consecutive semesters. Researchers visited students' physics 1 classrooms (virtually and in person) to present a brief overview of the study and provide a link to the screening survey, which contained the short-form version of the psychological threat survey, as well as contact information (e.g., name, email, phone number) and demographic questions. Students were given several minutes to complete the screening survey in class but were permitted to complete it anytime within three days of presentation. Three students from each class were randomly chosen to receive a $25 gift certificate for completing the survey to incentivize participation. To be eligible for the study, participants were required to be enrolled in calculus-based physics 1, have a psychological threat score greater than zero, be 18 years of age or older, and be fluent in English. Eligible participants were recruited by phone to participate in the study and could earn up to $150 for completing all study activities.

Baseline (Day 1, Monday)

Students began the study by joining a researcher-led online video conferencing session (using Zoom) in groups of up to 5 participants. In the group session, students were given an overview of the study and provided informed consent using an online form. Consenting participants were randomly assigned to one of two conditions and their phones were registered with the survey delivery platform (SurveySignal). Students were allowed to select times on Tuesday through Thursday to complete each of the ecological momentary assessment (EMA) surveys in accordance with their schedules. After enrollment and registration, students completed the remainder of the activities independently online using Qualtrics. Ninety-five percent of the study sample completed the full set of these activities at an average of 86.1 minutes (SD = 17.3 minutes). Baseline activities (in order) were a battery of self-report survey measures (self-paced; including the long-form Psychological Threat measure), physics tasks (described in more detail below), listening to an audio recording (20 minutes; mindfulness training or short story), and demographic questions (self-paced).

All physics task items were presented individually on separate pages. Open response explanations and item-level judgment questions were included on the same page as the question for every item, excluding a single open-ended quantitative problem where judgment ratings were provided on the following page. All questions were timed, and a countdown timer was visible at the top of the page. Participants were given 10 minutes for the quantitative problem for Part 1, 2 minutes for each of the problem categorization items for a total of 10 minutes for Part 2, and 2.5 minutes for each of the 4 qualitative items for a total of 10 minutes for Part 3. Responses were required for every item. For Parts 2 and 3, if the allotted time elapsed before responses were selected, students were prompted to respond to any unanswered questions and move on. If all responses were selected and text input boxes were not empty, the survey would automatically advance to the next page when the timer ran out. For the open-response quantitative problem in Part 1, students completed their work on a piece of scratch paper and were automatically advanced to the next page after 10 minutes, where they were prompted to upload a photo of their work.

Ecological Momentary Assessment (EMA, Days 2-4, Tuesday - Thursday)

On days 2, 3, and 4 of the study, students received links to complete the EMA surveys via text messages delivered to their phones twice per day (afternoon and evening). Students assigned to the mindfulness training condition additionally received the 20-minute audio for lessons 2-4 in the afternoon before answering self-report items. EMA surveys included the short-form of the psychological threat measure, as well as additional self-report survey items assessing momentary mindfulness, equanimity, affect, cognitive appraisals, emotional reactivity, positive reappraisal, and physics engagement. These items were self-paced.

Posttest (Day 5, Friday)

Participants joined another researcher-led Zoom session at the beginning of their activities on Friday. The researcher collected payment information and provided them with a link to begin the posttest activities. Posttest activities were completed independently, similar to baseline. The posttest was almost identical to the baseline activity with a few exceptions. First, activities were presented in reverse order: the listening activity (20 minutes) was followed by physics tasks, and then self-report survey items (self-paced). Second, there were no demographic questions. Finally, the physics tasks included the preparation for future learning (PFL) task. For the PFL, students were given three minutes to solve each of the problems, and four minutes to study the learning resource for a total of up to 10 minutes. Because of the addition of the PFL task, students were allotted 40 minutes total for the physics tasks at posttest (compared to 30 minutes at baseline).

Follow-up Assessments

Students completed follow-up surveys approximately two weeks and three months from their posttest session. These measures were not included in the current analysis.

Scoring Procedure[3]

Psychological Threat

Psychological threat scores were calculated for each participant at each timepoint by taking the mean rating on the resources items and subtracting it from the mean rating on the demands items, such that a score greater than zero indicates perceived demands outweighing perceived resources. Psychological threat scores could range from -5 to 5. For EMA, means were calculated for each participant across all completed surveys. The mean completion rate across students was 96.6% (127 students completed all 6 EMA surveys, 18 students completed five EMA surveys, three students completed four surveys, and one student completed baseline and posttest activities but did not complete any of the EMA surveys). Because compliance was so high, student means were calculated from the completed surveys (ignoring missing values), with the exception of the one participant who had no EMA survey data. Their data were removed from the mediation tests.

Statistical Analysis

All data wrangling and statistical tests were conducted using R version 4.2.1 (R Core Team, 2022), and the tidyverse family of packages (Wickham et al., 2019). Mixed-effects models were run using the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017). The tables displaying model results were generated using the sjPlot package (Lüdecke, 2022). For all statistical tests, we used a criterion of .05 to infer statistical significance. We additionally report p-values between .05 and .1 as marginally significant.

Descriptive Statistics and Preliminary Analyses

Before building our statistical models, we ran basic descriptive analyses and zero-order correlation tests for our primary study variables at baseline: item-level judgments of confidence, anxiety, and difficulty; accuracy scores on the problem-solving tasks; and physics psychological threat. The purpose of these analyses was to ensure that the variables were normally distributed and showing the expected patterns of association irrespective of the intervention and random assignment.

Covariates

Gender, study cohort, semester week, test version, item-level accuracy, and baseline psychological threat were included as covariates in the models. Each variable is defined below. Some coding/scoring details varied between the RQ1 and RQ2 models to match the specific goals of each

[3] A detailed description of the scoring procedures for the physics questions is provided in the supplemental materials.

analysis and are described separately. All of these variables were determined to be theoretically relevant to the research questions, motivating their inclusion in the models.

All covariates were pre-registered except item-level accuracy and baseline psychological threat. Since we expected confidence, anxiety, and difficulty judgments to covary with the accuracy of students' answers, we included item-level accuracy in the regression models as a covariate. Although there is a theoretical possibility that posttest item-level accuracy could be impacted by the mindfulness training, we found that (i) there was no main effect of mindfulness or gender × mindfulness interaction on item-level accuracy pre-post improvement and (ii) excluding all of the covariates from the regression model did not change the patterns of significance for effects including timepoint, condition, and gender.

We decided to include baseline psychological threat as a covariate to provide a more conservative test of our hypotheses, especially the mediation hypotheses in RQ2, since baseline levels of physics threat are expected to be associated with both the mediator (mid-week physics threat measured using EMA) and outcome variables (judgments of confidence, anxiety, and difficulty). Thus, controlling for physics threat at baseline helps isolate the variance in these focal variables that can be attributed to the experimental variable (mindfulness training).

Gender. Participants' self-reported gender identity was treated as a 2-level factor with women and non-binary students at one level and men as a second. We chose to combine women and non-binary identifying students since these gender identities have been historically underrepresented in physics and therefore students identifying with these genders may be more likely to face greater demands and threats relative to men. This, combined with the low number of non-binary identifying students (two), makes it difficult to analyze as a separate group within this statistical approach. Men served as the reference group because we were particularly interested in the outcomes for women and non-binary students.

Cohort. Data were collected across three semesters: Cohort 1 (n = 61), Cohort 2 (n = 16), and Cohort 3 (n = 72), during which there were many fluctuations in course structure due to data collection coinciding with the COVID-19 pandemic.

Semester Week. The week relative to the semester that students participated in the study was included in the models to capture how far along students were in their course (M = 9.1, SD = 2.79). This variable was included because students who were further along in the semester would have covered more relevant material and this could have reasonably affected judgments on the physics tasks.

Test Version. There were two equivalent but non-identical versions of the physics tasks balanced for each participant between baseline and posttest. Seventy-six students completed version A at baseline and version B at posttest, and 73 students completed the versions in reverse order. We did not apply weights to test version because this variable was experimentally manipulated and group sizes were almost identical.

