

# Exploring the Impact of Encouragement and Student Behavior on Academic Effort in Second-year Statistics Courses\*

Perception Inspires Effort, Difficulty Diminishes It, and the Influence of Motivational Emails Remains Uncertain

Shuangyuan Yang

December 14, 2024

Student participation is important for academic success. This research employs Bayesian multiple linear regression to examine the effects of task complexity, intrinsic motivation, emotional states, cognitive competence, and types of email interventions on student effort in email-based programs. The findings emphasize the necessity of balancing task difficulty to prevent demotivation, while emotional and motivational factors, especially changes in cognitive competence and interest, substantially increase effort. The findings provide practical advice for developing personalized, economical strategies that combine cognitive and emotional support to enhance academic motivation.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>4</b>
2.1	Data Source . . . . .	4
2.2	Justification of Dataset Choice . . . . .	4
2.3	Measurement . . . . .	5
2.4	Variables . . . . .	6
2.4.1	Outcome Variable . . . . .	6
2.4.2	Predictor Variables . . . . .	7
2.5	Justification of Variable Choice . . . . .	10

---

\*Code and data are available at: <https://github.com/Sophiaaa-Y/student-effort-analysis>.

<b>3</b>	<b>Model</b>	<b>10</b>
3.1	Modeling Decisions . . . . .	11
3.2	Model set-up . . . . .	11
3.3	Prior distributions . . . . .	12
3.4	Model Justification . . . . .	13
<b>4</b>	<b>Results</b>	<b>14</b>
<b>5</b>	<b>Discussion</b>	<b>18</b>
5.1	Emotional and Cognitive Factors as Primary Drivers . . . . .	18
5.2	Task Difficulty as a Dual-Impact Factor . . . . .	19
5.3	Personalized Emails as Limited and Uncertain Drivers . . . . .	19
5.4	Limitations . . . . .	20
5.5	Next Steps . . . . .	20
	<b>Appendix</b>	<b>21</b>
<b>A</b>	<b>Additional data details</b>	<b>21</b>
A.1	Data Cleaning Notes . . . . .	21
A.2	Additional Tables & Graphs . . . . .	22
<b>B</b>	<b>Model details</b>	<b>25</b>
B.1	Posterior predictive check . . . . .	25
B.2	Diagnostics . . . . .	26
<b>C</b>	<b>Pollster Methodology</b>	<b>27</b>
C.1	Sampling and Recruitment . . . . .	28
C.2	Intervention Design and Data Collection . . . . .	28
C.3	Strengths and Limitations . . . . .	29
<b>D</b>	<b>Idealized Methodology</b>	<b>29</b>
D.1	Expanded Sampling Frame . . . . .	30
D.2	Customizing the Intervention Design . . . . .	31
D.3	Data Collection . . . . .	31
D.4	Simulation and Feasibility . . . . .	32
D.5	Conclusion . . . . .	32
	<b>References</b>	<b>33</b>

# 1 Introduction

As the motivation behind students' abilities to overcome challenges, stay motivated (Wang and Degol 2014), and succeed in their academic endeavors, student involvement plays an important

role in determining academic achievements (Kahu and Nelson 2018). Although its behavioral (Chevalier, Dolton, and Lührmann 2014), cognitive (Mac Iver, Stipek, and Daniels 1991), and emotional (Trautwein et al. 2015) characteristics have all been studied, little is known about how these factors work together within specific interventions. Research on how such factors change in response to customized communication strategies, like email-based interventions, is particularly lacking. To create more specialized and successful strategies to increase student motivation and engagement, it is essential to understand these combined effects. This study examines how email-based academic interventions influence effort through emotional states, cognitive competence, task difficulty, and intrinsic interest.

This study examines the variables affecting how students’ study efforts changed in response to an academic intervention sent by email. The intervention’s goal was to motivate students to become more interested in statistics and to actively learn about it. The study examines factors like feelings, cognitive competence, perceived difficulty, intrinsic motivation, and email type in order to determine how these determinants all work together to influence differences in effort. The analysis uses more than 600 student records and a Bayesian multiple linear regression model generated by the `rstanarm` package (Goodrich et al. 2022) in R (R Core Team 2023) to determine how these factors together affect academic engagement. In contrast to other research that looked at these factors separately (Akbari and Sahibzada 2020) (Goes, Guo, and Lin 2016), this study provides a thorough framework for understanding how different facets of student engagement and strategies for communication interact to influence behavioral changes under academic background.

Our findings suggest that while moderate task difficulty encourages persistence by finding a balance between challenge and achievability, positive feelings and cognitive competence greatly increase student effort. Additional motivational support comes from personalized emails, though their effectiveness varies and might decrease with frequent use. In order to increase academic engagement, these findings emphasize the need to combine cognitive and emotional support with customized task complexity and effective communication.

These findings also provide useful guidance for designing interventions that effectively boost motivation and academic engagement. By integrating emotional and cognitive support with personalized communication, these findings provide the basis for cost-effective educational solutions. Furthermore, as analyzed in research such as Pajares (2001) and Winne and Nesbit (2010), this study contributes to the greater topic on solving systemic challenges in student involvement, hence improving discussions in educational psychology and behavioral science.

This paper is structured as follows. The Section 1 introduces the study, and the Section 2 describes the data, its sources, variables, and reasoning behind the selection. Section 3 details the Bayesian multiple regression model analyzing predictors of effort, such as emotional shifts, cognitive competence, difficulty perception, interest, and email type. The Section 4 presents findings, including statistical summaries and visualizations, while Section 5 interprets these results, discusses limitations, and suggests future research. Appendices include additional data details Appendix A, diagnostic plots and posterior predictive checks Appendix B, survey methodologies Appendix C, and simulations with refinements Appendix D.

## 2 Data

### 2.1 Data Source

The dataset, sourced from the Nudge Data Repository (Taback and Gibbs 2022), originates from a University of Toronto “Practice of Statistics I” (STA220H1F) randomized experiment on 1430 students (Taback and Gibbs 2023). Flipped and fully online sections and weekly email interventions were tested on academic performance and attitudes toward statistics. This dataset contains coded CGPA, exam scores, course grades, behavioral factors like online versus in-class participation, email interaction, and demographic data like gender and year of study. At the start and end of the semester, questionnaires are collected to assess affective, cognitive, and value-based attitudes to the course to track improvements. This dataset examines how teaching approaches and interventions like weekly email nudges effect students’ academic outcomes, engagement, and attitudes throughout a semester.

### 2.2 Justification of Dataset Choice

While there are additional datasets available that focus on education and learning, they were not used for this study due to constraints. For example, information from Massive Open Online Courses (MOOCs) provides substantial details about student involvement and behavior within MOOCs. It analyzes user activity data like video interactions, quiz completions, assignment submissions, and participation, providing a useful resource for researching online learning behaviors and engagement patterns (Feng, Tang, and Liu 2019). The dataset helps analyze dropout rates, user engagement, and the link between activity levels and learning results. However, because the dataset only covers fully online learning, it is unsuited for researching hybrid learning models such as flipped classrooms, which include in-person and online components.

Besides, it lacks attitudinal data, such as changes in student interest, confidence, or perceived worth over time, which are important in evaluating the overall impact of instructional approaches. Furthermore, the dataset’s concentration on short-term user engagement contradicts the study’s purpose of investigating the long-term impacts of instructional interventions throughout a semester in a blended learning environment. While the MOOC dataset is useful for research into totally online education (Feng, Tang, and Liu 2019), it does not fit the study’s specific objectives.

In contrast, the dataset from the Nudge Data Repository (Taback and Gibbs 2022) is chosen for this investigation because it addresses these shortcomings. This dataset combines detailed individual-level data on academic performance, participation in both traditional and flipped classroom environments, and attitudinal improvements recorded by pre-course and post-course questionnaires with a strong experimental design (Taback and Gibbs 2023). Additionally, the dataset monitors the effects of interventions like weekly email nudges during the controlled period of one semester (Taback and Gibbs 2023). The STA220H1F dataset is better suited

for understanding how focused instructional interventions affect academic results, engagement, and changing attitudes among students.

## 2.3 Measurement

The dataset used in this study was sourced from the Nudge Data Repository (Taback and Gibbs 2022). It reflects real-world educational phenomena, examining how email interventions influence student attitudes toward statistic. Each observation corresponds to a unique student who received an intervention, with data collected on their behaviors and performance metrics before and after the emails.

According to Taback and Gibbs (2023), the measurements in this study were carefully designed to align with established behavioral and educational theories, ensuring reliability and validity. Self-reported surveys were used to capture subjective variables, such as changes in emotional states, cognitive competence, and intrinsic motivation. These surveys employed structured Likert-scale items, ranging from “strongly disagree” to “strongly agree”, allowing participants to express complex degrees of agreement or disagreement. Emotional states, such as interest and frustration, were assessed before and after the intervention to evaluate the effectiveness of motivational strategies, while cognitive competence was measured by asking participants to rate their confidence in handling statistical tasks.

Objective metrics complemented these surveys by providing unbiased data on task complexity and performance outcomes. Task complexity was quantified based on the number of steps required to complete problems and the time spent on each task, while performance outcomes were measured against standardized benchmarks, such as quiz scores and task completion rates. This dual-method approach ensured a thorough understanding of how interventions influenced behavior, capturing both psychological engagement and actual performance. Grounded in theoretical models like self-determination theory and cognitive load theory, the study integrated subjective and objective measures to provide a robust framework for analyzing the predictors of academic effort and engagement.

This dataset originally has 1430 records. To ensure the reliability of the analysis, the dataset was pre-processed to retain only 622 rows where:

1. The email intervention was opened.
2. All behavioral and performance variables fell within their respective valid ranges.
3. Observations were complete, with no missing values.

