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

Supplementary Materials

Avital Pelakh Eric Kuo Melanie Good Michael Tumminia
Sara Jahanian Brian Galla Timothy Nokes-Malach

2025-10-15

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 reducing 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 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.

Table of contents

Preface

Format Notes

Due to formatting constraints, some portions of the web-based materials are not available in the PDF version (i.e., R code). Additionally, tables and images may appear out of order. We've done our best to provide reference links and documentation when this occurs. However, for the best experience, please refer to the web-based version of this Quarto book.

OSF Links

- Project Home
- Project Overview
- Materials
- Preregistration Plan

Part I

Supplementary Descriptive Text

1 Changes From the Preregistration

All changes below are in reference to **Preregistration 3 – Physics Task Outcomes** (https://doi.org/10.17605/OSF.IO/SA9W2).

1.1 Wording Adjustments

The preregistration refers to the physics tasks as "problem solving." We replaced this term with the phrase, "answering physics questions." While all of our tasks assessed processes involved in problem solving (i.e., problem categorization and conceptual understanding), the term "problem solving" is more commonly used among physics teachers in reference to straightforward quantitative problems. Therefore, we decided that "answering physics questions" was a more accurate description of our tasks.

1.2 Model Specifications

For RQ1, Hypotheses 1-3 in the main text: Instead of predicting outcomes at posttest while controlling for baseline (Hypotheses 4-6 in the preregistration), we used mixed-effects models that included *Timepoint* as a fixed effect, with baseline coded as 0 and posttest coded as 1. This allowed us to effectively test multiple hypotheses with fewer statistical models: baseline differences outlined in Preregistration, Aim 1, and effects of mindfulness training outlined in Preregistration, Aim 2. The advantage of this is that we were able to simultaneously reduce the likelihood of committing type 1 error (by running fewer individual tests) and lessen the burden on the reader. Both methods produced the same pattern of results (see Model 1 for each judgment type in Section 9.5 for results using linear models as described in the preregistration).

1.3 Gender Moderation

While we expected gender to be an important covariate, we did not preregister it as a moderator. Therefore, it is explicitly presented as exploratory in the main text. Results of the mixed-effects models without gender included as a moderator can be found here in Supplementary Table 8.2, Supplementary Table 8.3, and Supplementary Table 8.4. Results of the mediation

analyses without gender moderation can be found in Section 9.5. In both cases, the general pattern of results is the same. Including gender as a moderator helped bring our results into greater focus. For example, it revealed that overall effects of the intervention on judgments of confidence and anxiety were driven by women and non-binary students, while effects for judgements of difficulty were stable across genders.

1.4 Additional Control Variables

We added several additional control variables to the models that were not included in the preregistration. First, unlike what is specified in the preregistration, we included baseline psychological threat as a predictor in all the models, whereas it is only included in Aim 1: H1-H2 in the preregistration (testing baseline associations). Because our final models simultaneously tested the hypotheses in Aim 1 (baseline associations) and Aim 2 (effects of mindfulness on outcomes at posttest), we included it to control for overall effects of pre-intervention levels of psychological threat. We also included item-level accuracy as a covariate in the models to rule out potential confounding effects on perceptions. Including these variables had no meaningful impact on the results, as shown in Supplementary Table 8.2, Supplementary Table 8.3, and Supplementary Table 8.4.

1.5 Accuracy and Learning Outcomes

Finally, the analyses of accuracy performance and preparation for future learning outcomes described in the preregistration are not included in the main text. We did not find any effects of mindfulness training on these outcomes (preregistered hypotheses 1, 4, and 5) or any mediation effects (preregistered hypotheses 7 and 8). Therefore, we chose to remove the details of these analyses from the main text to narrow the scope of the paper. Those results are reported in these materials in Supplementary Table 10.1 and Supplementary Table 10.2.

2 Examples of Physics Task Items

2.1 Physics Tasks, Part 1: Quantitative Problem Solving

Students were provided with a general equation sheet for reference and asked to solve a single quantitative problem. They were instructed to use a blank sheet of paper to write out their work and final answer. The quantitative problems were developed for this study by Drs. Melanie Good and Eric Kuo (See Supplementary Figure 2.1).

2.2 Physics Tasks, Part 2: Problem Categorization

This part consisted of five items adapted from, and similar to, those used in Hardiman, Dufresne, and Mestre (1989). For each item, a model problem was presented along with two alternative problems. Students were instructed to select the alternative which is solved most like the model problem (deep structure match). Four out of five items contained surface feature distractors, meaning that there were surface features in the incorrect alternative which resembled the model problem (See Supplementary Figure 2.2).

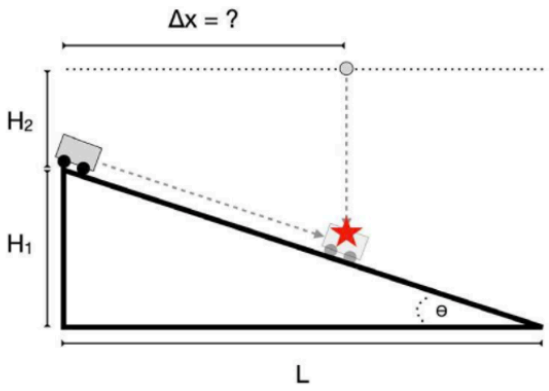
> **Note**
>
> Supplementary Figure 2.2: At first glance, the model problem and alternative 2 appear similar because they both involve projectiles, but this similarity is a surface-level distractor. The model problem and alternative 1 are asking for final velocity and can be solved using conservation of energy, while alternative 2 requires the use of kinematics to solve for time. Therefore, the correct response is alternative 1.

2.3 Physics Tasks, Part 3: Conceptual Questions (Qualitative Problem Solving)

This part consisted of three multiple-choice questions (See Supplementary Figure 2.3), followed by one open-ended problem (See Supplementary Figure 2.4). For each multiple-choice question, participants were also asked to provide a brief explanation for their choice. The open-ended problem was a word problem in which participants were asked to explain a solution.

Supplementary Figure 2.1: Sample Quantitative Physics Task Item

1. A cart starts at the top of a ramp of height H_1, length L, and angle θ. A ball is also held in place somewhere over the ramp, at a height H_2 above where the cart starts. At the same time, the cart and ball are released so that the cart rolls down the ramp and the ball falls straight down. If the ball lands right on top of the cart, what is the initial horizontal distance between the cart and the ball, Δx, in terms of the other given variables?

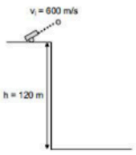


Please record your work on a separate sheet of blank paper *with your name on it.* When you have finished, or when the 10 minutes have elapsed, please circle your final answer. After you have finished, please click the "next" button to upload your work on the following page.

Supplementary Figure 2.2: Sample Categorization Physics Task Item

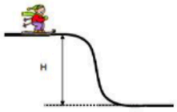
4. Model

A 4 kg shell is fired with an initial speed of 600 m/s from a cliff 120 m above level ground. What is its speed when it hits the ground?



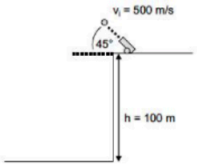
Alternative 1

A skier of mass *m* sliding along a horizontal region of snow at speed *v* suddenly slides down a slope to a lower horizontal region. If the lower horizontal region is a height *H* lower than the top horizontal region, what is the speed of the skier at the bottom? Assume there is negligible friction between the snow and the skis.



Alternative 2

A 2 kg projectile is fired with an initial velocity of 500 m/s at an angle of 45 degrees above the horizontal and height 100 m above level ground. Find the time needed for the projectile to reach the ground.

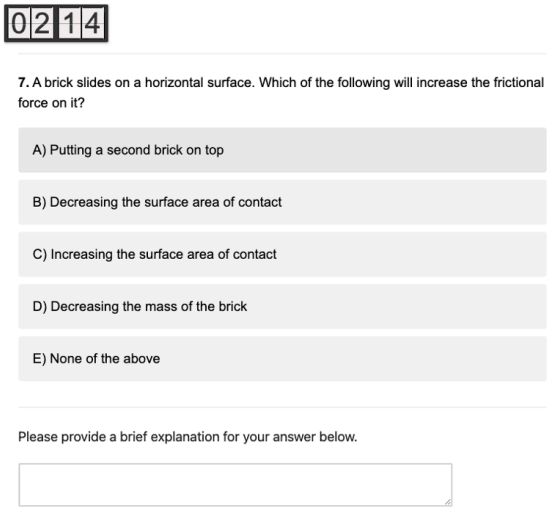


Please choose the alternative that is solved most like the model problem

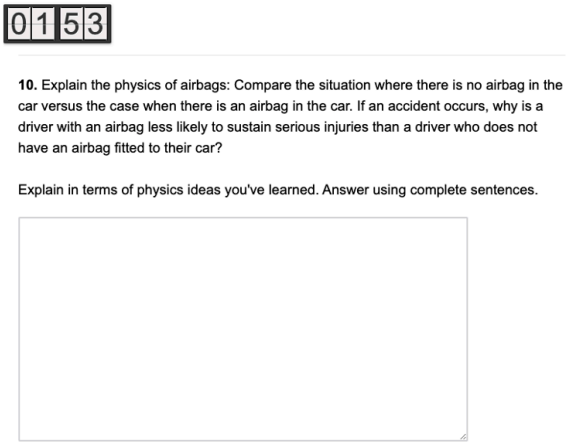
Alternative 1

Alternative 2

Supplementary Figure 2.3: Multiple-Choice Item



Supplementary Figure 2.4: Open-Ended Item



Note

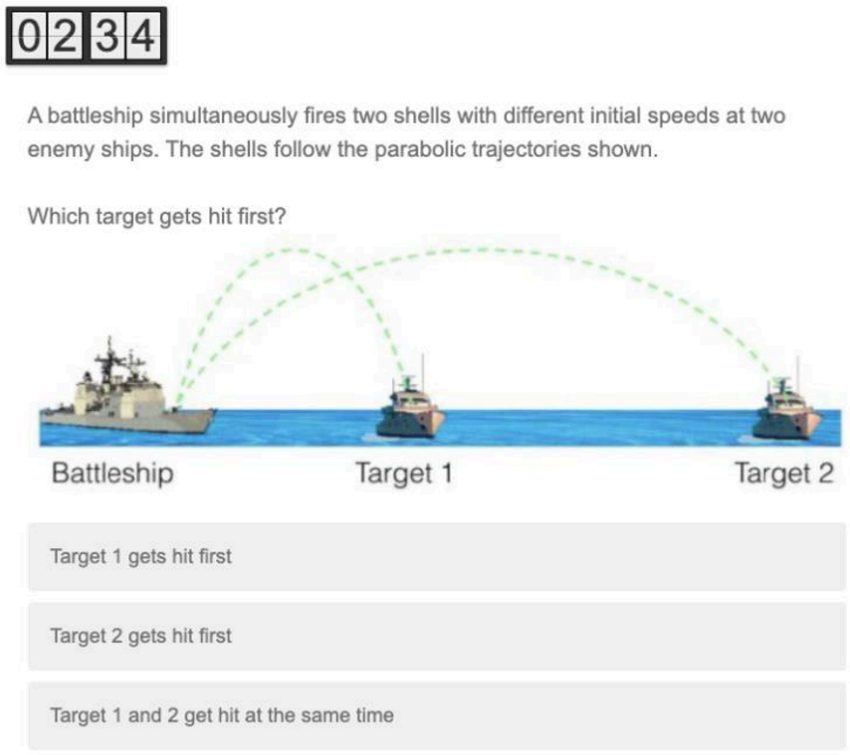
Supplementary Figure 2.3: Qualitative multiple-choice item: The correct answer is option A because the frictional force is proportional to the normal force, which is equal to the weight force. Therefore, the only way to increase the frictional force is to increase the mass; Supplementary Figure 2.4: Qualitative open response item: Newton's second law states that $F_{net} = \frac{\Delta p}{\Delta t}$. Airbags reduce the force (outcome) of a collision by increasing the time it takes for the collision to occur (mechanism). Half credit was given to responses that mentioned either force or time correctly, and full credit was given to responses that mentioned both force and time.