Item-Level Accuracy. To consistently account for accuracy at an item-level, we created a 2-level factor variable called *item-level accuracy*. Since two of the items on each test version were open-response and responses could receive partial credit, this variable reflected whether students were at least partially correct on a given item, compared to being completely incorrect.

Baseline Physics Psychological Threat. Self-reported psychological threat at baseline was included as a variable in all the models.

Multiple Comparisons

Corrections for multiple comparisons were considered, but ultimately not applied because our analysis consisted of a limited number of predefined, preregistered, directional hypotheses. Therefore, only unadjusted *p*-values are reported.

Research Question 1

Three linear mixed-effects models were used to test the effect of mindfulness training on item-level judgments of confidence (H1), anxiety (H2), and difficulty (H3). Condition (mindfulness vs. control), timepoint (baseline vs. posttest), as well as the interaction between condition and timepoint were included as fixed effects. Observations for these models were at the item response level, so item-level ratings of confidence, anxiety, and difficulty were not aggregated. A complete description of the fixed-effects variables included in the mixed models are provided in Table 3.

Each participant responded to 22 unique physics items across the two timepoints and test versions, resulting in a crossed random effects structure with random intercepts included for items and participants. Three responses were removed from the analyses using list-wise deletion due to missing accuracy data, resulting in $(149_{\text{Participant}} \times 22_{\text{Item}}) - 3_{\text{Missing}} = 3{,}275 \text{ observations per model.}$

Model Selection and Moderation by Gender. To determine whether to include gender as a moderator in the final models, we ran each of the 2-way interaction models specified in hypotheses 1-3 (timepoint × condition) and compared each of them to a model with a 3-way interaction (timepoint × condition × gender). We used multiple techniques to compare the models, including the *anova* function in R, which tests whether a more complex model provides a significantly better fit using a likelihood ratio test, as well as data visualization to determine which model was more consistent with the data.

Table 3

Characteristics of Fixed Effects Variables for Mixed Models

Variable	Type	Scale/Codes	Raw Mean (SD)
Dependent Variables			
Confidence	Continuous	Likert: 1 to 6	Baseline = 3.69 (1.44) Posttest = 3.88 (1.46)
Anxiety	Continuous	Likert: 1 to 6	Baseline = 3.31 (1.41) Posttest = 3.00 (1.37)
Difficulty	Continuous	Likert: 1 to 6	Baseline = 3.55 (1.26) Posttest = 3.43 (1.30)
Covariates			
Cohort 2	Weighted Contrasts	Cohort 1 = -0.11, Cohort 2 = +0.89, Cohort 3 = -0.11	
Cohort 3	Weighted Contrasts	Cohort 1 = -0.48, Cohort 2 = -0.48, Cohort 3 = +0.52	
Semester Week	Continuous	Mean-centered Centered min, max = -4.09, 4.91	9.1 (2.8)
Test Version [B]	Simple Contrasts	A = -0.5, B = +0.5	
Item-Level Accuracy	Weighted Contrasts	Incorrect = -0.44, (Partially) Correct = +0.56	Baseline = .43 (.50) Posttest = .44 (.50)
Baseline Threat	Continuous	Mean-centered Centered min, max = -3.1, 2.9	0.70 (1.3)
Main Independent Variables of Interest			
Timepoint [Posttest]	Binary	Baseline = 0, Posttest = 1	
Condition [Mindfulness]	Simple Contrasts	Control = -0.5, Mindfulness = +0.5	
Gender [Women or Non-binary]	Weighted Contrasts	Men = -0.56, Women or Non-binary = +0.44	

Note. The model intercepts reflect the mean value for each dependent variable at baseline at mean values of semester week and baseline threat, averaged across cohorts, test version, questions, gender, and condition.

Research Question 2

To test whether the effect of mindfulness training on judgments while answering physics questions was mediated by physics psychological threat (measured between baseline and posttest) we used the mediation package for R (Tingley et al., 2014), which provides simultaneous testing for total, direct, and indirect effects. The significance of the indirect effects was tested using 10,000 bias-corrected bootstrapped samples (Hayes, 2017). A random seed was set for each mediation test for replicability.

Psychological threat, measured using EMA, was included as the mediating variable (controlling for psychological threat measured at baseline) because these measurements were taken during the three days between the baseline and posttest assessments. For the purpose of these analyses, we used standard linear regression models instead of mixed-effects models, with the baseline measure of the dependent variable and baseline psychological threat included as covariates, in addition to cohort, semester week, test version, and accuracy.

Item-level judgment ratings and accuracy scores were aggregated for each participant at each timepoint so that participants had only one observation for each variable in the data. All continuous variables were converted to z-scores and standardized beta coefficients are reported as a measure of effect size and to allow for comparison between pathways. Categorical variables were dummy coded with reference levels consistent with the mixed models in RQ1. One participant did not respond to any of the EMA surveys, so their data were not included in any of the mediation analyses, leaving 148 observations in each model.

Moderated Mediation. If any of the effects of mindfulness training were determined to be moderated by gender, mediation analyses were reported separately for each level of the moderating variable as specified by Tingley et al. (2014).