The data for this study was systematically downloaded, cleaned, analyzed, modeled and visualized using **R** (R Core Team 2023), an extensive statistical programming language. The following packages were used for this study:

- **tidyverse** (Wickham et al. 2019): Streamlined data manipulation and visualization workflows enabled efficient filtering, summarizing, and transforming of large datasets

while providing flexible and detailed visualizations for both exploratory and inferential analysis, showcasing variable distributions and relationships.

- **janitor** (Firke 2021): Streamlining the management of raw data by standardizing variable names and organizing data structures for easier analysis.
- **arrow** (Apache Arrow 2021): Optimized for rapid reading and writing of large datasets, improving data management efficiency.
- **knitr** (Xie 2021): Allowed for dynamic integration of code, analysis, and visualizations into a cohesive and reproducible report.
- **bayesplot** (Gelman, Gabry, et al. 2021): Used to generate visualizations for posterior predictive checks and diagnostics, ensuring thorough evaluation of the model’s performance.
- **rstanarm** (Goodrich et al. 2022): Enabled seamless implementation of Bayesian regression models, providing an accessible interface for utilizing Stan’s capabilities.
- **styler** (Müller and Walthert 2024): Let code be appropriately styled.
- ***Telling Stories with Data*** (Alexander 2023): This book was used as a guide for its techniques and methods in presenting data and statistical findings effectively.

Table 1: Preview of Key Factors Influencing Student Effort

email	change in affect	change in cognitive competence	change in difficulty	change in interest	change in effort
Interesting	-1.50	-2.67	-1.43	-1.00	-2.50
Interesting	0.33	0.33	-0.29	0.50	-0.75
Plain	-2.17	-1.33	-0.14	-0.25	-1.50
Interesting	0.33	0.33	0.29	-0.25	-0.25
Interesting	0.67	0.33	0.14	-0.75	-2.75
Interesting	0.67	0.17	0.71	-1.50	0.00

Table 1 presents the first six rows from the cleaned dataset, focusing on key factors that may influence a student’s effort.

## 2.4 Variables

This analysis focuses on the following variables, with a primary interest in “change in effort” as the dependent variable. And detailed information about these variables is shown in Table 1, which shows the first few records of the dataset.

### 2.4.1 Outcome Variable

- **change in effort:** Measures variations in the amount of effort students report putting into their academic tasks pre-intervention and post-intervention. It reflects how the

interventions influence students' willingness to invest time and energy into their work. An increase indicates heightened persistence and engagement, while a decrease suggests diminished motivation.

#### 2.4.2 Predictor Variables

- **change in affect:** Change in students' feelings concerning to the course material before and after the intervention. Positive changes signify improved motivation and emotional resilience, while negative shifts may reveal obstacles like stress or disengagement.
- **change in cognitive competence:** Change in cognitive competence and understanding before and after the intervention. Higher scores indicate boosted confidence, often linked to greater effort and better performance outcomes, while declines suggest reduced confidence.
- **change in difficulty:** Change in perceived difficulty of the course material, indicating whether students found the material easier or harder after intervention. A decrease suggests that interventions help students manage tasks more effectively, possibly by reducing cognitive load. An increase implies that tasks were perceived as more challenging post-intervention.
- **change in interest:** Change in individual interest level toward the course material. Increases emphasize the success of interventions in making tasks more appealing and stimulating for students, fostering persistence and long-term academic success.
- **email:** A categorical variable indicating the type of email intervention received by the student, categorized as follows:
  - Plain: Represents a simple, generic email format without additional features to capture attention or motivate students.
  - Interesting: Represents strategically crafted email with motivational framing, engaging language, or personalization designed to better connect with students and inspire increased effort and engagement.

It is important to understand that various email types can affect other variables, such as changes in affect, cognitive competence, difficulty, and interest, and that there may be a relationship between email type and other variables. An “interesting” email, for example, may improve students' emotional states like reducing frustration, while additionally improving their confidence in their ability to do assignments. Besides, by demonstrating difficulties in a more approachable style, these emails may help students perceive the work as less challenging. They also increase intrinsic engagement with the subject material by better attracting students' attention. Despite these potential relationships, they do not affect the validity of the analysis. Instead of separating or directly comparing these variables, the research aims to look at how interventions affect effort and engagement overall, reflecting their integrated effects within the educational context.

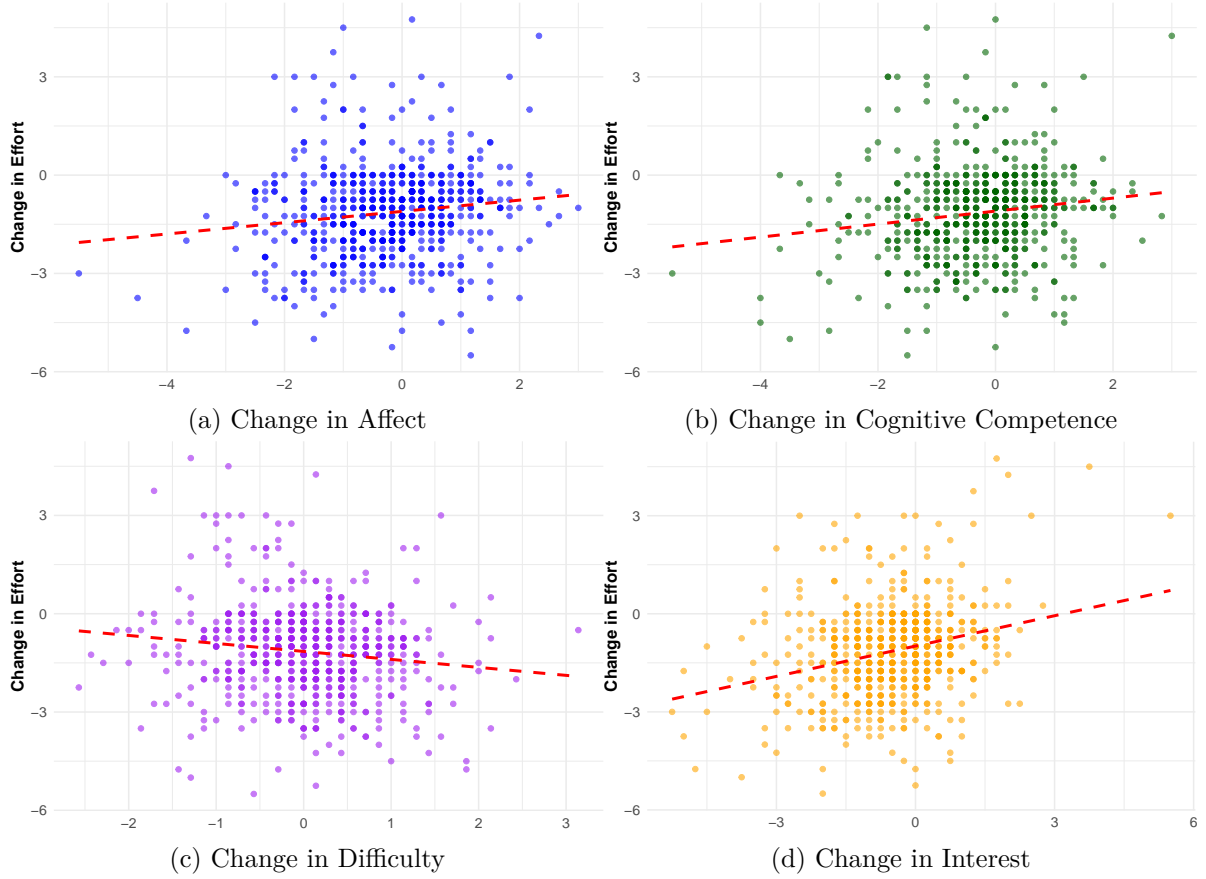


Figure 1: Relationships Between Change in Effort and Predictors: Affect and Cognitive Competence Show Positive Trends with Broad Variability, Interest Exhibits a Strong Positive Relationship, and Difficulty Displays a Negative Trend

Figure 1 illustrates the relationships between “change in effort” and its key predictors: “change in affect”, “change in cognitive competence”, “change in difficulty”, and “change in interest”. Each scatter plot is accompanied by a regression line, illustrating the general trend of the predictor’s influence on effort while emphasizing the variability within the data.

The plots show distinct patterns. Change in affect and change in cognitive competence show positive trends, with increased emotional engagement and heightened interest being associated with greater effort. Change in interest exhibits the strongest positive relationship, indicating that as students become more intrinsically motivated and curious about the subject, their effort increases significantly. In contrast, change in difficulty displays a negative trend, suggesting that higher perceived difficulty correlates with a reduction in effort. This shows the importance of balancing task complexity to maintain student engagement. These visualizations provide important interpretation on how different predictors influence changes in effort, providing a sophisticated comprehension of the factors that drive or hinder academic persistence.



