The Role of Information Frictions in Study Habit Choice and Academic Performance*

Agustina Affonso†
Universidad ORT Uruguay

Esteban M. Aucejo‡
Arizona State University

Stephanie El Khoury§
Arizona State University

Tomas Larroucau¶
Arizona State University

June 30, 2026

Abstract

We study whether information frictions about the relative effectiveness of study methods distort how college students allocate effort. Standard models of learning typically treat effort as a single input, with outcomes determined by its level rather than its composition. We relax this assumption and study effort as an allocation across methods with heterogeneous effectiveness. Combining panel surveys, administrative records, digital activity logs, and a randomized information intervention with over 2,000 undergraduates, we ask whether imperfect information about relative effectiveness leads to systematic misallocation and quantify the resulting consequences for performance. At baseline, students hold strikingly different beliefs about the effectiveness of common study methods, and many devote substantial time to passive strategies, such as rereading, despite evidence that active retrieval is more effective. These choices explain lower performance and larger errors in self-assessment. To distinguish biased beliefs from efficient heterogeneity, we randomize treated students into a general-feedback arm or a personalized-feedback arm that benchmarks their habits against peers and quantifies the gains from reallocation. General feedback has little effect. Personalized feedback widens the perceived effectiveness gap between active and passive methods by 42 percent, shifts time toward more effective practices, and raises subsequent exam scores by 0.05–0.08 standard deviations. To quantify the mechanisms linking belief updating to behavior, we develop and estimate a dynamic model of multidimensional study effort with learning-by-doing and convex allocation costs. We find that, for students with the weakest baseline beliefs, the personalized treatment is equivalent to a 16.5 percent reduction in the cost of adopting more effective methods. Together, these results identify information frictions about the effectiveness of multidimensional effort as a relevant determinant of academic performance.

*We thank seminar and conference participants at Arizona State University, University of Toronto, UC San Diego, Texas A&M, Universidad ORT Uruguay, University of Pennsylvania, and the Sixteenth International Workshop on Applied Economics of Education.

†Department of Economics, Universidad ORT Uruguay. affonso@ort.edu.uy

‡Department of Economics, Arizona State University, CEP, & NBER. eaucejo@asu.edu

§University College, Arizona State University. selkhou1@asu.edu

¶Department of Economics, Arizona State University. tlarrouc@asu.edu

“ (...) I’ve always simply memorized the material before an exam. Due to being in upper-level economics, this no longer slides. I really struggle with applying the information. Does anyone have any study tips to ace these tests?”

— ASU College Student on Reddit r/ASU, “How to study?”

1 Introduction

Course grades are among the most consequential measures of student progress in college. They determine credit accumulation, academic standing, major progression, scholarship eligibility, and shape students’ own beliefs about whether they can succeed. Poor performance early in college is also a leading predictor of longer-run setbacks. Students who struggle in their first terms are more likely to delay progress, switch paths, or leave school altogether (Stinebrickner and Stinebrickner, 2014).[1] How students arrive at those grades, however, depends on choices they often make without a clear sense of what works. A growing body of research identifies information frictions as an important determinant of students’ educational choices and outcomes (Altonji, 1993; Arcidiacono, 2004; Stange, 2012; Stinebrickner and Stinebrickner, 2012; Bettinger and Baker, 2014; Stinebrickner and Stinebrickner, 2014; Hoxby and Turner, 2015; Thomas, 2019; Arcidiacono et al., 2020; Larroucau and Rios, 2022; Aucejo et al., 2025; Larroucau et al., 2025).[2] Yet one friction with direct implications for academic performance remains understudied: biased beliefs about the effectiveness of different study habits—the concrete techniques and routines students use to learn.

This friction matters because study habits shape how effectively students learn, not merely how long they study. Much of the economics literature has treated effort primarily as a question of quantity—how much time students spend studying, working, attending class, or completing assignments—with Delavande et al. (2023) a notable exception.[3] But studying also has a quality dimension. Holding total study time fixed, students may allocate their effort across very different learning techniques: *how* they study, such as *retrieval practice*—explaining and applying concepts—rather than *rereading*; and *when* they study,

[1] Completion rates in U.S. higher education remain low: only 64 percent of first-time, full-time students at four-year institutions graduate within six years. National Center for Education Statistics. (2022). *Undergraduate Retention and Graduation Rates.* Condition of Education. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/ctr.

[2] Financial constraints, time pressures, and insufficient preparation also contribute to academic difficulties; see, among others, Barr (2019); Denning and Jones (2021); Black et al. (2020); Denning (2019, 2017); Stinebrickner and Stinebrickner (2003); Beffy et al. (2010); Kalenkoski and Pabilonia (2010); Scott-Clayton (2012); Darolia (2014); Neyt et al. (2019); Le Barbanchon et al. (2020); Aucejo and James (2019); Aucejo and Tobin (2021); Ketchen Lipson et al. (2019); Eisenberg et al. (2009).

[3] Delavande et al. (2023) randomize an intervention that promotes a growth mindset, framing the brain as malleable and encouraging the adoption of active learning strategies.

such as *spaced, distributed practice* rather than *last-minute cramming.* Economics research has rarely incorporated this allocation margin or measured students' beliefs about the relative effectiveness of these options. This gap is relevant because educational psychology consistently finds that these choices materially affect learning (Hartwig and Dunlosky, 2012). Substituting rereading for retrieval practice, for instance, can create an *illusion of mastery*: rereading feels fluent but yields weak retention, whereas retrieval practice strengthens learning and reveals knowledge gaps (Dunlosky et al., 2013). When biased beliefs about the effectiveness of each habit are combined with the costs of effort allocation, students can become locked into suboptimal methods, distorting the perceived payoff of studying, lowering performance, and amplifying vulnerability to academic setbacks.[4] Consistent with the importance of this channel, Affonso et al. (2025) construct survey-based indices of barriers to student success—work-life balance, social isolation, financial stress, and study habits—and find that poor study habits predict academic performance about as strongly as the other three, even conditional on prior academic skills.

This paper asks whether imperfect information about the relative effectiveness of alternative study methods leads students to misallocate effort, and quantifies the consequences for academic performance. Rather than treating studying as a single input, we model it as an allocation problem across distinct habits, where choices reflect both students' beliefs about each habit's effectiveness and the costs of allocating effort across them.[5] We analyze this question using panel survey data linked to administrative records for more than 2,000 undergraduates at Arizona State University (ASU). To capture the dynamics of student decision-making, we administered surveys at the start and end of the academic term, eliciting students' study-habit allocations, beliefs about the effectiveness of different habits, and expectations about academic performance. Midway through the semester, immediately after the first exam, we implemented a randomized information intervention designed to distinguish efficient heterogeneity in study choices from misallocation driven by information frictions. Students were assigned to one of three groups: a control group, a general feedback group receiving evidence-based recommendations, or a personalized feedback

[4]Students often begin college with overly optimistic academic expectations (Nowell and Alston, 2007; Zafar, 2011; Arcidiacono et al., 2012), which heightens their susceptibility to negative academic surprises.

[5]The relationship between study time and academic performance has been studied extensively. Using detailed time-use data and exogenous variation, Stinebrickner and Stinebrickner (2003, 2004, 2008) show that additional study time meaningfully improves college performance. We view our approach as complementary to these findings: rather than estimating the return to total study time, we focus on whether students allocate a given amount of study time across more or less effective methods.

group that paired those recommendations with peer comparisons and tailored suggestions.[6] By studying the joint response of beliefs, behavior, and achievement, we can assess whether baseline study allocations were distorted by information frictions. If students revise their beliefs, shift time toward more effective habits, and subsequently perform better, then their initial allocations cannot be rationalized solely by efficient heterogeneity in preferences or constraints. Instead, the evidence implies that biased beliefs about study-habit effectiveness contributed to the misallocation of effort. Finally, we develop and estimate a dynamic model of study-habit choice that links information signals, belief updating, learning-by-doing, and costly effort allocation. The model allows us to decompose the reduced-form response into belief and cost components, quantify how strongly costs limit the pass-through from information to behavior, and express the personalized treatment as a cost-equivalent reduction in the adoption of retrieval practice.

We first document substantial heterogeneity in students' beliefs about the effectiveness of specific study habits, as well as in their broader perceptions of the academic payoff to study time. Reported behavior mirrors this variation: students rely on a wide range of practices, and many struggle to identify the most effective ones, devoting considerable time to strategies that the educational psychology literature flags as less productive, such as rereading and cramming (i.e., concentrating study effort immediately before an exam). Students who report relying more heavily on these less effective habits tend to perform worse academically and to make larger errors when predicting their own outcomes. We also document a tight link between reported study habits and beliefs about the payoff to overall study time, suggesting that misperceptions about the effectiveness of particular habits may distort not only how students allocate effort, but also how they perceive the value of effort itself.

Building on these patterns, we move from description to inference by relying on our information intervention to assess the extent to which baseline choices were distorted by biased beliefs. In fact, we do find that information interventions meaningfully improve students' ability to distinguish between more and less effective study habits.[7] In the personalized feedback group, the perceived effectiveness gap between the more effective habit (retrieval practice) and the less effective habit (rereading) widens by 42% relative to the control group; this is followed by higher perceived payoffs to overall study time, a shift away from

[6]Rury and Carrell (2023) study a randomized information intervention in a large introductory economics course that provided students with feedback on the overall returns to study time; they find that it increased total effort but had no significant effect on exam performance. Similarly, Oreopoulos and Petronijevic (2023) evaluate online and follow-up coaching interventions focused on goal setting, mindset, and time management, finding improvements in weekly study hours and the likelihood of meeting with a tutor or instructor, but no statistically detectable gains in course grades. Instead of targeting the level of effort, we examine students' beliefs about the relative effectiveness of specific study strategies.

[7]We do not find a detectable updating in beliefs about cramming's effectiveness (i.e., when to study); accordingly, our main mechanism/modeling emphasis is on the retrieval-versus-rereading margin.

rereading, and greater effort devoted to active practice (class homework).[8] These behavioral shifts were accompanied by an average performance gain of 0.051 to 0.085 standard deviations. This joint adjustment of beliefs, behavior, and performance provides strong evidence that information frictions regarding the effectiveness of study habits generate misallocation of students' effort, with consequences for academic achievement.[9] In contrast, generalized feedback without personalized comparisons produced smaller and statistically weaker effects, underscoring the importance of how information is delivered.

Finally, to recover the primitives linking students' belief updates to changes in behavior, we develop and estimate a dynamic learning-by-doing model. In this framework, students allocate study time across habits based on their perceived effectiveness, update these beliefs after receiving information signals, and reoptimize their effort allocation subject to convex costs. The model formalizes a key wedge between information and behavior: even when students learn which methods are more effective, a different allocation of effort may be more costly. Using the estimated model, we conduct counterfactual decompositions showing that (i) effort allocation costs can substantially dampen the pass-through from belief updating to behavior; (ii) personalized information can nevertheless shift students toward retrieval practice, with effects that, depending on initial beliefs, are equivalent to reducing the cost of adopting retrieval methods by up to 16.5%; and (iii) gaps in beliefs about study-habit effectiveness are unlikely to close through learning-by-doing alone in the absence of targeted information, helping explain why many students may fail to "learn by themselves." These results suggest that the impact of personalized information could be substantially larger when paired with policies or technologies that lower the cost of acting on that information. Tools such as AI-based study assistants, structured practice platforms, or targeted coaching may make effective strategies easier to implement, thereby expanding the scope for interventions that combine personalized feedback with reductions in the cost of behavioral adjustment.

Taken together, our results suggest that the effectiveness of low-cost information interventions depends critically on their design: generic information may have limited effects, whereas personalized feedback targeted to specific margins of study behavior can improve outcomes by helping students recognize and act on more effective uses of their study time.[10] In addition, the intervention features also speak to

[8]In a similar vein, Ersoy (2023) shows that information interventions can influence students' beliefs about the effectiveness of effort and change their effort (i.e., students completed more lessons in Duolingo in response to an increase in the perceived effectiveness of effort).

[9]We use "misallocation" in a narrow behavioral sense: study-time allocations differ from those students would choose under correct beliefs about relative habit effectiveness. This is not a welfare claim, as information frictions may coexist with other constraints that students face.

[10]Oreopoulos and Petronijevic (2023) provides important lessons on when information interventions fail to improve outcomes in higher education.

external validity: the treatment was embedded in students' actual high-stakes courses at a large, broadly non-selective public university serving many students with diverse backgrounds; it relies on personalized but automatically generated feedback rather than intensive staff-delivered services; and it targets a non-rival behavior (the adoption of more effective study practices) that can plausibly be expanded without the capacity constraints that limit many education interventions.

The remainder of the paper is organized as follows. Section 2 describes the research setting and data. Section 3 provides descriptive evidence on the baseline heterogeneity in student beliefs and study habits. Section 4 outlines the experimental design and the implementation of the information treatments, while Section 5 reports the reduced-form effects of the intervention. To explore the underlying mechanisms, Section 6 develops a dynamic model of study-habit choice. Sections 7 and 8 discuss the model's identification and present the estimation results, respectively. Section 9 conducts counterfactual decompositions using the estimated model. Section 10 discusses the scalability and external validity of the intervention. Section 11 concludes.

2 Data

Our study examines students at ASU, the largest public university in the United States. As a broadly non-selective institution that admits most applicants, ASU serves a substantial share of students who are at the margin of college enrollment, a population of particular interest for efforts to improve educational outcomes.

We draw on four data sources: i) surveys of students enrolled in intermediate economics courses, our primary data source; ii) academic performance measures, including exam scores and final course grades; iii) administrative records, such as term-by-term credit hours and declared field of study;[11] and iv) digital weekly activity logs from ASU's learning platform "CANVAS," capturing the extent to which students engage in learning activities associated with retrieval practice (e.g., whether a student attempts practice

[11] We use administrative records from ASU for the 2023 academic year. These records are available at both the aggregate and individual levels. The aggregate statistics, provided by the University Office of Institutional Analysis, cover the entire ASU student body as well as the subset of business and economics majors. They are employed to benchmark the survey sample against the broader population (Table 1) and include demographic and background measures such as first-generation status, race and ethnicity, gender, high school GPA, majors, and household income. The individual-level data are linked to survey respondents who provided consent and for whom identifiers could be matched. These records are primarily used as control variables in the analysis, capturing academic progress and institutional engagement through measures such as total semesters enrolled at ASU (Fall and Spring), total credit hours completed, and cohort type (new student or transfer).

tests on the platform and homework performance).[12]

2.1 Sample Selection

We administered surveys at the beginning and end of the Fall 2022 and Spring 2023 semesters across 34 classes, many of which are high-stakes courses that economics and business students must complete with a minimum grade of C to satisfy graduation requirements.[13] A total of 2,523 students completed either the first survey only or both surveys. After excluding students who did not provide consent or had no valid administrative matches, the remaining 2,488 students comprise our *All Students* sample.[14]

From this group, we exclude 158 students who were stratification misfits. This yielded 2,328 students in the *Randomized* sample, the subset used for treatment assignment. For the analysis that involves using data from both surveys, we further excluded students who failed the survey attention checks or were missing key analytic variables, leaving 2,225 students.[15] Of these, 1,743 completed both surveys, forming the *Both Surveys* sample. Finally, usable CANVAS data are available for 1,683 of these students. Appendix Section A.1 provides the complete sample construction and attrition details.

2.2 Descriptive Statistics

Table 1 reports summary statistics for our survey sample and benchmarks them against the broader university populations. Columns (1) and (2) describe the survey sample, presenting the *All Students* sample and the *Randomized* sample, respectively. Columns (3) and (4) provide context, showing corresponding statistics for all students majoring in business and economics at ASU and the overall ASU student population.[16]

In Column (1), 31.5% of surveyed students are first-generation, 4.1% are Black, 14.1% are Hispanic, and 45.9% are women. The mean household income is $125,544, the average high school GPA is 3.55, and 77.3% of respondents major in business or economics. Column (2) indicates that the *Randomized* sample

[12]CANVAS is a learning management system that ASU uses for its courses. On this platform, students can receive messages from their teachers and teaching assistants in their classes. CANVAS data includes, among other information, submissions of problem sets and their grades, and attempts on practice tests pre- and post-intervention.

[13]Full survey available here.

[14]Three classes had no participation, mainly due to small enrollments. In this study, a class refers to all sections of a course taught by a given instructor and managed through a single CANVAS page. Instructors may group multiple sections of the same course on a single CANVAS course page; throughout, we treat that CANVAS page as the relevant class-level unit. In our dataset, only five classes include multiple sections: three classes have two sections each, and two have three. The average class size is 80 students, with a standard deviation of 136.

[15]Two-thirds of the randomized sample came from long-term (15-week) semesters, primarily in-person courses; the remainder were enrolled in short-term (7.5-week) semesters, primarily online. Of the 2,328 students, 960 were in online courses.

[16]The Department of Economics at ASU is part of the W. P. Carey School of Business.

Table 1: Summary Statistics

	Survey Samples		Bus. & Econ. Majors Pop.	ASU Population
	All Students (1)	Randomized (2)	All (3)	All (4)
First Generation	0.315	0.307	0.216	0.263
	(0.465)	(0.461)		
Black	0.041	0.040	0.043	0.056
	(0.198)	(0.197)		
Hispanic	0.141	0.142	0.221	0.248
	(0.348)	(0.349)		
Income (USD)	125,544	126,777	142,735	103,135
	(97,886)	(98,270)		
Female	0.459	0.457	0.445	0.543
	(0.498)	(0.498)		
High School GPA	3.548	3.552	3.42	3.39
	(0.439)	(0.437)		
Business and Economics Majors	0.773	0.773		0.198
	(0.419)	(0.419)		
Observations	2,488	2,328	22,655	114,484

Note: Columns (1)–(2) summarize different samples of students enrolled in intermediate economics courses: (1) All students completing the initial survey; (2) the *Randomized* subset (eligible for randomization). Column (3) reports population statistics for ASU students majoring in business and economics during 2023. Column (4) presents the same statistics for the full ASU student population. "First-generation" indicates neither parent completed college. "Income (USD)" is annual household income in U.S. dollars. High school GPA is measured on a 4.0 scale. Proportions are reported as fractions between 0 and 1. Standard deviations for columns (1)-(2) are in parentheses. Observations for each column appear in the final row.

closely mirrors the *All Students* survey sample.

Relative to the broader population of business and economics majors (Column (3)), our sample has a higher share of first-generation students and a lower share of Hispanic students, while other demographic characteristics are broadly similar. Compared to the overall ASU population (Column (4)), our sample has a lower share of Hispanic students, a slightly lower share of women, and a higher average household income.[17] Finally, Appendix Table C.1 presents summary statistics for the *Both surveys* sample, showing that its characteristics closely resemble those of the *All Students* and *Randomized* samples.[18]

[17] Aucejo et al. (2025) compares ASU students to the typical undergraduate at other large flagship universities (defined as the largest public university in each state). While ASU's gender composition is broadly similar to these institutions, its racial/ethnic composition differs: ASU enrolls a substantially higher share of Hispanic students (25% vs. 12%), a slightly lower share of Black students, and a markedly lower share of White students than the average flagship university. Finally, consistent with ASU's access-oriented admissions, ASU admits the vast majority of applicants (an acceptance rate of about 90% for Fall 2023) and has lower ACT percentile scores than the average flagship, implying that ASU undergraduates are closer to the margin of college enrollment.