2.4 Physics Tasks, Part 4: Preparation for Future Learning

The Preparation for Future Learning (PFL) task (Belenky and Nokes-Malach 2012; Schwartz and Martin 2004) is a 3-part learning activity that participants completed only at posttest. Physics problems used in the learning activity were taken from Weinlader et al. (2019). The first part consisted of solving a difficult multiple-choice problem which required comparing the trajectories of two projectiles and predicting which would hit their target first (see Supplementary Figure 2.5). In the second part, a learning resource explaining the first problem was provided. The resource contained an explanation of why the correct answer choice was valid, and why each incorrect answer was not. The third part consisted of a novel problem with

similar surface features to the first problem but had a different correct response. Answering correctly required applying the underlying principles of the first problem to a different set of conditions. Each of the two problems required both a multiple-choice answer as well as a short-answer explanation for the selected choice.

Supplementary Figure 2.5: Sample Preparation for Future Learning Task Item



3 Scoring Procedure For Physics Learning and Performance Outcomes

For all measures which required qualitative coding or scoring, the following procedure was used: First, a rubric was developed through group discussion with the research team. Then, a portion of the responses were coded independently by a minimum of two coders. Discrepancies from the first round of coding were reviewed by the team and the rubric was revised as needed. The remaining responses were coded again in full by two coders and all discrepancies were resolved in the presence of a third coder. Any responses that could not be easily resolved were brought to the research team for review. For each measure that was coded or scored, we report two inter-rater reliability statistics, one for each round of coding.

3.1 Physics Assessment, Part 1: Quantitative Problem Solving

The quantitative problems (one for each test version) were scored according to a rubric developed by the research team. Partial credit was given for incomplete or partially correct solutions. Responses from baseline and posttest were combined and randomized within each test version before scoring, and team members were blinded to condition. The intraclass correlation for the first round of coding was .74 for version A (20 responses) and .8 for version B (29 responses). The intraclass correlation for the remaining solutions were .7 for version A (129 responses), and .94 for version B (117 responses). Final scores were calculated by taking the proportion of points earned out of the total possible points. Three responses on the posttest quantitative problem-solving task were removed from analysis. In one case, it was clear there was additional work cropped out of the uploaded image which could not be scored; in another, the uploaded file was a duplicate of their baseline file; and in the last case the handwriting was deemed illegible by all of the raters.

3.2 Physics Assessment, Part 2: Problem Categorization

Individual items were scored dichotomously as 0 (incorrect) or 1 (correct). Final scores were calculated by taking the average accuracy across the five categorization items.

3.3 Physics Assessment, Part 3: Qualitative Problem Solving

The three multiple-choice questions were scored dichotomously as 0 (incorrect) or 1 (correct). For the purposes of the current work and questions we did not analyze the open-response explanations for the multiple-choice questions. The open-ended word problem was coded and scored by the research team. Responses were awarded a maximum of two points: one point for each component of the correct explanation. For example, the question in test version B described a scenario in which an elevator cable snaps, and the emergency friction brakes are engaged. The student was asked to describe the types of energy transfer which occur in this scenario. One point was given for describing gravitational potential energy being converted to kinetic energy, and another for describing kinetic energy being converted to thermal and sound energy due to work done by friction. Similar to the quantitative problem, all responses were combined and randomized across timepoints, and experimental condition was removed from the data before scoring. The weighted Kappa for the first round of coding was .36 for version A (50 responses), and .58 for version B (53 responses). The weighted Kappas for the second round of coding were .71 for version A (99 responses), and .74 for version B (95 responses). When calculating the final score, the open-ended word problem was weighted equal to the multiple-choice, such that two-point responses were scored as a 1 and one-point responses were scored as a .5. Final scores for qualitative problem solving were calculated by taking the mean score of all the items.

3.4 Physics Assessment, Part 4: Preparation for Future Learning (PFL)

There were two components to the response for each of the two PFL questions: multiple-choice selection and open-response explanation. We defined correctness on the PFL as selecting the correct multiple choice response option and reasoning correctly in the open-response explanation. For both questions, student reasoning was considered correct if they mentioned the relative maximum heights or initial y velocities of the two trajectories as a justification for their answer, or if they mentioned that the y component was most important for determining time in the air. These questions were the only ones for which we did not complete two separate rounds of coding because we build directly on our past work with a similar population (Weinlader et al. 2019). Otherwise, the coding procedure was identical to the others. Responses were randomized and coders were blinded to condition. Both PFL questions were coded in full by two coders as either correct or incorrect. The unweighted Kappa for both of the PFL questions was .88. All discrepancies were resolved through discussion in the presence of a third coder.

Supplementary Table 3.1: Summary of Item Types and Scoring Scales by Outcome

Outcome Measure	Items and Measurement	Scoring
Physics Problem Solving Performance		
Part 1: Quantitative Problem Solving	1 item, open response	Points earned/points possible, continuous from 0 to 1
Part 2: Problem Categorization	5 items, forced-choice	Mean score, continuous from 0 to 1
Part 3: Qualitative Problem Solving	3 multiple-choice, 1 open response	Mean score, continuous from 0 to 1
Preparation for Future Learning		
Part 4: Preparation for Future Learning	2 items, multiple-choice with open response explanation	Dichotomous, 0 or 1: 1 = Correct multiple choice response & attends to y component correctly in explanation
Momentary Item-Level Perceptions		
Confidence	1 item, measured repeatedly	Continuous from 1 to 6
Anxiety	1 item, measured repeatedly	Continuous from 1 to 6
Difficulty	1 item, measured repeatedly	Continuous from 1 to 6

4 Clustering Variables Examined but Not Included

We explored the particular class section that the participants were recruited from as a potential covariate. There were many aspects of the classes that were homogeneous: all instructors were required to adhere to the same set of pre-defined learning objectives, and beginning in cohort 3's semester (during which over half our sample participated), the department instituted measures which required classes to synchronize exam schedules, recitation quizzes, homework systems, and textbooks. Nevertheless, each class has its own variation in terms of the number of students enrolled, days and times they meet, the amount of synchronous vs. asynchronous activities, and the students that select to be in those classes. Furthermore, all instructors have idiosyncratic aspects to their teaching methods and can have different demands and resources in the class. The inclusion of class by instructor as a covariate was explored, but ultimately not included. Seven instructors taught the physics classes represented in our sample. Most instructors taught one class except one instructor who taught two, and another who taught 4 (total of eleven classes).

The number of students associated with each class ranged from 5 to 21 ($M = 13.5$, $SD = 5.01$). Theoretically, it made sense to include class and instructor as nested random intercept terms because we wanted to account for clustering by class and instructor, but we did not have any predictions about specific classes or professors. However, the ICC for class and instructor was at or very close to zero in all the models. This indicates that statistically, observations within classes and instructors were no more similar to each other than to observations from different classes and instructors. We also conducted a visual inspection of the all the focal study variables by instructor and did not detect any differences that appeared systematic. Based on these analyses, we did not include class or instructor in the reported models.

Part II

R Code and Analyses

5 Data and Environment Setup

5.1 Source Setup Script

In order to run the code included in this book, you first need to run "/R/setup-script.R" to install/load required packages and read the data. The data are located in "/data/preregistration_3_data_public.csv".

5.2 Session Info

```
R version 4.5.0 (2025-04-11)
Platform: aarch64-apple-darwin20
Running under: macOS Sonoma 14.6.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib;  L

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: America/New_York
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] lubridate_1.9.4    forcats_1.0.0      stringr_1.5.1      dplyr_1.1.4
 [5] purrr_1.0.4        readr_2.1.5        tidyr_1.3.1        tibble_3.2.1
 [9] tidyverse_2.0.0    ggpubr_0.6.0       rstatix_0.7.2      showtext_0.9-7
[13] showtextdb_3.0     sysfonts_0.8.9     patchwork_1.3.0    GGally_2.2.1
[17] ggplot2_3.5.2      sjtable2df_0.0.4   sjlabelled_1.2.0   sjmisc_2.8.10
[21] sjPlot_2.8.17      lmerTest_3.1-3     lme4_1.1-37        mediation_4.5.0
[25] sandwich_3.1-1     mvtnorm_1.3-3      Matrix_1.7-3       MASS_7.3-65
```

```
[29] gvlma_1.0.0.3      performance_0.15.1 kableExtra_1.4.0   pacman_0.5.1

loaded via a namespace (and not attached):
 [1] Rdpack_2.6.4        gridExtra_2.3       rlang_1.1.6
 [4] magrittr_2.0.3      compiler_4.5.0      systemfonts_1.2.2
 [7] vctrs_0.6.5         crayon_1.5.3        pkgconfig_2.0.3
[10] fastmap_1.2.0       backports_1.5.0     rmarkdown_2.29
[13] tzdb_0.5.0          nloptr_2.2.1        bit_4.6.0
[16] xfun_0.52           jsonlite_2.0.0      ggeffects_2.2.1
[19] parallel_4.5.0      broom_1.0.8         cluster_2.1.8.1
[22] R6_2.6.1            stringi_1.8.7       RColorBrewer_1.1-3
[25] car_3.1-3           boot_1.3-31         rpart_4.1.24
[28] numDeriv_2016.8-1.1 Rcpp_1.0.14         knitr_1.50
[31] zoo_1.8-14          base64enc_0.1-3     splines_4.5.0
[34] nnet_7.3-20         timechange_0.3.0    tidyselect_1.2.1
[37] rstudioapi_0.17.1   abind_1.4-8         yaml_2.3.10
[40] codetools_0.2-20    curl_6.2.2          lattice_0.22-7
[43] plyr_1.8.9          withr_3.0.2         evaluate_1.0.3
[46] foreign_0.8-90      archive_1.1.12      ggstats_0.9.0
[49] xml2_1.3.8          lpSolve_5.6.23      pillar_1.10.2
[52] carData_3.0-5       checkmate_2.3.2     reformulas_0.4.0
[55] insight_1.4.2       generics_0.1.3      vroom_1.6.5
[58] hms_1.1.3           scales_1.4.0        minqa_1.2.8
[61] glue_1.8.0          Hmisc_5.2-3         tools_4.5.0
[64] data.table_1.17.8   ggsignif_0.6.4      grid_4.5.0
[67] rbibutils_2.3       datawizard_1.2.0    colorspace_2.1-1
[70] nlme_3.1-168        htmlTable_2.4.3     Formula_1.2-5
[73] cli_3.6.5           viridisLite_0.4.2   svglite_2.1.3
[76] sjstats_0.19.0      gtable_0.3.6        digest_0.6.37
[79] htmlwidgets_1.6.4   farver_2.1.2        htmltools_0.5.8.1
[82] lifecycle_1.0.4     bit64_4.6.0-1
```

5.3 Datasets

The analyses in the main text are run using two forms of the data, one for each research question / analysis type. Summaries of the data charactaristics are described below.