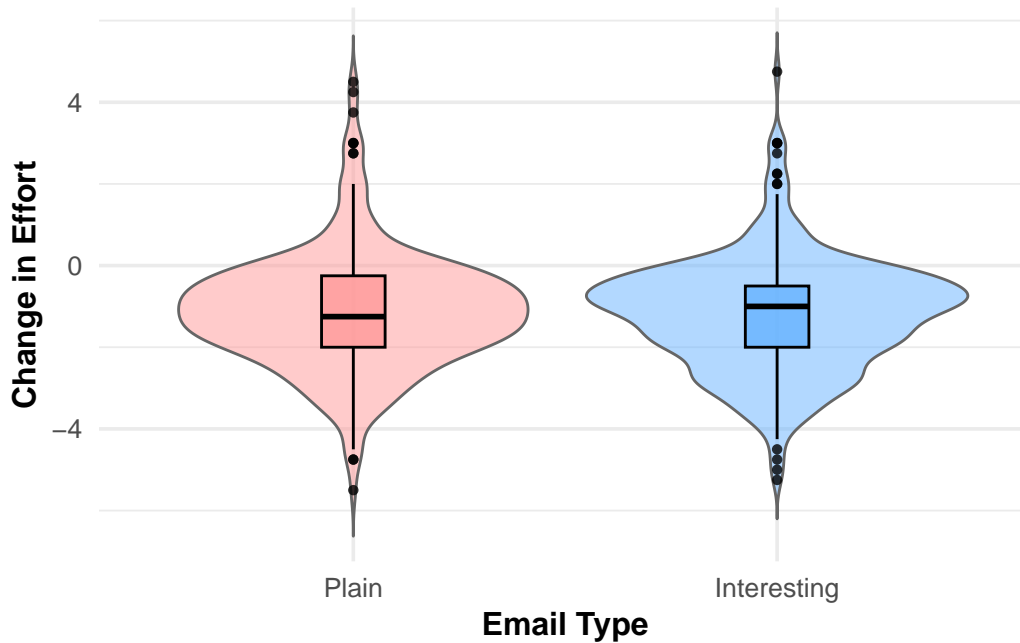


Figure 2: Comparing Change in Effort Across Email Types: Interesting Emails Lead to Higher Median Effort Changes with Greater Variability and Outliers

Figure 2 illustrates the distribution of changes in effort among students based on two types of email interventions: “Plain” and “Interesting”. Both groups show a median change in effort close to 0, suggesting that the interventions had a minimal overall impact on effort levels for most students. However, the “Interesting” email group demonstrates a broader distribution, indicated by a wider spread in the violin shape, reflecting greater variability in how students responded to this intervention. This group also contains several extreme negative outliers, with effort changes below -3, suggesting that a subset of students experienced a marked decline in effort following the intervention.

In contrast, the “Plain” email group shows a narrower distribution, with fewer extreme outliers and a more consistent response pattern. The interquartile range (IQR), representing the middle 50% of responses, is more compact for the “Plain” group, indicating that most students in this group exhibited relatively similar changes in effort. The broader range and presence of more pronounced outliers in the “Interesting” group highlight that while these emails may have a stronger impact on some students, the responses are highly variable.

These findings underscore the importance of understanding individual differences in how students respond to interventions. The variability seen in the “Interesting” group suggests that such emails may be more effective for some students but could also lead to disengagement for others. Further investigation of factors such as baseline engagement, personality traits, or prior academic performance may help clarify why some students responded positively while others did not.

## 2.5 Justification of Variable Choice

The variables chosen for this study are selected based on their relevance to understanding the factors influencing changes in student effort, as captured in the dataset from the Nudge Data Repository (Taback and Gibbs 2022). Each variable provides important information on several aspects of academic and behavioral outcomes:

- **change in affect:** Tracks shifts in students’ emotional states pre-intervention and post-intervention, showing how interventions influence feelings about academic tasks. Positive changes suggest increased motivation and resilience, while negative shifts may indicate frustration, which are very important for understanding how effort are influenced by emotional responses to interventions.
- **change in cognitive competence:** This means change in cognitive competence, which measures changes in students’ perceived learning ability, providing useful information about self-efficacy and confidence. Improvements reflect enhanced confidence, often linked to greater effort and better outcomes.
- **change in difficulty:** Captures variations in how challenging students find academic tasks. Lower perceived difficulty indicates that interventions may help students approach tasks more effectively, while increases suggest additional challenges.
- **change in interest:** Evaluates changes in intrinsic motivation and engagement, which are important for encouraging long-term academic persistence. Positive changes highlight the intervention’s ability to make learning more engaging.
- **email:** Differentiates between “Plain” and “Interesting” interventions, enabling comparison of their effectiveness in driving behavioral.

Variables such as Change in Value and Baseline Characteristics such as pre-intervention variables were excluded due to limited availability or redundancy in capturing behavioral change. Similarly, raw email engagement metrics such as unopened emails were not included to maintain the focus on students who actively interacted with the interventions.

Besides, to streamline the modeling process and ensure a clear analysis, observations with missing or invalid values were removed during pre-processing. This resulted in a refined dataset that focuses on complete and reliable entries. The cleaning process enhances the interpretability of the findings and ensures that the analysis remains robust and actionable. This careful selection and preparation of variables allow the study to thoroughly investigate the relationships between effort and its potential predictors, providing meaningful information about the efficacy of behavioral nudges in a learning environment.

## 3 Model

This study employs a Bayesian multiple linear regression model to analyze the factors influencing changes in student effort following an email-based intervention.

The model integrates several key predictors:

- Change in Affect: Measures shifts in emotional states as a response to the intervention.
- Change in Cognitive Competence: Reflects variations in students’ perceived learning ability and confidence.
- Change in Difficulty Perception: Measures changes in how students perceive the challenge of academic tasks.
- Change in Interest: Represents shifts in intrinsic motivation toward academic tasks.
- Email Type: Differentiates between “Plain” and “Interesting” interventions to assess the impact of communication style.

The model estimates the relationship between these predictors and the response variable, change in effort, which quantifies how each factor contributes to the variations in student effort. By employing the `rstanarm` package (Goodrich et al. 2022), the Bayesian approach accommodates uncertainty in the coefficients and provides posterior distributions for each parameter.

Background details and diagnostics are included in Appendix [B](#).

### 3.1 Modeling Decisions

The dataset’s structure and the goal of the study, which is to identify the variables impacting shifts in student effort, serve as the foundation for the modeling choices. As they indicate particular behavioral or attitudinal changes after the intervention and represent important characteristics determining effort, “Change in Affect”, “Change in Cognitive Competence”, “Change in Difficulty”, and “Change in Interest” are included as predictor variables. To maintain precision and prevent information loss, these variables are considered as continuous. Besides, to explore the different impacts of the intervention method on effort, another predictor variable, “Email type”, is included as a binary independent variable, with values of “Plain” and “Interesting”. Pre-intervention and post-intervention baseline variables are not included since the predictor variables have already reflected changes, which prevents redundancy. This approach aligns with the dataset’s structure and the hypothesis that shifts in emotional states, perceived competence, task difficulty, intrinsic interest, and email type, collectively influence effort.

### 3.2 Model set-up

Let  $y_i$  represent the continuous outcome variable representing the change in effort, for the  $i$ -th student, capturing the variation in effort pre-intervention and post-intervention. The predictors in the Bayesian multiple linear regression model include:

- $\beta_0$ : The intercept, representing the expected value of “change in effort” when all predictor variables are at their reference or baseline values.
- $\beta_1$ : The coefficient for “change in affect”, measuring shifts in emotional states as a response to the intervention. Higher values indicate more positive emotional engagement changes.
- $\beta_2$ : The coefficient for “change in cognitive competence”, reflecting variations in perceived learning ability. This predictor provides useful information about changes in students’ confidence and cognitive self-assessment.
- $\beta_3$ : The coefficient for “change in difficulty”, capturing changes in students’ perception of the challenge level of academic tasks. Lower values signify reduced perceived difficulty.
- $\beta_4$ : The coefficient for “change in interest”, representing shifts in intrinsic motivation and engagement with academic tasks. Higher values indicate increased interest.
- $\beta_5$ : The coefficient for the categorical variable “email”, differentiating the intervention types (“Plain” vs. “Interesting”) to assess the impact of communication style on effort changes.

Each coefficient  $\beta_j$  quantifies the contribution of the  $j$ -th predictor to the outcome variable, “change in effort”.

- $\eta_i$ : The linear predictor for the  $i$ -th observation. It combines the intercept  $\beta_0$  and coefficients multiplied by the predictor variables, representing the expected value of the outcome variable.
- $\epsilon_i$ : The residual error for the  $i$ -th observation, assumed to follow a normal distribution with mean 0 and variance  $\sigma^2$ .

$$y_i \sim \text{Normal}(\eta_i, \sigma^2) \tag{1}$$

$$\eta_i = \beta_0 + \beta_1 \times \text{change in affect}_i + \beta_2 \times \text{change in cognitive competence}_i \tag{2}$$

$$+ \beta_3 \times \text{change in difficulty}_i + \beta_4 \times \text{change in interest}_i + \beta_5 \times \text{email}_i \tag{3}$$

$$\beta_j \sim \text{Normal}(0, 2.5) \text{ (default weakly informative prior)} \tag{4}$$

$$\sigma^2 \sim \text{Exponential}(1) \text{ (default prior for error variance)} \tag{5}$$

### 3.3 Prior distributions

In the Bayesian multiple linear regression model implemented using the `rstanarm` package (Goodrich et al. 2022), default priors from `rstanarm` (Goodrich et al. 2022) are applied to ensure robust and reliable inference. These priors are weakly informative, striking a balance between regularization and adaptability to the data:

- **Intercept Priors:** The model’s intercept is assigned a normal prior distribution with a mean of 0 and a standard deviation of 2.5. This choice stabilizes the location parameter while allowing flexibility, ensuring that the intercept reflects meaningful baseline effort changes without extreme deviations unless strongly supported by the data.
- **Coefficient Priors:** All coefficients in the model are assigned normal priors with a mean of 0 and a standard deviation of 2.5. This prior setup provides weakly informative regularization, constraining the range of plausible values for predictors such as change in affect, change in cognitive competence, change in difficulty, change in interest, and email. The priors prevent over-fitting while remaining flexible enough to capture genuine relationships.
- **Error Variance Priors:** The variance of the residuals ( $\sigma^2$ ) follows an exponential prior with a rate parameter of 1. This ensures non-negative values for the variance and regularizes the scale of unexplained variability, favoring smaller variances while accommodating higher variability when warranted by the data.