[18] Detailed descriptive statistics for samples of students that are online and in-person are presented in Appendix Table C.2.

2.3 Survey Content and Administration

We fielded the baseline survey at the start of Fall 2022 and Spring 2023, before any exams. Mid-semester, we implemented the information intervention. The endline survey was administered during the final week of each course participating in the study. Students were told that the study examined study habits; each survey took approximately 15–25 minutes to complete. Instructors offered extra credit for completion at their discretion.

Survey content. The baseline survey collected students' background characteristics, academic expectations, study-habit allocations, and beliefs about the productivity of studying. We elicited study-habit allocations through a standardized exam-preparation scenario in which students reported how they would allocate study time across alternative methods and preparation windows. We elicited habit-specific beliefs by asking students to rate the effectiveness of each study method, and we elicited beliefs about the overall payoff to study time using vignettes that asked students to predict expected score gains per hour of study. The endline survey re-elicited the same key domains, allowing us to measure changes in beliefs, study allocations, and expectations after the intervention. Throughout, we use "perceived effectiveness" to refer to habit-specific beliefs and "payoff to overall study time" to refer to vignette-based score-per-study-hour beliefs.

Focal dimensions of study behavior. Our measures emphasize two dimensions highlighted in educational psychology: (i) retrieval versus rereading, which captures the study method used to engage with the material (the *how*), and (ii) distributed study versus cramming, which captures the temporal distribution of effort (the *when*). Retrieval practices involve active recall, such as self-quizzing, generating questions, and applying material. Rereading reflects more passive review, such as highlighting or skimming. We define *cramming* as intensive study within the last two days before an exam and *distributed study* as study occurring three or more days prior to the exam. These responses allow us to compute each student's reported share of study effort devoted to retrieval, rereading, cramming, and distributed study.

Ability proxy. Because a full IQ test was infeasible, we used a brief, relative ability measure. Students viewed a sample Raven item (Appendix Figure D.1) and rated their ability relative to classmates in four categories: top 25%, top 50%, bottom 50%, or bottom 25%. This framing captures self-perceived cognitive ability relative to peers in the class.[19]

Longitudinal design. Comparing baseline and endline responses provides a short panel around the

[19] Administrative data provide additional information on student achievement.

information intervention, enabling us to track changes in beliefs, reported study-habit allocations, and expectations over the semester.

3 Stylized Facts on Students' Beliefs about Achievement Production

Students' perceptions of the effectiveness of different study inputs play a central role in how they allocate limited study time. Undervaluing the effectiveness of high-yield habits can lead to misallocated effort and, ultimately, lower academic performance. To fix ideas, we first outline a simple framework that defines our main objects of interest. We then examine two belief objects: students' beliefs about the payoff to overall study time (our vignette-based score-per-hour measure) and their beliefs about the relative effectiveness of specific study habits highlighted in the learning sciences literature, and relate these beliefs to study-strategy choices and academic outcomes. The evidence presented in this section comes from student responses to the first survey at the beginning of the academic year, before any intervention was implemented. The precise wording of the survey questions used in this section is provided in Appendix Section B.1.

3.1 Simple Framework

We start by introducing a static framework that formalizes how beliefs about the effectiveness of different study habits influence students' effort allocation. In Section 6, we extend this model to a dynamic setting to characterize the learning mechanisms and allocation frictions.

Time allocation. Consider a student i endowed with total study time $\bar{E}_i \in \mathbb{R}_+$.[20] The student chooses how to allocate effort across J distinct study habits, indexed by $j \in \{1, \ldots, J\}$. Let $e_{ji} \in \mathbb{R}_+$ denote the effort allocated to habit j, subject to the time constraint:

$$\sum_{j=1}^{J} e_{ji} = \bar{E}_i. \tag{1}$$

[20] Because effort choices are made within the semester under tight time-budget constraints, we treat total study time as fixed and focus on how students allocate that effort across distinct study habits. This focus also aligns with the design of our information intervention, which targeted the quality of study time rather than its quantity. Consistent with that design, we find no evidence that the treatment increased total study time (see Appendix Table C.3).

It is convenient to express these choices as shares of total time, $Sh_{ji} \equiv e_{ji}/\bar{E}_i$, where $Sh_{ji} \in [0,1]$ and $\sum_{j=1}^{J} Sh_{ji} = 1$.

Preferences and objective. Let $\tilde{\alpha}_{ji}$ denote student i's *perceived effectiveness* of study habit j—that is, the perceived marginal benefit of allocating effort to that habit. The student derives utility from learning (academic achievement), net of the cost of effort allocation ($C_i(\cdot)$). The student maximizes the perceived net benefit of their effort allocation:

$$\tilde{U}_i \;=\; \sum_{j=1}^{J} \tilde{\alpha}_{ji}\, e_{ji} \;-\; C_i(\mathbf{e}_i) \tag{2}$$

Rewriting this in terms of shares yields:

$$\tilde{U}_i \;=\; \underbrace{\left(\sum_{j=1}^{J} Sh_{ji}\tilde{\alpha}_{ji}\right)}_{\tilde{\alpha}_{Ti}} \bar{E}_i \;-\; C_i(\mathbf{Sh}_i \bar{E}_i) \tag{3}$$

Here, $\tilde{\alpha}_{Ti}$ represents the perceived payoff of *total* effort, which is a weighted average of the habit-specific effectiveness parameters, conditional on the chosen allocation vector $\mathbf{Sh}_i = (Sh_{1i}, \ldots, Sh_{Ji})$.[21] This formulation highlights that a student's perceived payoff of aggregate effort depends crucially on their mix of study habits.

The student chooses the allocation $\mathbf{Sh}_i$ to maximize their expected utility:

$$\max_{\mathbf{Sh}_i} \;\; \mathbb{E}_i\left[\left(\sum_{j=1}^{J} Sh_{ji}\tilde{\alpha}_{ji}\right)\bar{E}_i\right] \;-\; C_i(\mathbf{Sh}_i \bar{E}_i) \quad \text{s.t.} \quad \sum_{j=1}^{J} Sh_{ji} = 1. \tag{4}$$

where $\mathbb{E}_i[\cdot]$ is the expectation over subjective beliefs.

In the remainder of this section, we provide descriptive evidence on these theoretical objects (Section 3.2). We begin by documenting students' beliefs about the payoff to overall study-time, i.e., perceived score gains per hour of study ($\tilde{\alpha}_{Ti}$), before unpacking these beliefs into their habit-specific components, Sh_{ji} and $\tilde{\alpha}_{ji}$. Finally, we examine how these subjective and heterogeneous beliefs correlate with academic achievement and expectation errors.

[21] In the data, $\tilde{\alpha}_{Ti}$ is elicited in "points per study-hour" from vignette-based predicted scores (predicted score divided by total study hours), whereas habit-level beliefs are elicited on a 1–10 effectiveness scale. Section 6 normalizes the units of (α_p, α_r) to this 1–10 scale in the structural analysis.

3.2 Descriptive Evidence

Beliefs about payoff to overall study-time. Our empirical analysis begins by examining students' perceived gains in score (payoff) per hour of study. We find that students display substantial heterogeneity in these beliefs ($\tilde{\alpha}_{Ti}$). To measure these beliefs, we presented each student with four randomized vignettes describing a peer with specified ability, start study time (2 days vs. 2 weeks before the exam), and total study hours (10 vs. 25). Students predicted the peer's final-exam score on a 0-100 scale with letter-grade anchors, and we computed perceived payoff per hour as the predicted score divided by total study hours.[22]

Figure 1 shows that perceived payoff varies widely across students within scenario–for example, in the first representative vignette (corresponding to one of the scenarios presented to the students) the interquartile range spans 4.0-6.6 points per hour (a 65% difference)–and a large dispersion persists even after residualizing for scenario and course fixed effects (see Appendix Figure D.2).[23] The fact that students hold markedly different views about how effort translates into academic performance raises an important question: why do students hold such divergent beliefs, and to what extent can differences in study habits explain them?

Defining and measuring study habits. We now turn to the specific behaviors underlying these aggregate beliefs. Educational psychology research (Roediger and Karpicke, 2006; Karpicke and Blunt, 2011; Wallace et al., 2022; Ekuni et al., 2022; Agarwal et al., 2021; Smith et al., 2019) categorizes study habits based on their effectiveness. Retrieval practices, which involve active recall of information, are widely recognized as more effective given that they enhance information retention and deeper understanding. These practices encompass a range of methods, including self-quizzing, explaining concepts to others, and formulating questions about the material, among others. In contrast, rereading practices, which involve passively reviewing material or highlighting text without active engagement, are typically deemed less effective. Along the timing dimension, cramming behavior, defined as intense, last-minute studying before exams, is also labeled as a less effective study habit compared to beginning study sessions early and spacing out study time. While it may provide short-term benefits for immediate memory assessments, research suggests that it can have negative consequences on long-term retention and learning efficacy (Fergus et al., 2021; Mark, 2024).

[22] Full vignette text and examples appear in Q30 of the survey. Hour-ability combinations were randomized to reduce survey length; without randomization, each student would have answered 16 scenarios.

[23] Appendix Figure D.3 reports results for all scenarios, showing that this substantial heterogeneity is present throughout.

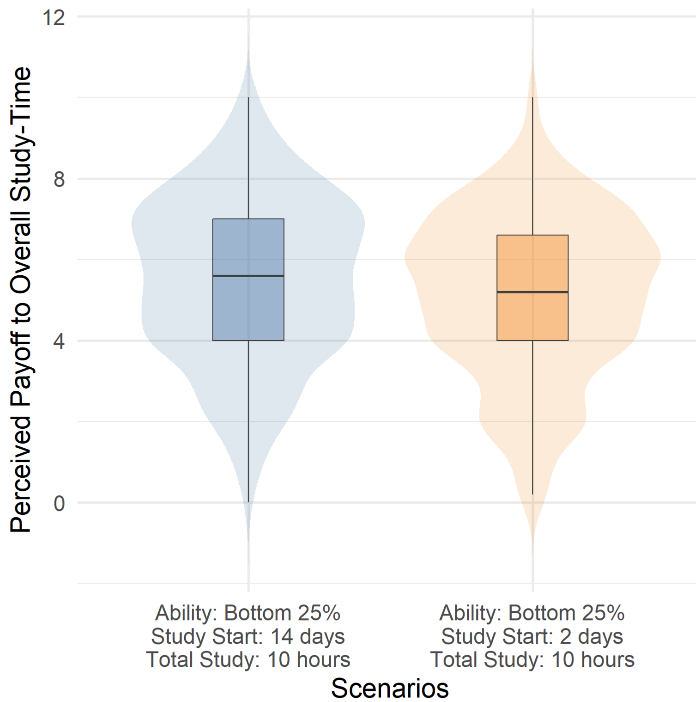

Figure 1: Payoff to Overall Study-Time (score-per-hour beliefs)

Note: This figure plots the distribution of students' perceived overall payoff to study-time for two illustrative scenarios. Students predicted a peer's final-exam score on a 0-100 scale with letter-grade anchors. In both scenarios, the peer is assigned to the Bottom 25% ability group and total study time is fixed at 10 hours; the study start time varies between 2 and 14 days before the exam. We compute perceived payoff per hour as the predicted score divided by total study hours. Violin plots show the full distribution of responses, and the overlaid boxplots report the median and interquartile range.

We measure study habits by asking students, within a standardized exam-preparation exercise, to allocate study time across different study methods and preparation windows. Specifically, students indicate how many hours they would devote to rereading versus retrieval practices when aiming for a target grade (B or A) on the final exam, with either two days or two weeks to prepare.

The four resulting responses for each student correspond to: $e_i = (e_i^{r,s}, e_i^{p,s}, e_i^{r,c}, e_i^{p,c})$, where r = rereading practices, p = retrieval practices, s = spaced study (more than 2 days before the exam), and c = cramming (within the last 2 days). From these, we construct two main measures used throughout the paper:

$$Sh_{reread,i} = \frac{e_i^{r,s} + e_i^{r,c}}{e_i^{r,s} + e_i^{p,s} + e_i^{r,c} + e_i^{p,c}} \tag{5}$$

$$Sh_{cram,i} = \frac{e_i^{r,c} + e_i^{p,c}}{e_i^{r,s} + e_i^{p,s} + e_i^{r,c} + e_i^{p,c}} \tag{6}$$

In addition to these measures, students were asked to provide their beliefs about the effectiveness of specific study practices in increasing their exam scores–such as quizzing, explaining material to others, asking questions, rereading, highlighting/underlining, and starting to study early–on a scale from 1 to 10.[24] These practices directly map onto the broader categories commonly studied in the literature: retrieval (e.g., quizzing, explaining, asking questions), rereading (e.g., rereading, highlighting/underlining), and cramming (e.g., last-minute intensive study).[25]

Allocation of study time. Figure 2 shows the median, interquartile ranges, and distributions of the share of study time students allocate to different study habits. Two insights stand out. First, many students seem to devote substantial time to habits the literature identifies as less effective: the median student allocates 45% of study time to rereading and 32% to cramming.[26] Second, there is wide dispersion in how time is allocated. For example, the 75th-percentile student assigns 46% more proportional time to rereading than the 25th-percentile student.[27] This significant heterogeneity in how students allocate time across study habits underscores the likely role of diverse beliefs about the relative effectiveness of different study methods.

Beliefs about effectiveness. Table 2 summarizes students' beliefs about the effectiveness of different study methods, rated on a scale from 1 to 10. Panel A reports perceptions of retrieval-based strategies (quizzing, explaining, asking questions), Panel B focuses on rereading practices (rereading, highlighting/underlining), and Panel C examines cramming behavior (starting to study early). Column (1) presents results for the full randomized sample, while columns (2) and (3) split the sample by exam performance–students scoring below a B versus those scoring a B or higher on their first exam prior to the intervention. Column (4) reports the mean difference between these groups and the statistical significance of those differences.

Three patterns stand out. First, relative to higher-achieving students, those scoring below a B assign

[24] The effectiveness ratings were elicited with respect to students' own expected learning outcomes, whereas the vignette-based measure used to construct perceived payoff per study hour was based on predictions about a hypothetical peer's performance. See Appendix B.1 for the exact wording of the survey questions.

[25] Quizzing refers to self-testing through practice exams, flashcards, or similar tools. Explaining refers to articulating the material to oneself or others. Asking questions refers to posing "how" and "why" questions about the material and seeking answers independently or with help. Rereading refers to reviewing material such as slides, textbooks, or handouts. Highlighting/underlining refers to marking key words, sentences, or sections of a text–typically with a highlighter or by drawing lines underneath–to emphasize important information. Starting to study early refers to beginning exam preparation well in advance and spreading study hours over multiple days. Appendix Table C.4 reports summary statistics for additional study habits that are less straightforward to classify within the retrieval, rereading, and cramming-studying early categories.

[26] Average rereading is similar during cramming and non-cramming periods (43.24% vs. 43.22%). In addition, we do not find evidence indicating that students with lower class attendance tend to report a high share of rereading time.

[27] Appendix Figure D.4 shows that the significant heterogeneity in study habits persists even after residualizing these study shares with respect to the full set of baseline control variables used in the analysis.

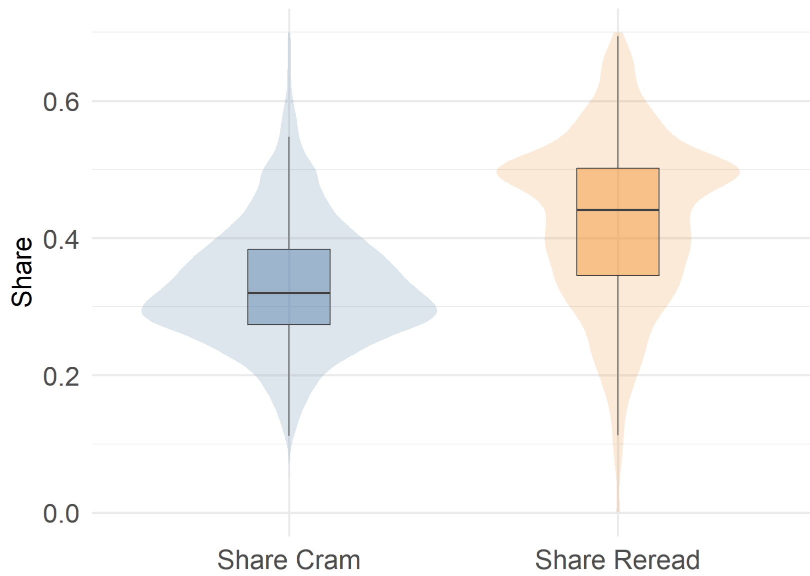

Figure 2: Shares Reread and Cramming

Note: This figure shows the distribution of students' reported time allocations across study habits. *Share Cram* and *Share Reread* denote, respectively, the fraction of total study time devoted to cramming and to rereading (values range from 0 to 1). Violin plots display the full distribution across students, and the overlaid boxplots report the median and interquartile range.

lower effectiveness to retrieval-based strategies: on average, their ratings are 0.29 points lower. Second, this pattern reverses for less effective practices: students below a B rate rereading and highlighting/underlining about 0.32 points higher than their higher-achieving peers.[28] Third, B or above students report, on average, that starting to study early is more effective than below B students. Taken together, these patterns suggest that struggling students not only underappreciate effective strategies (relative to high-achieving peers) but also overvalue less productive ones, likely enhancing gaps in academic performance.

Appendix Table C.5 provides reassuring evidence that students' reported beliefs are meaningfully connected to their reported study behaviors, examining the link between students' stated beliefs and their reported allocation of study time. The correlations show a clear pattern: students who consider rereading methods more effective report dedicating a larger share of their study time to rereading, whereas those who rate retrieval-based practices more effective report devoting relatively less time to rereading. The table also highlights how beliefs about specific study habits relate to students' broader views on the payoff of overall study-time ($\tilde{\alpha}_{Ti}$). In particular, students who judged retrieval practices and early studying to be highly effective are also more likely to perceive greater payoffs from overall study time. Conversely, those who reported more time allocated to rereading or cramming tend to report lower payoffs. This

[28] Appendix Figure D.5 illustrates the substantial heterogeneity in students' beliefs about the effectiveness of different practices.

pattern underscores a close alignment between having beliefs about effective habits more consistent with the literature and holding more optimistic views of the payoff to overall effort.