5.3.1 RQ1 Data

These data are in long format - (22 items x 149 participants - 3 NA values) x 3 judgments per item = 9825 observations.

> **Note**
>
> the variable `perception` is equivalent to "judgment" in the manuscript.

```
Rows: 9,825
Columns: 17
$ Participant           <chr> "mvU3yT4uTFpW58Z0", "mvU3yT4uTFpW58Z0", "mvU3yT4~
$ Condition             <fct> Control, Control, Control, Control, Control, Con~
$ Gender                <fct> Men, Men, Men, Men, Men, Men, Men, Men, Men, Men~
$ Cohort                <fct> Cohort 1, Cohort 1, Cohort 1, Cohort 1, Cohort 1~
$ Timepoint             <fct> Baseline, Baseline, Baseline, Baseline, Baseline~
$ Semester_Week         <dbl> 1.912752, 1.912752, 1.912752, 1.912752, 1.912752~
$ Test_Version          <fct> B, B, B, B, B, B, B, B, B, B, B, B, B, B, B, B, ~
$ Part                  <chr> "Quantitative", "Quantitative", "Quantitative", ~
$ Question              <dbl> 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, ~
$ Item                  <chr> "B01", "B01", "B01", "B02", "B02", "B02", "B03",~
$ Score                 <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, ~
$ `Item-Level Accuracy` <fct> Incorrect, Incorrect, Incorrect, Incorrect, Inco~
$ Accuracy_Raw          <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, ~
$ Baseline_Threat       <dbl> 1.63132, 1.63132, 1.63132, 1.63132, 1.63132, 1.6~
$ EMA_Threat            <dbl> 3.416667, 3.416667, 3.416667, 3.416667, 3.416667~
$ perception            <fct> Confidence, Anxiety, Difficulty, Confidence, Anx~
$ rating                <dbl> 1, 6, 6, 5, 6, 2, 4, 6, 3, 5, 6, 4, 5, 6, 4, 4, ~
```

Participant-Level Variables

```
 Participant             Condition                  Gender        Cohort
 Length:149         Control    :73   Men                :66   Cohort 1:61
 Class :character   Mindfulness:76   Women or Non-binary:83   Cohort 2:16
 Mode  :character                                            Cohort 3:72
```

variable	n	min	max	mean	sd
Semester_Week	149	-4.087	4.913	0	2.790
Baseline_Threat	149	-3.102	2.865	0	1.297

Item (Observation)-Level Variables

> **Note**
>
> `Score` includes values between 0 and 1 since 4 of the items (A01, B01, A10, and B10) were open-ended and could be awarded partial credit. `Item-Level Accuracy` (factor version) and `Accuracy_Raw` (numeric version) were derived from `Score` to compare incorrect responses to those that were at least partially correct (`Score` > 0).

```
    Item             Timepoint    Test_Version Item-Level Accuracy
 Length:9825        Baseline:4470   A:4908       Incorrect:5547
 Class :character   Posttest:5355   B:4917       Correct  :4278
 Mode  :character
     perception
 Confidence:3275
 Anxiety   :3275
 Difficulty:3275
```

variable	n	min	max	mean	sd
Score	9825	0	1	0.404	0.478
Accuracy_Raw	9825	0	1	0.435	0.496
rating	9825	1	6	3.474	1.408

5.3.2 RQ2 Data

Data are in wide format, so everything varies at the participant level. Judgment ratings and accuracy scores are averaged at baseline and posttest. Continuous variables are standardized and numeric dummy variables with contrast coding are created for `Cohort` because the `mediation` package does not accept factors with contrast coding, but this has no bearing on the results. One participant's data was removed because they did not complete any EMA surveys.

```
Rows: 148
Columns: 21
$ Participant           <chr> "mvU3yT4uTFpW58ZO", "bTvjXPvFUSNyAt5t", "hx3UKKU~
$ Condition             <fct> Control, Control, Control, Control, Control, Con~
$ Gender                <fct> Men, Men, Men, Women or Non-binary, Men, Women o~
$ Cohort                <fct> Cohort 1, Cohort 1, Cohort 1, Cohort 1, Cohort 3~
$ Semester_Week         <dbl> 0.6786886, 0.6786886, 0.6786886, 1.0374240, -0.7~
```

```
$ Baseline_Threat       <dbl> 1.25100904, -0.75116043, -1.05918650, 1.89273003~
$ EMA_Threat            <dbl> 1.55096412, 1.29550795, -0.95250627, 1.29550795,~
$ Baseline_n_items      <int> 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, ~
$ Posttest_n_items      <int> 12, 12, 12, 12, 12, 12, 11, 12, 12, 12, 12, 12, ~
$ Baseline_Score        <dbl> -0.31125932, 0.33654914, -0.49120612, -1.6068762~
$ Posttest_Score        <dbl> -1.34709242, 1.43972775, -0.78972839, -2.4618204~
$ Baseline_Confidence   <dbl> -0.37485943, 0.01220473, 0.39926888, -2.18115882~
$ Posttest_Confidence   <dbl> 0.23891063, 0.53917812, 0.53917812, -3.36429935,~
$ Baseline_Anxiety      <dbl> 2.49202355, -0.85159304, -0.57295832, 1.65611940~
$ Posttest_Anxiety      <dbl> 2.15254688, -0.25189270, -0.99809809, 2.40128201~
$ Baseline_Difficulty   <dbl> 1.15470752, -1.16207408, -0.20810518, 2.24495769~
$ Posttest_Difficulty   <dbl> 1.4257825, -0.7405502, -0.0184393, 2.4573694, 0.~
$ Baseline_Test_Version <chr> "B", "A", "B", "A", "B", "A", "A", "B", "A", "A"~
$ Posttest_Test_Version <fct> A, B, A, B, A, B, B, A, B, B, A, B, B, B, B, B, ~
$ Cohort_2              <dbl> -0.3469771, -0.3469771, -0.3469771, -0.3469771, ~
$ Cohort_3              <dbl> -0.9569993, -0.9569993, -0.9569993, -0.9569993, ~
```

```
 Participant              Condition                  Gender        Cohort
 Length:148         Control    :73   Men                :66   Cohort 1:61
 Class :character   Mindfulness:75   Women or Non-binary:82   Cohort 2:16
 Mode  :character                                             Cohort 3:71
 Baseline_Test_Version Posttest_Test_Version
 A:75                  A:73
 B:73                  B:75
```

variable	n	min	max	mean	sd
Semester_Week	148	-1.474	1.755	0.00	1.000
Baseline_Threat	148	-2.394	2.201	0.00	1.000
EMA_Threat	148	-2.230	2.420	0.00	1.000
Baseline_n_items	148	10.000	10.000	10.00	0.000
Posttest_n_items	148	11.000	12.000	11.98	0.141
Baseline_Score	148	-2.111	3.511	0.00	1.000
Posttest_Score	148	-2.462	2.957	0.00	1.000
Baseline_Confidence	148	-3.213	2.206	0.00	1.000
Posttest_Confidence	148	-3.464	1.840	0.00	1.000
Baseline_Anxiety	148	-2.152	2.492	0.00	1.000
Posttest_Anxiety	148	-1.993	2.650	0.00	1.000
Baseline_Difficulty	148	-2.389	2.381	0.00	1.000
Posttest_Difficulty	148	-2.701	2.457	0.00	1.000
Cohort_2	148	-0.347	2.863	0.00	1.000

variable	n	min	max	mean	sd
Cohort_3	148	-0.957	1.038	0.00	1.000

6 Characteristics of Fixed Effects Variables for Mixed Models

The following code reproduces the factor coding and descriptive statistics reported in Table 3 of the main text.

6.1 Dependent Variables

Confidence

Timepoint	variable	n	min	max	median	iqr	mean	sd	se	ci
Baseline	rating	1490	1	6	4	2	3.69	1.44	0.037	0.073
Posttest	rating	1785	1	6	4	2	3.88	1.46	0.035	0.068

Anxiety

Timepoint	variable	n	min	max	median	iqr	mean	sd	se	ci
Baseline	rating	1490	1	6	3	2	3.31	1.41	0.036	0.071
Posttest	rating	1785	1	6	3	2	3.00	1.37	0.033	0.064

Difficulty

Timepoint	variable	n	min	max	median	iqr	mean	sd	se	ci
Baseline	rating	1490	1	6	4	1	3.55	1.26	0.033	0.064
Posttest	rating	1785	1	6	4	2	3.43	1.30	0.031	0.060

6.2 Covariates

6.2.1 Cohort

Factor Levels	Cohort 2	Cohort 3
Cohort 1	-0.11	-0.48
Cohort 2	0.89	-0.48
Cohort 3	-0.11	0.52

6.2.2 Semester Week

variable	n	min	max	median	iqr	mean	sd	se	ci
Semester_Week	149	-4.087	4.913	-1.087	5	0	2.79	0.229	0.452

6.2.3 Test Version

Factor Levels	B
A	-0.50
B	0.50

6.2.4 Item-Level Accuracy

Factor Coding:

Factor Levels	Correct
Incorrect	-0.44
Correct	0.56

Baseline and Posttest Means and Standard Deviations:

Timepoint	variable	n	min	max	median	iqr	mean	sd	se	ci
Baseline	Accuracy_Raw	1490	0	1	0	1	0.43	0.50	0.013	0.025
Posttest	Accuracy_Raw	1785	0	1	0	1	0.44	0.50	0.012	0.023

6.2.5 Baseline Threat

variable	n	min	max	median	iqr	mean	sd	se	ci
Baseline_Threat	149	-3.102	2.865	-0.035	1.933	0	1.297	0.106	0.21