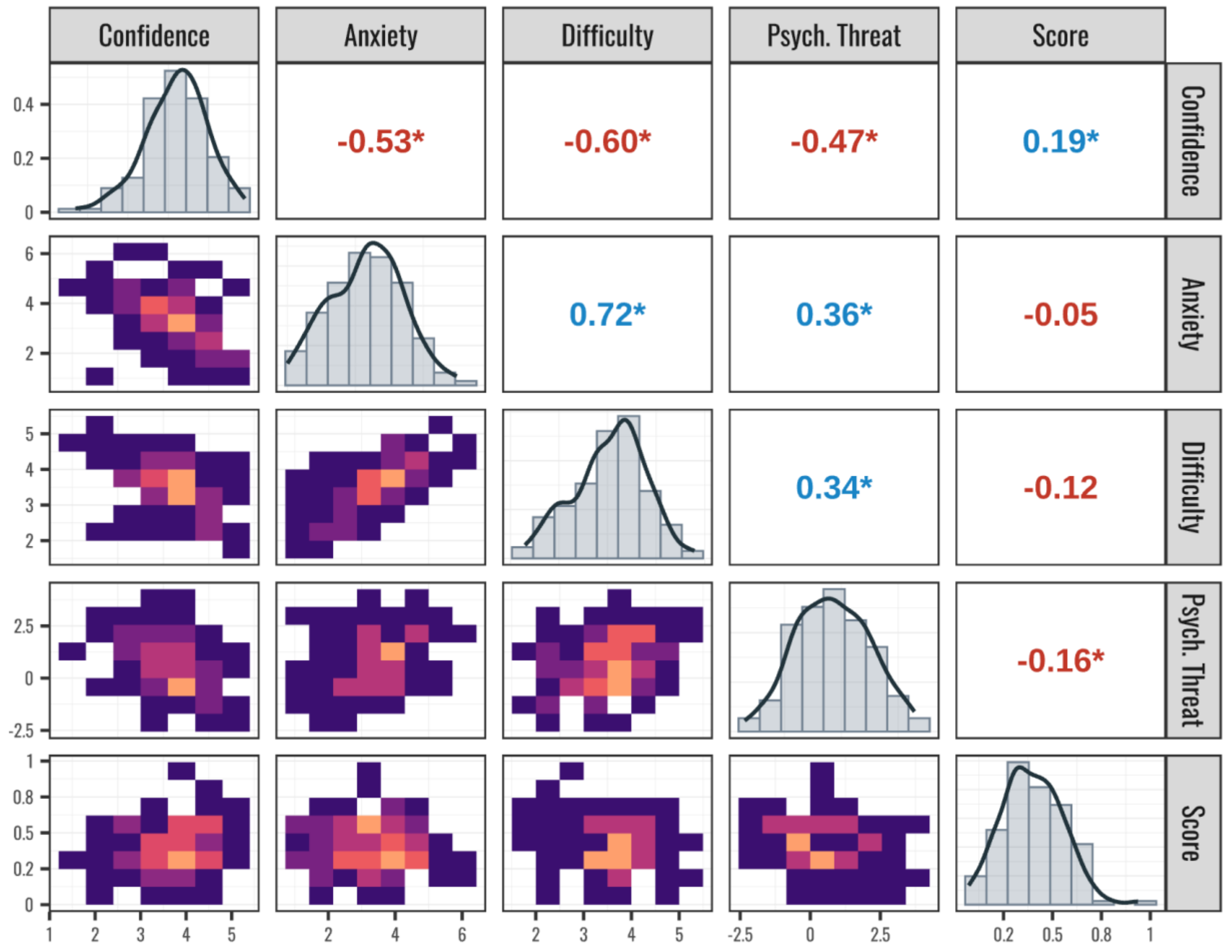
Results

Distributions and Correlations of Study Variables Before the Intervention

Initial descriptive analyses of the study variables at baseline showed that our measures were normally distributed and followed expected patterns of association (see Table 3 for means and standard deviations). Figure 3 shows that judgments of anxiety, difficulty, and psychological threat were all positively associated at baseline, with mean judgments of anxiety and difficulty showing the strongest association. Judgments of item confidence were negatively associated with anxiety, difficulty, and physics psychological threat. Accuracy scores on the physics items were positively associated with confidence and negatively associated with physics threat, though the strength of these associations were weaker than the self-reported measures with each other. Correlations between score with anxiety and difficulty were in the negative direction as expected, but these associations were not statistically significant.

Figure 3

Correlations and Variable Distributions at Baseline



Note. Two-dimensional histograms are provided in the lower-left triangle to show the joint distributions of the study variables. The lighter (yellow/orange) colored squares represent a greater concentration of observations, while the darker (purple) squares contain fewer observations. *Score* reflects the mean item accuracy scores for each participant. For the correlation coefficients in the upper-right triangle, *all $p < .05$.

Research Question 1: What is the effect of mindfulness training on item-level judgments of confidence, anxiety, and difficulty?

Hypotheses 1-3 predicted that students randomly assigned to the mindfulness training group would report greater confidence, and lower anxiety and difficulty, at posttest compared to students assigned to the control group. The complete results of the three mixed-effects models used to test these hypotheses are reported in Table 4. For all three models, timepoint is coded as a two-level factor (baseline = 0, posttest = 1). Therefore, the intercepts of the models reflect the mean judgment ratings for each judgment type at baseline, and the estimates for timepoint reflect the average change in judgment ratings for each judgment type from baseline to posttest, and the timepoint by condition interaction estimates represent the difference in change in judgment ratings between the mindfulness group and the control group at posttest above and beyond any baseline differences by condition and average change from baseline to posttest. Participants' mean judgment ratings at baseline and posttest as a function of condition and gender are depicted in **Error! Reference source not found.**. In the following sections, effects of involving the main independent variables of interest at baseline are reported first, followed by effects at posttest (includes all interactions with timepoint). Lastly, we report effects of the covariates.

Table 4

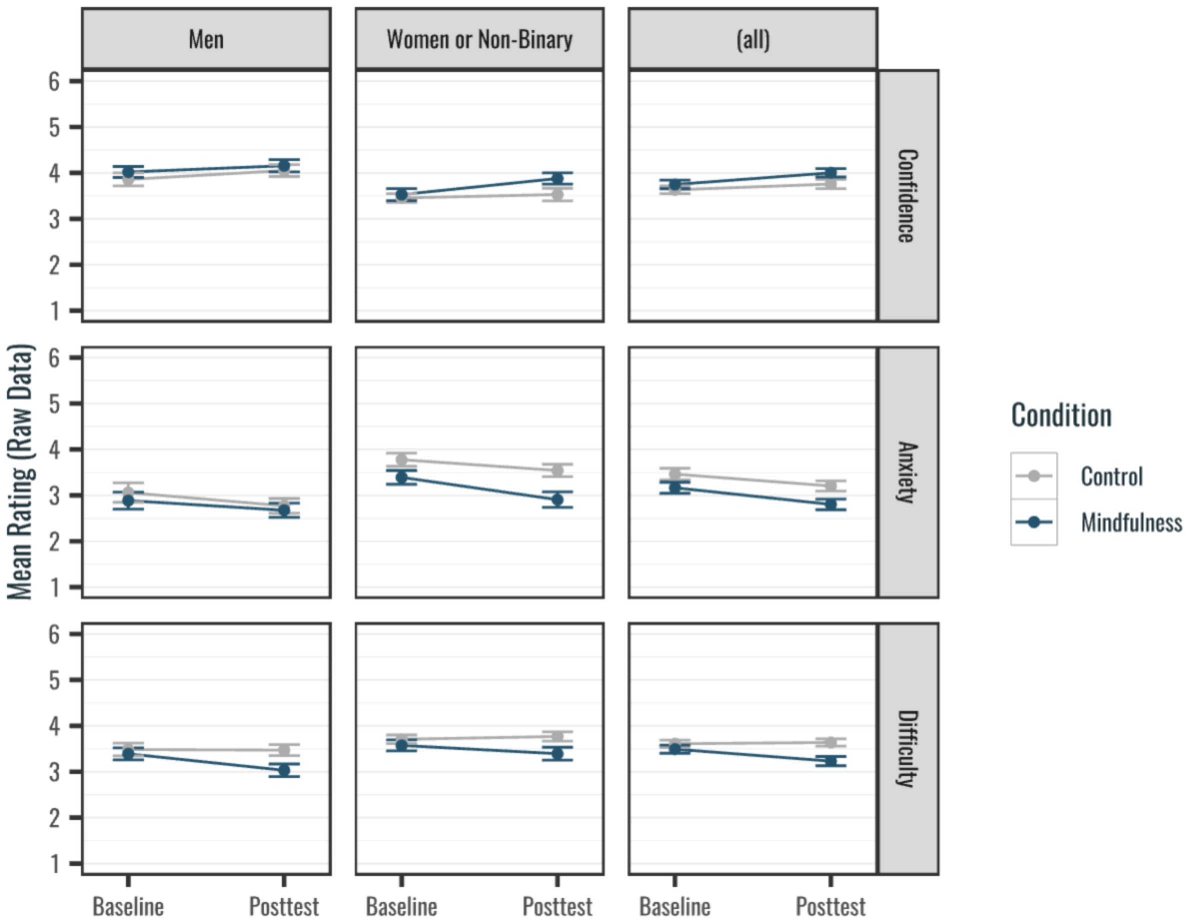
Results from Mixed Effects Models Testing Hypotheses 1-3: Effects of Mindfulness Training on Item-Level Judgments While Answering Physics Questions