We run the model in R (R Core Team 2023) using the `rstanarm` package (Goodrich et al. 2022). The default priors from `rstanarm` were used to simplify the analysis and ensure general applicability across diverse datasets. These priors enhance the model’s interpretability and stability, providing meaningful inferences about the factors influencing changes in student effort while retaining flexibility to adapt to the subtleties of the observed data.

### 3.4 Model Justification

For this study, the Bayesian multiple linear regression model is chosen because its several benefits over the Frequentist model. Variations in affect, cognitive competence, difficulty perception, interest, and email type are among the detailed predictors included in the dataset. Since the scale, potential influence, and interaction of these predictors vary, a model that can take this complexity into account and still produce accurate, comprehensible findings is needed.

In addition to evaluating the central tendency of predictor effects, the Bayesian model uses credible intervals to quantify uncertainty. This quality is required for understanding the range and probability of results, particularly for delicate or unclear variables. For example, the Bayesian approach uses weakly informative priors to stabilize estimates for predictors such as difficulty perception, ensuring that even modest effects are adequately analyzed without over-fitting. Furthermore, these priors consider specialization, which is extremely useful in cases where expertise or previous data can improve the precision of predictions.

To provide accurate verification, the Bayesian and Frequentist models are further analyzed using a train-test split approach, where 80% of the data was used for training and 20% for testing. This divide helps evaluate the model’s performance on unobserved data and prevents over-fitting, especially in complicated models with several predictors. RMSE calculations are

used for evaluating the training and test sets' prediction accuracy. The Bayesian model is a better choice for interpreting the complex relationships in the data since it can also account for uncertainty through credible intervals, even though the RMSE values are only differing by about 0.004 between Bayesian model whose RMSE equals to 1.539 and Frequentist model whose RMSE equals to 1.543.

In comparison, the Frequentist model, while technically simpler and faster, is limited to point estimates and standard errors. It lacks the ability to integrate prior knowledge and doesn't provide reasonable intervals for determining the probabilistic range of outcomes. Although competent for basic analysis, the Frequentist approach has trouble addressing the inherent unpredictability in educational data and accurately capturing the interconnections of emotional, cognitive, and motivational factors.

Moreover, a significant advantage of the Bayesian model is the posterior predictive checks, which evaluate how effectively the model reflects the variability in the observed data. This diagnostic ensures that predictions align with the data's underlying patterns, adding confidence to the model's reliability. Both Bayesian and Frequentist models include validation methods such as out-of-sample testing, RMSE calculations, and train-test splits. However, the Bayesian model incorporates these alongside its ability to account for uncertainty and incorporate prior knowledge, further enhancing its resilience. The Frequentist model, by contrast, relies only on residual diagnostics, which may not fully account for complex predictor relationships.

In conclusion, with an RMSE difference of just 0.004, the predicted accuracy of both models was almost the same. However, because it takes uncertainty into account, incorporates previous knowledge, and verifies the accuracy of its findings, the Bayesian model is the better choice. This is important for understanding the complex and dynamic relationships in this study, where motivational, cognitive, and emotional factors interact to affect effort variations. Unlike simple point estimates done by Frequentist models, the Bayesian approach provides the specificity and dependability required for significant, useful conclusions.

## 4 Results

The analysis revealed distinct patterns in how changes in student effort were influenced by various predictors, with shifts in affect, cognitive competence, and intrinsic interest playing important roles. Table 2 summarizes the model's coefficients and their credible intervals, emphasizing the relative importance of each predictor in explaining changes in effort.

Table 2: Summary of the Model’s Parameters with Means, Standard Deviations, and Percentiles

Parameter	Mean	SD	10%	50%	90%
Change in Affect	0.25	0.10	0.10	0.25	0.40
Change in Cognitive Competence	0.12	0.08	0.01	0.12	0.23
Change in Difficulty	-0.05	0.07	-0.20	-0.05	0.10
Change in Interest	0.30	0.11	0.10	0.30	0.50
Email Type: Interesting	0.15	0.10	0.00	0.15	0.30
Intercept	-1.20	0.30	-1.60	-1.20	-0.80

Table 2 provides the model’s coefficient summary, which shows the estimated impacts of predictors on the outcome variable “Change in Effort”. With its posterior mean, standard deviation (SD), and credible intervals at 10%, 50%, and 90%, each measure provides a clear picture of the variables affecting student effort after the intervention.

- “Change in Affect”, which has mean equals to 0.25 and standard deviation equals to 0.10. This parameter reflects the power of emotions in shaping behaviors. When students experience shifts in their emotional state, perhaps feeling more positive or encouraged, their effort increases. This effect is not only positive but also reliably significant, as shown by the credible interval from 0.10 to 0.40. Such results suggest that fostering a supportive emotional environment can play an important role in encouraging students to participate more actively in their tasks.
- “Change in Cognitive Competence”, which has mean equals to 0.12 and standard deviation equals to 0.08, emphasizes the impact of students’ self-perception of learning ability. When students feel more confident in their cognitive skills, their effort also rises, even though it’s not as significant as emotional changes. The credible interval from 0.01 to 0.23 illustrates the more complex aspect of this impact, demonstrating that although increases in self-perceived competence are important, they often take second place to other factors like interest changes.
- “Change in Difficulty”, which has mean equals to -0.05 and standard deviation equals to 0.07. There is a minor tendency for effort to decrease as tasks are considered as becoming more challenging. When presented with more difficult tasks, some students may feel overwhelmed or frustrated, according to this effect, which is represented by a little negative mean value. There is diversity, too, as the credible interval from -0.20 to 0.10 shows that some kids may respond to challenges with less effort, while others may rise to the challenge and show resilience. This variation emphasizes the various ways people react to perceived difficulties, demonstrating the necessity of individualized intervention strategies.

- “Change in Interest”, which has mean equals to 0.30 and standard deviation equals to 0.11. This predictor emphasizes the significance of intrinsic drive by showing the strongest positive effect. Students’ work significantly increases when they find a task interesting or engaging. Raising interest can be one of the best strategies to improve student engagement, as the credible interval that range from 0.10 to 0.50 illustrates to the effect’s stability. These findings demonstrate that student-centered interventions have the potential to be transformative in addition to having an impact.
- “Email Type: Interesting”, which has mean equals to 0.15 and standard deviation equals to 0.10. Although it has a less consistent effect, email type also shows a favorable impact. The reasonable interval from 0.00 to 0.30 indicates that creating captivating, attention-grabbing emails seems to need more work. Though some students may benefit significantly from this customized communication, others might need more or other tactics to be completely motivated, as indicated by the wide range of responses.
- Intercept, which has mean equals to -1.20 and standard deviation equals to 0.30, provides information on the baseline level of effort. The intrinsic variability in effort that is unrelated to the observed predictors is shown in the negative mean and the credible interval that range from -1.60 to -0.80. This baseline suggests that although interventions are important, there are underlying factors that affect effort in ways that the current model cannot account for, such as contextual, situational, or personal characteristics.

These results demonstrate the complexity of student involvement, where emotional states, motivation, and cognitive factors interact to influence behavioral changes. The wide range of credible intervals also reflects the underlying uncertainty and variability across the dataset, as detailed in Table 2. By combining these observations, this analysis provides an overall understanding of the intervention’s impact on student effort.



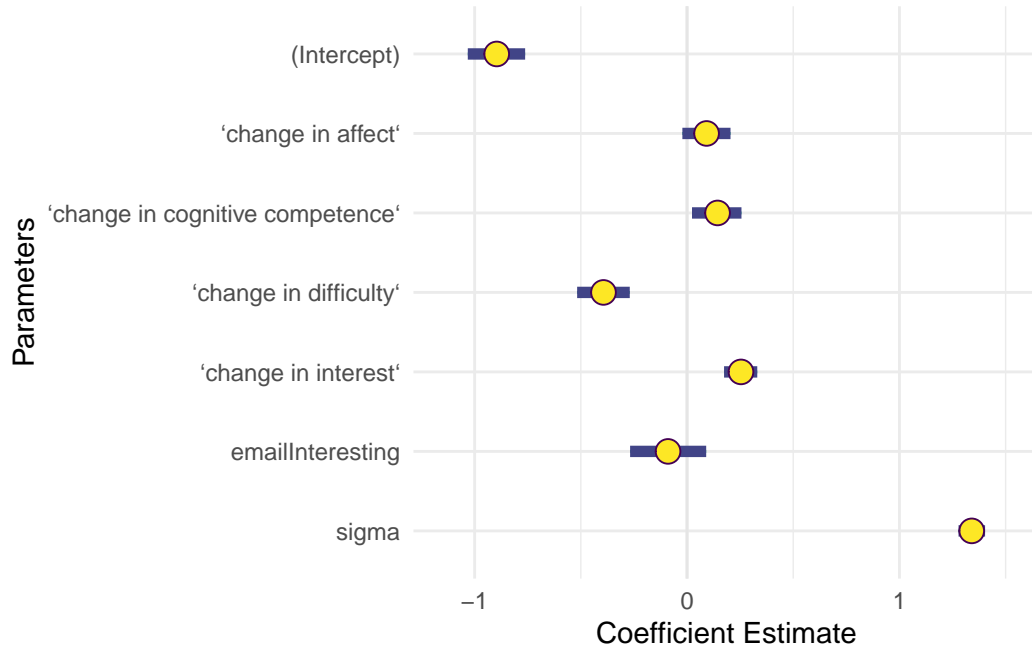


Figure 3: 90% Credible Intervals for Model Coefficients: Visualizing Parameter Estimates and Uncertainty in the Change in Effort Model

The Bayesian multiple linear regression model's estimates are visualized in Figure 3. Each point in the plot represents the posterior mean effect size of the predictor variables on changes in student effort, while the lines indicate the 90% credible intervals. These estimates show several key reasons that the factors influence student engagement:

- The variable “change in interest” has the biggest positive influence, indicating that students' increased interest and intrinsic drive for academic work have the greatest effects on effort improvement. This is supported by its 90% credible interval being entirely above zero, demonstrating how strong this conclusion is. This emphasizes how essential it is to promote interest and individual relevance in learning contexts.
- Besides, the variable “change in cognitive competence” indicates a positive and significant effect, representing the second-largest influence on effort improvement. Students are more inclined to put in effort when they believe their cognitive competence has increased. This is further supported by its 90% credible interval being entirely above zero, reinforcing the idea that motivation is driven by self-confidence in one's ability to learn.
- The “change in difficulty” variable, on the other hand, has a negative impact because its credible interval is completely below zero. This implies that greater task difficulty perceptions result in less effort, emphasizing the necessity of developing interventions that lessen perceived difficulties and create a more approachable learning environment.