6.3 Main Independent Variables of Interest

6.3.1 Timepoint

Factor Levels	Posttest
Baseline	0
Posttest	1

6.3.2 Condition

Factor Levels	Mindfulness
Control	-0.5
Mindfulness	0.5

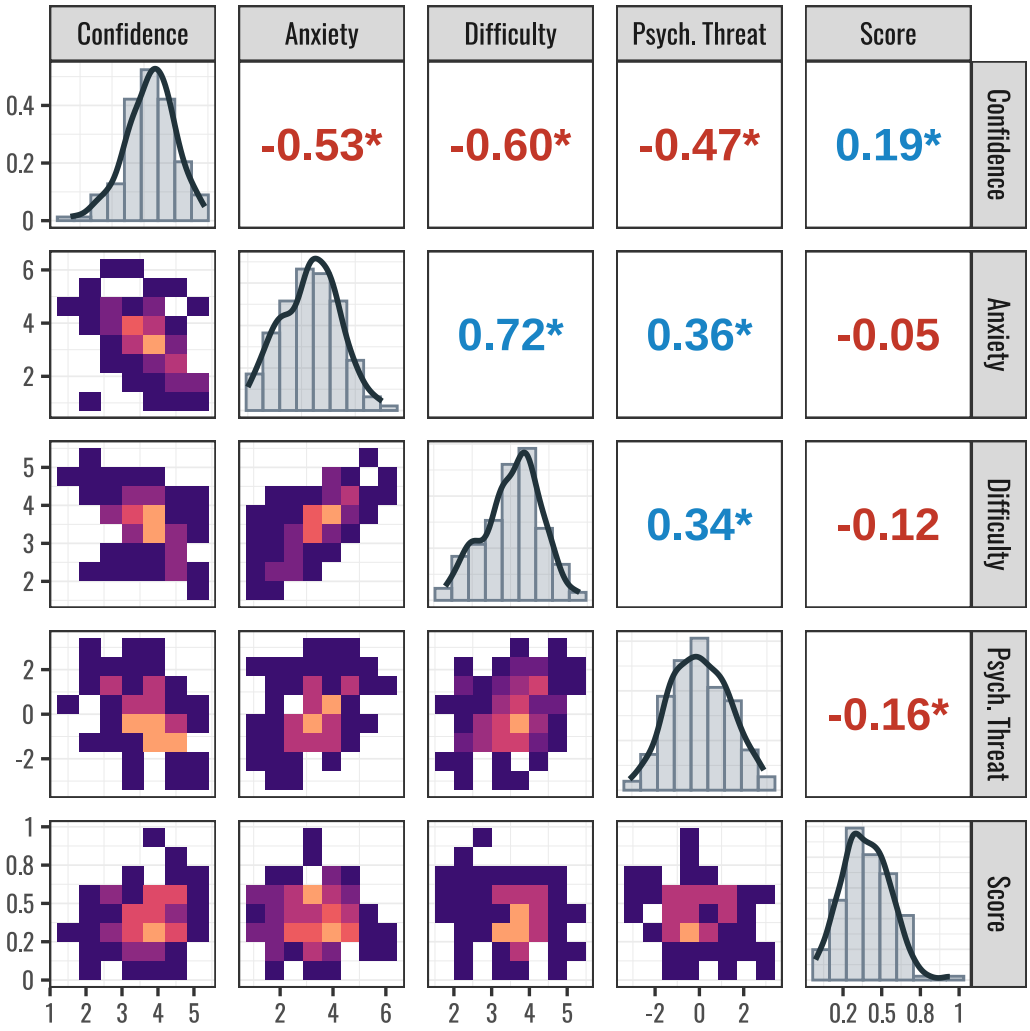
6.3.3 Gender

Factor Levels	Women or Non-binary
Men	-0.56
Women or Non-binary	0.44

7 Correlations of Variables at Baseline

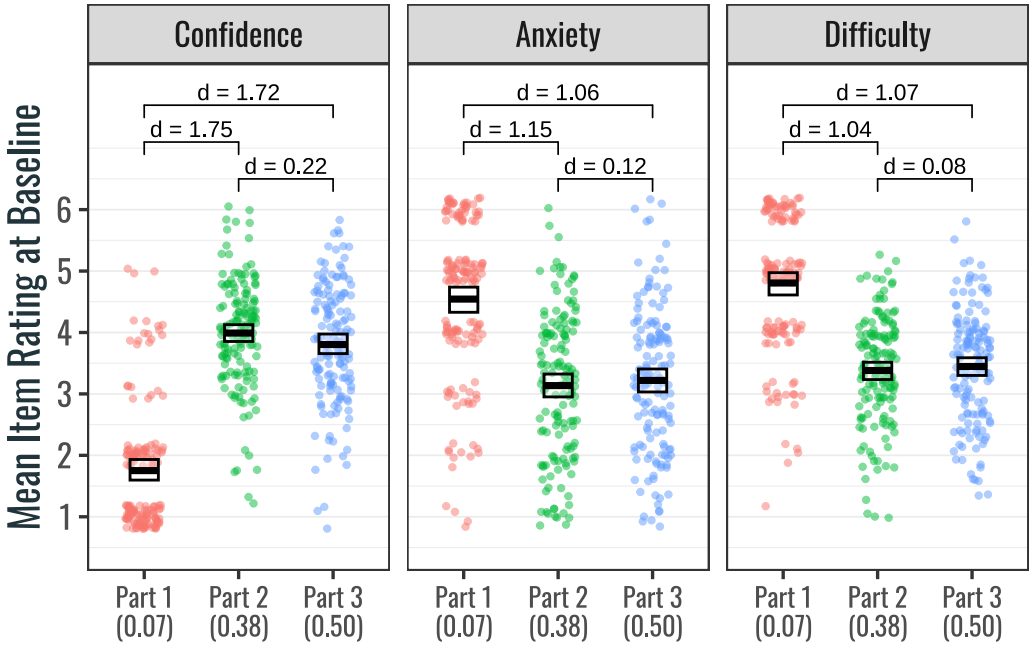
7.1 Reproduction of Figure 3

Supplementary Figure 7.1: Correlations and Variable Distributions at Baseline (Main Text, Figure 3)



7.2 Association Between Item-Level Judgments by Task

Supplementary Figure 7.2: Mean Judgments by Physics Task Type at Baseline



Note

Supplementary Figure 7.2: Cohen's d values for paired samples were obtained by calculating a mean perception rating for each participant by task and perception type and then dividing the mean difference by the standard deviation of the difference for each comparison. Crossbars show means and bootstrapped 95% confidence intervals. Mean accuracy for each task is shown in parentheses along the x axis.

Perception ratings were sensitive to fluctuations in performance between the physics tasks, indicating construct validity (see Supplementary Figure 7.2). For example, there was a large effect size difference in perceptions of confidence, anxiety, and difficulty between ratings on the quantitative problem solving item compared to mean ratings on both the problem categorization items and the qualitative problem solving items (Cohen's d = 1.04 - 1.75). These differences make sense given that students performed near floor on the quantitative problem, and considerably better on the other two tasks. There was a small effect size difference between confidence ratings on the problem categorization items compared to the qualitative items (Cohen's $d = 0.22$), with greater confidence reported on the problem categorization items. A likely explanation is that some of the categorization items were designed to appear simpler than they are in reality, so students may have been overconfident on those items. The difference between anxiety and difficulty ratings on the problem categorization items compared to the qualitative

items was marginal (Cohen's $d = 0.12, 0.08$). Overall, students' perceptions varied with mean accuracy on the different types of items.

8 Research Question 1

8.1 Model Specification

Code for model specification not available in PDF format.

8.2 Reproduction of Table 4

See Supplementary Table 8.1

8.3 Confidence Judgments: Model Comparison

See Supplementary Table 8.2

8.4 Anxiety Judgments: Model Comparison

See Supplementary Table 8.3

8.5 Difficulty Judgments: Model Comparison

See Supplementary Table 8.4

8.6 Reproduction of Figure 4

See Supplementary Figure 8.1

Supplementary Table 8.1: 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		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	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 [Correct]	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		

Supplementary Table 8.2: Comparison of Models Predicting Confidence Judgments

	Accuracy and Baseline Threat Removed			2-Way Interaction			3-Way Interaction		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	3.69	0.16	**<0.001**	3.69	0.15	**<0.001**	3.69	0.15	**<0.001**
Cohort [Cohort 2]	0.28	0.34	0.409	0.38	0.31	0.221	0.38	0.31	0.222
Cohort [Cohort 3]	0.01	0.30	0.977	0.07	0.27	0.786	0.07	0.27	0.785
Semester Week	0.09	0.05	0.082	0.10	0.05	**0.035**	0.10	0.05	**0.036**
Test Version [B]	-0.14	0.12	0.235	-0.14	0.12	0.253	-0.13	0.12	0.268
Gender [Women or Non-binary]	-0.51	0.12	**<0.001**	-0.32	0.11	**0.003**	-0.35	0.12	**0.003**
Timepoint [Posttest]	0.19	0.04	**<0.001**	0.19	0.04	**<0.001**	0.19	0.04	**<0.001**
Condition [Mindfulness]	0.10	0.12	0.394	0.07	0.11	0.543	0.07	0.11	0.545
Timepoint [Posttest] × Condition [Mindfulness]	0.13	0.08	0.086	0.13	0.08	0.090	0.13	0.08	0.089
Item-Level Accuracy [Correct]				0.12	0.04	**0.005**	0.13	0.04	**0.005**
Baseline Threat				-0.24	0.04	**<0.001**	-0.24	0.04	**<0.001**
Timepoint [Posttest] × Gender [Women or Non-binary]							0.06	0.08	0.451
Condition [Mindfulness] × Gender [Women or Non-binary]							-0.19	0.22	0.391
(Timepoint [Posttest] × Condition [Mindfulness]) × Gender [Women or Non-binary]							0.32	0.15	**0.034**
Random Effects									
σ^2	1.17			1.17			1.17		
τ_{00}	$0.42_{Participant}$			$0.33_{Participant}$			$0.33_{Participant}$		
	0.46_{Item}			0.45_{Item}			0.45_{Item}		
ICC	0.43			0.40			0.40		
N	$149_{Participant}$			$149_{Participant}$			$149_{Participant}$		
	22_{Item}			22_{Item}			22_{Item}		
Observations	3275			3275			3275		
Marginal R^2 / Conditional R^2	0.057 / 0.463			0.099 / 0.460			0.099 / 0.461		

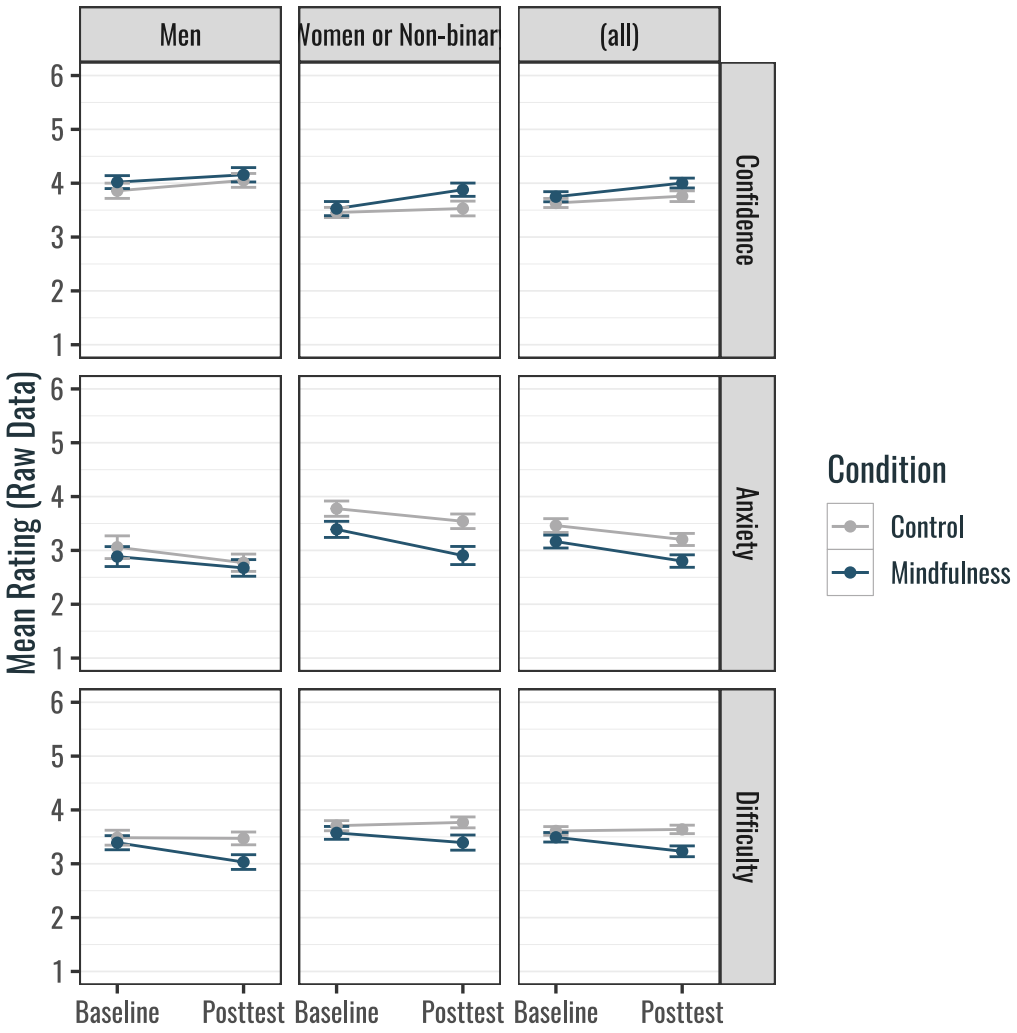
Supplementary Table 8.3: Comparison of Models Predicting Anxiety Judgments