	H1: Confidence			**H2: Anxiety**			**H3: Difficulty**		
Predictors	*Estimates*	*SE*	*p*	*Estimates*	*SE*	*p*	*Estimates*	*SE*	*p*
(Intercept)	3.69	0.15	**<0.001**	3.33	0.12	**<0.001**	3.56	0.12	**<0.001**
Cohort [Cohort 2]	0.38	0.31	0.222	-0.16	0.43	0.709	-0.52	0.32	0.106
Cohort [Cohort 3]	0.07	0.27	0.785	0.06	0.38	0.868	-0.33	0.28	0.235
Semester Week	0.10	0.05	**0.036**	-0.04	0.07	0.537	-0.12	0.05	**0.019**
Test Version [B]	-0.13	0.12	0.268	-0.08	0.10	0.385	-0.07	0.10	0.518
Item-Level Accuracy	0.13	0.04	**0.005**	-0.08	0.04	**0.026**	-0.03	0.04	0.426
Baseline Threat	-0.24	0.04	**<0.001**	0.22	0.06	**<0.001**	0.15	0.04	**<0.001**
Timepoint [Posttest]	0.19	0.04	**<0.001**	-0.34	0.03	**<0.001**	-0.12	0.04	**<0.001**
Condition [Mindfulness]	0.07	0.11	0.545	-0.25	0.15	0.087†	-0.08	0.11	0.479
Gender [Women or Non-binary]	-0.35	0.12	**0.003**	0.49	0.16	**0.002**	0.22	0.11	**0.048**
Timepoint [Posttest] × Condition [Mindfulness]	0.13	0.08	0.089†	-0.11	0.07	0.102	-0.29	0.07	**<0.001**
Timepoint [Posttest] × Gender [Women or Non-binary]	0.06	0.08	0.451	-0.12	0.07	0.071†			
Condition [Mindfulness] × Gender [Women or Non-binary]	-0.19	0.22	0.391	-0.11	0.30	0.721			
Timepoint [Posttest] × Condition [Mindfulness] × Gender [Women or Non-binary]	0.32	0.15	**0.034**	-0.31	0.13	**0.018**			
Random Effects									
σ^2	1.17			0.87			0.93		
τ_{00}	0.33 Participant			0.72 Participant			0.37 Participant		
	0.45 Item			0.17 Item			0.27 Item		
ICC	0.40			0.51			0.41		
N	149 Participant			149 Participant			149 Participant		
	22 Item			22 Item			22 Item		
Observations	3275			3275			3275		
Marginal R^2 / Conditional R^2	0.099 / 0.461			0.123 / 0.567			0.068 / 0.448		

Note. The estimates for the intercept represent the overall mean ratings and standard errors for each of the item-level judgment outcomes at baseline. The estimate for timepoint represents the change in the dependent variable from baseline to posttest across conditions. The estimate for condition represents the difference on the dependent variable for mindfulness compared to control at baseline. The estimate for the timepoint × condition interaction represents the difference in average change from baseline to posttest between conditions, above and beyond any overall effect of timepoint and holding all other variables constant. Additional interaction terms are reported for judgments of confidence and anxiety because the effect of condition on change in ratings differed significantly by gender. *P*-values below .05 are in indicated by bold font. †indicates marginally significant *p*-values ($p < .1$).

Figure 4

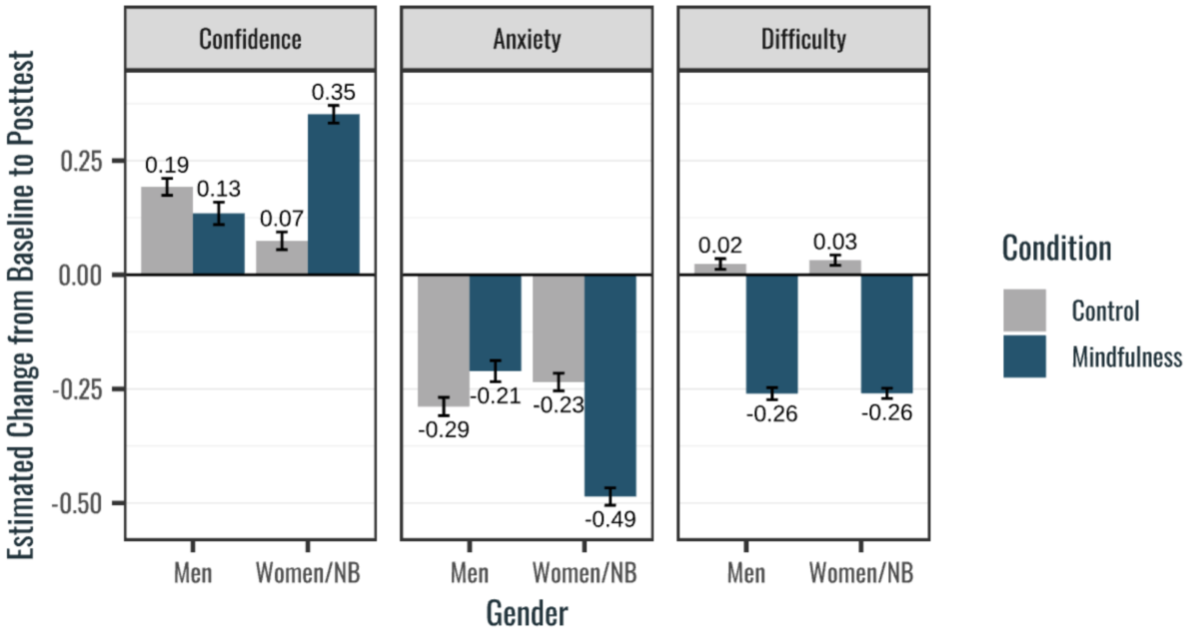
Participants' Mean Judgment Ratings at Baseline and Posttest by Experimental Condition and Gender



Note. Error bars reflect standard errors. The effect of mindfulness on judgments of difficulty was consistent across gender. Effects on confidence and anxiety are moderated by gender.

Figure 5

Estimated Marginal Means for Effects of Main Variables of Interest



Note. Group means were calculated by taking the mean of the model-predicted judgment ratings for each participant across items at each timepoint and then calculating the average change from baseline to posttest for each group. Error bars represent 95% confidence intervals from a paired t-test between participants' baseline and posttest model-predicted ratings, equal variance assumed.

H1: Confidence

For judgments of confidence, we report results from the model that includes a 3-way interaction between time, condition, and gender. Aligned with the goals of random assignment, there was no significant difference in confidence by condition at baseline. There was a significant baseline effect of gender on confidence (B = -.35 SE = .12, p = .003), indicating that women and non-binary identifying students reported confidence ratings that were .35 units lower than men. This baseline gender difference was similar for the control and mindfulness groups, as reflected by the non-significant condition by gender interaction.

There were experimental effects of mindfulness on confidence seen at posttest. Specifically, there was a marginally significant effect of condition (B = .13, SE = .08, p = .089) and a significant interaction between condition and gender (B = .32, SE = .15, p = .034). Figure 4 and Figure 5 show that there was a greater increase in confidence among women and non-binary students in the mindfulness group compared to women and non-binary students in the control group. The difference between experimental groups among men was negligible, with the control group having a slightly greater increase. Overall, the average change in judgments of confidence from baseline to posttest did not differ between gender groups (*B* = .06, *SE* = .08, *p* = .451).

Of the covariates, there were only significant effects of semester week (*B* = .10, *SE* = .05, *p* =.036), item-level accuracy (*B* = .13, *SE* = .04, *p* = .005), and baseline psychological threat (*B* = -.24, *SE* = 0.04, *p* < .001). These mean that: a one week progression in the semester was associated with a .10 unit increase in judgment of confidence, confidence ratings were .13 units higher for answers that were at least partially correct compared to incorrect, and a one unit increase in baseline psychological threat was associated with a .24 unit decrease in item-level judgments of confidence.

H2: Anxiety

For judgments of anxiety, we report results from the model that includes a 3-way interaction between time, condition, and gender. There was a marginally significant effect of condition at baseline, with students in the mindfulness training group reporting judgments of anxiety that were .25 units lower than the control group before the intervention (*B* = -.25, *SE* = .15, *p* = .087), showing that, counter to what was intended, the groups were not completely balanced on anxiety to begin with. There was also a significant effect of gender (*B* = .49, *SE* = .16, *p* = .002) on item-level judgments of anxiety at baseline, with women and non-binary identifying students reporting anxiety ratings that were .49 units higher than men. The baseline gender difference in anxiety did not differ by condition as shown by the non-significant condition by gender interaction (*B* = -.11, *SE* = .30, *p* = .721).

There were experimental effects of mindfulness training on judgments of anxiety at posttest. The effect of condition at posttest was not significant, indicating no overall effect of mindfulness training on changes in judgments of anxiety (*B* = -.11, *SE* = .07, *p* = .102), however, there was a significant interaction of condition by gender described below. There was a marginally significant effect of gender at posttest (*B* = -.12, *SE* = .07, *p* = .071), with women and non-binary students reporting greater reduction in anxiety on average at posttest compared to men. There was also a significant condition by gender interaction at posttest (*B* = -.31, *SE* = .13, *p* = .018). Figure 4 and Figure 5 show that among women and non-binary students, those who received mindfulness training had a substantially greater reduction in judgments of anxiety compared to control. Among men, students in the control group showed a slightly greater reduction in anxiety compared to those in the mindfulness group.