Table 2: Perceived Effectiveness of Study Habitss

	Randomized Sample (1)	Below B (2)	B or Above (3)	Diff (2)-(3)
Panel A: Perceived Effectiveness: Retrieval Practices (Scale 1-10)				
Quizzing	8.015	7.923	8.197	-0.273***
	(0.044)	(0.054)	(0.075)	
Explaining	7.058	6.970	7.231	-0.261**
	(0.050)	(0.063)	(0.083)	
Asking questions	6.692	6.581	6.914	-0.333***
	(0.049)	(0.060)	(0.086)	
Retrieval practice average	7.255	7.158	7.447	-0.289***
	(0.035)	(0.043)	(0.059)	
Panel B: Perceived Effectiveness: Rereading Practices (Scale 1-10)				
Rereading	6.263	6.375	6.039	0.337***
	(0.054)	(0.067)	(0.092)	
Highlighting/underlining	6.175	6.281	5.964	0.317***
	(0.053)	(0.064)	(0.094)	
Rereading practice average	6.219	6.328	6.001	0.327***
	(0.044)	(0.053)	(0.076)	
Gap Retrieval - Rereading	1.036	0.830	1.446	-0.616***
	(0.045)	(0.054)	(0.078)	
Panel C: Perceived Effectiveness: Cramming Practices (Scale 1-10)				
Start studying early	7.306	7.240	7.437	-0.197*
	(0.050)	(0.062)	(0.085)	
Observations	2328	1550	778	2328

This table reports the descriptive statistics of 3 groups of students: the randomized sample in column 1, the subsample of students that have scored below a B on their first exam in column 2, and finally the subsample students that have score B or above on their first exam in column 3. The effectiveness of study habits is reported by students on a scale of 1 to 10, where 1 denotes low effectiveness and 10 high effectiveness. The last column of the table looks at the difference of the mean from "Below B" and the mean of "B or above". Standard errors in parentheses. Table C.4 of the appendix reports all the study habits students were asked about in the survey. A t-test was conducted for the differences and significance is reported as follows: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

Study habits and performance. Finally, we examine in more detail the potential implications of ineffective study habits on academic performance. Our analysis focuses on students' performance on the first exam taken in the course (i.e., before any intervention). For comparability, the scores have been standardized. Column (1) of Table 3 presents Ordinary Least Squares (OLS) coefficients estimating the impact of the proportion of study hours students believe they should devote to cramming and rereading practices on our performance measure. This analysis controls for a comprehensive set of variables, including minority status, gender, class fixed effects, self-reported ability, high school GPA, employment status, family income level, total study hours per week, field of study, number of courses taken during the semester, and average study hours for the exam, among other controls. Our findings align with the existing literature, which underscores the positive relationship between effective study habits and academic performance. Specifically, a 25 percentage point increase in the proportion of study time students report spending on rereading (a priori, an ineffective practice) is associated with a decrease of 0.21 standard deviations in the first exam score. Similarly, a 25 percentage point increase in reported cramming is associated with a 0.24 standard deviation decrease in the first exam score.[29]

Table 3: Impact of Study Habits on Performance and Misprediction of Classroom Outcomes

	Std. First Exam (1)	Exp.Grade-First Exam (2)	Abs(Exp.Grade-First Exam) (3)
Share Cram	-0.964***	10.598***	13.130***
	(0.229)	(3.762)	(2.697)
Share Reread	-0.832***	9.038**	10.913***
	(0.226)	(3.843)	(3.602)
Female	-0.296***	3.851***	1.940***
	(0.031)	(0.884)	(0.607)
Black/African American	-0.458***	7.242***	7.659***
	(0.111)	(1.824)	(1.952)
Hispanic	-0.069	0.712	0.762
	(0.077)	(1.406)	(1.079)
Observations	2328	2328	2328
Controls	X	X	X
Randomized Sample	Yes	Yes	Yes
Both Surveys	No	No	No

Standard errors in parentheses are clustered at the class level. Grades are standardized at the class level. "Std. First Exam" refers to the standardized score students received in the first exam of the semester. The mean first exam grade is 70. "Exp.Grade-First Exam" represents the difference between the students' expected performance in the class and the actual first exam score, while "Abs(Exp.Grade-First Exam)" is the absolute value of the former variable. The means of these variables are 15.6 and 19.8. Controls include: indicator for students taking the survey more than once, average hours studied per week, self-reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income (students were not forced to respond on these demographic questions in the survey). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, columns (2) and (3) of Table 3 explore whether study habits influence students' ability to predict their own performance (i.e., expected grade minus performance in the first exam). Consistent with prior work, most students are overly optimistic: 77.4% expected a higher score than they actually earned. However, we find that students who report dedicating more time to rereading and cramming display larger gaps between expected and realized scores, suggesting that less effective habits both hinder preparation and undermine accurate self-assessment.[30]

[29]The mean proportion of study time dedicated to cramming and rereading is 0.336 and 0.434, respectively.

[30]Appendix Table C.6 illustrates the relationship between beliefs about the payoff to overall study time and academic performance/expectations. As expected, we find that students who believe in higher payoffs to effort not only tend to achieve better academic outcomes, but they also predict their scores more accurately.

Taken together, the framework and evidence presented in this section provide a convincing decomposition of the perceived payoff of total effort into practice-level beliefs and allocations: students who, on average, overweight rereading and cramming tend to both underperform and mispredict academic outcomes, revealing a wedge between perceived and actual grade-production functions. Because the student-specific study habit optimum is unobserved, we move from description to causal inference by using responses to an information intervention to assess whether baseline allocations were distorted by biased beliefs: if information realigns beliefs, shifts study time away from less effective habits, and improves grades, the joint response implies that pre-intervention allocations were suboptimal and shaped by mistaken beliefs about relative effectiveness.

4 Information Intervention Design

We pursue three main goals by combining an information intervention on study habits with student survey data. First, we evaluate whether providing students with information about effective study strategies improves academic performance. Second, we examine whether the intervention changes students' beliefs about the relative effectiveness of different study habits and their allocation of study effort across those habits. Third, we assess whether the delivery mode of the information (personalized feedback versus non-personalized recommendations) matters for students' academic performance and belief updating.

4.1 Implementation

Figure 3 presents the timeline of the surveys, the information intervention, and academic outcomes. After completing the baseline survey and receiving their first-exam scores, students who had provided consent were randomized within class into one of three groups: control, general feedback treatment, or personalized feedback treatment. Stratification was based on first-exam score, gender, class, and first-generation status.[31] Treatment messages were delivered via CANVAS within one week of the first exam grade release, ensuring that students received the intervention while their exam performance was still salient. To reinforce exposure, a reminder email was sent to all consenting students approximately two weeks before the final exam.[32] Toward the end of the semester, participants completed an endline survey designed to capture the mechanisms through which the intervention operated (e.g., beliefs, perceived payoffs, and study-time

[31]The pre-analysis plan was registered at the AEA RCT Registry: AEARCTR-0009980.

[32]The reminder mirrored the initial intervention, with treated students receiving their assigned messages and control students receiving a thank-you note.

allocation). We then linked survey responses to administrative records on course outcomes, including final exam scores and final grades.

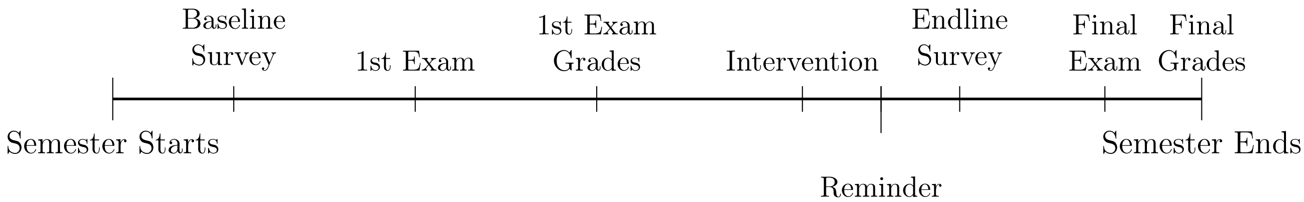

Figure 3: Timeline of Events

Note: The figure illustrates the timeline of the surveys and the information intervention. The exact dates vary depending on the course structure and the term in which the class participates in the study.

To account for the possible effects of receiving an email via CANVAS, we also sent messages to the control group. This group received a simple "thank you" note acknowledging their survey completion without any additional information regarding their performance or scores.[33]

The second group, labeled as the general feedback treatment, not only received a "thank you" note but also obtained comprehensive information on effective study habits, proven by the educational psychology literature to enhance academic performance. This guidance encompassed recommended study practices and additional resources for further insights into developing successful study habits. More specifically, we provided information on the four key groups of study habits we have focused on: *rereading practices*, *retrieval practices*, *cramming behavior*, and *spaced studying*. The objective was to encourage students to incorporate more retrieval practices into their study sessions and decrease the proportion of time spent on rereading study methods. Additionally, students were advised to start studying early and avoid last-minute cramming before exams. In essence, our goal was to promote effective study habits, emphasizing not the quantity but the quality of study time utilization through the adoption of proper study techniques.

The third group, denoted as the personalized feedback treatment, received guidance similar to that of the general feedback treatment, along with personalized insights into their study habits compared to their peers. This tailored feedback included assessments of their cramming tendencies and engagement in retrieval practices relative to the median student in their class. Furthermore, leveraging our collected class-level data, we informed these students about the potential impact of increasing the share of hours allocated to retrieval practices and reducing the share of hours allocated to cramming on their exam scores, all else being equal. This information was based on coefficients derived from OLS regression analyses, indicating the percent increase in the likelihood of scoring above the S^{th} percentile of students in their

[33]Appendix E.1 replicates the email that the students in the control group received.

class if they reallocate R percentage points of their study time to retrieval practices or non-cramming periods.[34] Consequently, the objective was to push students to increase the proportion of hours allocated to retrieval practices ($Sh_{retrieval,i}$) while diminishing the proportion of hours spent cramming ($Sh_{cram,i}$). Importantly, this feedback was specific to each student's class context. Our hypothesis posits that providing personalized, relevant information could motivate students to adopt effective study habits. Appendix Table E.1 summarizes the information shared with students.[35]

We assessed the validity of the randomization by conducting covariate balance tests in two samples used in our analysis. The first, the *Randomized Sample*, includes all students assigned to either the treatment or control group. The second is the subset of students who completed *both* the baseline and endline surveys. Appendix Table C.7 reports the results, showing that the treatment and control groups are well balanced across these samples.[36]

4.2 Empirical Model

We estimate the effects of the intervention on student outcomes using the following OLS specification:

$$y_{i,t+1} = \alpha_0 + \alpha_1 y_{i,t} + \lambda \mathbf{T_i} + \alpha_2 X_i + \zeta_{i,t+1}, \tag{8}$$

where $y_{i,t}$ and $y_{i,t+1}$ denote lagged and current outcome measures, respectively, including exam scores, beliefs, and study habits. The vector $\mathbf{T_i}$ contains treatment indicators reflecting the type of intervention a student was assigned, while X_i includes stratification variables and student-specific time-invariant controls to improve precision. The error term is given by $\zeta_{i,t+1}$.

[34]We conducted the following regression to obtain class-specific coefficients:

$$Above\ S^{th}\ perc_{im} = \beta^s_{0m} + \beta^s_{1m}\, Sh_{retrieval,im} + \beta^s_{2m}\, Sh_{cram,im} + X_i + \epsilon^s_{im}, \tag{7}$$

where $Above\ S^{th}\ perc_{im}$ is a dummy that equals 1 if student i scores above the S^{th} percentile of students in class m (i.e., is in the top S percentile). We report the percent change in the probability of being above the S^{th} percentile induced by reallocating R_m percentage points of study time, relative to the baseline probability $S_m/100$: $\beta^s_{1m}\frac{R_m}{S_m/100}$ and $\beta^s_{2m}\frac{R_m}{S_m/100}$, where β, S, and R are class-specific. In practice, $S \in \{25, 50\}$ and $R \in \{25, 50\}$. Appendix Table E.2 summarizes the numbers reported to students.

[35]Detailed messages sent to students as well as screenshots of the messages are available in Appendix E.2 for general treatment and Appendix E.3 for personalized treatment email.

[36]The only variable that appears "not balanced" in the standard covariance test is perceived "Payoff to Study-time," but only in the general feedback treatment group. To assess whether this difference is meaningful, we computed the normalized difference proposed by Imbens and Rubin (2015) for this variable. The normalized differences in the payoff to study-time between the control group and the general feedback group are 0.063 in the *Randomized Sample* and 0.069 in the *Both-Surveys Sample*. Both values are well below the 0.25 threshold suggested by Imbens and Rubin (2015), indicating that the observed imbalance is unlikely to be consequential.

Because treatment was delivered via the CANVAS platform, we cannot directly verify whether students opened or read the intervention emails. Accordingly, the baseline specification identifies the intention-to-treat (ITT) effect. In the endline survey, however, students reported whether they recalled receiving an email with study habit tips. Leveraging this information, we estimate additional specifications that replace $\mathbf{T_i}$ with indicators for self-reported receipt of treatment, instrumented by the original assignment. This instrumental variable approach allows us to recover local average treatment effects (LATE), capturing the causal impact among compliers.[37]

5 Information Intervention Results

5.1 Impact on Students' Academic Performance

We begin by examining whether the intervention improved students' test scores. Because some courses allow students to drop their lowest exam grade, which does not count toward the final course grade, our analysis is restricted to exams that contribute to final performance, which we refer to as *valid exams.*[38]

Table 4 reports ITT estimates of the treatment effect on two outcomes: (i) the first valid exam following the intervention and (ii) the average of all subsequent valid exams. The analysis is conducted on the randomized sample, ensuring that attrition does not impact the results. The key regressors are indicators for assignment to the general feedback treatment (T1) or the personalized treatment (T2), along with baseline values of the dependent variable and additional controls to improve precision.

Two key findings emerge. First, the general feedback treatment has no detectable effect on student performance. Second, the personalized feedback treatment raises performance by 0.052 to 0.085 standard

[37]The LATE analysis, as well as any analysis using outcomes measured in the survey data, requires relying on the *Both Surveys* sample, which involves a smaller number of observations. Nevertheless, as shown in Appendix Table C.7, this sample remains balanced when comparing the treatment and control groups. Consistent with this evidence, the probability of completing the second survey is not correlated with our treatment assignment.

[38]In a subset of classes, students are allowed to drop one exam so that it does not contribute to their final grade. In these settings, students who are already satisfied with their standing in the course often choose not to sit for the last exam, effectively earning a zero that is subsequently dropped. Because we observe all exam scores and the grading rules, we can identify which exam is dropped (the lowest score under the course's policy) and, consequently, which exams are retained for the final grade calculation. In our data, 17 classes (86% of our sample) allow students to drop any one of four exams they have, one class (1% of the sample) restricts the drop option to the first three exams, and 13 classes (13% of the sample) do not permit dropping any exams. Among students who earned an "A" (90%) in the course and were eligible to drop one of the four exams they took, 56% received a zero on the final exam, consistent with the interpretation that many of these students skipped the final once they were confident in their grade. To avoid bias arising from such selective non-participation in later exams, our main analysis also focuses on the first retained exam after treatment. That is, the earliest post-treatment exam score that is not dropped under the course's grading policy. Importantly, we find no evidence that treatment assignment affected which exam was dropped under the course's grading policy, indicating that the intervention did not alter students' strategic use of the drop option.

deviations, with significant effects on both the first valid post-intervention exam and the average of subsequent valid exams. These magnitudes are economically meaningful (roughly one-half of the Black–white achievement gap reported in Table 4) and are robust across alternative samples and specifications.[39] To limit concerns about multiple testing across outcomes and treatment arms, we report Romano–Wolf adjusted p-values using the stepwise procedure of Romano and Wolf (2005), following List et al. (2019): for the personalized treatment, the adjusted p-values are 0.030 for Std. Valid Score$_{t+1}$ and 0.079 for Mean Std. Valid Scores$_{t+1}$. Cluster wild-bootstrap inference yields the same qualitative conclusions, with p-values of 0.030 and 0.097, respectively.[40] Taken together, these results indicate that information can improve academic performance, but only when tailored to the individual student.

5.2 Mechanisms

5.2.1 Impact on Students' Beliefs

To shed light on the mechanisms underlying the intervention's effects on academic outcomes, we first examine how it influenced students' beliefs about the payoffs to overall study-time and the relative effectiveness of different study habits.[41] This analysis is restricted to students who completed both baseline and endline surveys. We report ITT and IV estimates, where the latter instrument students' self-reports of having opened the study-habit email with their randomized treatment assignment. All specifications control for baseline measures of the respective beliefs and include a vector of additional covariates.

Columns (1) and (2) of Table 5 show that both the ITT and LATE estimates point to a positive shift in perceived payoff to total study-time. Students in both treatment arms report that the payoff to an additional hour of study is about 3% higher relative to the control-group mean.[42] While modest in

[39]Appendix Table C.8 shows that the personalized-treatment estimates are similar when we restrict the analysis to the *Both Surveys Sample*. The estimates are also stable when we omit additional controls or include only lagged outcomes, and Wald tests do not reject equality with our preferred specification.

[40]Appendix Table C.9 reports heterogeneous treatment effects across the stratification dimensions used in the randomization (first-exam score, gender, and first-generation status) and reveals no consistent pattern. However, we find that the treatment effects on academic performance are concentrated among students whose baseline beliefs were most misaligned with the educational psychology evidence (those who viewed rereading as relatively more effective than retrieval practice). In particular, students in the bottom quartile of the retrieval-minus-rereading effectiveness gap who were assigned to the personalized treatment scored 0.205 standard deviations higher than comparable control students. Because this subgroup analysis was not pre-registered, we interpret it as suggestive rather than confirmatory. Finally, Appendix Tables C.10 and C.11 estimate the main specification separately for in-person and online classes: while the personalized-treatment effect is statistically significant in the in-person sample but not in the online sample, the point estimates are not statistically distinguishable across modalities ($p-value$ of the difference in the coefficients is 0.54 for the Std. Valid Score$_{t+1}$ and 0.84 for the Mean Std. Valid Scores$_{t+1}$).

[41]Students were asked four parallel questions on the payoff to study-time; we stack these responses to increase the effective sample size fourfold.

[42]Quantitatively, treated students perceive that the payoff to an extra study hour is about 0.130 points higher on a 100-point scale.

Table 4: Treatment Effect on Student Classroom Outcomes

	Std. Valid Score$_{t+1}$ (1)	Mean Std. Valid Scores$_{t+1}$ (2)
T1	0.015	0.007
	(0.030)	(0.027)
T2	0.085***	0.052*
	(0.027)	(0.027)
Std. First Exam$_t$	0.452***	0.371***
	(0.057)	(0.048)
Std. First Exam$_t^2$	0.044***	0.036***
	(0.010)	(0.010)
Female	-0.115^{***}	-0.076^{***}
	(0.037)	(0.027)
Black/African American	-0.197^{*}	-0.113^{*}
	(0.106)	(0.065)
Hispanic	-0.070^{*}	-0.037
	(0.037)	(0.032)
Observations	2328	2328
Randomized Sample	Yes	Yes
Both Surveys	No	No

Standard errors in parentheses. Controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken in the semester, first generation status, income (students were not forced to respond on these demographic questions in the survey), initially reported study habits effectiveness scale (0 to 10), B or above indicator for the first exam. $Std.\ Valid\ Score_{t+1}$ is the class-standardized value of the first exam score valid toward the final grade after treatment, while $Mean\ Std.\ Valid\ Scores_{t+1}$ is the mean of the standardized scores of all valid post-treatment exams. $Std.\ First\ Exam_t$ is the pre-treatment standardized score on the first exam, and $Std.\ First\ Exam_t^2$ is its squared term. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

magnitude, this effect is notable given that the intervention consisted of a few emails emphasizing study strategies rather than explicitly stressing the marginal benefit of additional effort.