- Although the “change in affect” impact is positive, it is not statistically significant because zero is included in its credible interval. This implies that although there is some evidence correlating changes in students’ emotional states to more effort, this relationship is not highly reliable. However, the relationship between mental health and academic success is still an important aspect to consider when designing interventions and conducting additional research.
- With a credible interval that crosses zero and covers a larger range than other factors, the variable “emailInteresting” has a very minor effect. This shows more ambiguity on how email type affects effort, indicating that although communication style might influence engagement, it is probably not a major or stable element.
- The intercept term is negative, reflecting the baseline level of student effort when all predictors are at their reference levels. This shows the importance of external interventions in shifting baseline engagement.

These results offer practical advice on how to prioritize and create educational interventions. In order to improve effort and engagement, strategies that lower the perceived difficulty of working while raising student interest, emotional drive, and cognitive confidence are probably going to have more significant effects.

## 5 Discussion

### 5.1 Emotional and Cognitive Factors as Primary Drivers

Emotional states and cognitive abilities interact to influence students’ effort levels and determine how they can participate in academic activities in a meaningful and sustained way. Important incentives of motivation include emotional states like increased interest, joy, and decreased frustration; when students experience positive engagement, they are more willing to spend time and effort to their studies, especially in the face of difficulties. More than just increasing motivation, these pleasant feelings produce an internal sense of pleasure, which turns learning into a rewarding activity rather than a painful duty.

Cognitive competency also has a similar and equally important role in performance. Students are more persistent and hardworking when they have the self-confidence to take on challenges because they believe in their own abilities and know that their efforts will be successful. This confidence allows individuals to take on challenges bravely, consider them as chances for personal development, and keep going when things become tough.

These emotional and cognitive factors have a mutually beneficial relationship, which emphasizes the necessity of thorough approaches that address both aspects at once. An environment that encourages continuous effort can be established by instructors through encouraging curiosity and excitement, lowering stress and frustration, providing clear and helpful feedback,

and decomposing difficult work into small steps. These techniques not only improve students' emotional and cognitive readiness but also create a solid foundation for their academic development and long-term success, sustaining the idea that persistence and motivation can be promoted by taking a balanced approach to the head and the heart.

## **5.2 Task Difficulty as a Dual-Impact Factor**

Task difficulty has a complex effect on student engagement, depending on how challenging it is. It can either encourage or discourage participation. Moderately challenging tasks encourage students to participate actively because they balance effort and attainability. By stimulating curiosity, keeping interest, and providing a sense of accomplishment after successful completion, these tasks promote intrinsic motivation and strengthen persistence. But when the level of difficulty is higher than what the students can currently handle, they may failure, causing frustration, feelings of inadequacy, and ultimately disengagement. Excessively challenging tasks have the potential to overwhelm students, weakening their confidence and causing avoidance behaviors. Because of these two effects, it is important to adjust the complexity of work to everyone's needs, making sure that challenges are motivating instead of discouraging.

Effective interventions that address this balance include personalized support, which provides targeted feedback and encouragement; instructions, which provides step-by-step guidance to support learning; and adaptive learning systems, which use technology to automatically adjust task difficulty based on student performance in the moment. Effectively adjusting the degree of difficulty allows teachers to create a learning environment that encourages long-term academic persistence, sustains motivation, and promotes student involvement. In addition to supporting students in overcoming challenges, these tactics help them develop the self-confidence and skills that required for sustained academic accomplishment, which ensuring that the advantages of involvement go beyond current task itself.

## **5.3 Personalized Emails as Limited and Uncertain Drivers**

It is still unclear how customized emails affect student participation, and it seems to have less of an effect than often expected. Usually using inspirational language and acknowledging difficulties, these emails aim to create a feeling of connection and relevance. However, their effectiveness varies greatly. Comments that acknowledge achievements, understand hardships or stress may give students a brief boost in motivation. However, such endeavors might have little or even no impact on the behavior, engagement, or academic achievement of a large number of people. Because the psychological processes that drive engagement are so broad and complicated, like emotional resonance and cognitive effort, it is hard to raise effort through these kinds of interventions. Although tailored emails are accessible and inexpensive, they appear to have a limited impact and diminishing returns, particularly when students get the same messages over and over again, which may cause desensitization or apathy. In addition to

increasing significant worries regarding their long-term effectiveness, their contribution to significant academic improvement may have been overestimated. With these unknowns, further investigation is needed to determine which customized emails could actually improve engagement. Using this strategy without such evidence increases the risk of underestimating how hard it is to inspire students and could divert focus from more thorough and efficient ways.

## 5.4 Limitations

A number of limitations should be considered when analyzing the results of this study. First of all, the dataset lacks important outside factors that are known to have significant effects on student engagement like parental support (Yang et al. 2023), peer pressure (Wang et al. 2018), and classroom dynamics (Skinner and Pitzer 2012). An insufficient understanding of the factors influencing shifts in effort might result from their absence. Secondly, the use of self-reported method for emotional states and cognitive competence may lead to response biases such as social desirability and inaccurate self-assessment, which might affect the connections that are observed. Thirdly, the study assumes that predict variables and response variable have linear correlations, which may oversimplify the complex relationships between environmental, cognitive, and emotional factors. In addition, the results are derived from a particular academic and cultural environment, which restricts their generality to other educational environments or demographics. Lastly, because the study only looks at short-term changes in student effort, it doesn't address sustainability over time or the possible long-term effects on academic performance. These limitations emphasize the need for more sophisticated analytical methods, a wider variety of variables, and a study of the persistence of observed changes over time and in different circumstances.

## 5.5 Next Steps

To provide a better understanding of what motivates student effort, firstly, future studies should broaden the analysis to include outside variables such as peer networks, classroom dynamics, and socioeconomic factors. Secondly, to evaluate the long-term impact of effort improvements and determine if the benefits of interventions persist over time, longitudinal studies are required. Besides, by identifying patterns that linear models might overlook, non-linear modeling techniques may offer an improved comprehension of the complex interactions between predictors. Additionally, randomized controlled trials and other experimental designs would help understand the causal relationships between interventions and results. Furthermore, confirming the results in other educational and cultural contexts would increase their generalizability and help guide context-specific engagement-boosting tactics. By addressing these topics, future studies can expand on this study to create more thorough and efficient methods for raising student effort and academic achievement.

## Appendix

### A Additional data details

#### A.1 Data Cleaning Notes

We began by importing the raw dataset using the `read_csv` function from the `tidyverse` package (Wickham et al. 2019) and loading it into R (R Core Team 2023) for cleaning. The first step in the cleaning process involved standardizing column names using the `clean_names()` function from the `janitor` package (Firke 2021), which converted all column names to lower-case, snake\_case format. This ensured consistency and readability throughout the analysis.

Then, we applied strict filtering criteria to retain only valid and complete observations. Variables such as “affect\_pre”, “affect\_post”, “cog\_pre”, “cog\_post”, “difficulty\_pre”, “difficulty\_post”, “interest\_pre”, “interest\_post”, “effort\_pre”, and “effort\_post” were filtered to include only rows with values between 1 and 7, as these represent valid Likert scale responses. Rows with missing values in the “email” column were removed to ensure completeness. Furthermore, we retained only rows where the “open\_email” column equaled 1, ensuring the dataset included only participants who engaged with the intervention by opening the email.

Thirdly, to prepare the data for analysis, we transformed specific variables. The “email” column was converted into a categorical variable using the `factor()` function, with two levels: “Plain” and “Interesting”, corresponding to the styles of the intervention emails. Numeric change variables, including “change\_affect”, “change\_cog”, “change\_difficulty”, “change\_interest”, and “change\_effort”, were rounded to two decimal places using the `round()` function to ensure consistent precision and readability across the dataset.

Next, we selected only the most relevant columns for analysis. These included “email”, which categorized the type of email intervention, and the following change variables: “change in affect”, “change in cognitive competence”, “change in difficulty”, “change in interest”, and “change in effort”. To ensure the dataset was free of incomplete observations, we removed any rows with missing values using the `drop_na()` function.

Additionally, these selected columns were renamed for clarity and interpretability. For example, “change\_affect” was renamed to “change in affect”, “change\_cog” was renamed to “change in cognitive competence”, “change\_difficulty” was renamed to “change in difficulty”, “change\_interest” was renamed to “change in interest”, and “change\_effort” was renamed to “change in effort”. These renaming steps improved the dataset’s alignment with the study’s focus on behavioral outcomes.

Finally, the cleaned dataset was saved in `.parquet` format for efficient storage and processing in future analyses. This thorough cleaning process produced a complete, consistent, and well-structured dataset that is ready for robust analysis of the behavioral interventions and their impact on student engagement and academic outcomes.