	Accuracy and Baseline Threat Removed			2-Way Interaction			3-Way Interaction		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	3.33	0.12	**<0.001**	3.33	0.12	**<0.001**	3.33	0.12	**<0.001**
Cohort [Cohort 2]	-0.08	0.45	0.867	-0.16	0.43	0.703	-0.16	0.43	0.709
Cohort [Cohort 3]	0.12	0.40	0.764	0.06	0.38	0.879	0.06	0.38	0.868
Semester Week	-0.03	0.07	0.639	-0.04	0.07	0.538	-0.04	0.07	0.537
Test Version [B]	-0.08	0.10	0.408	-0.08	0.10	0.399	-0.08	0.10	0.385
Gender [Women or Non-binary]	0.60	0.16	**<0.001**	0.42	0.15	**0.006**	0.49	0.16	**0.002**
Timepoint [Posttest]	-0.34	0.03	**<0.001**	-0.34	0.03	**<0.001**	-0.34	0.03	**<0.001**
Condition [Mindfulness]	-0.29	0.15	0.063	-0.25	0.15	0.087	-0.25	0.15	0.087
Timepoint [Posttest] × Condition [Mindfulness]	-0.11	0.07	0.102	-0.11	0.07	0.105	-0.11	0.07	0.102
Item-Level Accuracy [Correct]				-0.08	0.04	**0.028**	-0.08	0.04	**0.026**
Baseline Threat				0.23	0.06	**<0.001**	0.22	0.06	**<0.001**
Timepoint [Posttest] × Gender [Women or Non-binary]							-0.12	0.07	0.071
Condition [Mindfulness] × Gender [Women or Non-binary]							-0.11	0.30	0.721
(Timepoint [Posttest] × Condition [Mindfulness]) × Gender [Women or Non-binary]							-0.31	0.13	**0.018**
Random Effects									
σ^2	0.87			0.87			0.87		
τ_{00}	$0.80_{Participant}$			$0.72_{Participant}$			$0.72_{Participant}$		
	0.18_{Item}			0.17_{Item}			0.17_{Item}		
ICC	0.53			0.51			0.51		
N	$149_{Participant}$			$149_{Participant}$			$149_{Participant}$		
	22_{Item}			22_{Item}			22_{Item}		
Observations	3275			3275			3275		
Marginal R^2 / Conditional R^2	0.078 / 0.565			0.120 / 0.565			0.123 / 0.567		

Supplementary Table 8.4: Comparison of Models Predicting Difficulty Judgments

	Accuracy and Baseline Threat Removed			2-Way Interaction			3-Way Interaction		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	3.56	0.13	**<0.001**	3.56	0.12	**<0.001**	3.56	0.12	**<0.001**
Cohort [Cohort 2]	-0.46	0.33	0.166	-0.52	0.32	0.106	-0.52	0.32	0.107
Cohort [Cohort 3]	-0.29	0.29	0.315	-0.33	0.28	0.235	-0.33	0.28	0.234
Semester Week	-0.11	0.05	**0.032**	-0.12	0.05	**0.019**	-0.12	0.05	**0.019**
Test Version [B]	-0.07	0.10	0.528	-0.07	0.10	0.518	-0.07	0.10	0.507
Gender [Women or Non-binary]	0.34	0.11	**0.003**	0.22	0.11	**0.048**	0.16	0.12	0.196
Timepoint [Posttest]	-0.12	0.04	**<0.001**	-0.12	0.04	**<0.001**	-0.12	0.04	**<0.001**
Condition [Mindfulness]	-0.10	0.12	0.376	-0.08	0.11	0.479	-0.08	0.11	0.478
Timepoint [Posttest] × Condition [Mindfulness]	-0.29	0.07	**<0.001**	-0.29	0.07	**<0.001**	-0.29	0.07	**<0.001**
Item-Level Accuracy [Correct]				-0.03	0.04	0.426	-0.03	0.04	0.436
Baseline Threat				0.15	0.04	**<0.001**	0.15	0.04	**<0.001**
Timepoint [Posttest] × Gender [Women or Non-binary]							0.12	0.07	0.076
Condition [Mindfulness] × Gender [Women or Non-binary]							0.03	0.23	0.910
(Timepoint [Posttest] × Condition [Mindfulness]) × Gender [Women or Non-binary]							0.11	0.14	0.420
Random Effects									
σ^2	0.93			0.93			0.93		
τ_{00}	$0.41_{Participant}$			$0.37_{Participant}$			$0.38_{Participant}$		
	0.27_{Item}			0.27_{Item}			0.27_{Item}		
ICC	0.42			0.41			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	0.046 / 0.448			0.068 / 0.448			0.068 / 0.449		

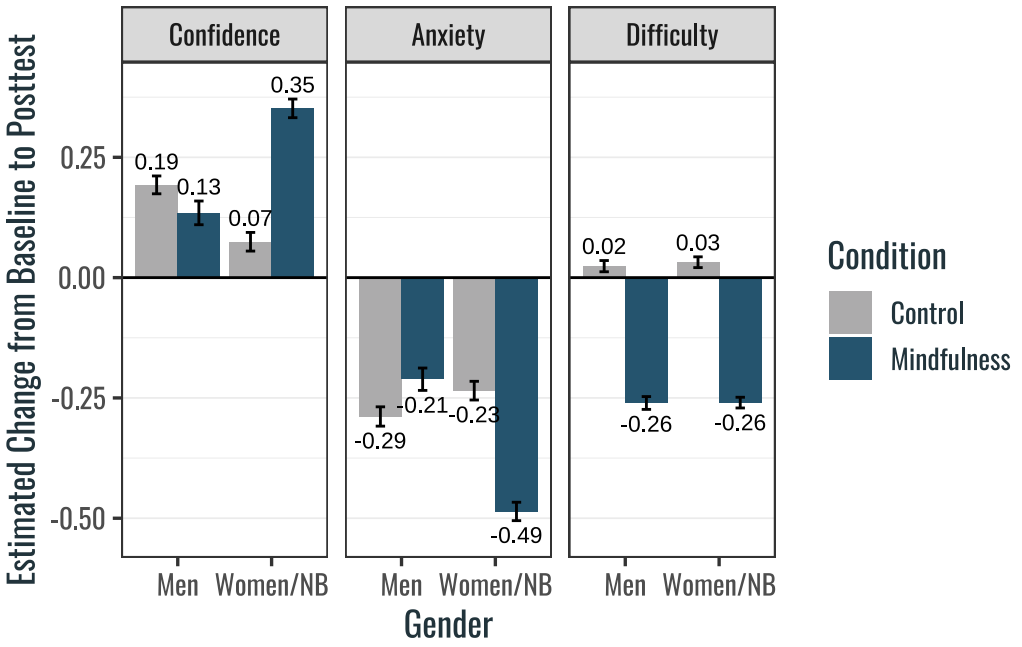
Supplementary Figure 8.1: Participants' Mean Judgment Ratings at Baseline and Posttest by Experimental Condition and Gender



8.7 Reproduction of Figure 5

See Supplementary Figure 8.2

Supplementary Figure 8.2: Estimated Marginal Means for Effects of Main Variables of Interest



9 Research Question 2

9.1 Model Specification

Mediation tests were conducted using `mediation::mediate()` using bias-corrected bootstrapped confidence intervals.

Code for model specification not available in PDF format.

9.2 Confidence

9.2.1 Men Results

Linear Model Results

See Supplementary Table 9.1

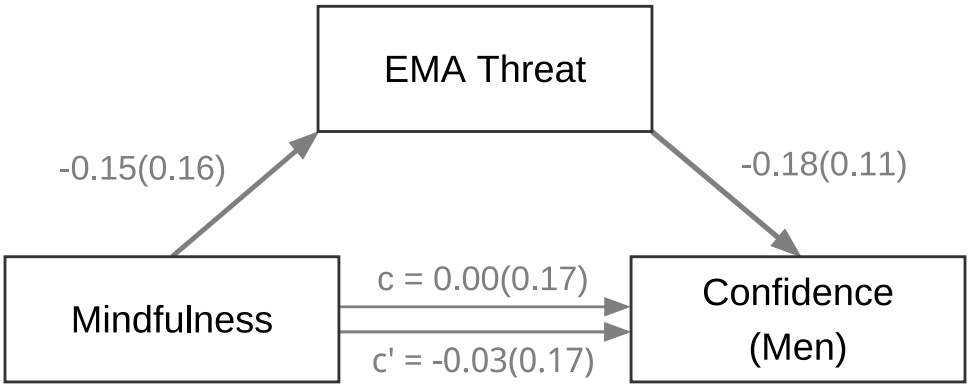
Supplementary Table 9.1: Mediation Analysis for Confidence at Posttest: Men

	Model 1 DV: Posttest Confidence			Model 2 DV: EMA Threat			Model 3 DV: Posttest Confidence		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.07	0.13	0.586	0.11	0.13	0.388	0.08	0.13	0.545
Condition [Mindfulness]	-0.00	0.17	0.985	-0.15	0.16	0.353	-0.03	0.17	0.867
Gender [Women or Non-binary]	-0.23	0.17	0.175	0.02	0.17	0.886	-0.25	0.17	0.151
Baseline Score	0.09	0.06	0.110	0.04	0.06	0.443	0.10	0.06	0.087
Cohort 2	0.09	0.10	0.364	-0.04	0.10	0.726	0.08	0.10	0.423
Cohort 3	-0.01	0.14	0.935	-0.10	0.14	0.485	-0.03	0.14	0.826
Semester Week	0.03	0.15	0.838	-0.02	0.14	0.895	0.01	0.15	0.919
Posttest Test Version [B]	-0.08	0.11	0.479	0.09	0.11	0.392	-0.07	0.11	0.540
Baseline Threat	-0.05	0.07	0.479	0.65	0.06	**<0.001**	0.05	0.09	0.594
Baseline Confidence	0.69	0.07	**<0.001**	-0.17	0.07	**0.015**	0.67	0.07	**<0.001**
Condition [Mindfulness] × Gender [Women or Non-binary]	0.35	0.22	0.115	-0.35	0.22	0.123	0.35	0.23	0.128
EMA Threat							-0.18	0.11	0.115
Gender [Women or Non-binary] × EMA Threat							0.10	0.12	0.420
Observations	148			148			148		
R2 / R2 adjusted	0.589 / 0.559			0.590 / 0.561			0.597 / 0.561		