We then turn to students' beliefs about the relative effectiveness of different study practices. The first outcome of interest is the perceived gap between the effectiveness of the strategies highlighted in the intervention as more effective (quizzing, explaining material, and asking questions) and those emphasized as less effective (rereading). Columns (3) and (4) of Table 5 indicate that students in the personalized treatment group (T2) perceived the relative advantage of retrieval-based practices over rereading to be approximately 42% larger than in the control group.[43] We also examine students' beliefs about the effectiveness of avoiding cramming (i.e., starting to study early). As shown in columns (5) and (6), we find no statistically significant effects of the intervention on this domain.[44]

Therefore, our findings suggest that personalized information meaningfully increased students' perceived effectiveness of retrieval-based practices over rereading (bringing beliefs closer to evidence from the learning sciences) and also increased perceived overall payoff to study-time (score-per-hour).

5.2.2 Impact on Study Habits

A central objective of the intervention was to encourage students to reallocate their study time toward more effective strategies. While belief updating is an important step, prior work emphasizes that translating revised perceptions into concrete behavioral change is often more demanding (Fuster and Zafar, 2023). We therefore examine whether students adjusted their study behavior along two dimensions. First, we use self-reported time allocations to measure the share of study time devoted to the two less effective habits emphasized in the intervention: rereading and cramming.[45] Second, we use more objective proxies for retrieval practice: homework performance and the use of practice exams.

Table 6 presents ITT and IV estimates for the post-treatment reported shares of time devoted to rereading ($\textit{share reread}_{t+1}$) and cramming ($\textit{share cram}_{t+1}$). All specifications control for baseline values of the respective outcomes as well as additional covariates, and the analysis is restricted to the *Both Surveys* sample, since it requires both baseline and follow-up measures. The results show that the general feedback treatment (T1) affected study habits only marginally, reaching significance in one specification for rereading.

[43] Appendix Table C.12 shows that this effect is mostly driven by changes in the perceived effectiveness of rereading rather than by changes in the perceived effectiveness of practicing.

[44] We cannot compute an analogous gap to the practicing–rereading gap because we did not elicit beliefs about the effectiveness of cramming.

[45] Time allocation across study habits reported in the second survey was elicited in the same way as described in Section 3.2.

Table 5: Treatment Effect on Student Beliefs

	Payoff to Study-time$_{t+1}$		Eff. Gap$_{t+1}$		Eff. Non Cramming$_{t+1}$	
	(1)	(2)	(3)	(4)	(5)	(6)
T1	0.121**	0.141**	0.128	0.149	0.129	0.151
	(0.057)	(0.066)	(0.089)	(0.096)	(0.126)	(0.139)
T2	0.129**	0.143**	0.269*	0.299**	-0.083	-0.093
	(0.059)	(0.065)	(0.137)	(0.145)	(0.085)	(0.091)
Payoff to Study-time$_t$	0.112***	0.112***				
	(0.014)	(0.014)				
Eff. Gap$_t$			0.363***	0.364***		
			(0.024)	(0.023)		
Eff. Non Cramming$_t$					0.309***	0.309***
					(0.018)	(0.017)
Std. First Exam$_t$	0.177***	0.177***	0.231**	0.229**	0.107	0.108
	(0.042)	(0.042)	(0.107)	(0.100)	(0.098)	(0.092)
Std. First Exam$_t^2$	0.021***	0.020***	0.035	0.035	0.028*	0.028**
	(0.008)	(0.008)	(0.037)	(0.035)	(0.015)	(0.014)
Female	-0.087^{*}	-0.090^{*}	0.343	0.341	-0.042	-0.047
	(0.050)	(0.050)	(0.222)	(0.208)	(0.102)	(0.094)
Black/African American	-0.004	-0.003	-0.243	-0.241	-0.308	-0.310
	(0.124)	(0.124)	(0.187)	(0.176)	(0.288)	(0.271)
Hispanic	-0.005	-0.006	-0.186	-0.189^{*}	0.127	0.128
	(0.072)	(0.072)	(0.110)	(0.102)	(0.138)	(0.129)
Observations	6972	6972	1743	1743	1743	1743
IV (received)		X		X		X
First-stage F (T1 received)		1823.05		1200.64		1200.64
First-stage F (T2 received)		2639.02		1051.68		1051.68
Randomized Sample	Yes	Yes	Yes	Yes	Yes	Yes
Both Surveys	Yes	Yes	Yes	Yes	Yes	Yes
Dep. Var. Mean (Control)	4.74	4.74	0.71	0.71	6.79	6.79

Standard errors in parentheses clustered at the student (class) level for payoff to study-time (effectiveness gap and effectiveness non-cramming). Effectiveness gap is the difference between the average effectiveness of quizzing, explaining class material, and asking questions and the effectiveness of rereading. Effectiveness non cramming captures the effectiveness of starting to study early (i.e., avoiding cramming). Columns (2), (4), and (6) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. For column (2) in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.014) and the coefficient on the $T2$ assignment indicator is 0.002 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.899 (s.e. 0.012) and the coefficient on the $T1$ assignment indicator is 0.000 (s.e. 0.003). For columns (4) and (6) in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.019) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.899 (s.e. 0.021) and the coefficient on the $T1$ assignment indicator is 0.000 (s.e. 0.003). Because these outcomes are observed only for students who completed both surveys, rather than for the full randomized sample, we focus on specifications that include baseline outcomes and additional covariates to account for potential bias in survey completion and improve precision in this restricted sample. Controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income, B or above indicator for the first exam, reported study habits effectiveness scale (0 to 10), indicator for scenario type in the first and second survey, as well as an interaction term between the effort level that students received in the first and second survey. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In contrast, the personalized treatment (T2) produced consistent and statistically significant reductions: roughly 4% less time allocated to rereading and 4% less to cramming, relative to the control mean.

To assess whether these treatment-induced effects on self-reported study habits are consistent with proxy measures of actual academically productive behavior, we next examine homework performance and the use of practice exams (i.e., study habits more consistent with retrieval practices).[46] To this end, we merged the students' CANVAS activities with the randomized treatment information and the *Both Surveys* samples. Because this pull occurred after courses had ended, some student grade pages were no longer accessible and could not be matched. In addition, several courses did not assign homework. As a result, of the 1,743 students on the original rosters, 1,683 are retained for the homework analyses. Finally, practice (review) exams were available only in two courses, leaving 141 students for the practice-exam outcomes. Nevertheless, we find that the observable characteristics of the treatment and control groups remain balanced in both samples.[47]

Appendix Table C.13 reports ITT and IV estimates for these effort outcomes. Columns (1) and (2) present results where the dependent variables corresponds to the standardized post-treatment average homework score as the dependent variable. The estimates indicate that the general feedback treatment (T1) has small and statistically insignificant effects on homework performance. In contrast, the personalized treatment (T2) increases standardized homework scores by about 0.13-0.15 standard deviations relative to the control group, and these effects are statistically significant across both the IV and OLS specifications. Columns (3) to (6) report practice-exam taking behavior. In columns (3) and (4), *Practice Exams (Valid)* equals one if a student completes the practice exam corresponding to the first valid graded exam after the intervention. Treatment increases this extensive-margin outcome by about 8–10 percentage points, with statistically significant estimates. Columns (5) and (6) examine *Practice Exams Share (Valid)*, defined as the share of completed practice exams among those counting toward the final grade post-treatment. The coefficients are qualitatively similar, though estimated less precisely and not always statistically significant.

Taken together, Tables 6 and C.13 show that personalized feedback changes students' study behavior in the direction predicted by the intervention. Students reduce their reliance on less effective strategies and

[46] Appendix Table C.14 provides validation evidence for the self-reported time-allocation measure. Across the baseline sample, the endline sample restricted to students who completed both surveys, and a pooled baseline-endline specification, students who report allocating more study time to rereading earn lower standardized homework scores. In the pooled specification, a 10 p.p. higher rereading share is associated with roughly a 0.025 SD lower homework score. These patterns suggest that the survey measures are indeed capturing meaningful differences in study behavior.

[47] Covariate balance tests for these subsamples are presented in Appendix Table C.15. The fact that the randomizations were conducted within each class ensures balance between the treatment and control groups for the courses that provide practice exams.

Table 6: Treatment Effect on Study Habits

	Share Reread$_{t+1}$		Share Cram$_{t+1}$	
	(1)	(2)	(3)	(4)
T1	-0.009	-0.010^{*}	-0.004	-0.005
	(0.006)	(0.006)	(0.005)	(0.006)
T2	-0.016^{***}	-0.018^{***}	-0.015^{**}	-0.016^{**}
	(0.005)	(0.005)	(0.006)	(0.007)
Share Reread$_t$	0.279^{***}	0.279^{***}	-0.011	-0.011
	(0.024)	(0.023)	(0.016)	(0.015)
Share Cram$_t$	0.013	0.012	0.120^{***}	0.119^{***}
	(0.015)	(0.014)	(0.026)	(0.024)
Std. First Exam$_t$	-0.009^{**}	-0.009^{**}	-0.008	-0.008^{*}
	(0.004)	(0.004)	(0.005)	(0.005)
Std. First Exam$_t^2$	-0.002	-0.002	-0.002^{***}	-0.002^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.004		-0.002	
	(0.009)		(0.006)	
Black/African American	-0.007		0.012	
	(0.011)		(0.012)	
Hispanic	-0.014		0.006	
	(0.009)		(0.009)	
Observations	1737	1737	1737	1737
IV (received)		X		X
First-stage F (T1 received)		1290.19		1290.19
First-stage F (T2 received)		1229.62		1229.62
Randomized Sample	Yes	Yes	Yes	Yes
Both Surveys	Yes	Yes	Yes	Yes
Dep. Var. Mean (Control)	0.45	0.45	0.36	0.36

Clustered by class standard errors in parentheses. Six individuals had a denominator of 0 while computing the post-treatment shares. Therefore, they are not included in the analysis and the number of observations drops from 1,743 to 1,737. Columns (2) and (4) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. First-stage coefficients are as follows: in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.020) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.901 (s.e. 0.019) and the coefficient on the $T1$ assignment indicator is 0.000 (s.e. 0.003). Because these outcomes are observed only for students who completed both surveys rather than for the full randomized sample, we focus on specifications that include baseline outcomes and additional covariates to improve precision in this restricted sample. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, dummies for missing values for high school GPA, major, number of courses taken, first generation status, income, reported study habits effectiveness scale (0 to 10), B or above indicator for the first exam. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

increase engagement with retrieval/recall forms of effort, including homework and practice-test use.[48]

Overall, the intervention provides credible evidence that students initially misallocated study effort because they held inaccurate beliefs about the relative productivity of different study strategies. Personalized feedback revised these beliefs, shifted behavior toward more effective study habits, and improved academic performance. The results, therefore, support the interpretation that misperceptions about study habits distorted pre-intervention allocations.

6 Model

6.1 Motivation

The reduced-form results show that information frictions matter; however, they leave open quantitative questions that require additional structure. In particular, they do not show how allocation costs mediate the pass-through from belief updating to study-habit choices. Students may revise their beliefs about the relative productivity of different methods, yet adjust their behavior more modestly if adopting retrieval-based strategies is costly. The reduced-form evidence also does not indicate how much students can learn from their own experience in the absence of a strong targeted information signal. Because using a study method today generates feedback that can shape beliefs tomorrow, study choices and belief formation are jointly determined over time. Thus, the reduced-form estimates identify the net effect of the intervention on beliefs, choices, and performance, but they do not separately identify the roles of belief updating, allocation costs, and endogenous learning-by-doing. To recover these primitives and explain why students exposed to the same informational shock may respond differently, we develop and estimate a dynamic model of study-habit choice.

[48]A potential concern is that the intervention affects outcomes mainly by increasing the salience of study strategies rather than by changing beliefs about their relative effectiveness. The comparison between the two treatment arms helps address this concern: both arms received study-strategy content and were therefore similarly prompted to think about studying. If salience alone were central, the general-feedback arm should have generated comparable changes in perceived effectiveness gaps and study-time allocations. Two additional patterns help distinguish belief updating from other possible mechanisms. First, the intervention does not increase overall effort relative to the control group (see Appendix Table C.3); instead, students shift time away from rereading and cramming and toward retrieval-based activities. This pattern is more consistent with updating about relative productivity than with a generic pressure to "work more." Second, Appendix Table C.16 provides little evidence of simple conformity to the class median: students above the baseline class median do not move systematically toward it (these heterogeneity checks should be interpreted cautiously, however, because message content was not randomly assigned: receiving an "above median" or "below median" message is mechanically related to baseline study behavior).

In the model, students allocate study time across activities based on beliefs about their effectiveness.[49] Accordingly, we model students as expected-utility maximizers who trade off expected learning against the costs of effort. Because effort choices are made within the semester with limited flexibility, we treat total study time as fixed and focus on the allocation of that effort across study habits, specifically practicing versus rereading.[50][51] This approach allows us to abstract from extensive-margin substitutions between studying and leisure and focus instead on the intensive margin of substitution across study habits, which was the spirit of the intervention.

Importantly, effort allocations also affect the information students acquire. Time spent using a given study habit increases the precision of the signal students receive about that habit's effectiveness. The model therefore captures a dynamic feedback loop: study choices shape belief updating, and updated beliefs shape subsequent effort allocations.

This dynamic structure is useful because it translates the reduced-form results into quantitative statements about the primitives that govern study-habit choice. By linking information signals, belief updating, and effort allocation, the model serves three purposes. First, it estimates how allocation costs shape the pass-through from belief updating to changes in study habits. Second, it places the personalized treatment on a common behavioral scale by expressing the treatment-induced response as a cost-equivalent reduction in the effort of adopting retrieval practice. Third, it evaluates whether learning by doing is likely to close belief gaps among students in the absence of strong targeted information interventions. These decompositions/counterfactuals are directly relevant to policy: if students revise their beliefs but still face costs to adopting different study methods, then personalized feedback is likely to be most effective when paired with tools or supports that make retrieval practice easier to implement (e.g., AI tools could help in this regard).

[49] In estimation, the model's allocation variables are mapped to the survey-elicited effort shares described in Section 3.2. This is a maintained measurement assumption, but one supported by the significant correlation between changes in these shares and changes in objective academic outcomes documented in Section 5.

[50] Section 7 describes our estimation sample, which includes the personalized treatment and control groups. We validate the fixed-effort assumption in this sample in Appendix Table C.3, showing that the intervention had no statistically significant effect on total study hours for students in the personalized treatment group relative to those in the control group.

[51] While the reduced-form analysis reveals a treatment effect on the reported share of effort dedicated to cramming, we find no evidence of a corresponding shift in students' subjective beliefs regarding its effectiveness. Given that our structural framework is designed to rationalize behavioral changes specifically through the mechanism of belief updating, we omit the cramming and spaced-study dimensions to maintain a transparent link between the experimental signals and the resulting behavioral responses.

6.2 Environment

We denote the set of students $i \in \{1, \ldots, N\}$ by $\mathcal{I}$. The model spans two periods $t \in \{1, 2\}$. In each period, students decide how to allocate a fixed time endowment $\bar{E}_i > 0$ across two activities (study habits): practicing and rereading. We denote the share of time devoted to practicing by $Sh^p_{it} \in [0, 1]$, and the share devoted to rereading by $Sh^r_{it} = 1 - Sh^p_{it}$. Actual time in activity $j \in \{p, r\}$ is $e^j_{it} = Sh^j_{it}\, \bar{E}_i$.

Students differ in their total study-time endowments, $\bar{E}_i$, and in their subjective beliefs about the effectiveness of practicing and rereading, $(\tilde{\alpha}^p_i, \tilde{\alpha}^r_i)$. These beliefs are characterized by heterogeneous period-1 prior means, $(\mu^{\alpha_p}_{i1}, \mu^{\alpha_r}_{i1})$, and known prior variances, $(\sigma^2_{\alpha_p}, \sigma^2_{\alpha_r})$, which are assumed to be common across students.[52] The true effectiveness of each study habit, (α_p, α_r), are unknown but common across students. Students form expectations about these effectiveness parameters and choose how to allocate effort accordingly, updating their beliefs over time as new information becomes available. The environment is further characterized by habit-specific effort cost parameters (γ_p, γ_r), signal noise variances $(\sigma^2_{\varepsilon_p}, \sigma^2_{\varepsilon_r})$, and learning intensities (k_p, k_r), all of which are common across students and assumed to be common knowledge.

Students make their effort choices to maximize their expected utility, based on their preferences and beliefs about the effectiveness of each study habit. The timing within the semester is as follows. At the beginning of period 1, given priors, the student chooses Sh^p_{i1} (and thus Sh^r_{i1}); period-1 utilities realize, and at the end of the period the student privately observes signals s^p_i and s^r_i and updates beliefs. In period 2 (terminal), given posteriors, the student chooses Sh^p_{i2} (and thus Sh^r_{i2}) and period-2 utilities realize. The next subsections formalize the flow payoff and the corresponding dynamics.