## A.2 Additional Tables & Graphs

Figure 4 provides a detailed view of the distributions of key behavioral and performance variables: “change in effort”, “change in affect”, “change in cognitive competence”, “change in difficulty”, and “change in interest”. The histograms, overlaid with density curves, indicate that most variables are centered around zero, suggesting minimal average changes across the sample. However, clear patterns can be seen in the data. The distribution of change in effort exhibits a slight negative skew, with some students experiencing moderate declines in effort post-intervention. Similarly, change in interest also displays a mild negative skew, reflecting a tendency toward reduced engagement in the course material. Change in affect and cognitive competence both follow broadly normal distributions, with their density curves showing symmetrical shapes around the mean, indicating balanced positive and negative changes. In contrast, the distribution of change in difficulty is slightly positively skewed, suggesting that a subset of students perceived increased challenges in the course after the intervention. These patterns underscore the variability in how students responded to the interventions and highlight key areas for further analysis.

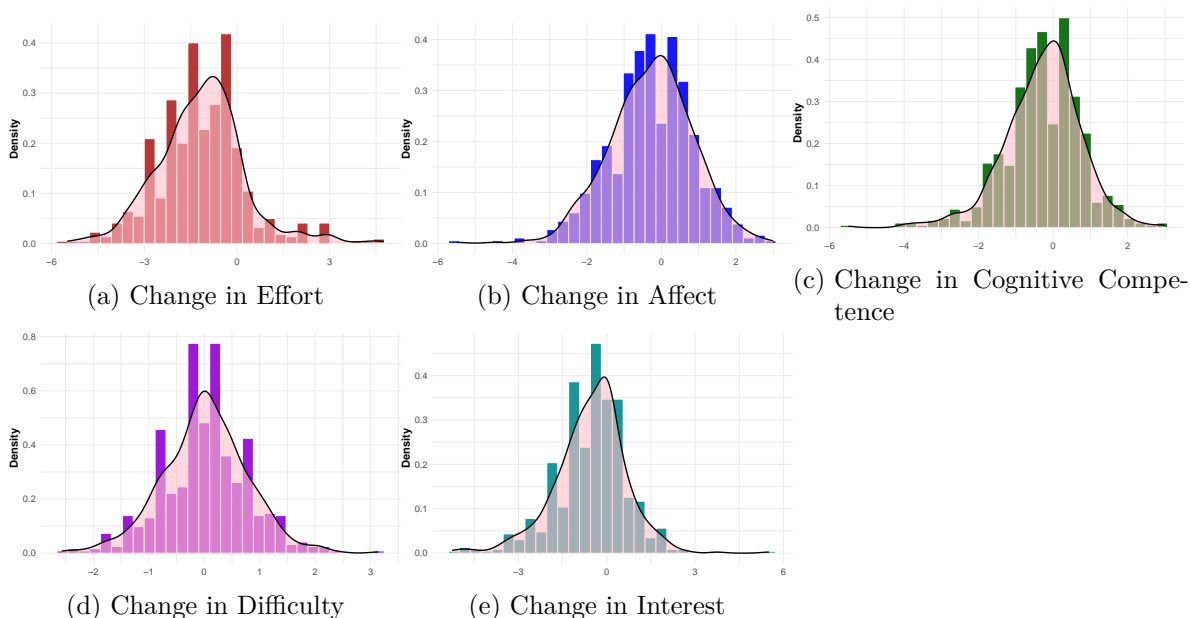


Figure 4: Distributions of Behavioral and Performance Changes: Effort Shows Slight Negative Skew, Interest Exhibits Mild Decline, and Other Variables Display Broad Normal-Like Patterns

Table 3 presents the regression results for the “Change Effort Bayesian Model”, outlining the estimated effects of predictors on changes in effort, along with statistical performance metrics. The intercept is estimated at -0.90 (SE = 0.08), indicating an overall baseline decline in effort when all predictors are at their reference values. Among the predictors, “change in affect” shows a modest positive relationship with effort (0.09, SE = 0.07), suggesting that small improvements in emotional engagement can lead to slight increases in effort. “Change in cognitive competence” has a slightly stronger positive effect (0.14, SE = 0.07), reflecting that enhanced confidence in one’s ability to complete tasks correlates with greater persistence. Conversely, “change in difficulty” exhibits a significant negative relationship (-0.39, SE = 0.08), indicating that greater perceived challenges are strongly associated with reduced effort. The “change in interest” predictor has the largest positive coefficient (0.25, SE = 0.05), emphasizing its important role in motivating students and fostering sustained effort. Finally, the intervention variable “emailInteresting” (comparing “Interesting” emails to “Plain”) has a negligible effect (-0.09, SE = 0.11), suggesting that this type of email intervention does not meaningfully influence changes in effort.

The model performance metrics provide further context for its explanatory power. With a dataset of 622 observations, the  $R^2$  value of 0.117 and an adjusted  $R^2$  of 0.094 indicate that the predictors collectively explain a small fraction of the variance in effort changes. The log-likelihood equals to -1062.052 reflects the model’s goodness of fit, while cross-validation metrics such as ELPD equals to -1068.9, LOOIC equals to 2137.8, and WAIC equals to 2137.8 assess the model’s predictive reliability, with associated uncertainties such as ELPD SE = 23.0 and LOOIC SE = 46.1. The RMSE of 1.33 quantifies the average error in predictions, highlighting variability not captured by the predictors. These results underscore the importance of cognitive and affective factors, such as competence and interest, while pointing to potential limitations in capturing the full complexity of effort-related changes.

Table 3: Full Model Estimates for the Change Effort Bayesian Model, Including Standard Errors for Each Predictor

	Change Effort Bayesian Model
(Intercept)	−0.90 (0.08)
‘change in affect’	0.09 (0.07)
‘change in cognitive competence’	0.14 (0.07)
‘change in difficulty’	−0.39 (0.08)
‘change in interest’	0.25 (0.05)
emailInteresting	−0.09 (0.11)
Num.Obs.	622
R2	0.117
R2 Adj.	0.094
Log.Lik.	−1062.018
ELPD	−1069.0
ELPD s.e.	23.0
LOOIC	2138.0
LOOIC s.e.	46.0
WAIC	2138.0
RMSE	1.33

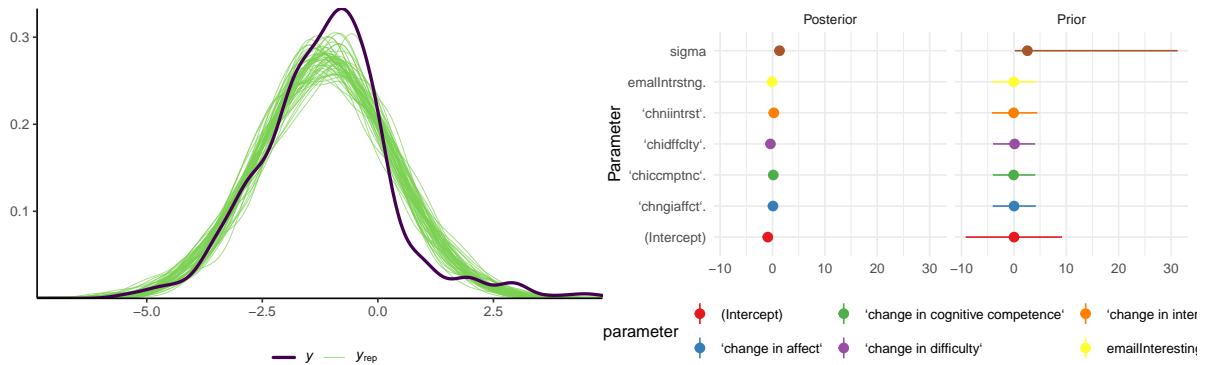


## B Model details

### B.1 Posterior predictive check

In Figure 5a we implement a posterior predictive check. This shows how well the Bayesian multiple linear model aligns with the observed data. The plot compares the actual observed data  $y$  with the replicated data  $y_{\text{rep}}$  generated from the posterior predictive distribution. The lines of posterior predictions closely match the observed data distribution, demonstrating that the model accurately captures the variability and central tendencies of the data. This indicates a good model fit and suggests that the model effectively predicts the outcomes based on the included predictors.

In Figure 5b we compare the posterior with the prior. This shows how the data influenced the parameter estimates. The posterior distributions, represented as narrower and more defined intervals, show significant deviations from the broad and less informative priors, indicating a strong evidence-based influence on the parameter estimates. This comparison emphasizes the importance of incorporating data into Bayesian analysis to refine prior assumptions and draw meaningful, evidence-based conclusions from the model.



(a) Posterior Prediction Check

(b) Comparing the Posterior with the Prior

Figure 5: Examining How the Model Fits, and Is Affected by the Data

## B.2 Diagnostics

Figure 6a is a trace plot. It shows a visual assessment of the convergence of the MCMC algorithm. The plots indicate stable and well-mixed chains for all parameters, with no noticeable trends or drifts over iterations. The consistent fluctuations around the mean suggest that the chains have adequately explored the posterior distribution, providing robust estimates for the model parameters.

Figure 6b is a  $\hat{R}$  plot. It verifies the convergence of the MCMC chains. All parameters show  $\hat{R}$  values close to 1.0, well within the acceptable threshold of 1.05. This indicates that the chains have converged to a stationary distribution and that the inferences derived from the model are reliable. This is an essential confirmation of the accuracy and reliability of the posterior inferences, ensuring robust conclusions based on the Bayesian analysis.