Path Diagram

See Supplementary Figure 9.1

Supplementary Figure 9.1: Mediation Analysis for Confidence at Posttest: Men



Mediation Test

See Supplementary Table 9.2

Supplementary Table 9.2: Mediation Analysis for Confidence at Posttest: Men

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.03	-0.02	0.10	0.3546
Avg. Direct Effect	-0.03	-0.32	0.27	0.8776
Total Effect	0.00	-0.30	0.30	0.9940
Proportion Mediated	-77.28	-0.29	47.14	0.8886

Note. Sample Size Used: 148; Simulations: 10000

9.2.2 Women or Non-binary Results

Linear Model Results

See Supplementary Table 9.3

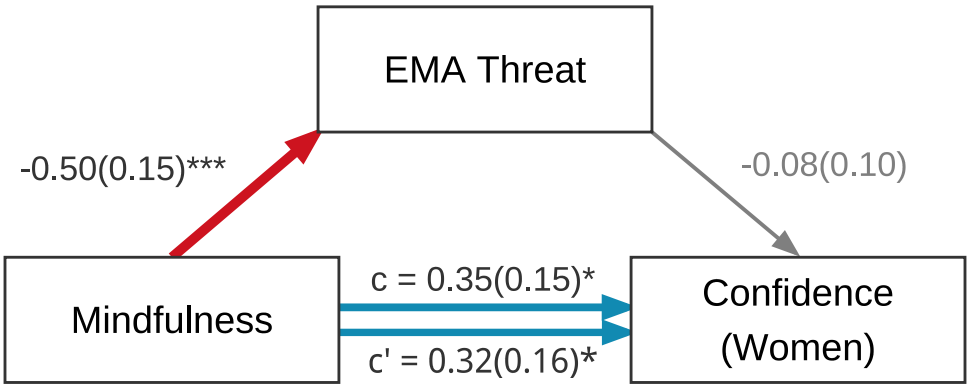
Path Diagram

See Supplementary Figure 9.2

Supplementary Table 9.3: Mediation Analysis for Confidence at Posttest: Women

	Model 1 DV: Posttest Confidence			Model 2 DV: EMA Threat			Model 3 DV: Posttest Confidence		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	-0.16	0.12	0.203	0.14	0.12	0.275	-0.17	0.13	0.196
Condition [Mindfulness]	0.35	0.15	**0.019**	-0.50	0.15	**0.001**	0.32	0.16	**0.044**
Gender [Men]	0.23	0.17	0.175	-0.02	0.17	0.886	0.25	0.17	0.151
Baseline Score	0.09	0.06	0.110	0.04	0.06	0.443	0.10	0.06	0.087
Cohort 2	0.09	0.10	0.364	-0.04	0.10	0.726	0.08	0.10	0.423
Cohort 3	-0.01	0.14	0.935	-0.10	0.14	0.485	-0.03	0.14	0.826
Semester Week	0.03	0.15	0.838	-0.02	0.14	0.895	0.01	0.15	0.919
Posttest Test Version [B]	-0.08	0.11	0.479	0.09	0.11	0.392	-0.07	0.11	0.540
Baseline Threat	-0.05	0.07	0.479	0.65	0.06	**<0.001**	0.05	0.09	0.594
Baseline Confidence	0.69	0.07	**<0.001**	-0.17	0.07	**0.015**	0.67	0.07	**<0.001**
Condition [Mindfulness] × Gender [Men]	-0.35	0.22	0.115	0.35	0.22	0.123	-0.35	0.23	0.128
EMA Threat							-0.08	0.10	0.401
Gender [Men] × EMA Threat							-0.10	0.12	0.420
Observations	148			148			148		
R2 / R2 adjusted	0.589 / 0.559			0.590 / 0.561			0.597 / 0.561		

Supplementary Figure 9.2: Mediation Analysis for Confidence at Posttest: Women



Mediation Test

See Supplementary Table 9.4

Supplementary Table 9.4: Mediation Analysis for Confidence at Posttest: Women

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.04	-0.03	0.19	0.3432
Avg. Direct Effect	0.32	0.01	0.63	0.0444
Total Effect	0.36	0.07	0.67	0.0140
Proportion Mediated	0.11	-0.03	3.84	0.3500

Note. Sample Size Used: 148; Simulations: 10000

9.3 Anxiety

9.3.1 Men Results

Linear Model Results

See Supplementary Table 9.5

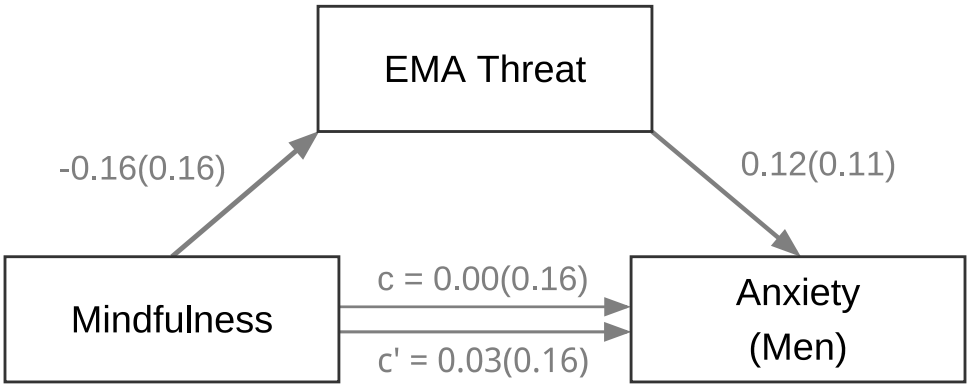
Supplementary Table 9.5: Mediation Analysis for Anxiety at Posttest: Men

	Model 1 DV: Posttest Anxiety			Model 2 DV: EMA Threat			Model 3 DV: Posttest Anxiety		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	-0.08	0.13	0.551	0.12	0.13	0.337	-0.11	0.13	0.392
Condition [Mindfulness]	0.00	0.16	0.991	-0.16	0.16	0.334	0.03	0.16	0.872
Gender [Women or Non-binary]	0.24	0.17	0.155	-0.01	0.17	0.935	0.22	0.17	0.193
Baseline Score	-0.08	0.06	0.169	0.03	0.06	0.624	-0.08	0.06	0.145
Cohort 2	-0.03	0.10	0.751	-0.05	0.10	0.631	-0.03	0.10	0.756
Cohort 3	-0.04	0.14	0.773	-0.12	0.14	0.390	-0.03	0.14	0.849
Semester Week	-0.10	0.14	0.465	-0.07	0.14	0.611	-0.11	0.14	0.455
Posttest Test Version [B]	0.09	0.11	0.439	0.09	0.11	0.427	0.07	0.11	0.538
Baseline Threat	0.07	0.06	0.263	0.67	0.06	**<0.001**	-0.05	0.08	0.558
Baseline Anxiety	0.68	0.06	**<0.001**	0.18	0.06	**0.004**	0.65	0.06	**<0.001**
Condition [Mindfulness] × Gender [Women or Non-binary]	-0.37	0.22	0.100	-0.29	0.22	0.190	-0.26	0.22	0.254
EMA Threat							0.12	0.11	0.268
Gender [Women or Non-binary] × EMA Threat							0.13	0.12	0.261
Observations	148			148			148		
R2 / R2 adjusted	0.589 / 0.559			0.597 / 0.568			0.610 / 0.575		

Path Diagram

See Supplementary Figure 9.3

Supplementary Figure 9.3: Mediation Analysis for Anxiety at Posttest: Men



Mediation Test

See Supplementary Table 9.6

Supplementary Table 9.6: Mediation Analysis for Anxiety at Posttest: Men

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	-0.02	-0.13	0.01	0.4680
Avg. Direct Effect	0.03	-0.29	0.33	0.8690
Total Effect	0.01	-0.31	0.31	0.9702
Proportion Mediated	-2.98	0.09	2292.06	0.9298

Note. Sample Size Used: 148; Simulations: 10000

9.3.2 Women or Non-binary Results

Linear Model Results

See Supplementary Table 9.7

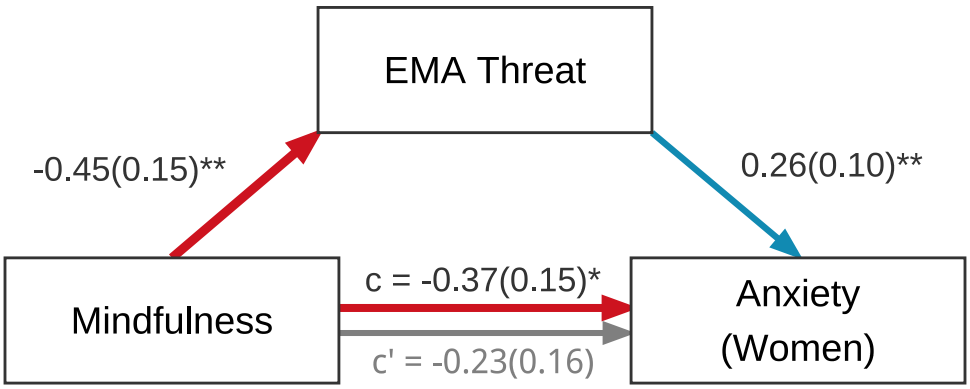
Path Diagram

See Supplementary Figure 9.4

Supplementary Table 9.7: Mediation Analysis for Anxiety at Posttest: Women

	Model 1 DV: Posttest Anxiety			Model 2 DV: EMA Threat			Model 3 DV: Posttest Anxiety		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.17	0.13	0.188	0.11	0.12	0.377	0.11	0.13	0.384
Condition [Mindfulness]	-0.37	0.15	**0.015**	-0.45	0.15	**0.003**	-0.23	0.16	0.140
Gender [Men]	-0.24	0.17	0.155	0.01	0.17	0.935	-0.22	0.17	0.193
Baseline Score	-0.08	0.06	0.169	0.03	0.06	0.624	-0.08	0.06	0.145
Cohort 2	-0.03	0.10	0.751	-0.05	0.10	0.631	-0.03	0.10	0.756
Cohort 3	-0.04	0.14	0.773	-0.12	0.14	0.390	-0.03	0.14	0.849
Semester Week	-0.10	0.14	0.465	-0.07	0.14	0.611	-0.11	0.14	0.455
Posttest Test Version [B]	0.09	0.11	0.439	0.09	0.11	0.427	0.07	0.11	0.538
Baseline Threat	0.07	0.06	0.263	0.67	0.06	**<0.001**	-0.05	0.08	0.558
Baseline Anxiety	0.68	0.06	**<0.001**	0.18	0.06	**0.004**	0.65	0.06	**<0.001**
Condition [Mindfulness] × Gender [Men]	0.37	0.22	0.100	0.29	0.22	0.190	0.26	0.22	0.254
EMA Threat							0.26	0.10	**0.009**
Gender [Men] × EMA Threat							-0.13	0.12	0.261
Observations	148			148			148		
R2 / R2 adjusted	0.589 / 0.559			0.597 / 0.568			0.610 / 0.575		