6.3 Preferences and Objective

We assume that students derive flow utility in each period from learning (academic achievement), net of the cost of effort allocation. The flow utility in period t is given by:

$$U_{it} = \alpha_p Sh^p_{it}\bar{E}_i + \alpha_r Sh^r_{it}\bar{E}_i - C(Sh^p_{it}\bar{E}_i, Sh^r_{it}\bar{E}_i)$$

[52] The habit-specific belief measures used in the structural analysis, $(\mu^{\alpha_p}_{it}, \mu^{\alpha_r}_{it})$, are elicited from students' 1–10 effectiveness ratings for practicing and rereading. In Sections 6–9, we treat these ratings as a *cardinal proxy* for perceived marginal effectiveness and normalize the units of (α_p, α_r) to align with the 1–10 scale. As a result, all cost and learning parameters are estimated in these normalized units. Consequently, a one-point change in $\mu^{\alpha_j}_{it}$ should not be interpreted as a one-point change in exam scores. Because effort shares depend only on the *difference* $\mu^{\alpha_p}_{it} - \mu^{\alpha_r}_{it}$ (see $\mathcal{M}_{it}$ below), any student-specific level shift—whether from heterogeneity in how students use the 1–10 rating scale or from unobserved differences in the valuation of learning—drops out of the optimality condition. Our survey measures therefore need only capture the *relative* perceived effectiveness across habits, not their absolute levels.

where α_p and α_r denote the true effectiveness of practicing and rereading, respectively, that is, the marginal contribution of effort in each habit to the student's flow utility from learning; and $C(Sh^p_{it}\bar{E}_i, Sh^r_{it}\bar{E}_i) = \gamma_p(Sh^p_{it}\bar{E}_i)^2 + \gamma_r(Sh^r_{it}\bar{E}_i)^2$ captures effort-allocation costs.[53]

Since the true effectiveness parameters are unknown, we define the *subjective flow utility* $\tilde{U}_{it}$ by replacing (α_p, α_r) with the belief random variables $(\tilde{\alpha}^p_i, \tilde{\alpha}^r_i)$, the student's perceived effectiveness of each habit, whose distributions are specified in Section 6.4:

$$\tilde{U}_{it} = \tilde{\alpha}^p_i \, Sh^p_{it} \, \bar{E}_i + \tilde{\alpha}^r_i \, Sh^r_{it} \, \bar{E}_i - C(Sh^p_{it} \, \bar{E}_i, \, Sh^r_{it} \, \bar{E}_i).$$

Students maximize expected subjective utility over both periods. Let $\mathcal{F}_{it}$ denote the information available to student i at the beginning of period t, which includes the history of prior beliefs, effort choices, and observed signals up to period $t-1$. Throughout, $\mathbb{E}_i\big[\cdot \mid \mathcal{F}_{it}\big]$ denotes the expectation taken under student i's subjective beliefs, integrating over the perceived effectiveness parameters $(\tilde{\alpha}^p_i, \tilde{\alpha}^r_i)$ and, where relevant, over the information yet to be observed, conditional on what the student knows at the time of the choice. The student's problem can then be written recursively as:

$$\max_{Sh^p_{i1} \in [0,1]} \mathbb{E}_i\left[\tilde{U}_{i1} + \underbrace{\max_{Sh^p_{i2} \in [0,1]} \mathbb{E}_i\Big[\tilde{U}_{i2} \,\Big|\, \mathcal{F}_{i2} \Big]}_{\text{continuation value } V_{i2}(\mathcal{F}_{i2})} \,\Bigg|\, \mathcal{F}_{i1} \right] \tag{9}$$

Equation (9) separates the student's objective into current flow utility and the continuation value. The period-2 share is a contingent plan chosen after the information set has been updated to $\mathcal{F}_{i2}$, while the period-1 share is chosen anticipating two effects: its direct contribution to current utility, and its influence on the information the student will hold when re-optimizing in period 2. The next subsection specifies the signal structure and updating rule that determine how $\mathcal{F}_{i2}$, and hence $V_{i2}(\mathcal{F}_{i2})$, depends on period-1 effort.

[53]The proposed linear-quadratic specification can also be interpreted as a local approximation to a more general allocation problem in which students value learning and face convex costs of reallocating effort across study habits. Around a reference allocation, the learning benefit of effort is locally linear in habit-specific effort, while convex implementation costs generate the quadratic cost terms.

6.4 Beliefs and Signals

We assume the true effectiveness parameters α_j, $j \in \{p, r\}$, are unknown and common across students. Let $\tilde{\alpha}_i^j$ denote the random variable representing student i's subjective belief about the true effectiveness α_j. Student i's prior at $t = 1$ is given by:

$$\tilde{\alpha}_i^j \sim N\big(\mu_{i1}^{\alpha_j}, \sigma_{\alpha_j}^2\big), \quad j \in \{p, r\}.$$

At the end of period 1, student i observes signals:

$$s_i^j = \tilde{\alpha}_i^j + \varepsilon_i^j, \quad \varepsilon_i^j \sim N(0, \sigma_{\varepsilon_j}^2).$$

independent across j, with $\sigma_{\varepsilon_j}^2$ known to students. Students are unaware of our intervention and take this as the signal DGP. Assumption A1 summarizes students' updating rule:

Assumption A1 (Effort-dependent learning)**.** *For each activity $j \in \{p, r\}$, posterior means satisfy the following updating rules:*

$$\mu_{i2}^{\alpha_p} = \mu_{i1}^{\alpha_p} + Sh_{i1}^p\, \bar{E}_i\, \frac{k_p}{\sigma_{\varepsilon_p}^2}\, \big(s_i^p - \mu_{i1}^{\alpha_p}\big),$$
$$\mu_{i2}^{\alpha_r} = \mu_{i1}^{\alpha_r} + (1 - Sh_{i1}^p)\, \bar{E}_i\, \frac{k_r}{\sigma_{\varepsilon_r}^2}\, \big(s_i^r - \mu_{i1}^{\alpha_r}\big).$$

The updating rules in Assumption A1 capture the idea that students learn more about an activity when they devote more time to it. Specifically, the impact of the innovation term $(s_i^j - \mu_{i1}^{\alpha_j})$ on posterior beliefs is scaled by $\big(k_j/\sigma_{\varepsilon j}^2\big)\, Sh_{i1}^j \bar{E}_i$. This implies that, holding fixed $k_j/\sigma_{\varepsilon j}^2$ (i.e., the updating speed parameter and signal precision), greater effort in activity j increases the influence of newly realized information on beliefs about that activity, which has implications for students' dynamic considerations.[54]

6.5 Model Solution

We solve the student's problem by backward induction. In period 2 the choice trades off differential perceived effectiveness (given posteriors) against quadratic costs; in period 1 the choice internalizes that

[54] In Appendix F.1, we show that this updating rule arises as a first-order approximation of a standard Bayesian learning model in which effort increases signal precision by generating more independent feedback draws.

effort affects learning via Assumption A1. To state the result, define the static payoff wedge, the allocation cost, and the value of experimentation for activity $j \in \{p, r\}$ as

$$\mathcal{M}_{it} \equiv \mu_{it}^{\alpha_p} - \mu_{it}^{\alpha_r} + 2\gamma_r \bar{E}_i,$$

$$\mathcal{K}_i \equiv 2\bar{E}_i(\gamma_p + \gamma_r),$$

$$\mathcal{V}_{ji} \equiv \frac{\bar{E}_i}{2(\gamma_p + \gamma_r)} k_j^2 \frac{\sigma_{\alpha_j}^2 + \sigma_{\varepsilon_j}^2}{(\sigma_{\varepsilon_j}^2)^2},$$

and let $Sh_{it}^{p,\text{stat}} \equiv \mathcal{M}_{it}/\mathcal{K}_i$ denote the associated static share. Here $\mathcal{M}_{it}$ captures the net marginal benefit of practicing over rereading at current beliefs, inclusive of cost asymmetries; $\mathcal{K}_i$ represents the convexity of the effort-allocation cost, governing how steeply marginal costs rise as effort is reallocated; and $\mathcal{V}_{ji}$ measures the expected payoff gain from better information about activity j, and is larger when prior uncertainty is high or signal noise is low. The resulting optimal shares are summarized in the following proposition, with the complete proof provided in Appendix F.2.

Proposition 1 (Optimal effort shares)**.** *Under Assumption A1,*

$$Sh_{i2}^{p*} = \underbrace{Sh_{i2}^{p,\text{stat}}}_{\textit{static share}} = \frac{\mu_{i2}^{\alpha p} - \mu_{i2}^{\alpha r} + 2\gamma_r \bar{E}_i}{2\bar{E}_i(\gamma_p + \gamma_r)},$$

$$Sh_{i1}^{p*} = \underbrace{Sh_{i1}^{p,\text{stat}}}_{\textit{static share}} + \underbrace{\frac{\mathcal{V}_{pi}\, Sh_{i1}^{p,\text{stat}} - \mathcal{V}_{ri}\big(1 - Sh_{i1}^{p,\text{stat}}\big)}{\mathcal{K}_i - \mathcal{V}_{pi} - \mathcal{V}_{ri}}}_{\textit{forward-looking experimentation tilt}}.$$

The period-2 share is the static benchmark: students trade off differential perceived effectiveness against relative costs. The period-1 share adds a net experimentation tilt because current effort affects future information. The term $\mathcal{V}_{pi} Sh_{i1}^{p,\text{stat}} - \mathcal{V}_{ri}(1 - Sh_{i1}^{p,\text{stat}})$ in the numerator represents the net informational incentive to deviate from the static benchmark; higher values of $\mathcal{V}_{pi}$ increase the marginal return to practicing, while higher values of $\mathcal{V}_{ri}$ incentivize tilting toward rereading. Second, the denominator $\mathcal{K}_i - \mathcal{V}_{pi} - \mathcal{V}_{ri}$ represents the effective marginal penalty for adjusting the effort share. By subtracting the total value of experimentation from the structural convexity $\mathcal{K}_i$, this term demonstrates that forward-looking students perceive a lower adjustment penalty—a flatter effective cost—when tilting their effort, as they internalize the future informational returns generated by current choices.

7 Identification

In this section, we describe our identification strategy and the mapping between the observed data and the model's primitives. First, we define the components of the model that are directly recovered from the data described in Section 3.2. We observe students' total study time ($\bar{E}_i$) and their reported allocation of effort across habits (Sh^j_{it}) from our survey. Additionally, we elicit students' subjective expectations about the effectiveness of different habits ($\mu^{\alpha_j}_{it}$) at both the start and end of the semester.

The availability of these data allows us to identify the remaining structural parameters by linking observed belief updates and behavioral responses to our experimental variation. We organize these parameters into three categories: (i) cost parameters, γ_j; (ii) signal precision and updating speed parameters, k_j, and $\sigma^2_{\varepsilon_j}$; and (iii) the prior variance, σ^2_α, which we restrict to be common across activities for identification reasons.

Given that the strongest results in our reduced-form analysis were found among students in the personalized group, we leverage the variation induced by this specific treatment as the primary exogenous shifter for identification. In the following paragraphs, we describe the specific identification argument for each group of parameters.

Signal precision and updating speed ($\lambda_j = \dfrac{k_j}{\sigma^2_{\varepsilon_j}}$, $\sigma^2_{\varepsilon_j}$). To identify the parameters governing the learning process, we first define the signals received by students. We denote the experimental groups as $g \in \{C, PT\}$, where C represents the control group and PT the personalized treatment group. We assume the following reduced form for the signals:

$$s^j_i = \bar{S}^g_j + \varepsilon^j_i \quad \text{for} \quad g \in \{C, PT\} \tag{10}$$

where $\bar{S}^g_j$ is the mean signal for habit j in group g, and $\varepsilon^j_i \sim N(0, \sigma^2_{\varepsilon_j})$ represents the idiosyncratic noise in signal perception. This signal structure can be rewritten as:

$$s^j_i = \pi^{1j} + D^j_{PT} PT_i + \varepsilon^j_i \tag{11}$$

where π^{1j} captures the baseline signal for the control group and D^j_{PT} represents the exogenous shift in information provided by the personalized treatment. The variable PT_i is an indicator equal to 1 if student i was assigned to the personalized treatment group and 0 otherwise.

Substituting s_i^j into the updating rule from Assumption A1, and defining the updating speed as $\lambda_j = \dfrac{k_j}{\sigma^2_{\varepsilon_j}}$, we obtain the empirical specification for the change in beliefs:

$$\mu_{i2}^{\alpha_j} - \mu_{i1}^{\alpha_j} = \lambda_j \cdot Sh_{i1}^j \bar{E}_i \cdot \left(\pi^{1j} + D_{PT}^j PT_i - \mu_{i1}^{\alpha_j} + \varepsilon_i^j\right) \tag{12}$$

Because treatment assignment is randomized, the variation in updating is exogenously generated by the different signals received across arms. To recover the structural primitives, we estimate the following equations:

$$\underbrace{\frac{\mu_{i2}^{\alpha_j} - \mu_{i1}^{\alpha_j}}{Sh_{i1}^j \bar{E}_i}}_{y_i^j} = \beta_0^j + \beta_1^j PT_i + \beta_2^j \mu_{i1}^{\alpha_j} + u_i^j \tag{13}$$

The updating speed λ_j is identified by the coefficient on baseline beliefs, $\beta_2^j = -\lambda_j$. The mean signals for each group are subsequently recovered from the remaining coefficients, where $\beta_0^j = \lambda_j \pi^{1j}$ and $\beta_1^j = \lambda_j D_{PT}^j$. Finally, since the regression error term is defined as $u_i^j = \lambda_j \varepsilon_i^j$, the noise variance is identified by $Var(u_i^j) = \lambda_j^2 \sigma^2_{\varepsilon_j}$.

Prior variance (σ^2_α). The prior variance parameters $\sigma^2_{\alpha_j}$ are identified from students' optimal decisions through their period-1 effort shares. From Proposition 1, the period-1 share Sh_{i1}^{p*} is a non-linear function of the prior variance, as the variance directly affects the option value of learning captured in the denominator through the terms $\dfrac{\sigma^2_{\alpha_j} + \sigma^2_{\varepsilon_j}}{(\sigma^2_{\varepsilon_j})^2}$. This creates variation in observed period-1 shares that depends on students' prior uncertainty about effectiveness. However, since we have only one moment condition (the period-1 share) to identify two activity-specific prior variances, we impose the normalization $\sigma^2_{\alpha_p} = \sigma^2_{\alpha_r} = \sigma^2_\alpha$, restricting the prior variance to be common across activities. In estimation, we exploit the full moment condition linking optimal shares to this common variance parameter, using the fact that students with higher prior uncertainty optimally choose different effort allocations to manage their learning across periods.

Cost parameters (γ_r and γ_p). We identify the cost parameters by leveraging the experimental variation in effort shares across treatment arms in period 2. Let $\mathcal{N}_{PT}$ and $\mathcal{N}_C$ denote the sets of students assigned to the personalized treatment and control groups, respectively, and let N_{PT} and N_C represent the number of students in each group. Using the optimal-share condition from Proposition 1, we can express the average effort share for the personalized treatment group ($\bar{Sh}_2^{PT}$) and the control group ($\bar{Sh}_2^{C}$) as follows:

$$\bar{Sh}_2^{PT} = \frac{1}{N_{PT}} \sum_{i \in \mathcal{N}_{PT}} \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r}}{2\bar{E}_i(\gamma_p + \gamma_r)} + \frac{2\gamma_r}{2(\gamma_p + \gamma_r)} = \frac{1}{2(\gamma_p + \gamma_r)} \underbrace{\frac{1}{N_{PT}} \sum_{i \in \mathcal{N}_{PT}} \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r}}{\bar{E}_i}}_{\bar{\Delta}_2^{PT}} + \frac{2\gamma_r}{2(\gamma_p + \gamma_r)}. \tag{14}$$

$$\bar{Sh}_2^{C} = \frac{1}{N_{C}} \sum_{i \in \mathcal{N}_{C}} \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r}}{2\bar{E}_i(\gamma_p + \gamma_r)} + \frac{2\gamma_r}{2(\gamma_p + \gamma_r)} = \frac{1}{2(\gamma_p + \gamma_r)} \underbrace{\frac{1}{N_{C}} \sum_{i \in \mathcal{N}_{C}} \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r}}{\bar{E}_i}}_{\bar{\Delta}_2^{C}} + \frac{2\gamma_r}{2(\gamma_p + \gamma_r)}. \tag{15}$$

Because beliefs ($\mu_{it}^{\alpha_j}$), survey effort shares (Sh_{it}^{j}), and total effort ($\bar{E}_i$) are all observed in our data, the group averages $\bar{\Delta}_2^g \equiv \frac{1}{N_g} \sum_{i \in \mathcal{N}_g} \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r}}{\bar{E}_i}$ and $\bar{Sh}_2^g$ are known quantities. This system of two equations allows us to uniquely solve for the two cost parameters:

$$\gamma_r = \frac{1}{2} \frac{\bar{Sh}_2^C \, \bar{\Delta}_2^{PT} - \bar{Sh}_2^{PT} \, \bar{\Delta}_2^{C}}{\bar{Sh}_2^{PT} - \bar{Sh}_2^{C}}, \qquad \gamma_p = \frac{1}{2} \frac{\bar{\Delta}_2^{PT}\big(1 - \bar{Sh}_2^{C}\big) + \bar{\Delta}_2^{C}\big(\bar{Sh}_2^{PT} - 1\big)}{\bar{Sh}_2^{PT} - \bar{Sh}_2^{C}}. \tag{16}$$

Identification thus relies on the wedge between the treatment-induced change in subjective beliefs and the corresponding change in survey effort shares. If a large shift in beliefs results in only a small shift in effort shares, the model recovers high values for the cost parameters γ_j.

8 Estimation

We estimate the model parameters following the identification strategy described above. In this section, we aggregate the practicing methods–quizzing, explaining, and asking questions–into a single composite practice, such that $\mu_{it}^{\alpha_p}$ represents the average perceived effectiveness of these methods, while $\mu_{it}^{\alpha_r}$ denotes the effectiveness of rereading.

Sample and data preparation. To align the estimation with our identification strategy, we focus on students in the control and personalized treatment groups. The survey-elicited effort shares and belief measures described in Section 3.2 serve as the empirical counterparts of the model's allocation and belief variables. We restrict the sample to individuals with strictly interior effort shares and non-zero total study time in the baseline, as the model's first-order conditions for effort allocation and the equations for belief updating are well-defined for students who engage in both activities.[55]

To isolate the learning specifically induced by the experimental signals, we account for potential drift in beliefs and behavior—such as general time-trends or common shocks—that are unrelated to our intervention. We center period-2 outcomes by subtracting the mean changes observed in the control group between the two periods.[56] This normalization ensures that the control group's mean signal ($\bar{S}_j^C$) is anchored to students' initial priors. Consequently, the estimated updating speed (λ_j) is identified solely as the marginal response to the exogenous information provided in the personalized reports, rather than being confounded by broader changes in the academic environment.

Estimation procedure. Following the identification argument, we estimate the model in two steps. First, we estimate the signal precision and updating-speed parameters ($\lambda_j, \sigma^2_{\varepsilon j}$) and the mean signals ($\bar{S}_j^{PT}, \bar{S}_j^C$) via OLS. Second, holding the OLS estimates fixed, we estimate the remaining structural parameters via GMM by matching model-implied and empirical moments of period-1 and period-2 optimal shares across treatment groups.

Step 1: OLS

Belief updating is characterized by Equation 13, which relates the change in beliefs about activity $j \in \{p, r\}$—normalized by effort—to baseline beliefs and treatment status. Specifically, for each activity we estimate

$$y_i^j = \beta_0^j \; + \; \beta_1^j \, PT_i \; - \; \lambda_j \, \mu_{i1}^{\alpha_j} \; + \; \nu_i^j, \tag{17}$$

where $y_i^j = \dfrac{\mu_{i2}^{\alpha_j} - \mu_{i1}^{\alpha_j}}{Sh_{i1}^j \bar{E}_i}$ is the belief updating normalized by effort allocated to activity j in period one, PT_i is the personalized treatment indicator, and $\mu_{i1}^{\alpha_j}$ is the baseline belief about the effectiveness of activity j, $\mu_{i2}^{\alpha_j}$ is the re-centered period-2 belief about the effectiveness of activity j. We run OLS separately for each activity and obtain $(\hat{\beta}_0^j, \hat{\beta}_1^j, \hat{\lambda}_j)$ and $\widehat{\mathrm{Var}}(\hat{\nu}_i^j)$. From these, we construct the signal's noise variances $\hat{\sigma}^2_{\varepsilon j} = \widehat{\mathrm{Var}}(\hat{\nu}_i^j)/\hat{\lambda}_j^2$ and the group-specific mean signals $\hat{\bar{S}}_j^C = \dfrac{\hat{\beta}_0^j}{\hat{\lambda}_j}$ and $\hat{\bar{S}}_j^{PT} = \hat{\bar{S}}_j^C + \dfrac{\hat{\beta}_1^j}{\hat{\lambda}_j}$. Because $\mu_{i2}^{\alpha_j}$ is re-centered using the control group's mean change between periods, these mean signals are recovered on the same normalized (detrended) scale as y_i^j. Importantly, this normalization removes common drift but does not mechanically fix $\bar{S}_j^C$; $\bar{S}_j^C$ is identified from the intercept of (17) in that normalized scale. This

[55] These filters result in the exclusion of 15 observations, representing 1.3% of the sample of students who completed both surveys and were assigned to either the control or the personalized treatment group.