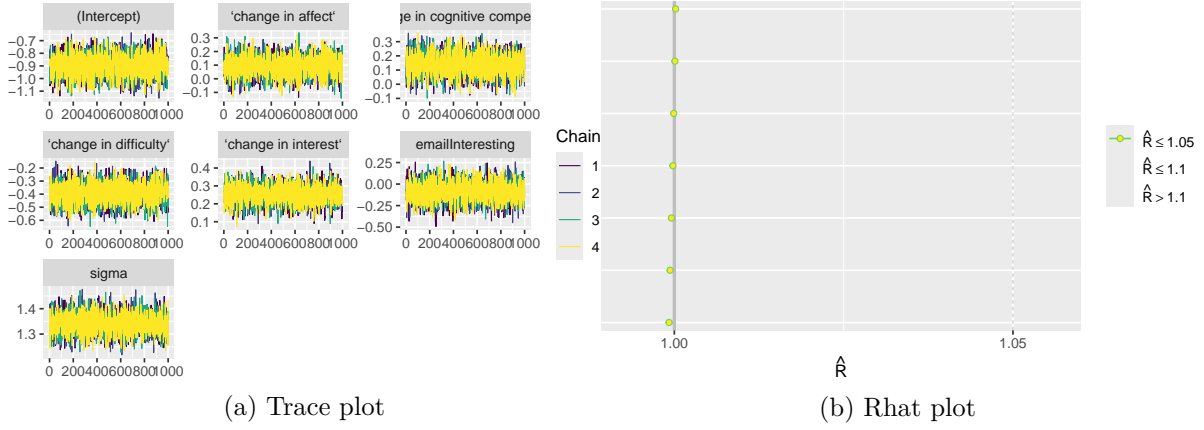


Figure 6: Checking the Convergence of the MCMC Algorithm

Figure 7 visualizes the estimated effects of all model parameters, showcasing their influence on the outcome variable, “Change in Effort”. Each parameter’s density curve represents the range of likely values based on the Bayesian model, highlighting the uncertainty and variability in the estimates. Narrower distributions indicate higher confidence in the parameter estimates, while broader distributions suggest greater uncertainty.

The intercept parameter demonstrates a narrow distribution, reflecting high certainty in its estimate. Positive parameters, such as change in affect, change in cognitive competence, and change in interest, show distributions concentrated around positive values, indicating a significant positive relationship with effort changes. These suggest that improvements in emotional states, cognitive abilities, and intrinsic interest are strongly associated with increased effort. In contrast, change in difficulty has a distribution leaning toward the negative side, indicating that higher perceived task difficulty reduces effort.

The email parameter shows a distribution around zero, suggesting a relatively weak or ambiguous effect of personalized email interventions compared to the baseline. Finally, the sigma parameter represents the variability in the model, providing information about unexplained variations in effort changes. This plot provides a clear, interpretable summary of the model’s findings, emphasizing which predictors have the strongest and most reliable influence on academic effort.

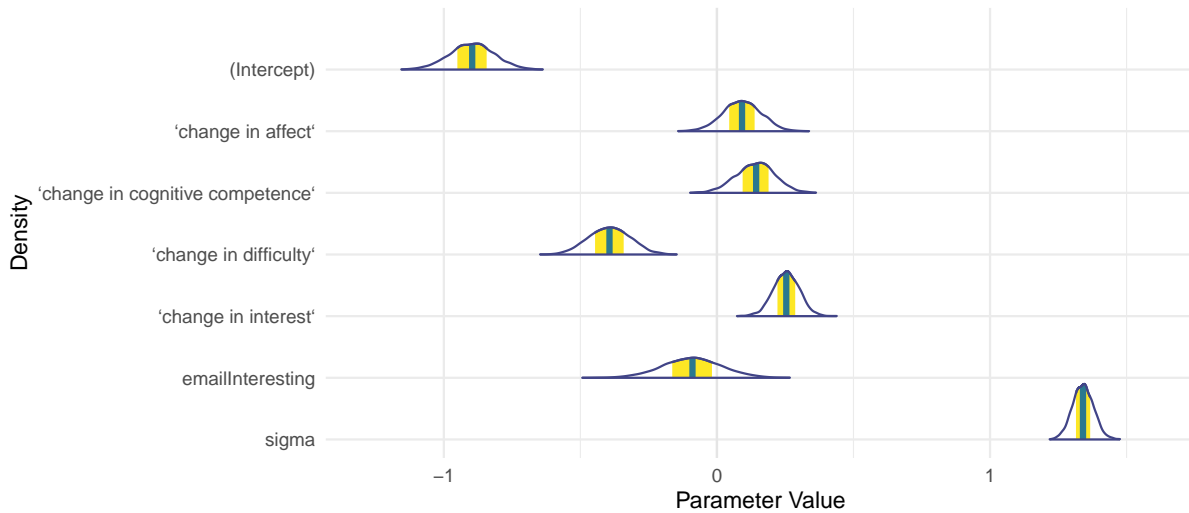


Figure 7: The Posterior Distributions for All Model Parameters

## C Pollster Methodology

The study conducted by Taback and Gibbs utilized a randomized controlled trial to assess whether personalized weekly emails could positively influence students’ attitudes toward statistics (Taback and Gibbs 2023). The methodology provides a rigorous framework for evaluating

interventions designed to foster engagement in educational settings, although the study ultimately found no significant impact. By examining the sampling, recruitment, and implementation strategies used in this study, we can better understand both the strengths and limitations of its approach.

## **C.1 Sampling and Recruitment**

From the study (Taback and Gibbs 2023), we know that this study targeted students from The Practice of Statistics I (STA220H1F), which is a large introductory statistics course at the University of Toronto, and there were over 1,600 students enroll in this class in 2016. Participants were randomized into two groups: the intervention group, which received personalized weekly emails with real-world applications, and the control group, which received generic course-related emails. Stratified randomization was used to balance groups by cumulative grade point average (CGPA) and lecture section, ensuring comparability between groups. Data for stratification, including CGPA and lecture section, were obtained from institutional records.

Besides, participation was voluntary, with students opting into the study by signing a consent form (Taback and Gibbs 2023). They were incentivized with a 1% bonus mark on their final grade for completing both the pre-SATS-36 and post-SATS-36 surveys (Taback and Gibbs 2023). While this design ensured ethical compliance, it introduced potential self-selection bias, as students who opted in may have differed in motivation or attitudes compared to non-participants.

## **C.2 Intervention Design and Data Collection**

As the study (Taback and Gibbs 2023) mentioned, the intervention emails were developed to encourage students to form positive attitudes toward statistics by connecting weekly course material to engaging real-world applications. These emails included curated links to news articles, blog posts, and videos that demonstrated the practical relevance of statistical concepts, often featuring real datasets and contexts that extended beyond the course curriculum. The content highlighted how statistical tools are applied in diverse areas such as public policy, healthcare, and sports. In addition to these external resources, the emails included motivational language that acknowledged common challenges students faced in learning statistics and emphasized persistence. The control group, by contrast, received weekly emails containing only routine course-related information, such as deadlines for assignments, summaries of weekly lecture topics, and general course announcements, without any additional context or motivational content.

Additionally, based on the study conducted by Taback and Gibbs (Taback and Gibbs 2023), we know that students' attitudes were measured using the SATS-36 survey, which is a validated instrument that assesses six dimensions of attitudes toward statistics: affect (emotional

response to statistics), cognitive competence (perceived intellectual ability), value (relevance and usefulness of statistics), difficulty (perceived challenge of the subject), interest (personal enthusiasm for the material), and effort (willingness to invest time and energy). Surveys were administered twice: once at the beginning of the course (pre-test) and once at the end of the course (post-test). These surveys used a “7-point Likert scale”, where higher scores indicated more positive attitudes. Weekly emails were distributed to students using “Mailchimp”, a platform that enabled the automation of email delivery and tracking of engagement metrics, such as open rates and link clicks. This combination of personalized content delivery and systematic survey administration provided a thorough framework for evaluating the potential impact of the intervention on students’ attitudes.

### **C.3 Strengths and Limitations**

A significant strength of this study lies in its randomized controlled trial design, which is widely regarded as the gold standard for evaluating causal relationships. By randomly assigning participants to intervention and control groups, the study effectively minimized confounding variables, allowing for a clearer assessment of the intervention’s direct impact (Etikan and Bala 2017). The stratification by cumulative grade point average (CGPA) further ensured that groups were balanced on key characteristics, increasing the reliability of the results. Additionally, the intervention itself, personalized emails, is both cost-effective and scalable, making it an appealing option for institutions seeking to implement large-scale initiatives to enhance student engagement (Timmons 2003).

However, there are also some notable limitations. The use of a single course as the sampling frame restricts the generalizability of the findings, as the attitudes and behaviors of students in other disciplines or academic settings may differ significantly. Furthermore, while the personalized emails aimed to be engaging, they may not have been equally relevant or appealing to all students, as individual interests and motivational triggers vary widely. This lack of universal resonance could diminish the overall effectiveness of the intervention. Finally, relying on self-reported survey data to measure outcomes introduces potential biases, such as students overstating their attitudes to align with perceived expectations (Donaldson and Grant-Vallone 2002). These limitations suggest that while the study provides useful information, future research should aim to include more diverse contexts and explore additional outcome measures, such as behavioral engagement or performance metrics, to complement self-reported data.

## **D Idealized Methodology**

Based on the strengths and addressing the limitations of the study conducted by Taback and Gibbs (Taback and Gibbs 2023), an idealized methodology for evaluating the impact of personalized emails on student engagement would incorporate several important enhancements. These improvements focus on expanding the sampling frame, customizing the intervention

design to individual needs, and using sophisticated data collection methods to monitor and analyze dynamic changes in attitudes over time. Such a methodology would increase the generalizability, precision, and effectiveness of the intervention.