Of the covariates, there were only significant effects of item-level accuracy (*B* = -.08, *SE* = .04, *p* = .026) and physics psychological threat (*B* = .22, *SE* = .06, *p* < .001) on judgments of anxiety. These mean that: anxiety ratings were .08 units lower for answers that were at least partially correct and that a one unit increase in psychological threat at baseline was associated with a .22 unit increase in anxiety ratings on average.

H3: Difficulty

We did not see any evidence of gender moderation for judgments of difficulty; therefore, we report the results of the simpler 2-way (Timepoint × Condition) interaction model. Aligned with the goals of random assignment, there was no significant difference in difficulty by condition at baseline. There was a significant overall effect of gender (B = .22, SE = .11, p = .048) on difficulty, indicating that women and non-binary identifying students reported difficulty ratings that were .22 units higher than men. This gender difference was similar for the control and mindfulness groups, as reflected by the non-significant condition by gender interaction (B = -.08, SE = .11, p = .479).

There was an experimental effect of mindfulness training on judgments of difficulty at posttest, as shown by the significant timepoint by condition interaction (B = -.29, SE = .07, p < .001). Students who received mindfulness training reported an average decrease in difficulty ratings from baseline to posttest that was .29 units greater than students in the control group.

Of the covariates, there were only significant effects of semester week (B = -.12, SE = .05, p = .019) and baseline threat (B = .15, SE = .04, p < .001) on judgments of difficulty. This means that a one-week progression in the semester was associated with a .12 unit decrease in judgments of difficulty and higher baseline threat was associated with increased difficulty ratings, with a one unit increase in physics threat corresponding to a .15 unit increase in item-level difficulty judgments.

Summary of Research Question 1 Results

We found some evidence consistent with our predictions about the effects of mindfulness training on item-level judgments while answering physics questions. The intervention led to an increase in confidence judgments and a decrease in anxiety judgments for women and non-binary students. Confidence and anxiety judgments were unaffected by the intervention among men. For difficulty judgments, the mindfulness intervention showed clear overall effectiveness, with all participants in the mindfulness condition experiencing significantly greater reductions in judgments of item-level difficulty compared to those assigned to the control condition[4].

The results of the models testing hypotheses 1-3 also reveal consistent gender differences at baseline across all three outcome measures, with women and non-binary students reporting significantly lower confidence, higher anxiety, and higher difficulty judgments compared to men, holding

[4] To ensure consistency across task type, each of the physics subtasks included at baseline and posttest (Parts 1-3) were also analyzed separately and showed the same pattern of results.

baseline physics threat constant. Additionally, physics psychological threat emerged as a robust predictor across models, with higher threat at baseline predicting lower confidence and higher anxiety and perceived difficulty, holding all other predictors constant. Alternatively, semester progression showed beneficial effects on confidence and difficulty judgments, suggesting that additional time in the course has a positive effect on student judgments. There also appeared to be a practice effect across all judgment outcomes, with all students reporting greater confidence and lower anxiety and difficulty at posttest compared to baseline on average.

Research Question 2: Does domain-level physics psychological threat mediate the association between mindfulness training and item-level judgments of confidence, anxiety, and difficulty?

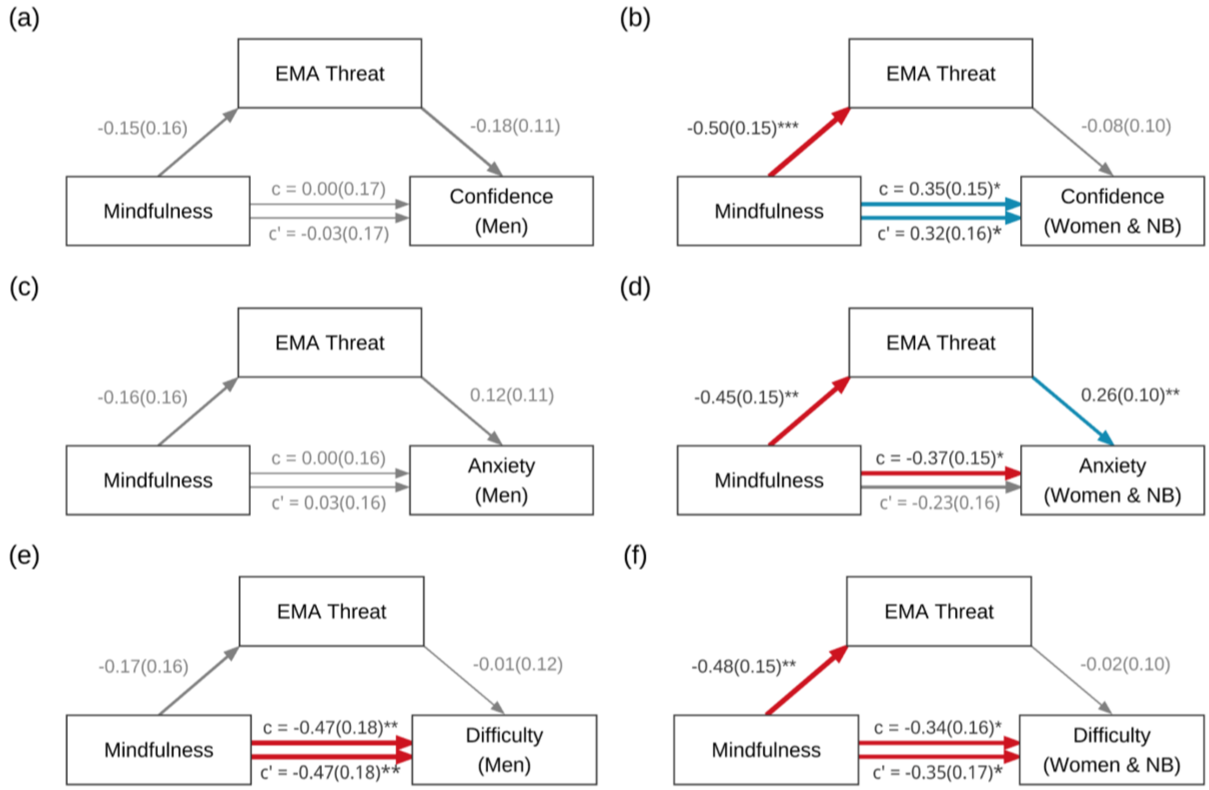
Hypothesis 4 predicted that psychological threat, measured using EMA, would mediate the relationship between mindfulness training and item-level judgments of confidence, anxiety, and difficulty. As a reminder, we used a linear regression-based approach to test for mediation. To do this, we took the average of each judgment rating at baseline and posttest and predicted posttest ratings while controlling for baseline ratings.

Despite the fact that we found no evidence of moderation by gender for judgments of difficulty in RQ1, we decided to split our mediation analyses by gender for all three judgment outcomes[5]. We chose to do this because our preliminary tests indicated some difference in the effect of mindfulness on our mediator (EMA physics threat) by gender. Therefore, performing a moderated mediation test shows whether this effect has any bearing on the gender-neutral effect of mindfulness training on judgments of difficulty. We followed the procedure outlined in Tingley et al. (2014, p. 10) for our moderated mediation tests, which included two mediation tests per outcome: one estimating the indirect effect for women and non-binary students, and the other, for men. Path diagrams for all of the mediation models tested for hypothesis 4 are provided in Figure 6.

[5] Mediation analyses without gender moderation are provided in the supplemental materials.

Figure 6

Path Diagrams Showing the Association Between Mindfulness Training and Item-Level Judgments as Mediated by Psychological Threat, Measured using EMA.