[56] This is applied to period-2 beliefs about practicing and rereading, and to the period-2 share of practicing.

yields the first-step parameter vector:

$$\hat{\theta}^{(1)}_{\text{OLS}} = \big(\hat{\lambda}_j,\ \hat{\sigma}^2_{\varepsilon j},\ \hat{\bar{S}}^{C}_j,\ \hat{\bar{S}}^{PT}_j\big)_{j\in\{p,r\}}.$$

Step 2: GMM

Holding $\hat{\theta}^{(1)}_{\text{OLS}}$ fixed, we estimate the remaining parameters by matching moments of optimal effort shares in the model to those observed in the data. Let $g \in \{C, PT\}$ denote the control and treatment groups. Define the group-specific empirical moment vector

$$m_{0,g} = \big(\overline{Sh^p_{1,g}},\ \overline{Sh^p_{2,g}}\big)',$$

and stack across groups to obtain $m_0 = (m'_{0,C},\ m'_{0,PT})'$. For a given vector θ_{GMM}, the model implies optimal shares $Sh^p_{t,g}(\theta_{GMM})$ (see Proposition 1), which we stack analogously to obtain the moment vector $m(\theta_{GMM})$. The estimation procedure minimizes the weighted distance between the model-implied moments $m(\theta_{GMM})$ and the empirical moments m_0:

$$\hat{\theta}_{GMM} = \arg\min_{\theta_{GMM}} \big(m(\theta_{GMM}) - m_0\big)' W \big(m(\theta_{GMM}) - m_0\big), \tag{18}$$

where W is the efficient weighting matrix.

Parameter estimates and model fit. Table 7 reports the structural estimates and the corresponding measures of model fit. The estimates for the mean signals ($\bar{S}^g_j$) capture the information shock received by students in each treatment arm. Consistent with the intervention design, students in the personalized treatment group received signals emphasizing the high relative effectiveness of practicing over rereading ($\bar{S}^{PT}_p = 7.569$ vs. $\bar{S}^{PT}_r = 5.985$), whereas signals for the control group remain centered near baseline priors.[57]

The estimated cost parameters are $\gamma_r = 1.663$ [1.100, 4.820] and $\gamma_p = 1.365$ [0.930, 3.803], where brackets denote 95% bootstrap confidence intervals. These parameters govern the curvature of the effort-cost function rather than cost levels. The economically relevant quantity is the marginal cost evaluated at the observed allocation, $MC_j = 2\hat{\gamma}_j Sh^j_1 \bar{E}$. Because students already devote, on average, more time to practicing ($Sh^p_1 = 0.565$ vs. $Sh^r_1 = 0.435$), the marginal hour of practicing is in fact 7% more costly than

[57] For comparison, the baseline means for the perceived effectiveness of rereading and practicing are 6.44 and 7.36, respectively.

the marginal hour of rereading.[58] The presence of these costs can substantially limit the behavioral response to new information regarding study-habit effectiveness. In Section 9, we quantify the extent to which the cost of allocating effort scales the mapping from belief revisions to changes in study habits and assess the resulting implications for learning outcomes.

Finally, in terms of model fit, the right panel of Table 7 compares model predictions to the survey-based data for key moments: period-1 practice shares ($\bar{Sh}_1^p$), period-2 posterior beliefs ($\bar{\mu}_2^p, \bar{\mu}_2^r$), and period-2 practice shares ($\bar{Sh}_2^p$). The model closely replicates these moments both overall and separately by treatment arm, giving support to the model's ability to rationalize students' learning and effort-allocation responses to the intervention.

Table 7: Model Estimates and Fitting Results

Panel A: Model Estimates

		Rereading	Practicing
λ	Weight on signal in belief updating.	0.137 [0.123, 0.153]	0.120 [0.108, 0.133]
$\bar{S}^{\text{PT}}$	Mean signal (treated/PT).	5.985 [5.567, 6.359]	7.569 [7.322, 7.782]
$\bar{S}^{\text{C}}$	Mean signal (control).	6.637 [6.197, 7.000]	7.387 [7.207, 7.598]
σ_ε^2	Signal noise variance.	28.567 [23.201, 36.583]	9.639 [8.068, 11.824]
σ_α^2	Prior belief variance (effectiveness).	13.342 [3.136, 17.525]	13.342 [3.136, 17.525]
γ	Effort cost parameter.	1.663 [1.100, 4.820]	1.365 [0.930, 3.803]

Panel B: Model Fitting

Variable	Data	Model
$\bar{e}_1^p$	0.565	0.564
$\bar{\mu}_2^p$	7.372	7.356
$\bar{\mu}_2^r$	6.339	6.123
$\bar{e}_2^p$	0.570	0.572
$\bar{e}_1^{p,C}$	0.564	0.565
$\bar{\mu}_2^{p,C}$	7.351	7.371
$\bar{\mu}_2^{r,C}$	6.418	6.304
$\bar{e}_2^{p,C}$	0.561	0.569
$\bar{e}_1^{p,PT}$	0.567	0.563
$\bar{\mu}_2^{p,PT}$	7.394	7.341
$\bar{\mu}_2^{r,PT}$	6.258	5.935
$\bar{e}_2^{p,PT}$	0.580	0.576

Notes: Panel A reports parameter estimates from the structural model across Rereading and Practicing study habits. In brackets, we report empirical confidence intervals obtained via a bootstrap procedure that replicates the stratification used in the treatment assignment. Panel B compares model predictions to data moments for the full sample, control group (C), and personalized treatment group (PT).

9 Counterfactuals

The dynamic model provides a disciplined framework for interpreting the intervention by separating two forces: (i) *learning* from signals about the relative effectiveness of study habits (belief updating), and (ii) the *curvature of the effort-cost function* (γ_p, γ_r), which governs how steeply marginal costs rise as effort is reallocated across study habits. A common concern with information interventions is that, even when beliefs shift, the behavioral response may be limited by the presence of effort allocation costs (Fuster and Zafar, 2023). We therefore use the estimated model to run counterfactual decompositions that quantify:

[58] This difference is statistically different from zero.

(a) how the cost-curvature parameters γ_p and γ_r scale the mapping from belief revisions into study-habit choices, and (b) the magnitude of a reduction in the curvature parameter γ_p—and hence in the marginal cost of practicing—that would be required to generate behavioral effects equivalent to those induced by the information treatment.

Moreover, the model's learning channel has two distinct components: an immediate, signal-driven revision of beliefs, and an endogenous learning-by-doing mechanism, whereby belief updating depends on students' effective exposure to the study habits they implement. In the exercises below, we use this structure to decompose heterogeneity in belief updating into differences in effective exposure and differences in initial priors, clarifying how much cross-student variation in updating is attributable to learning-by-doing versus prior beliefs.

9.1 Effort Allocation Costs and a Cost-Equivalent Benchmark

A useful implication of the model is that the convexity of effort costs governs how strongly a given belief revision maps into changes in study behavior. Let Sh_{i2}^{p*} denote student i's optimal share of practice in period 2 under the treatment. Let $Sh_{i2}^{p,f*}$ denote the corresponding counterfactual optimal share in period 2 in the absence of the treatment—i.e., evaluated holding student i's beliefs fixed at their period-1 values. Using the model's optimality conditions, the model-implied individual treatment effect on the practice share can be written as:[59]

$$\widehat{TE}_i \equiv Sh_{i2}^{p*} - Sh_{i2}^{p,f*} = \frac{\Delta\mu_i^p - \Delta\mu_i^r}{2\bar{E}_i(\gamma_p + \gamma_r)}, \tag{19}$$

where $\Delta\mu_i^p$ and $\Delta\mu_i^r$ are the treatment-induced changes in beliefs about the effectiveness of practicing and rereading, $\bar{E}_i$ is total study effort, and the γ's are the curvature parameters of the effort-cost function associated with each type of effort. Averaging over students yields:

$$\overline{\widehat{TE}} = \frac{1}{2(\gamma_p + \gamma_r)} \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta\mu_i^p - \Delta\mu_i^r}{\bar{E}_i}. \tag{20}$$

Equations (19)–(20) make the role of the cost curvature transparent: holding belief revisions fixed, the magnitude of the behavioral response is inversely proportional to the sum of the cost parameters,

[59] Because the structural parameters (γ_j, k_j, $\sigma^2_{\varepsilon_j}$, σ^2_{α}) are estimated as common across students, heterogeneity in $\widehat{TE}_i$ is driven by variation in initial beliefs ($\mu_{i1}^{\alpha_j}$) and study-time endowments ($\bar{E}_i$). This parsimonious specification maximizes statistical power while still generating rich cross-student variation in predicted responses through observed differences in baseline conditions.

$\gamma_p + \gamma_r$. Specifically, the term $1/[2(\gamma_p + \gamma_r)]$ governs the "pass-through" from belief updating to study-habit reallocation.

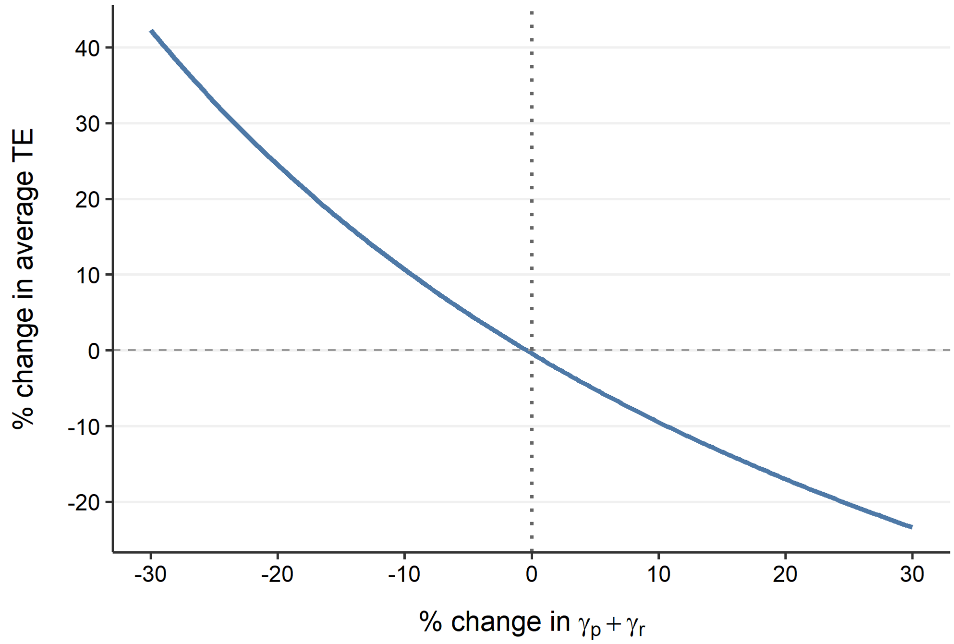

Figure 4: Sensitivity of the Treatment Effect to Effort Allocation Parameters

Notes: The figure reports a sensitivity exercise that proportionally varies the sum of the effort-cost curvature parameters, $(\gamma_p + \gamma_r)$, holding the treatment-induced belief changes fixed at their estimated values. The y-axis reports the percentage change in the model-implied average treatment effect on the period-2 share of practicing, relative to the baseline estimate. The vertical dashed line indicates the baseline structural estimate, $\widehat{\gamma}_p + \widehat{\gamma}_r = 3.02$. The horizontal dashed line corresponds to zero percent change, so points above or below this line indicate larger or smaller model-implied treatment effects than under the baseline estimate.

In this regard, Figure 4 reports a sensitivity exercise that isolates the role of the effort-cost curvature in translating belief revisions into behavior. We proportionally vary the estimated sum of the cost parameters, $(\gamma_p + \gamma_r)$, while holding the treatment-induced belief revisions fixed at their estimated values, and plot the resulting proportional change in the model-implied average treatment effect on the period-2 share of practicing. The vertical line corresponds to the estimated value, $\widehat{\gamma}_p + \widehat{\gamma}_r = 3.02$. The figure shows that the behavioral response is highly sensitive to this allocation friction. For example, moving from a value of $(\gamma_p + \gamma_r)$ that is 10% below the estimate to one that is 10% above it reduces $\overline{\widehat{TE}}$ by 17.5%. Thus, even moderate variation in the convexity of the reallocation cost can materially change the behavioral response to the same informational shock. Put differently, the behavioral "pass-through" of information is mediated by this allocation friction: when $(\gamma_p + \gamma_r)$ is high, belief updating translates into limited reallocation of study time, whereas when it is low, the same belief revision induces a larger shift in effort shares. Accordingly, information interventions are likely to scale most effectively when paired with tools that directly reduce the marginal cost of effort. For example, an increasingly plausible channel is the use of AI-based study assistants that lower the time and cognitive costs of implementing more effective study strategies (e.g., generating targeted practice questions), and translate course content into formats that are easier to consume and review.

The sensitivity analysis in Figure 4 asks a comparative-static question: holding the treatment-induced belief revisions fixed, how does varying the cost curvature ($\gamma_p + \gamma_r$) change the implied reallocation toward practicing? We now complement this perspective by inverting the exercise. Instead of varying the curvature parameters and reading off the implied $\widehat{\widehat{TE}}$, we take the treatment-implied increase in the practice share as the target object and ask: in the absence of new information, what reduction in the cost parameter γ_p would be required for the model to generate the same increase in practicing? In doing so, it provides a benchmark for comparing information-based policies to interventions that directly lower the marginal cost of effort. To implement this benchmark, Appendix Table C.17 contrasts three scenarios: (a) baseline (no new information), (b) information treatment (students receive the estimated average treatment signals), and (c) a cost counterfactual (no new information, but the cost parameter associated with practicing is reduced so that behavior matches the treatment).

The first panel of Appendix Table C.17 represents the average student, serving as a benchmark. For the mean student, baseline beliefs about the effectiveness of practicing and rereading are already relatively close to the personalized-treatment signals: $\mu_1^{\alpha_p} = 7.355$ compared with $\bar{S}_p^{PT} = 7.569$, and $\mu_1^{\alpha_r} = 6.443$ compared with $\bar{S}_r^{PT} = 5.985$. As a result, there is relatively less scope for large belief revisions for this student. The more informative comparison comes from students whose baseline beliefs are more misaligned with the treatment signal.

Students in the first quartile of the baseline effectiveness-gap distribution (second panel of Appendix Table C.17) initially view rereading as much more effective than practicing ($\mu_1^{\alpha_p} = 6.605$ versus $\mu_1^{\alpha_r} = 8.751$) and allocate only half of their study time to practicing ($Sh_1^p = 0.500$). For this subgroup, the personalized signal generates much larger belief revisions: $\mu_2^{\alpha_p}$ rises by 8.5%, while $\mu_2^{\alpha_r}$ falls by 25.1%. These revisions translate into an 8.1% increase in the practice share. Replicating this behavioral response through costs alone would require a 16.5% reduction in γ_p—nearly five times the reduction needed for the average student, and cost-equivalent to decreasing total study time by 4.2%.[60] Thus, the value of the information intervention scales with the gap between students' priors and the treatment signals: students whose initial beliefs are most distorted experience the largest belief revisions and, consequently, the largest behavioral response.

Together with Figure 4, the cost-equivalent benchmark provides a unified way to interpret magnitudes. Figure 4 highlights how the cost curvature governs the pass-through from belief revisions to behavior, while the cost-equivalent exercise complements this by expressing the personalized treatment on the same scale. Reproducing the treatment-implied increase in practicing would require a reduction of 16.5% in the cost parameter γ_p for students in the first quartile of the beliefs distribution. Therefore, to the best of our knowledge, this paper is among the few in economics–and, to our knowledge, the first in the study-habits context–to structurally quantify the effort-allocation costs that prevent experimentally induced belief changes from fully translating into behavioral adjustment.

9.2 Learning-by-doing and Heterogeneous Belief Updating

The previous exercises show that the personalized treatment moves both beliefs and behavior for the average student, and that the cost of allocating effort can tightly govern the behavioral response. We now turn to complementary questions: why does *belief updating* itself differ across student types? Can belief gaps across students close in the absence of a strong information intervention? To address this, we compare a representative student in the first quartile of the baseline effectiveness-gap index with the mean student, distinguishing two channels: (i) differences in effective exposure to an activity that generates different learning-by-doing, and (ii) differences in initial priors that make the same information signal more or less "surprising".

9.2.1 Learning-by-doing versus prior beliefs: decomposing belief-updating gaps

Let $\mu^j_{1,\theta}$ and $\mu^j_{2,\theta}$ denote period-1 and period-2 beliefs about the effectiveness of activity $j \in \{p, r\}$ (practicing and rereading) for student type θ. Let $e_{j,\theta}$ denote the *effective exposure* to activity j in period 1, defined as the share of time devoted to j times total study time. In the model, belief updating under the personalized

[60] Using the cost function defined in Section 6.3, we define the hours-equivalent study time $\bar{E}_i^{EV}$ as the value that, evaluated at the baseline parameters (γ_p, γ_r), yields the same cost as applying (γ'_p, γ'_r) to baseline study time $\bar{E}_i$ holding shares fixed at baseline. For each student i, we solve $C_i(\bar{E}_i^{EV}; \gamma_p, \gamma_r) = C_i(\bar{E}_i; \gamma'_p, \gamma'_r)$, which implies $\bar{E}_i^{EV} = \sqrt{\dfrac{\gamma'_p(Sh^p_{it})^2 + \gamma'_r(Sh^r_{it})^2}{\gamma_p(Sh^p_{it})^2 + \gamma_r(Sh^r_{it})^2}}\,\bar{E}_i, \qquad \Delta\bar{E}_i \equiv \bar{E}_i^{EV} - \bar{E}_i$. We then report the sample average change in study time, $\Delta\bar{E} = \frac{1}{N}\sum_i \Delta\bar{E}_i$.

treatment can be written as

$$\mu^j_{2,\theta} - \mu^j_{1,\theta} = \hat{\lambda}_j \, e_{j,\theta} \left(\bar{S}^{PT}_j - \mu^j_{1,\theta} \right), \tag{21}$$

where $\bar{S}^{PT}_j$ is the (average) signal about activity j and $\hat{\lambda}_j$ is the learning sensitivity, both estimated in Section 8. Consider the gap in belief updating for activity j between a representative student in the first quartile of the effectiveness-gap distribution ($\theta = Q1$) and the mean student ($\theta =$ mean).[61] Define $\Delta L_j \equiv (\mu^j_{2,Q1} - \mu^j_{1,Q1}) - (\mu^j_{2,\text{mean}} - \mu^j_{1,\text{mean}})$. Using (21) and defining $\Delta e_j \equiv e_{j,Q1} - e_{j,\text{mean}}$, $\Delta \mu^j_1 \equiv \mu^j_{1,Q1} - \mu^j_{1,\text{mean}}$, $\bar{e}_j \equiv (e_{j,Q1} + e_{j,\text{mean}})/2$, and $\bar{\mu}_j \equiv (\mu^j_{1,Q1} + \mu^j_{1,\text{mean}})/2$, we can rewrite the updating gap between these two representative students as the sum of two components:

$$\Delta L_j = \underbrace{\hat{\lambda}_j \, (\bar{S}^{PT}_j - \bar{\mu}_j) \, \Delta e_j}_{\text{learning-by-doing}} + \underbrace{\left(-\hat{\lambda}_j \, \bar{e}_j \, \Delta \mu^j_1 \right)}_{\text{prior beliefs}}. \tag{22}$$

The first term captures differences in updating that arise because one type of student engages more in activity j in period 1 (hence learns more from the same signal), corresponding to the learning-by-doing mechanism. The second term captures differences in updating that arise because students start from different priors; holding exposure fixed, the same signal induces larger revisions when initial beliefs are farther from $\bar{S}^{PT}_j$.