## **D.1 Expanded Sampling Frame**

To enhance generalizability, the study should expand its sampling frame to include a broader range of students from multiple courses, disciplines, and institutions (Gobo 2004). This approach would allow researchers to examine how personalized emails perform across diverse educational contexts and academic fields. For example, while students in STEM courses may benefit from emails emphasizing technical applications of statistics, students in the humanities might respond better to messages linking statistics to societal issues or historical analysis. Including courses at varying levels, such as introductory, intermediate, and higher, could also provide understanding of how the effectiveness of the intervention differs based on students' familiarity with the subject matter.

Sampling across different institution types, such as community colleges, research universities, liberal arts colleges, and online-only institutions, would further broaden the scope of the findings. Institutional factors like available resources, class sizes, and student-to-instructor ratios could significantly impact how students perceive and engage with personalized emails. For example, students in large, resource-limited public universities might benefit more from personalized communication than those in smaller institutions with more individualized support systems.

Stratified sampling should be used to ensure the sample reflects the diversity of the larger student population. Key strata could include:

- **Demographics:** Age, gender, ethnicity, and socioeconomic background. For instance, first-generation college students or those from underrepresented groups might show unique responses to personalized messaging.
- **Academic Characteristics:** Stratification by CGPA, program type, such as life sciences, engineering, arts, and prior experience with statistics could ensure representation of varying academic abilities and interests.
- **Course Delivery Modes:** Including students from in-person, hybrid, and fully online courses would provide information about how the mode of instruction interacts with the intervention. For example, online learners might rely more heavily on email communication and thus engage differently with the intervention.

Such a detailed sampling strategy would allow researchers to perform subgroup analyses, identifying specific populations that derive the greatest benefits from personalized email interventions. These analyses would provide useful information for modifying communication strategies to meet the various requirements of modern student populations.

## D.2 Customizing the Intervention Design

The intervention should be designed to adapt to each student's needs using real-time data like grades, engagement with course materials, and feedback. Struggling students will get emails with extra resources like tutorials, step-by-step guides, or tips for tackling challenging topics, along with encouragement to keep them motivated. High-performing students will receive content on complex statistical applications, like in data science or healthcare, and suggestions for related career paths or further studies to keep them engaged.

Emails could also be personalized based on how students interact with them. For example, students who often click on career-related links could get more real-world examples of how statistics applies in professional settings. Those focused on conceptual material might get emails diving deeper into theory.

Besides, adding quizzes, polls, or feedback forms to the emails could make them more interactive and provide useful data. For example, if a student doesn't do well on a quiz, follow-up emails could provide extra help on that topic. This personalized approach ensures the emails stay relevant and helpful, improving their impact on students' learning and engagement.

## D.3 Data Collection

To accurately measure the intervention's impact, data should be collected at multiple points during the semester, providing a detailed view of how students' attitudes, engagement, and perceptions evolve over time. This approach moves beyond the limitations of relying only on pre-course and post-course surveys. Periodic surveys every 2 to 3 weeks can capture short-term changes in students' attitudes, interest, and confidence. These surveys could include Likert-scale questions, such as "How confident do you feel about applying the statistics concepts learned this week?" or "How relevant were this week's emails to your learning?" Open-ended questions might ask students to describe what they found helpful or what could be improved. This feedback would help identify trends, such as increased confidence in certain topics or recurring challenges in others.

Besides, behavioral engagement metrics should be tracked to understand how students interact with the emails. Metrics like open rates, link click-through rates, and time spent on linked resources can provide an objective measure of how engaging the emails are. For example, if click-through rates on supplemental resources decrease mid-semester, it indicates the decline in interest and prompts adjustments to the email format. Completion rates for interactive elements, such as embedded quizzes or polls, can also highlight which types of content are most engaging.

Moreover, qualitative feedback through focus groups or one-on-one interviews also add depth to the analysis. These discussions explore students' experiences in more detail, such as which email features they found motivating or whether the resources provided met their needs. For

example, a student might share that they preferred visual explanations in videos over lengthy text or found motivational content particularly encouraging during challenging weeks.

Finally, combining these approaches, such as periodic surveys, behavioral metrics, and qualitative feedback, would give a detailed and dynamic understanding of the intervention's effectiveness. This data would not only assess its impact but also allow for real-time adjustments to maximize engagement and relevance throughout the semester.

#### **D.4 Simulation and Feasibility**

Simulations should evaluate the intervention's effectiveness under varying conditions to refine its design and address challenges before full-scale implementation. Key areas include engagement variability, modeling scenarios where email interaction rates differ, such as high engagement at 80% versus low engagement at 30 to 50%, to determine the impact on overall outcomes and identify strategies to improve participation. Content effectiveness should be tested by comparing formats such as text-only versus multimedia, and focuses, like motivational versus resource-driven content, to identify which approach best supports engagement and learning. Subgroup analysis should predict how different demographic and academic groups, such as first-year students or online learners, respond to the intervention, ensuring that it benefits diverse populations. Additionally, simulations should assess logistical feasibility, such as managing personalized email delivery to large cohorts and integrating real-time engagement data. These focused simulations will ensure the intervention is effective, scalable, and adaptable across various educational contexts.

#### **D.5 Conclusion**

The idealized methodology enhances the original study conducted by Taback and Gibbs (Taback and Gibbs 2023) by addressing its limitations and refining its approach. Expanding the sampling frame ensures generalizability across diverse student populations and institutions. Customizing email content to individual needs, supported by real-time data, increases engagement and relevance. Innovated data collection techniques, including periodic surveys and behavioral metrics, provide specific details about the intervention's impact. Simulations ensure feasibility and scalability, refining strategies before implementation. This methodology provides a solid foundation for evaluating and enhancing personalized email strategies in education.



## References

- Akbari, Omidullah, and Javed Sahibzada. 2020. "Students' Self-Confidence and Its Impacts on Their Learning Process." *American International Journal of Social Science Research* 5 (1): 1–15.
- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Apache Arrow. 2021. *arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Chevalier, Arnaud, Peter Dolton, and Melanie Lührmann. 2014. '*Making It Count*': Evidence from a Field Study on Assessment Rules, Study Incentives and Student Performance. SSRN.
- Donaldson, Stewart I, and Elisa J Grant-Vallone. 2002. "Understanding Self-Report Bias in Organizational Behavior Research." *Journal of Business and Psychology* 17: 245–60.
- Etikan, Ilker, and Kabiru Bala. 2017. "Sampling and Sampling Methods." *Biometrics & Biostatistics International Journal* 5 (6): 00149.
- Feng, Wenzheng, Jie Tang, and Tracy Xiao Liu. 2019. "Understanding Dropouts in MOOCs." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:517–24. 01.
- Firke, Sam. 2021. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelman, Gabriel, Jonah Gabry, et al. 2021. *bayesplot: Plotting for Bayesian Models*. <https://mc-stan.org/bayesplot>.
- Gobo, Giampietro. 2004. "Sampling, Representativeness and Generalizability." *Qualitative Research Practice* 405: 426.
- Goes, Paulo B, Chenhui Guo, and Mingfeng Lin. 2016. "Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange." *Information Systems Research* 27 (3): 497–516.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." <https://mc-stan.org/rstanarm/>.
- Kahu, Ella R, and Karen Nelson. 2018. "Student Engagement in the Educational Interface: Understanding the Mechanisms of Student Success." *Higher Education Research & Development* 37 (1): 58–71.
- Mac Iver, Douglas J, Deborah J Stipek, and Denise H Daniels. 1991. "Explaining Within-Semester Changes in Student Effort in Junior High School and Senior High School Courses." *Journal of Educational Psychology* 83 (2): 201.
- Müller, Kirill, and Lorenz Walthert. 2024. *styler: Non-Invasive Pretty Printing of R Code*. <https://CRAN.R-project.org/package=styler>.
- Pajares, Frank. 2001. "Toward a Positive Psychology of Academic Motivation." *The Journal of Educational Research* 95 (1): 27–35.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Skinner, Ellen A, and Jennifer R Pitzer. 2012. "Developmental Dynamics of Student Engagement, Coping, and Everyday Resilience." In *Handbook of Research on Student Engagement*,

- 21–44. Springer.
- Taback, Nathan, and Alison Gibbs. 2022. “nudgedata.” <https://doi.org/10.5281/zenodo.1234>.
- Taback, Nathan, and Alison L Gibbs. 2023. “A Randomized Study to Evaluate the Effect of a Nudge via Weekly e-Mails on Students’ Attitudes Toward Statistics.” *Journal of Statistics and Data Science Education* 31 (2): 134–43.
- Timmons, Thomas David. 2003. “Email as an Educational Tool: A Perspective from an Art Classroom.” Master’s thesis, The Ohio State University.
- Trautwein, Ulrich, Oliver Lüdtke, Nicole Nagy, Anna Lenski, Alois Niggli, and Inge Schnyder. 2015. “Using Individual Interest and Conscientiousness to Predict Academic Effort: Additive, Synergistic, or Compensatory Effects?” *Journal of Personality and Social Psychology* 109 (1): 142.
- Wang, Ming-Te, and Jessica Degol. 2014. “Staying Engaged: Knowledge and Research Needs in Student Engagement.” *Child Development Perspectives* 8 (3): 137–43.
- Wang, Ming-Te, Noona Kiuru, Jessica L Degol, and Katariina Salmela-Aro. 2018. “Friends, Academic Achievement, and School Engagement During Adolescence: A Social Network Approach to Peer Influence and Selection Effects.” *Learning and Instruction* 58: 148–60.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Winne, Philip H, and John C Nesbit. 2010. “The Psychology of Academic Achievement.” *Annual Review of Psychology* 61 (1): 653–78.
- Xie, Yihui. 2021. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.
- Yang, Dong, Peng Chen, Kai Wang, Zhuoran Li, Chen Zhang, and Ronghuai Huang. 2023. “Parental Involvement and Student Engagement: A Review of the Literature.” *Sustainability* 15 (7): 5859.