Supplementary Figure 9.4: Mediation Analysis for Confidence at Posttest: Women



Mediation Test

See Supplementary Table 9.8

Supplementary Table 9.8: Mediation Analysis for Confidence at Posttest: Women

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	-0.11	-0.32	-0.02	0.0176
Avg. Direct Effect	-0.23	-0.55	0.11	0.1858
Total Effect	-0.35	-0.66	-0.04	0.0284
Proportion Mediated	0.33	0.02	2.11	0.0436

Note. Sample Size Used: 148; Simulations: 10000

9.4 Difficulty

9.4.1 Men Results

Linear Model Results

See Supplementary Table 9.9

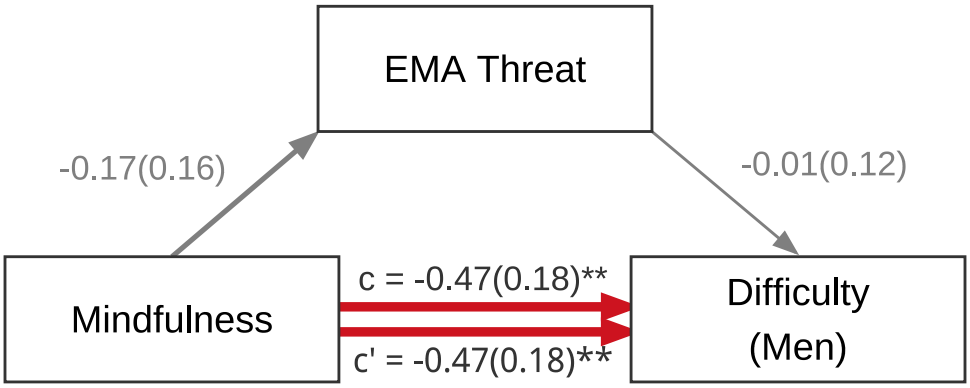
Supplementary Table 9.9: Mediation Analysis for Difficulty at Posttest: Men

	Model 1 DV: Posttest Difficulty			Model 2 DV: EMA Threat			Model 3 DV: Posttest Difficulty		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.09	0.14	0.519	0.10	0.13	0.440	0.09	0.14	0.516
Condition [Mindfulness]	-0.47	0.18	**0.009**	-0.17	0.16	0.298	-0.47	0.18	**0.009**
Gender [Women or Non-binary]	0.16	0.18	0.377	0.05	0.17	0.763	0.16	0.18	0.380
Baseline Score	-0.04	0.06	0.465	0.04	0.06	0.496	-0.04	0.06	0.474
Cohort 2	-0.15	0.11	0.166	-0.03	0.10	0.747	-0.15	0.11	0.170
Cohort 3	-0.18	0.15	0.243	-0.09	0.14	0.527	-0.18	0.16	0.246
Semester Week	-0.12	0.15	0.446	-0.02	0.15	0.880	-0.12	0.16	0.454
Posttest Test Version [B]	0.05	0.12	0.669	0.09	0.11	0.426	0.05	0.12	0.664
Baseline Threat	-0.01	0.07	0.891	0.68	0.06	**<0.001**	0.00	0.09	0.995
Baseline Difficulty	0.66	0.07	**<0.001**	0.15	0.06	**0.018**	0.67	0.07	**<0.001**
Condition [Mindfulness] × Gender [Women or Non-binary]	0.12	0.24	0.609	-0.31	0.22	0.165	0.12	0.25	0.639
EMA Threat							-0.01	0.12	0.916
Gender [Women or Non-binary] × EMA Threat							-0.00	0.13	0.976
Observations	148			148			148		
R2 / R2 adjusted	0.535 / 0.501			0.590 / 0.560			0.535 / 0.493		

Path Diagram

See Supplementary Figure 9.5

Supplementary Figure 9.5: Mediation Analysis for Difficulty at Posttest: Men



Mediation Test

See Supplementary Table 9.10

Supplementary Table 9.10: Mediation Analysis for Difficulty at Posttest: Men

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.00	-0.03	0.09	0.8944
Avg. Direct Effect	-0.47	-0.80	-0.13	0.0070
Total Effect	-0.47	-0.79	-0.13	0.0064
Proportion Mediated	0.00	-0.31	0.06	0.8940

Note. Sample Size Used: 148; Simulations: 10000

9.4.2 Women or Non-binary Results

Linear Model Results

See Supplementary Table 9.11

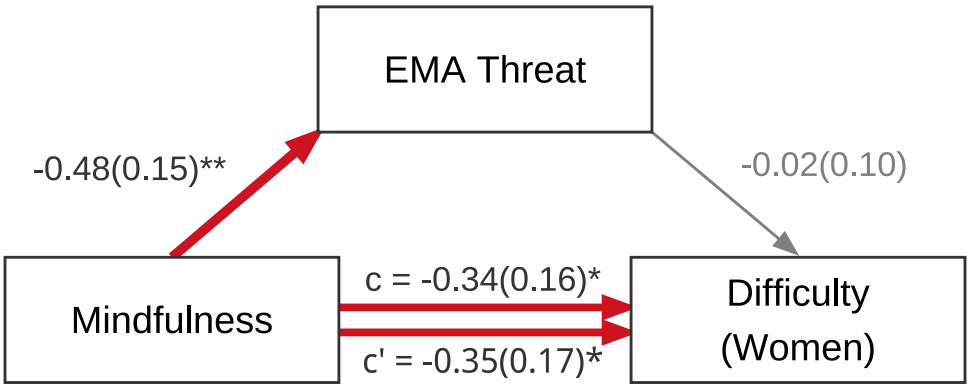
Path Diagram

See Supplementary Figure 9.6

Supplementary Table 9.11: Mediation Analysis for Difficulty at Posttest: Women

	Model 1 DV: Posttest Difficulty			Model 2 DV: EMA Threat			Model 3 DV: Posttest Difficulty		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.25	0.13	0.063	0.15	0.12	0.225	0.25	0.14	0.070
Condition [Mindfulness]	-0.34	0.16	**0.031**	-0.48	0.15	**0.001**	-0.35	0.17	**0.040**
Gender [Men]	-0.16	0.18	0.377	-0.05	0.17	0.763	-0.16	0.18	0.380
Baseline Score	-0.04	0.06	0.465	0.04	0.06	0.496	-0.04	0.06	0.474
Cohort 2	-0.15	0.11	0.166	-0.03	0.10	0.747	-0.15	0.11	0.170
Cohort 3	-0.18	0.15	0.243	-0.09	0.14	0.527	-0.18	0.16	0.246
Semester Week	-0.12	0.15	0.446	-0.02	0.15	0.880	-0.12	0.16	0.454
Posttest Test Version [B]	0.05	0.12	0.669	0.09	0.11	0.426	0.05	0.12	0.664
Baseline Threat	-0.01	0.07	0.891	0.68	0.06	**<0.001**	0.00	0.09	0.995
Baseline Difficulty	0.66	0.07	**<0.001**	0.15	0.06	**0.018**	0.67	0.07	**<0.001**
Condition [Mindfulness] × Gender [Men]	-0.12	0.24	0.609	0.31	0.22	0.165	-0.12	0.25	0.639
EMA Threat							-0.02	0.10	0.875
Gender [Men] × EMA Threat							0.00	0.13	0.976
Observations	148			148			148		
R2 / R2 adjusted	0.535 / 0.501			0.590 / 0.560			0.535 / 0.493		

Supplementary Figure 9.6: Mediation Analysis for Difficulty at Posttest: Women



Mediation Test

See Supplementary Table 9.12

Supplementary Table 9.12: Mediation Analysis for Difficulty at Posttest: Women

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.01	-0.13	0.13	0.9034
Avg. Direct Effect	-0.35	-0.70	-0.03	0.0320
Total Effect	-0.34	-0.66	-0.05	0.0254
Proportion Mediated	-0.02	-0.66	0.55	0.9068

Note. Sample Size Used: 148; Simulations: 10000

9.5 Unmoderated Mediation Results (preregistered analyses)

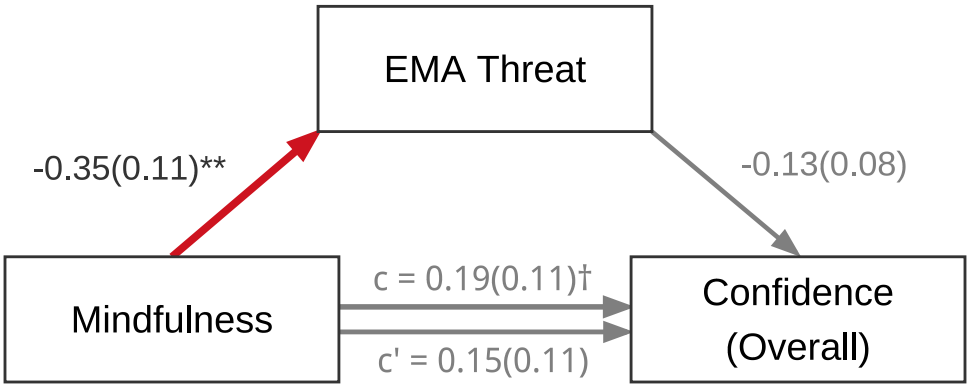
9.5.1 Confidence

Linear Models

Supplementary Table 9.13: Mediation Analysis for Confidence at Posttest without Gender Moderation

	Model 1 DV: Posttest Confidence			Model 2 DV: EMA Threat			Model 3 DV: Posttest Confidence		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	-0.03	0.12	0.828	0.21	0.12	0.076	0.00	0.12	0.982
Condition [Mindfulness]	0.19	0.11	0.083	-0.35	0.11	**0.002**	0.15	0.11	0.200
Gender [Women or Non-binary]	-0.05	0.12	0.708	-0.16	0.12	0.210	-0.07	0.12	0.587
Baseline Score	0.09	0.06	0.116	0.04	0.06	0.435	0.10	0.06	0.093
Cohort 2	0.10	0.10	0.351	-0.04	0.10	0.705	0.09	0.10	0.375
Cohort 3	-0.01	0.15	0.966	-0.11	0.14	0.464	-0.02	0.14	0.887
Semester Week	0.03	0.15	0.812	-0.02	0.15	0.868	0.03	0.15	0.828
Posttest Test Version [B]	-0.09	0.11	0.406	0.11	0.11	0.329	-0.08	0.11	0.483
Baseline Threat	-0.06	0.06	0.349	0.67	0.06	**<0.001**	0.03	0.09	0.734
Baseline Confidence	0.68	0.07	**<0.001**	-0.16	0.07	**0.021**	0.66	0.07	**<0.001**
EMA Threat							-0.13	0.08	0.114
Observations	148			148			148		
R2 / R2 adjusted	0.581 / 0.554			0.583 / 0.556			0.589 / 0.559		

Supplementary Figure 9.7: Mediation Analysis for Confidence at Posttest without Gender Moderation



Supplementary Table 9.14: Mediation Analysis for Confidence at Posttest without Gender Moderation

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.05	0.00	0.13	0.0610
Avg. Direct Effect	0.15	-0.07	0.37	0.1890
Total Effect	0.19	-0.02	0.42	0.0816
Proportion Mediated	0.24	0.16	25.32	0.1370