Note. For all plots, standardized betas and standard errors estimated using linear regression are shown along the pathways. Statistically significant ($p < .05$) pathways are colored red for negative associations or blue for positive associations, and the strength of each path is represented by linewidth. Plots (a), (c), and (e) show the mediation pathways on confidence and anxiety for men. Plots (b), (d), and (f) show mediation pathways on confidence and anxiety for women and non-binary (NB) students. The pathway coefficients and error estimates for each level of gender were obtained by running the regression models with gender interaction twice, once with men coded as 1 (panels (b), (d), and (f)) and once with women and non-binary students coded as 1 (panels (a), (c), and (e)).

Confidence

For the women and non-binary group, there was a positive total effect of mindfulness on confidence (c = .35, SE = .15, p = .019). Although mindfulness had a negative effect on EMA threat (a = -.50, SE = .15, p = .001), EMA threat was not associated with confidence when controlling for condition (b = -.08, SE = .10, p = .401). Therefore, the indirect effect was not significant (-0.50 × -0.08 = 0.04, bootstrapped 95% CI [-.03, .19]), and the direct effect reflected the bulk of the total effect (c' = .32, SE = .16, p = .044).

For the men, there was no significant total, direct, or indirect effect (-0.15 × -0.18 = .03, bootstrapped 95% CI [-.02, .1]) of mindfulness on confidence. Also, unlike for the women and non-binary group, mindfulness had no effect on EMA threat (*a* = -0.15, SE = 0.16, p = .353).

Anxiety

We tested the significance of the moderated indirect effect of mindfulness training on judgments of anxiety at posttest, as mediated by psychological threat, measured using EMA, and found a significant effect for women and non-binary students (-.45 × .26 = -.12, bootstrapped 95% CI [-.32, -.02]) but not for men (-.16 × .12= -.02, bootstrapped 95% CI [-.13, -.01]).

For women and non-binary students, mindfulness training had a significant negative total effect on judgments of anxiety (*c* = -.37, *SE* = .15, *p* = .015) and a significant negative association with physics threat measured using EMA (*a* = -.45, *SE* = .15, *p* = .003). EMA threat had a significant positive association with anxiety (*b* = .26, *SE* = .10, *p* = .009). However, the direct effect of mindfulness training on anxiety was not significant (*c'* = -.23, *SE* = .16, *p* = .140), suggesting that the relationship between mindfulness training and reduced anxiety was mediated by reductions in physics psychological threat. None of the pathways were significant for men.

Difficulty

Consistent with the results from the RQ1 model, there was a significant total effect of mindfulness training on judgments of difficulty at posttest for men (*c* = -.47, *SE* = .18, *p* = .009) and women and non-binary students (*c* = -.34, *SE* = .16, *p* = .031) alike. Though mindfulness training had a significant negative effect on physics threat measured using EMA only for women and non-binary students (*a* = -.48, *SE* = .15, *p* = .001), EMA threat was not associated with judgments of difficulty when controlling for condition in either group. The direct effects were nearly identical to the total effects for men (*c'* = -.47, *SE* = .18, *p* = .009) as well as women and non-binary students (*c'* = -.35, *SE* = .17, *p* = .040). Therefore, the indirect effects were also non-significant for men (-.17 × -.01 = .00, bootstrapped 95% CI [-.03, .09]) and women and non-binary students as well (-.48 × -.02 = .01, bootstrapped 95% CI [-.12, .13]). The consistency across genders with regard to judgments of difficulty despite the moderated effect on EMA threat shows that the effect is quite robust to gender differences.

Summary of Research Question 2 Results

We found some support for physics psychological threat as a mechanism for the effects of mindfulness on judgment outcomes. The mediation analyses revealed distinct patterns in how mindfulness training affects item-level judgments while answering physics questions across gender and judgment type.

Across all three judgment outcomes, the effects of mindfulness training on physics threat, as measured midweek using EMA, were statistically significant for women and non-binary students. There was also a significant total effect of mindfulness training on the three judgment outcomes for women and non-binary students (when EMA threat was not included in the model). However, the mindfulness-based reduction in physics threat explained the change in judgment ratings for anxiety but not for confidence or difficulty.

The effects of mindfulness training on EMA physics threat for men were in the expected direction (negative), but the associated standard errors were quite large and the pathways were non-significant. Mindfulness training had a significant total and direct effect on difficulty judgments for men, but there were no significant indirect effects of EMA threat on the association between mindfulness training and judgment ratings for any of the three judgment types.

Discussion

We set out to better understand the relationship between psychological appraisals of stress at the domain level (physics threat) and judgments of confidence, anxiety, and difficulty while answering physics questions. We tested the effects of a 5-day mindfulness training intervention targeting meaning-making processes around stress on judgments while answering physics questions. We further tested whether domain-level threat mediates the association between mindfulness training and judgments and explored gender as a moderator of the intervention effects.

To our knowledge, ours is the first study to test the association between challenge/threat and item-level judgments during problem solving tasks. Probing student experiences in the moment—as opposed to prospectively and/or retrospectively creates opportunities to test and refine process models of the challenge-threat framework and make specific predictions for specific kinds of problems and experiences. This also helps connect abstract theoretical constructs (i.e., resources, demands, challenge, threat) to specific dimensions of student experiences and problem solving activities which can be further investigated and tested at the item level (e.g., physics threat and momentary anxiety among women and non-binary students).

Baseline Associations

We found that physics threat and gender both predicted item-level judgments of confidence, anxiety, and difficulty while answering physics questions at baseline (before the intervention), holding accuracy performance constant. The robust effect of physics threat is consistent with our theoretical framework which supposes that challenge/threat states affect how performance situations (like answering physics questions) are experienced, with greater threat predicting more negative experiences. The fact that gender effects persisted when holding threat and accuracy constant, with women and non-binary students reporting lower confidence and greater anxiety and difficulty, suggests that there are aspects of experience related to social identity that are not entirely explained by challenge-threat model. Our results somewhat contradict the those from several studies which found that women reported greater math anxiety compared to men when measured at the domain-level, but found no gender differences in momentary math anxiety when measured before and after a specific task (Bieg et al., 2015; Goetz et al., 2013). Whereas we found gender differences in momentary judgments for every judgment type during our tasks. However, these discrepancies may be explained by differences in sample characteristics (high school vs. college) or measurement approach (task-level vs. item-level).

Effects of Mindfulness Training

Despite the fact that judgments of confidence, anxiety, and difficulty were strongly intercorrelated and also associated with appraisals of baseline physics threat, they showed distinct patterns in response to the intervention with respect to both gender identity and domain-level threat. All of our judgment measures were affected, at least in part, by the mindfulness intervention. Consistent with hypothesis 3, we observed the most consistent effect of the intervention on judgments of item difficulty, with students of all genders reporting lower difficulty after receiving mindfulness training. It could be that mindfulness training helped buffer negative side effects of mental struggle (i.e., self-doubt), or helped them see moments of struggle as temporary, making the whole experience feel less effortful.

In contrast, increases in confidence and decreases anxiety (consistent with hypotheses 1 and 2) were specific to women and non-binary students, so potential explanations need to take this interaction into account. Though there is little prior work looking specifically at the effects of mindfulness training on momentary judgments during problem solving, this result is consistent with other work that found a stronger reduction anxiety in women enrolled in STEM-focused college courses compared to men as a result of an intervention that combined mindfulness with growth mindset principles (Samuel et al., 2022).

The results of our mediation models help us to further understand the outcome effects. Overall, the EMA data show that the intervention effectively reduced domain-level physics threat, with students in the mindfulness group reporting lower mid-week threat overall, though this effect was only statistically significant for women and non-binary students. Consistent with hypothesis 4, physics threat emerged as a mechanism of anxiety reduction for women and non-binary students during the physics activities. This is also consistent with the Interpretation Account of math anxiety (Ramirez et al., 2018), which emphasizes the connection between psychological appraisals (threat) and anxiety.