Table 8 applies eq. (22) under the personalized treatment. For rereading, first-quartile students revise their beliefs downward more than the mean student ($\Delta L_r = -1.811$), primarily because their baseline beliefs are more optimistic about rereading, so the same signal induces a larger downward correction (-1.678).[62] Learning-by-doing contributes around 7% of the total updating gap, consistent with first-quartile ($Q1$) students having higher effective rereading exposure in period 1.[63]

For practicing, first-quartile students also update their beliefs more than the mean student ($\Delta L_p = 0.521$),

[61] A representative student in the first quartile of the effectiveness-gap distribution is constructed by setting $(\mu^p_{1,Q1}, \mu^r_{1,Q1})$ and total study time equal to the corresponding sample means in that quartile and then computing the model-implied effective exposures $e_{j,Q1}$ given these primitives. If instead we worked directly with the average updating $\mathbb{E}[\mu^j_{2,i,\theta} - \mu^j_{1,i,\theta}]$ within each group θ, applying the individual-level updating equation and then taking expectations would generate an additional term involving the covariance between $e_{j,i,\theta}$ and $\mu^j_{1,i,\theta}$ within each group. To keep the expressions and decomposition as simple as possible, we avoid carrying this covariance term explicitly and work with representative students.

[62] For context, the baseline gap in perceived effectiveness for rereading between the representative student in the first quartile and the mean student is 2.31. This implies that students in the first quartile initially believe rereading is 2.31 points more effective than the mean student does.

[63] As described in Section 8, our estimation procedure nets out common drift in beliefs and behavior by centering period-2 outcomes on the mean change observed in the control group between periods. As a robustness check, we recompute the learning-by-doing decomposition using an augmented signal for each study habit, defined as the personalized-treatment signal plus the corresponding control-group trend, allowing the signal to also incorporate systematic within-semester evolution in beliefs common to untreated students. The results are qualitatively unchanged: learning by doing remains quantitatively small, accounting for only 8.7% of the rereading updating gap and 0.4% of the practicing updating gap.

Table 8: Decomposition of belief-updating gaps between $\theta = Q1$ and the mean student

	Rereading ($j = r$)	Practicing ($j = p$)
	(1)	(2)
ΔL_j	-1.811	0.521
Learning-by-doing	-0.133	-0.030
Prior-beliefs component	-1.678	0.551

Notes: This table reports the decomposition of the belief-updating gap between a representative student in the first quartile of the effectiveness-gap distribution ($\theta = Q1$) and the mean student ($\theta =$ mean), defined as $\Delta L_j \equiv (\mu^j_{2,Q1} - \mu^j_{1,Q1}) - (\mu^j_{2,\text{mean}} - \mu^j_{1,\text{mean}})$. Following Equation (22), the *Learning-by-doing component*, $\hat{\lambda}_j(\bar{S}^{PT}_j - \bar{\mu}_j)\Delta e_j$, captures the portion of the gap attributable to across-group differences in period-1 exposure (Δe_j). The *Prior-beliefs component*, $-\hat{\lambda}_j \bar{e}_j \Delta\mu^j_1$, captures the portion of the gap driven by across-group differences in baseline priors ($\Delta\mu^j_1$). All components are evaluated using the recovered structural parameters and the mean baseline values for each respective group.

and this gap is almost entirely driven by the prior-beliefs component (0.551), reflecting that first-quartile students start from more pessimistic priors about practicing relative to the mean student, so the same signal induces a larger upward correction. The learning-by-doing component is small and slightly negative (-0.030), consistent with first-quartile students having lower effective practice exposure in period 1. Because these students devote less time to practicing, they accumulate less experiential feedback about its effectiveness, which mechanically attenuates their belief updating through the learning-by-doing channel relative to the mean student.

Overall, these decompositions indicate that the personalized treatment is especially powerful for first-quartile students because it directly corrects their most distorted priors–pessimism about practicing and optimism about rereading–whereas learning-by-doing alone would be too weak to close these belief gaps quickly. This might also help explain why struggling students may fail to catch up by "learning by themselves" about effective practices in the absence of strong, targeted signals.

10 Discussion: Scalability and External Validity of the Intervention

Many education interventions that show promise in small-scale trials fail to replicate when implemented more broadly (Al-Ubaydli et al., 2017; List, 2022). Al-Ubaydli et al. (2017) trace this "scale-up problem" to three main sources: false positives, unrepresentativeness of the experimental population, and unrepresentativeness of the situation; to which Davis et al. (2017) add quality drops and diseconomies of scale in implementation costs. We briefly discuss how our design addresses each of these concerns.

First, scaling up a policy based on false positive findings is a leading explanation for failed replications (Maniadis et al., 2014). Our design and analysis were pre-registered, and, as discussed in Section 5, we adjust inference for the multiplicity of outcomes and treatment arms using the stepwise procedure of Romano and Wolf (2005), following List et al. (2019), complemented with cluster wild-bootstrap inference. The effects of the personalized treatment on academic performance remain significant under both procedures.

Second, our experiment was conducted at ASU, the largest public university in the United States, a broadly non-selective institution serving a large share of students at the margin of college enrollment, arguably the most policy-relevant population for interventions targeting academic performance (Aucejo et al., 2025). Moreover, the intervention is a natural field experiment: randomization occurred within students' actual courses, feedback was delivered through the learning management system students already use, and outcomes are administrative records of high-stakes exams. Both the population and the situation thus closely mirror those of a policy implementation in the STEM/Business-related fields.

Third, in contrast to "boots-on-the-ground" programs such as tutoring or coaching, which face quality drops and diseconomies of scale as the supply of qualified personnel is stretched (Davis et al., 2017), our personalized feedback is generated automatically from a short survey and first-exam scores, data that institutions already collect or could collect at minimal cost. Implementation costs are essentially fixed, rendering the marginal cost of treating an additional student virtually zero and preserving the quality of the feedback at scale (Kremer et al., 2013; Angrist et al., 2020).

Finally, treatment effects could shrink at scale due to equilibrium externalities. This concern is central for information policies in centralized admissions, where capacity constraints imply that part of the gains are redistributive (Larroucau et al., 2025). Effective study habits, in contrast, are not a rival resource: one student adopting retrieval practice does not crowd out learning by another. The main caveat is that, in courses graded on a curve, part of a treated student's improvement in relative standing would be positional under full rollout; our outcomes, however, measure raw exam performance, that is, learning, which is not mechanically zero-sum. Taken together, these considerations suggest that our estimates are informative about the policy's likely effectiveness at scale.

11 Conclusion

Students' academic outcomes depend not only on how much they study, but also on how they allocate study time across alternative learning strategies. This paper shows that imperfect information about the relative effectiveness of those strategies can distort effort allocation and lower academic performance. By treating study effort as a multidimensional input, we highlight a margin of student decision-making that is central to learning but largely absent from standard models of educational investment.

We combine panel survey data, administrative records, digital activity measures, and a randomized information intervention with more than 2,000 undergraduate students at Arizona State University. At baseline, students hold heterogeneous beliefs about the effectiveness of common study methods, and many devote substantial time to strategies that the educational psychology literature identifies as less effective, such as rereading and cramming. These reported allocations are strongly associated with lower exam performance and larger score prediction errors, suggesting that students' beliefs about study methods are meaningfully connected to both their choices and their outcomes.

The randomized intervention provides causal evidence that these baseline patterns partly reflect information frictions rather than only efficient heterogeneity in preferences or constraints. Generic study advice has little effect. In contrast, personalized feedback (which combines evidence-based recommendations with information about students' own study habits relative to peers) shifts beliefs, changes behavior, and improves performance. The personalized treatment increases the perceived effectiveness gap between practicing and rereading by 42 percent, reduces reliance on less effective strategies, increases engagement with retrieval-based effort, and raises subsequent exam scores by approximately 0.051 to 0.085 standard deviations. These joint movements in beliefs, behavior, and achievement indicate that mistaken beliefs about study-habit effectiveness generate a meaningful misallocation of academic effort.

The structural model clarifies why the behavioral response to information is both incomplete and heterogeneous. Belief updating does not translate one-for-one into changes in study habits because allocating effort is costly. Estimated effort-allocation costs substantially dampen the pass-through from information to behavior. At the same time, the value of personalized information is largest for students whose initial beliefs are most misaligned with the signal. For students in the first quartile of the baseline effectiveness-gap distribution, the personalized treatment produces changes in study behavior equivalent to a 16.5 percent reduction in the cost of adopting retrieval-based strategies. The model also shows that learning-by-doing alone is unlikely to close these belief gaps quickly: students who initially underuse effective strategies receive less experiential feedback about their returns, limiting their ability to learn their way out of distorted beliefs.

These findings have two broader implications. First, general advice may be too coarse to change behavior, whereas personalized feedback can make the relevant margin more credible and actionable by linking recommendations to students' own choices. Second, personalized information is likely to be more effective when paired with tools that lower the cost of implementing effective strategies. Structured practice platforms, targeted coaching, and AI-based study assistants can help students translate information into action by generating practice questions, organizing review sessions, and converting course material into retrieval-based exercises. More generally, the results suggest that improving academic performance may require not only encouraging students to study more, but helping them understand which forms of effort are more effective and making those forms of effort easier to adopt.

Finally, our findings are informative beyond the experimental setting. The intervention was evaluated in a natural, high-stakes environment within a population that resembles the typical U.S. undergraduate, its automated delivery implies a near-zero marginal cost, and the gains from improved study habits are non-rival. These features position personalized study-habit feedback, embedded directly in learning management systems, as a strong candidate for large-scale implementation.

References

Affonso, A., Aucejo, E. M., El Khoury, S., Fruehwirth, J. C., and Troop, M. (2025). Navigating college: Personalized information and student success.

Agarwal, P., Nunes, L., and Blunt, J. (2021). Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33:1409–1453.

Al-Ubaydli, O., List, J. A., and Suskind, D. L. (2017). What can we learn from experiments? understanding the threats to the scalability of experimental results. *American Economic Review*, 107(5):282–286.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, 11(1):48–83.

Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, F. H., and Sabarwal, S. (2020). How to improve education outcomes most efficiently?: A comparison of 150 interventions using the new learning-adjusted years of schooling metric. Technical report, World Bank.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Arcidiacono, P., Aucejo, E. M., and Spenner, K. (2012). What happens after enrollment? an analysis of the time path of racial differences in gpa and major choice. *IZA Journal of Labor Economics*, 1(1):1–24.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, 128(12):4475–4522.

Aucejo, E. and Tobin, Z. (2021). Assessing racial disparities in postsecondary education. *Federal Reserve Bank of Boston*, 5.

Aucejo, E. M., French, J., Araya, P. U., and Zafar, B. (2025). Understanding gaps in college outcomes by first-generation status. Technical report, National Bureau of Economic Research.

Aucejo, E. M. and James, J. (2019). Catching up to girls: Understanding the gender imbalance in educational attainment within race. *Journal of Applied Econometrics*, 34(4):502–525.

Barr, A. (2019). Fighting for Education: Veterans and Financial Aid. *Journal of Labor Economics*.

Beffy, M., Fougère, D., and Maurel, A. (2010). The effect of part-time work on post-secondary educational attainment: new evidence from french data.

Bettinger, E. P. and Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1):3–19.

Black, S., Denning, J. T., Dettling, L. J., Goodman, S., and Turner, L. J. (2020). Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well being. *NBER Working Paper*, (27658).

Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, 38:38–50.

Davis, J. M., Guryan, J., Hallberg, K., and Ludwig, J. (2017). The economics of scale-up. Technical report, National Bureau of Economic Research.

Delavande, A., Bono, E. D., Holford, A., Sen, S., and Lesic, V. (2023). Expectations about the productivity of effort and academic outcomes: Evidence from a randomized information intervention. *Available at SSRN 4483970*.

Denning, J. (2019). Born Under A Lucky Star: Financial Aid, College Completion, Labor Supply, and Credit Constraints. *Journal of Human Resources*, 54(3).

Denning, J. T. (2017). College on the cheap: Consequences of community college tuition reductions. *American Economic Journal: Economic Policy*, 9(2):155–188.

Denning, J. T. and Jones, T. R. (2021). Maxed Out? The Effect of Larger Student Loan Limits on Borrowing and Education Outcomes. *Journal of Human Resources*, 56(4).

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013). What works, what doesn't. *scientific american mind*, 24(4):46–53.

Eisenberg, D., Golberstein, E., and Hunt, J. B. (2009). Mental health and academic success in college. *The BE journal of economic analysis & policy*, 9(1).

Ekuni, R., Macacare, O. T., and Pompeia, S. (2022). Reducing the negative effects of multitasking on online or distance learning by using retrieval practice. *Scholarship of Teaching and Learning in Psychology.*

Ersoy, F. (2023). Effects of perceived productivity on study effort: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, 207:376–391.

Fergus, S., Heelan, A., Ibrahim, S., Oyman, H., Diaz-de Mera, Y., and Notario, A. (2021). Insights into study strategies and habits: a study with undergraduate students in spain and the u.k. *Journal of Chemical Education*, 98:3084–3089.

Fuster, A. and Zafar, B. (2023). Survey experiments on economic expectations. In *Handbook of Economic Expectations*, pages 107–130. Elsevier.

Hartwig, M. K. and Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic bulletin & review*, 19:126–134.

Hoxby, C. M. and Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, 105(5):514–517.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press.

Kalenkoski, C. M. and Pabilonia, S. W. (2010). Parental transfers, student achievement, and the labor supply of college students. *Journal of Population Economics*, 23(2):469–496.

Karpicke, J. and Blunt, J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331:772–775.

Ketchen Lipson, S., Abelson, S., Ceglarek, P., Phillips, M., and Eisenberg, D. (2019). Investing in student mental health: Opportunities & benefits for college leadership.

Kremer, M., Brannen, C., and Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130):297–300.

Larroucau, T. and Rios, I. (2022). Dynamic college admissions. *Working Paper.*

Larroucau, T., Rios, I. A., Fabre, A., and Neilson, C. (2025). College application mistakes and the design of information policies at scale. NBER Working Paper 34164, National Bureau of Economic Research.

Le Barbanchon, T., Ubfal, D., and Araya, F. (2020). The effects of working while in school: evidence from uruguayan lotteries. *Available at SSRN 3398385*.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Currency.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.

Maniadis, Z., Tufano, F., and List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1):277–290.

Mark, I. (2024). Integration of neuroscience into psychiatric training and practice: suggestions for implementation. *Bjpsych Bulletin*, pages 1–7.

Neyt, B., Omey, E., Verhaest, D., and Baert, S. (2019). Does student work really affect educational outcomes? a review of the literature. *Journal of Economic Surveys*, 33(3):896–921.

Nowell, C. and Alston, R. M. (2007). I thought i got an a! overconfidence across the economics curriculum. *The Journal of Economic Education*, 38(2):131–142.

Oreopoulos, P. and Petronijevic, U. (2023). The promises and pitfalls of using (mostly) low-touch coaching interventions to improve college student outcomes. *The Economic Journal*, 133(656):3034–3070.

Roediger, H. L. and Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Rury, D. and Carrell, S. E. (2023). Knowing what it takes: The effect of information about returns to studying on study effort and achievement. *Economics of Education Review*, 94:102400.

Scott-Clayton, J. (2012). What explains trends in labor supply among us undergraduates? *National Tax Journal*, 65(1):181–210.

Smith, A. M., Race, E., Davis, F. C., and Thomas, A. K. (2019). Using practice testing, public speaking, and source monitoring to examine the influences of learning strategies and stress on episodic memory. *Journal of Visualized Experiments*.

Stange, K. M. (2012). An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4(1):49–84.

Stinebrickner, R. and Stinebrickner, T. (2014). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644.

Stinebrickner, R. and Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics*, 21(2):473–491.

Stinebrickner, R. and Stinebrickner, T. R. (2004). Time-use and college outcomes. *Journal of Econometrics*, 121(1-2):243–269.

Stinebrickner, R. and Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Stinebrickner, T. and Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4):707–748.

Thomas, J. (2019). The signal quality of grades across academic fields. *Journal of Applied Econometrics*, 34(4):566–587.

Wallace, A., Elliot, A., and Rogge, R. (2022). Spontaneous use of retrieval and rereading: relation to achievement goals and exam performance. *Journal of Educational Psychology*, 114:1412–1426.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348.

A Appendix for Section 2

A.1 Sample construction

We administered surveys at the beginning and end of the Fall 2022 and Spring 2023 semesters across 34 classes. Three classes had no participation, primarily due to small enrollments. Among participating classes, 11 were online, and the remainder were in person. A total of 2,523 students completed either the first survey only or both surveys. We excluded 24 students who did not provide consent to participate in the study. In addition, the second survey did not include a consent form; students who took only the second survey are therefore excluded. We also excluded 11 students whose identifying information could not be matched to ASU administrative data. The remaining 2,488 students comprise the *All Students* sample.

For treatment assignment, we stratified on pre-specified variables and removed 158 students due to missing stratification variables or to maintain balance across treatment groups. Demographic disclosure was optional; stratification variables potentially missing include gender, first-generation college status, and earning a grade of B or higher on the first exam. Students who withdrew before completing the first exam were excluded from treatment assignment due to missing exam scores. This yields 2,328 students in the *Randomized* sample. Two-thirds of this sample came from long-term (15-week) semesters, primarily in-person courses; the remainder were enrolled in short-term (7.5-week) semesters, primarily online. Of the 2,328 students, 976 were in online courses.

For analyses that require responses in both surveys, we further excluded 104 students who failed attention checks and one student missing key analytic variables, leaving 2,225 students. Of these, 1,743 completed both surveys (the *Both Surveys* sample). Finally, usable CANVAS data are available for 1,683 of the 1,743 students who completed both surveys; the shortfall largely reflects course-level differences in grading practices (e.g., some courses do not assign homework or do not provide the inputs required for our CANVAS-based measures).