Note. Sample Size Used: 148; Simulations: 10000

Path Diagram

Mediation Test

9.5.2 Anxiety

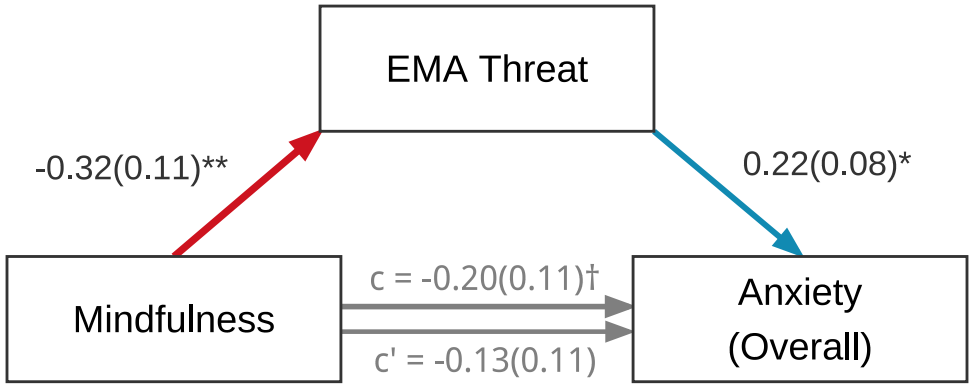
Linear Models

Supplementary Table 9.15: Mediation Analysis for Anxiety at Posttest without Gender Moderation

	Model 1 DV: Posttest Anxiety			Model 2 DV: EMA Threat			Model 3 DV: Posttest Anxiety		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.02	0.12	0.836	0.20	0.11	0.075	-0.02	0.11	0.857
Condition [Mindfulness]	-0.20	0.11	0.073	-0.32	0.11	**0.004**	-0.13	0.11	0.244
Gender [Women or Non-binary]	0.05	0.12	0.697	-0.17	0.12	0.169	0.08	0.12	0.487
Baseline Score	-0.08	0.06	0.179	0.03	0.06	0.611	-0.08	0.06	0.139
Cohort 2	-0.03	0.10	0.738	-0.05	0.10	0.621	-0.02	0.10	0.816
Cohort 3	-0.05	0.15	0.748	-0.13	0.14	0.376	-0.02	0.14	0.894
Semester Week	-0.11	0.14	0.462	-0.07	0.14	0.606	-0.09	0.14	0.524
Posttest Test Version [B]	0.10	0.11	0.363	0.10	0.11	0.365	0.08	0.11	0.467
Baseline Threat	0.08	0.06	0.195	0.68	0.06	**<0.001**	-0.07	0.08	0.405
Baseline Anxiety	0.68	0.06	**<0.001**	0.18	0.06	**0.004**	0.65	0.06	**<0.001**
EMA Threat							0.22	0.08	**0.011**
Observations	148			148			148		
R2 / R2 adjusted	0.581 / 0.554			0.592 / 0.566			0.601 / 0.571		

Path Diagram

Supplementary Figure 9.8: Mediation Analysis for Anxiety at Posttest without Gender Moderation



Mediation Test

Supplementary Table 9.16: Mediation Analysis for Anxiety at Posttest without Gender Moderation

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	-0.07	-0.18	-0.01	0.0202
Avg. Direct Effect	-0.13	-0.36	0.10	0.2546
Total Effect	-0.20	-0.42	0.02	0.0784
Proportion Mediated	0.35	-0.44	3.16	0.0930

Note. Sample Size Used: 148; Simulations: 10000

9.5.3 Difficulty

Linear Models

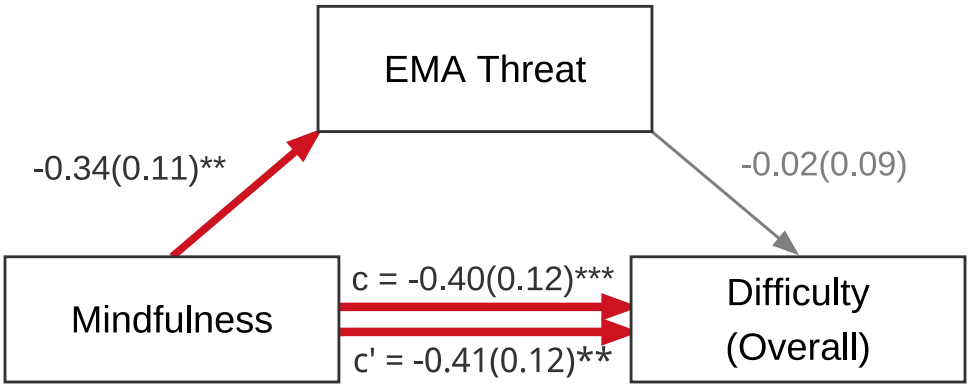
Supplementary Table 9.17: Mediation Analysis for Difficulty at Posttest without Gender Moderation

	Model 1 DV: Posttest Difficulty			Model 2 DV: EMA Threat			Model 3 DV: Posttest Difficulty		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.06	0.12	0.647	0.19	0.11	0.108	0.06	0.12	0.629
Condition [Mindfulness]	-0.40	0.12	**0.001**	-0.34	0.11	**0.002**	-0.41	0.12	**0.001**
Gender [Women or Non-binary]	0.22	0.13	0.082	-0.11	0.12	0.346	0.22	0.13	0.087
Baseline Score	-0.04	0.06	0.459	0.04	0.06	0.484	-0.04	0.06	0.469
Cohort 2	-0.15	0.11	0.167	-0.04	0.10	0.733	-0.15	0.11	0.167
Cohort 3	-0.18	0.15	0.246	-0.10	0.15	0.508	-0.18	0.15	0.243
Semester Week	-0.12	0.15	0.448	-0.02	0.15	0.869	-0.12	0.15	0.448
Posttest Test Version [B]	0.05	0.12	0.699	0.10	0.11	0.362	0.05	0.12	0.688
Baseline Threat	-0.01	0.06	0.842	0.69	0.06	**<0.001**	0.00	0.09	0.991
Baseline Difficulty	0.67	0.07	**<0.001**	0.15	0.06	**0.020**	0.67	0.07	**<0.001**
EMA Threat							-0.02	0.09	0.823
Observations	148			148			148		
R2 / R2 adjusted	0.534 / 0.503			0.584 / 0.557			0.534 / 0.500		

Path Diagram

Mediation Test

Supplementary Figure 9.9: Mediation Analysis for Difficulty at Posttest without Gender Moderation



Supplementary Table 9.18: Mediation Analysis for Difficulty at Posttest without Gender Moderation

Statistic	Estimate	CI Lower	CI Upper	p
Avg. Causal Mediation Effect	0.01	-0.06	0.08	0.8550
Avg. Direct Effect	-0.41	-0.65	-0.17	0.0012
Total Effect	-0.40	-0.63	-0.18	0.0004
Proportion Mediated	-0.02	-0.26	0.17	0.8550

Note. Sample Size Used: 148; Simulations: 10000

10 Physics Task Accuracy

The tables below show that there was no effect of mindfulness training on physics task accuracy (Supplementary Table 10.1) or learning on the PFL (Supplementary Table 10.2).

10.1 Model Specification

10.2 Preregistered Hypotheses 1 and 4

Supplementary Table 10.1: Results from Mixed Effects Models Testing Preregistered Hypotheses 1 and 4: Effects of Mindfulness Training on Problem Solving Accuracy

	Part1: Quantitative			Part2: Categorization			Part3: Qualitative		
Predictor	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
(Intercept)	0.07	0.01	**<0.001**	0.38	0.02	**<0.001**	0.50	0.02	**<0.001**
Cohort [Cohort 2]	0.01	0.06	0.902	-0.11	0.08	0.178	-0.10	0.08	0.222
Cohort [Cohort 3]	0.07	0.05	0.213	-0.03	0.07	0.630	-0.08	0.07	0.288
Semester Week	0.01	0.01	0.339	-0.00	0.01	0.750	0.01	0.01	0.675
Test Version [B]	0.03	0.02	0.092	-0.10	0.02	**<0.001**	0.21	0.02	**<0.001**
Baseline Threat	-0.04	0.01	**<0.001**	-0.00	0.01	0.766	-0.03	0.01	**0.008**
Gender [Women or Non-binary]	0.02	0.02	0.344	-0.01	0.03	0.774	-0.06	0.03	**0.033**
Timepoint [Posttest]	0.01	0.02	0.545	-0.03	0.02	0.205	0.07	0.02	**0.003**
Condition [Mindfulness]	0.02	0.03	0.385	0.04	0.04	0.292	-0.06	0.04	0.083
Timepoint [Posttest] × Condition [Mindfulness]	-0.02	0.04	0.558	-0.05	0.04	0.277	0.04	0.05	0.369
Random Effects									
σ^2	0.03			0.04			0.04		
τ_{00}	$0.00_{Participant}$			$0.01_{Participant}$			$0.01_{Participant}$		
ICC	0.04			0.20			0.15		
N	$149_{Participant}$			$149_{Participant}$			$149_{Participant}$		
Observations	295			298			298		
Marginal R^2 / Conditional R^2	0.095 / 0.135			0.072 / 0.259			0.274 / 0.380		

> **Note**
>
> Supplementary Table 10.1: The estimates for the intercept represent the overall mean score (percent correct) and standard errors for each of the problem solving performance outcomes at baseline. The estimate for timepoint represents the change in the dependent variable from baseline to posttest across

10.3 Preregistered Hypothesis 5

Supplementary Table 10.2: Results from Logistic Mixed Effects Model Testing Preregistered Hypothesis 5: Effects of Mindfulness Training on Learning During the Preparation for Future Learning Task

	PFL Correctness		
Predictor	Odds Ratios	SE	p
(Intercept)	0.23	0.06	**<0.001**
Cohort [Cohort 2]	1.13	1.04	0.894
Cohort [Cohort 3]	0.51	0.41	0.403
Semester Week	1.10	0.16	0.521
Baseline Threat	0.86	0.11	0.234
Gender [Women or Non-binary]	0.37	0.13	**0.005**
Question [2]	4.87	1.59	**<0.001**
Condition [Mindfulness]	1.42	0.64	0.433
Question [2] × Condition [Mindfulness]	0.91	0.52	0.873
Random Effects			
σ^2	3.29		
τ_{00}Participant	0.62		
ICC	0.16		
NParticipant	149		
Observations	298		
Marginal R^2 / Conditional R^2	0.235 / 0.356		

> **Note**
>
> Supplementary Table 10.2: The odds ratio for the intercept term represents the odds of getting question 1 correct compared to incorrect. The odds ratio for the question × condition interaction term represents the difference in odds of getting question 2 correct between conditions, above and beyond any condition differences on question 1 and overall differences on question 2, compared to question 1. P-values below .05 are indicated by

> bold font.

References

Belenky, Daniel M, and Timothy J Nokes-Malach. 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.

Hardiman, Pamela Thibodeau, Robert Dufresne, and Jose P Mestre. 1989. "The Relation Between Problem Categorization and Problem Solving Among Experts and Novices." *Memory & Cognition* 17 (5): 627–38.

Schwartz, Daniel L, and Taylor Martin. 2004. "Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction." *Cognition and Instruction* 22 (2): 129–84.

Weinlader, Nolan K., Eric Kuo, Benjamin M. Rottman, and Timothy J. Nokes-Malach. 2019. "A New Approach for Uncovering Student Resources with Multiple-Choice Questions." In, 621–26. American Association of Physics Teachers. https://doi.org/10.1119/perc.2019.pr.Weinlader.