Mediation results for judgments of confidence and difficulty suggest the presence of alternative pathways. Effects of the intervention on these measures were not mediated by changes in physics threat for any gender group. It's possible that the aspects of training focused on attention and awareness in the present moment influenced confidence and difficulty judgements directly.

Gender and Mindfulness Training

Despite consistent associations between gender and all three types of judgments at baseline, effects of mindfulness training differed by gender only for judgments of confidence and anxiety. Of these, only anxiety was related to changes in mid-week physics threat. Here, we propose some possible interpretations for these findings.

One possibility comes from looking at the differences in the wording of the items measuring confidence and anxiety vs. difficulty. The statements used in the confidence and anxiety measures were self-focused (e.g., "I am confident…", "I felt anxious…"), while difficulty was question-focused ("This item was difficult"). Therefore, our confidence and anxiety measures could have been more sensitive to identity-related differences that intersect with gender.

Some research suggests that women are more likely to notice bodily sensations and connect them to emotions compared to men (Grabauskaitė et al., 2017). Therefore, women may have been more receptive to aspects of the intervention that were focused on the interface between bodily awareness and emotional response. This explanation may be especially relevant for the connection between EMA physics threat and judgments of anxiety among women and non-binary students, since both threat and anxiety are defined by both somatic and cognitive components.

Some additional potential explanations for the gender-moderated effects are that improvements in mindful attention could be more beneficial for students experiencing stereotype threat specifically, with women experiencing more negative stereotypes in physics (Cwik & Singh, 2021a), or that a training focus on non-judgmental acceptance could boost self-compassion for students more prone to self-criticism, with women historically underestimating their competence relative to men (Bench et al., 2015; Goetz et al., 2013; Marshman, Kalender, Nokes-Malach, et al., 2018). We cannot speak to the validity of any of these explanations, but our results highlight the importance of measuring multiple facets of experience; effects of an intervention on judgments can vary in both magnitude and mechanism, regardless of how similar the different judgment types appear at the outset.

Limitations

There are several points to keep in mind when interpreting the results reported in this study. First, we address our decision to sample from the subset of students who reported experiencing physics threat during the screening survey. This was an intentional choice, consistent with other studies of mindfulness and stress that use perceived stress as a selection criterion (Lindsay et al., 2018). In this case, we wanted to know if we could help threatened students approach physics as more of a challenge: a research objective that assumes the presence of threat to begin with. Therefore, though it was the right choice for our study, we cannot assume that mindfulness training would have the same effects for non-threatened students, or that these effects would even be desirable for a student with resources far outweighing demands (Binning & Browman, 2020).

Second, there were several variables that influenced changes in judgments from baseline to posttest irrespective of the intervention. Significant effects of timepoint showed overall improvement in all three judgment measures regardless of condition, suggesting task familiarity benefits. This assumption is reinforced by modest effects of semester week, with students participating later in the semester showing more favorable outcomes. This pattern is consistent with the challenge-threat model, which identifies uncertainty as a demand and prior experience as a resource, but could have masked some effects of the intervention. For example, the intervention could have seemed less relevant to a student with lower uncertainty demands, dampening the potential impact of the training. However, the fact that training effects were present after controlling for task familiarity and course progression only speaks to the unique contribution of mindfulness training.

Finally, though our gender categories included non-binary students, the low number of students identifying as non-binary (n = 2) means that we cannot really make any assumptions about this group. Our decision to group them with women was theoretical and other studies interested in women specifically have made different choices (Hazari et al., 2022).

Future Directions

By connecting the challenge-threat framework to fine-grained measures of students' subjective experiences answering physics questions, this study lays the groundwork for the development of process models that more accurately represent these types of experiences.

Additionally, this work opens several new avenues of questioning. One question that can be addressed is: would any stress management or reappraisal training benefit students or is there something unique about mindfulness training that is especially impactful, or impactful in certain ways? For example, future studies could test a mindfulness training intervention against other types of established stress-related interventions to better understand the specific contributions of mindfulness principles.

Another potential avenue for research is to expand the range of outcome measures to include more aspects of experiences students regularly encounter during learning and problem solving. For example, now that we have evidence that mindfulness training affects judgments of confidence, anxiety, and difficulty, we can ask: do these effects extend to other antecedents of learning? Future studies could include tasks that incorporate and measure multiple indicators of behavioral engagement (e.g., giving students more autonomy and recording their decisions). This could provide a more realistic simulation of homework and study activities, and help us understand which of these indicators are affected by training.

Conclusion

We presented a study that draws on multiple lines of work to make both theoretical and practical contributions. We hope to advance understanding of challenge and threat processes in academic contexts, and especially at a fine-grained item-by-item level during learning and problem-solving activities. Further, we introduced brief mindfulness training as an accessible, scalable intervention to help students experiencing threat in competitive university environments and as a potential avenue to help address gender disparities in confidence and anxiety in college-level physics.

Acknowledgements

We thank the many introductory physics teachers who helped us recruit participants over several semesters and our research assistants, Brennon Paik and Alexi Zukas, who spent many hours coding and scoring the physics task responses. We are also grateful to our National Science Foundation advisory board members (David Creswell, Zahra Hazari, Jeremy Jamieson) for their helpful feedback and support during this project. We also want to thank the many other people who contributed to this larger project, including our postdoctoral fellow, Tessa Benson-Greenwald, PhD students (Angela Griffo, Nabila Jamal-Orozco, Therese Schnettler, Shana DeVlieger, Tyree Langley), master's students (Kristin Hinshaw, Sai Aishwarya Karanam), project coordinators (Mike Diamond, McNoah Vinoya, Dalia Maeroff, Tyree Langley), and undergraduate research assistants (Amy Adelman, Jordann Antoan, Alyssa Chomitzky, Jovan Corrales, Brandon Baun, Nicole El Hayek, Jessica Giuffrida, Irvin Li, Keely Lombardi, Gillian Murphy, Connor McGrath, Austin Rosenkranz, Kayla Tupper).

References

Bai, S., Elavsky, S., Kishida, M., Dvořáková, K., & Greenberg, M. T. (2020). Effects of mindfulness training on daily stress response in college students: Ecological momentary assessment of a randomized controlled trial. *Mindfulness*, *11*(6), 1433–1445. https://doi.org/10.1007/S12671-020-01358-X/FIGURES/2

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Worth Publishers.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Beilock, S. L., Schaeffer, M. W., & Rozek, C. S. (2017). Understanding and addressing performance anxiety. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation* (pp. 155–172). The Guilford Press.

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *The Journal of the Learning Sciences*, *21*(3), 399–432. https://doi.org/10.1080/10508406.2011.651232

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender Gaps in Overestimation of Math Performance. *Sex Roles*, *72*(11–12), 536–546. https://doi.org/10.1007/S11199-015-0486-9/TABLES/7

Bieg, M., Goetz, T., Wolter, I., & Hall, N. C. (2015). Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Frontiers in Psychology*, *6*, 158430. https://doi.org/10.3389/FPSYG.2015.01404/BIBTEX

Binning, K. R., & Browman, A. S. (2020). Theoretical, ethical, and policy considerations for conducting social–psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*, *14*(1), 182–216. https://doi.org/10.1111/sipr.12066

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, *11*(3), 230–241. https://doi.org/10.1093/CLIPSY.BPH077

Blascovich, J. (2008). Challenge and threat. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 431–445). Psychology Press. https://psycnet.apa.org/record/2008-10051-025

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In *Advances in Experimental Social Psychology* (Vol. 28, Issue C, pp. 1–51). Academic Press. https://doi.org/10.1016/S0065-2601(08)60235-X

Brach, T. (2019, February 26). *Guided Meditation: The Practice of RAIN with Tara Brach [Video]*. https://www.youtube.com/watch?v=W8e_tAEM80k&t=19s

Brach, T. (2013). *True refuge: Finding peace and freedom in your own awakened heart*. Bantam Books.

Brewer, J. (2017). The craving mind: From cigarettes to smartphones to love—why we get hooked and how we can break bad habits. In *Journal of Nervous & Mental Disease* (Issue 4). Yale University Press.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *10*(6), 1691–1730. https://doi.org/10.1086/321299

Cwik, S., & Singh, C. (2021a). Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men. *Physical Review Physics Education Research*, *17*(2), 020138. https://doi.org/10.1103/PhysRevPhysEducRes.17.020138

Cwik, S., & Singh, C. (2021b). How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses. *Physical Review Physics Education Research*, *17*(2), 020143. https://doi.org/10.1103/PhysRevPhysEducRes.17.020143

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). *Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit*. https://doi.org/10.1371/journal.pone.0157447

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., & Boone, H. (2016). *More Comprehensive and Inclusive Approaches to Demographic Data Collection*. http://docs.lib.purdue.edu/enegs/60

Galante, J., Friedrich, C., Tomfohr-Madsen, L., Tang, F. C., Speckens, A. E. M., Singh, N. N., Siebelink, N. M., Schellekens, M. P. J., Rosenkranz, M. A., Medvedev, O. N., Jones, P. B., Hwang, Y. S., Hunsinger, M., Huang, S. L., Hirshberg, M. J., Greven, C. U., Goldberg, S. B., Errazuriz, A., Davidson, R. J., … White, I. R. (2023). Systematic review and individual participant data meta-analysis of randomized controlled trials assessing mindfulness-based programs for mental health promotion. *Nature Mental Health 2023 1:7*, *1*(7), 462–476. https://doi.org/10.1038/s44220-023-00081-5

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/https://doi.org/10.1177/0956797613486989

Grabauskaitė, A., Baranauskas, M., & Griškova-Bulanova, I. (2017). Interoception and gender: What aspects should we pay attention to? *Consciousness and Cognition*, *48*, 129–137. https://doi.org/10.1016/j.concog.2016.11.002

Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. In *the Guilford Press* (Issue 3). Guilford publications.

Hazari, Z., Dou, R., Sonnert, G., & Sadler, P. M. (2022). *Examining the relationship between informal science experiences and physics identity: Unrealized possibilities*. https://doi.org/10.1103/PhysRevPhysEducRes.18.010107

Hazari, Z., Tai, R. H., & Sadler, P. M. (2007). Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors. *Science Education*, *91*(6), 847–876. https://doi.org/10.1002/sce.20223

Jamieson, J. P., Black, A. E., Pelaia, L. E., & Reis, H. T. (2020). *The Impact of Mathematics Anxiety on Stress Appraisals, Neuroendocrine Responses, and Academic Performance in a Community College Sample*. https://doi.org/10.1037/edu0000636.supp

Jamieson, J. P., Hangen, E. J., Lee, H. Y., & Yeager, D. S. (2018). Capitalizing on appraisal processes to improve affective responses to social stress. *Emotion Review*, *10*(1), 30–39. https://doi.org/10.1177/1754073917693085

Jamieson, J. P., Peters, B. J., Greenwood, E. J., & Altose, A. J. (2016). Reappraising stress arousal improves performance and reduces evaluation anxiety in classroom exam situations. *Social Psychological and Personality Science*, *7*(6), 579–587. https://doi.org/10.1177/1948550616644656

Jenifer, J. B., Levine, S. C., & Beilock, S. L. (2023). Studying while anxious: mathematics anxiety and the avoidance of solving practice problems during exam preparation in college calculus. *ZDM – Mathematics Education*, *55*(2), 359–369. https://doi.org/10.1007/s11858-022-01456-1

John-Henderson, N. A., Rheinschmidt, M. L., & Mendoza-Denton, R. (2015). Cytokine responses and math performance: The role of stereotype threat and anxiety reappraisals. *Journal of Experimental Social Psychology*, *56*, 203–206. https://doi.org/10.1016/j.jesp.2014.10.002

Kabat-Zinn, J. (2003). Mindfulness-Based Interventions in Context: Past, Present, and Future. *Clinical Psychology: Science and Practice*, *10*(2), 144–156. https://doi.org/10.1093/CLIPSY.BPG016

Kornfield, J. (2008). *The wise heart: A guide to the universal teachings of Buddhist psychology.* Bantam Books.

Kuo, E. (2023). Two perspectives on physics problem solving and their relation to adaptive expertise. In M.F. Taşar & P. R. L. Heron (Eds.), *The International Handbook of Physics Education Research: Learning Physics* (pp. 1–26). AIP Publishing. https://doi.org/10.1063/9780735425477_010

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer Publishing Company.

Lindsay, E. K., Young, S., Smyth, J. M., Brown, K. W., & Creswell, J. D. (2018). Acceptance lowers stress reactivity: Dismantling mindfulness training in a randomized controlled trial. *Psychoneuroendocrinology*, *87*, 63–73. https://doi.org/10.1016/j.psyneuen.2017.09.015

Luberto, C. M., Hall, D. L., Park, E. R., Haramati, A., & Cotton, S. (2020). A perspective on the similarities and differences between mindfulness and relaxation. *Global Advances in Health and Medicine*, *9*. https://doi.org/10.1177/2164956120905597

Lüdecke, D. (2022). *Data visualization for statistics in social science [R package sjPlot version 2.8.12]*. https://CRAN.R-project.org/package=sjPlot

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marshman, E., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm? *Physical Review Physics Education Research*, *14*(2), 020123. https://doi.org/10.1103/PhysRevPhysEducRes.14.020123

Marshman, E., Kalender, Z. Y., Schunn, C., Nokes-Malach, T., & Singh, C. (2018). A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences. *Canadian Journal of Physics*, *96*(4), 391–405. https://doi.org/10.1139/cjp-2017-0185

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological Science*, *24*(5), 776–781. https://doi.org/10.1177/0956797612459659

Pelakh, A., Good, M. L., Kuo, E., Tumminia, M., Jamal-Orozco, N., Adelman, A., Antoan, J., Galla, B., & Nokes-Malach, T. J. (2025). *Thematic analysis of student perceptions of resources and demands experienced in introductory physics*. http://arxiv.org/abs/2502.14692

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Rojiani, R., Santoyo, J. F., Rahrig, H., Roth, H. D., & Britton, W. B. (2017). Women benefit more than men in response to college-based meditation training. *Frontiers in Psychology*, *8*(APR), 254657. https://doi.org/10.3389/FPSYG.2017.00551/BIBTEX

Rumi, J. (2004). *The essential Rumi* (C. Barks, Ed.; new expanded edition). Harper Collins.

Samuel, T. S., Buttet, S., & Warner, J. (2022). *"I Can Math, Too!": Reducing Math Anxiety in STEM-Related Courses Using a Combined Mindfulness and Growth Mindset Approach (MAGMA) in the Classroom*. https://doi.org/10.1080/10668926.2022.2050843

Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Seery, M. D. (2013). The biopsychosocial model of challenge and threat: Using the heart to measure the mind. *Social and Personality Psychology Compass*, *7*(9), 637–653. https://doi.org/10.1111/SPC3.12052

Singh, C., Maries, A., Heller, K., & Heller, P. (2023). Instructional strategies that foster effective problem-solving. In *The International Handbook of Physics Education Research: Learning Physics* (pp. 17-1-17–28). AIP Publishing LLC. https://doi.org/10.1063/9780735425477_017

Tingley, D., Yamamoto, H. T., Hirose, K., Keele, L., & Princeton, K. I. (2014). *mediation: R package for causal mediation analysis*. http://CRAN.R-project.org/package=mediation

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Article in Journal of Personality and Social Psychology*, *73*(1), 63–72. https://doi.org/10.1037/0022-3514.73.1.63

Whitcomb, K. M., Doucette, D., & Singh, C. (2022). Comparing Major Declaration, Attrition, Migration, and Completion in Physics With Other STEM Disciplines: A Sign of Inequitable Physics Culture? *Journal of Higher Education Theory and Practice*, *22*(17), 63–81. https://doi.org/10.33423/jhetp.v22i17.5658

Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/JOSS.01686