B Appendix for Section 3

B.1 Survey Questions

In this subsection, we describe the main questions used in the analysis.

- **Payoff to Overall Study-time:** *What do you think would be the score of a student on the final exam (assuming it's 30% of their course grade) if they started studying 2 (**or 2 weeks**) days before* *the exam and studied 10 hours (**or 25 hours**) in total considering that they ranked between their classmates on an IQ test in the: top 25%, top %50, bottom 50% or bottom 25%?*

- **Study habits:** *In the following questions, assume that your grade in the course before taking the final exam is 75/100; we are trying to see how many hours you think you should study for the final exam (assuming it's 30% of your course grade) in order to achieve the grade we specify (where 75/100 is a C, 85/100 is a B, and 95/100 is an A). Let's say it's 2 days (**or 2 weeks**) before the final exam, how many hours should you spend on the following tasks per day (per week excluding the last 2 days) before the exam if you want to get an 85/100 (**or 95/100**) on the final exam? a) [Insert Hours] Tasks that involve re-reading the material, taking notes, and highlighting, b) [Insert Hours] Tasks that involve recalling the material through discussing the material with others, trying to do assignments alone first, and doing practice exams.*
- **Effectiveness of study practices**: *How effective do you think each of the following study habits is at increasing your score on the exam? (scale from 1 to 10):*
  - *Summarizing the study material.*
  - *Asking "how" and "why" questions about the material and finding the answers on your own or by asking for help.*
  - *Using imagery for text learning such as graphs and diagrams.*
  - *Starting early for an exam and distributing study hours across multiple days.*
  - *Highlighting/underlining important points in a chapter.*
  - *Testing yourself through practice tests, flashcards, or quizzing yourself on the material.*
  - *Explaining the material to yourself.*
  - *Using mnemonics to memorize keywords and information (e.g., songs, poems, or similar strategies to retain the material).*
  - *Mixing up different types of questions and topics, rather than studying only one type of question and topic at a time.*
  - *Re-reading the material.*
- **Total Study Time:** *How many hours on average per week are you studying for this class?*

C Appendix Tables

Table C.1: Summary Statistics

	All Students (1)	Randomized (2)	Both Surveys (3)	Diff = (2) - (3) (4)
First generation	0.315	0.307	0.300	-0.007
	(0.465)	(0.461)	(0.458)	
Black	0.041	0.040	0.036	0.005
	(0.198)	(0.197)	(0.185)	
Hispanic	0.141	0.142	0.147	-0.005
	(0.348)	(0.349)	(0.354)	
Income (USD)	125,544	126,777	125,729	1,048
	(98,000)	(98,000)	(97,000)	
Female	0.459	0.457	0.480	-0.022
	(0.498)	(0.498)	(0.500)	
High school GPA	3.548	3.552	3.558	-0.006
	(0.439)	(0.437)	(0.443)	
Business and Economics Majors	0.773	0.773	0.767	0.005
	(0.419)	(0.419)	(0.423)	
Observations	2488	2328	1743	

Note: Entries are means with standard deviations in parentheses. Column (1) reports summary statistics for all the students who completed the first survey. Column (2) corresponds to the randomized sample, which is the subset used for treatment assignment. Column (3) reports the sample of students who completed both surveys and passed the attention check. Column (4) reports differences between the randomized and both-survey samples; none of the differences are statistically significant. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.2: Summary Statistics with Online and In-person Samples

	Survey Samples				Bus. & Econ. Majors Pop.			ASU Population		
	All Students (1)	Randomized (2)	Online (3)	In-person (4)	All (5)	Online (6)	In-person (7)	All (8)	Online (9)	In-person (10)
First Generation	0.315	0.307	0.365	0.266	0.216	0.311	0.170	0.263	0.324	0.217
	(0.465)	(0.461)	(0.482)	(0.442)						
Black	0.041	0.040	0.052	0.032	0.043	0.065	0.032	0.056	0.075	0.042
	(0.198)	(0.197)	(0.222)	(0.176)						
Hispanic	0.141	0.142	0.155	0.132	0.221	0.228	0.216	0.248	0.236	0.257
	(0.348)	(0.349)	(0.362)	(0.339)						
Income (USD)	125,544	126,777	105,438	141,979	142,735	82,079	180,959	103,135	64,858	138,300
	(97,886)	(98,270)	(87,286)	(102,763)						
Female	0.459	0.457	0.541	0.399	0.445	0.549	0.394	0.543	0.627	0.481
	(0.498)	(0.498)	(0.499)	(0.490)						
High School GPA	3.548	3.552	3.461	3.615	3.42	3.21	3.49	3.39	3.18	3.50
	(0.439)	(0.437)	(0.531)	(0.344)						
Business and Economics Majors	0.773	0.773	0.721	0.809				0.198	0.151	0.234
	(0.419)	(0.419)	(0.449)	(0.394)						
Observations	2,488	2,328	960	1,368	22,655	7,427	15,228	114,484	49,310	65,174

Note: Columns (1)–(4) summarize different samples of students enrolled in intermediate economics courses: (1) All students completing the initial survey; (2) the *Randomized* subset (eligible for randomization); (3) online respondents; and (4) in-person respondents. Columns (5)–(7) report population statistics for ASU students majoring in business and economics during 2023 (5) all, (6) online, (7) in-person. Columns (8)–(10) present the same statistics for the full ASU student population - (8) all, (9) online, (10) in-person. "First-generation" indicates neither parent completed college. "Income (USD)" is annual household income in U.S. dollars. High school GPA is measured on a 4.0 scale. Proportions are reported as fractions between 0 and 1. Standard deviations for columns (1)-(4) are in parentheses. Observations for each column appear in the final row.

Table C.3: Treatment Effect on Total Effort

	Total Study Time$_{t+1}$		Total Seconds Online$_{t+1}$	
	(1)	(2)	(3)	(4)
T1	-1.422^{***}	-1.655^{***}	-178.720	-200.832
	(0.283)	(0.296)	(165.214)	(165.913)
T2	-0.353	-0.387	-97.387	-102.837
	(0.399)	(0.420)	(149.779)	(141.602)
Total Study Time$_t$	0.427^{***}	0.429^{***}		
	(0.048)	(0.045)		
Total Seconds Online$_t$			0.502^{***}	0.503^{***}
			(0.055)	(0.050)
Std. First Exam$_t$	-0.268	-0.268	124.705	123.447
	(0.322)	(0.307)	(103.761)	(92.949)
Std. First Exam$_t^2$	-0.110	-0.107	15.604	15.649
	(0.072)	(0.068)	(17.582)	(15.816)
Female	0.152	0.193	91.690	95.137
	(0.380)	(0.360)	(283.333)	(251.443)
Black/African American	0.002	0.008	-564.342	-572.205^{*}
	(1.057)	(0.979)	(343.654)	(306.975)
Hispanic	0.237	0.246	-242.237	-248.929
	(0.395)	(0.387)	(271.055)	(236.657)
Observations	1743	1743	817	817
IV (received)		X		X
First-stage F (T1 received)		1292.19		1758.89
First-stage F (T2 received)		1036.60		2372.33
Restricted Sample	Yes	Yes	Yes	Yes
Both Surveys	Yes	Yes	Yes	Yes
Dep. Var. Mean	11.88	11.88	4972	4972

Standard errors in parentheses clustered at the class level. Total study time is measured as the self-reported number of hours per week the student studies for this class. The precise wording of the survey questions is provided in Appendix Section B.1. Total Seconds Online is the average weekly time, in seconds, that a student spends on CANVAS. This variable is defined only for online courses. Columns (2) and (4) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. For column (2) in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.020) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.900 (s.e. 0.020) and the coefficient on the $T1$ assignment indicator is 0.001 (s.e. 0.003). For column (4) in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.886 (s.e. 0.015) and the coefficient on the $T2$ assignment indicator is 0.008 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.931 (s.e. 0.015) and the coefficient on the $T1$ assignment indicator is 0.008 (s.e. 0.005). Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income, B or above indicator for the first exam, reported study habits effectiveness scale (0 to 10), indicator for scenario type in the first and second survey, an interaction term between the effort level that students received in the first and second survey, Female, Black/African American indicator, Hispanic indicator, and lagged outcome variables. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.4: Perceived Effectiveness of Study Strategies (Randomized Sample)

Panel A: Perceived Effectiveness: Retrieval Practices (Scale 1–10)	
Quizzing	8.015
	(0.044)
Explaining	7.058
	(0.050)
Asking questions	6.692
	(0.049)
Retrieval practice average	7.255
	(0.035)
Panel B: Perceived Effectiveness: Rereading Practices (Scale 1–10)	
Rereading	6.263
	(0.054)
Highlighting/underlining	6.175
	(0.053)
Rereading practice average	6.219
	(0.044)
Panel C: Perceived Effectiveness: Cramming Practices (Scale 1–10)	
Start studying early	7.306
	(0.050)
Panel D: Perceived Effectiveness: Other Strategies (Scale 1–10)	
Summarizing the study material	6.597
	(0.044)
Imagery	6.351
	(0.051)
Mnemonics	5.005
	(0.058)
Mixing topics	6.322
	(0.053)
Observations	2328

Notes: This table reports the perceived effectiveness of different study strategies for the randomized sample. Students rated each strategy on a scale from 1 to 10, where 1 denotes low effectiveness and 10 denotes high effectiveness. Standard errors are reported in parentheses.

Table C.5: Study Habits and Beliefs in Terms of Perceived Effectiveness of Habits

	Share Reread (1)	Share Cram (2)	Payoff to Study-time (3)	Payoff to Study-time (4)
Perceived Effectiveness: Retrieval Practices				
Quizzing	-0.010^{***}		0.064^{***}	
	(0.002)		(0.012)	
Explaining	-0.002		0.040^{***}	
	(0.002)		(0.010)	
Asking questions	-0.004^{***}		0.059^{***}	
	(0.001)		(0.011)	
Perceived Effectiveness: Rereading Practices				
Rereading	0.014^{***}		0.007	
	(0.002)		(0.010)	
Highlighting/underlining	0.011^{***}		0.010	
	(0.001)		(0.010)	
Perceived Effectiveness: Cramming Practices				
Start studying early		-0.001	0.036^{***}	
		(0.001)	(0.011)	
Study Habits				
Share Reread				-0.782^{***}
				(0.188)
Share Cram				-0.514^{*}
				(0.281)
Observations	2328	2328	9312	9312
Controls	X	X	X	X
Randomized Sample	Yes	Yes	Yes	Yes
Both Surveys	No	No	No	No

Note: Standard errors (in parentheses) are clustered at the classroom level in columns (1)–(2) and at the individual level in columns (3)–(4). Each reported coefficient comes from a separate regression that includes the same set of controls and uses the indicated outcome variable. The outcomes are: (1) *Share Reread*, the share of reported study hours devoted to rereading practices; (2) *Share Cram*, the share of reported study hours allocated to the two days before the exam; and (3)–(4) *Payoff to Study-time*, defined as the reported score divided by the effort level, interpreted as the average score increase from one additional hour of studying. *Payoff to Study-time* has four observations per student, corresponding to the four scenarios each student answered. The perceived-effectiveness measures are the independent variables in the top panels and take values from 0 to 10. Sample means are: Share Cram = 0.336, Share Reread = 0.434, and Payoff to Study-time = 4.55. All regressions include the following controls: gender; an indicator for taking the survey more than once; average hours studied per week; self-reported ability indicator; high school GPA; classroom fixed effects; household income; STEM major indicator; indicator for working full time; number of courses taken during the survey semester; first-generation status; total semesters enrolled at ASU (Fall and Spring); total hours completed at ASU; LEAD status (ASU Success program); cohort type (new student or transfer); race indicators; survey completion time (seconds); and indicators for missing values in high school GPA, major, number of courses taken, first-generation status, and income (demographics were optional). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.6: Impact of Payoff to Overall Study-time on Performance and Misprediction of Classroom Outcomes

	Std. First Exam (1)	Exp.Grade-First Exam (2)	Abs(Exp.Grade-First Exam) (3)
Payoff to Study-time	0.076^{***}	-0.499^{**}	-0.727^{***}
	(0.012)	(0.221)	(0.174)
Female	-0.249^{***}	3.417^{***}	1.395^{*}
	(0.049)	(0.888)	(0.728)
Black/African American	-0.493^{***}	7.668^{***}	8.185^{***}
	(0.140)	(2.593)	(2.034)
Hispanic	-0.080	0.960	0.965
	(0.079)	(1.307)	(1.063)
Observations	9312	9312	9312
Controls	X	X	X
Randomized Sample	Yes	Yes	Yes
Both Surveys	No	No	No

Note: Standard errors in parentheses clustered at the individual level. Grades are standardized at the classroom level. "Std. First Exam" refers to the standardized score students received in the first exam of the semester. The mean first exam grade is 70. "Exp.Grade-First Exam" represents the difference between the expected score of students and the actual first exam score and "Abs(Exp.Grade-First Exam)" is the absolute value of the former variable. The means of these variables are 15.56 and 19.78. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, classroom indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income (students were not forced to respond on these demographic questions in the survey), and scenario type. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.7: Covariate Balance Test: Two-Sample t-test between Treatment and Control Group

	General Treatment			Personalized Treatment		
Covariate	Diff (1)	SE (2)	Observations (3)	Diff (4)	SE (5)	Observations (6)
Panel A: Randomized Sample						
B or Above	0.003	(0.024)	1558	0.007	(0.024)	1553
Female	-0.001	(0.026)	1558	-0.002	(0.026)	1553
Minority	0.009	(0.022)	1558	0.011	(0.022)	1553
HSGPA	-0.022	(0.022)	1558	0.003	(0.025)	1553
Hrs Study per Week	0.149	(0.339)	1558	0.243	(0.337)	1553
Share Reread	0.005	(0.007)	1558	0.006	(0.007)	1552
Share Cram	-0.001	(0.005)	1558	0.003	(0.005)	1552
Std. First Exam	0.052	(0.061)	1558	0.036	(0.060)	1553
Payoff to Study-time	-0.224^{***}	(0.064)	6232	0.018	(0.062)	6212
Panel B: Both Surveys Sample						
B or Above	0.009	(0.028)	1177	0.015	(0.028)	1157
Female	0.008	(0.029)	1177	0.011	(0.030)	1157
Minority	0.003	(0.025)	1177	0.014	(0.025)	1157
HSGPA	-0.019	(0.025)	1177	0.005	(0.029)	1157
Hrs Study per Week	0.211	(0.389)	1177	0.452	(0.395)	1157
Share Reread	0.004	(0.008)	1177	0.005	(0.008)	1157
Share Cram	0.002	(0.006)	1177	0.007	(0.006)	1157
Std. First Exam	0.025	(0.068)	1177	0.025	(0.068)	1157
Payoff to Study-time	-0.267^{***}	(0.073)	4708	-0.029	(0.072)	4628

Note: This table reports covariate balance tests based on two-sample t-tests comparing the control group to the general treatment group (columns 1–3) and to the personalized treatment group (columns 4–6). Columns (1) and (4) report the difference in mean outcomes (Treatment mean − Control mean) for each covariate; columns (2) and (5) report the associated standard errors; and columns (3) and (6) report the number of observations used in each comparison. Panel A is the full randomized sample, and Panel B is the subsample of treated students who passed the attention check and completed both surveys. For all covariates except "Payoff to Study-time" there is no statistically significant difference, indicating good balance. "Payoff to Study-time" ranges from 0 to 10 with a mean of 4.53 and a standard deviation of 2.48, so the estimated differences are small in magnitude. We further assess balance using the normalized difference test by Imbens and Rubin (2015). The normalized differences in the Payoff to Study-time between the control group and the general information group are 0.063 in the *Randomized Sample* and 0.069 in the *Both-Surveys Sample*. Both values are well below the 0.25 threshold suggested by Imbens and Rubin (2015), indicating that the observed imbalance is unlikely to be consequential. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.8: Treatment Effect on Student Classroom Outcomes

	Std. Valid Score$_{t+1}$		Mean Std. Valid Scores$_{t+1}$	
	(1)	(2)	(3)	(4)
T1	-0.004	-0.005	-0.016	-0.018
	(0.022)	(0.025)	(0.020)	(0.022)
T2	0.076**	0.084**	0.046	0.052*
	(0.032)	(0.033)	(0.028)	(0.030)
Std. First Exam$_t$	0.412***	0.411***	0.321***	0.321***
	(0.060)	(0.057)	(0.046)	(0.044)
Std. First Exam$_t^2$	0.059***	0.059***	0.045***	0.045***
	(0.008)	(0.008)	(0.009)	(0.009)
Female	−0.106***	−0.105***	−0.073**	−0.072**
	(0.031)	(0.030)	(0.030)	(0.029)
Black/African American	-0.117	-0.116	-0.005	-0.005
	(0.112)	(0.106)	(0.064)	(0.060)
Hispanic	−0.142***	−0.143***	−0.095**	−0.096***
	(0.046)	(0.045)	(0.037)	(0.036)
Observations	1743	1743	1743	1743
IV (received)		X		X
First-stage F (T1 received)		1200.64		1200.64
First-stage F (T2 received)		1051.68		1051.68
Randomized Sample	Yes	Yes	Yes	Yes
Both Surveys	Yes	Yes	Yes	Yes

Note: Clustered by class standard errors in parentheses. *Std. Valid Score*$_{t+1}$ is the class-standardized value of the first exam score valid toward the final grade after treatment, while *Mean Std. Valid Scores*$_{t+1}$ is the mean of the standardized scores of all valid post-treatment exams. *Std. First Exam*$_t$ is the pre-treatment standardized score on the first exam, and *Std. First Exam*$_t^2$ is its squared term. Columns (2) and (4) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. First-stage coefficients are as follows: in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.020) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.899 (s.e. 0.021) and the coefficient on the $T1$ assignment indicator is 0.000 (s.e. 0.003). Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken in the semester, first generation status, income (students were not forced to respond on these demographic questions in the survey), initially reported study habits effectiveness scale (0 to 10), B or above indicator for the first exam. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.9: Treatment Effect on Student Classroom Outcomes - Heterogeneity

	Standardized Valid Score$_{t+1}$		
	(1)	(2)	(3)
T1	-0.017	-0.006	-0.003
	(0.039)	(0.029)	(0.032)
T2	0.060	0.091**	0.070
	(0.038)	(0.043)	(0.041)
T1 × First Gen	0.105		
	(0.070)		
T2 × First Gen	0.082		
	(0.074)		
T1 × Female		0.045	
		(0.074)	
T2 × Female		-0.014	
		(0.090)	
T1 × B or above			0.055
			(0.049)
T2 × B or above			0.045
			(0.061)
Female	-0.115^{***}	-0.126^{**}	-0.115^{***}
	(0.036)	(0.050)	(0.037)
First Gen	-0.122^{***}	-0.059^{**}	-0.060^{**}
	(0.041)	(0.024)	(0.024)
B or above	-0.006	-0.003	-0.035
	(0.076)	(0.076)	(0.077)
Std. First Exam$_t$	0.453***	0.452***	0.452***
	(0.057)	(0.057)	(0.057)
Std. First Exam$_t^2$	0.044***	0.044***	0.044***
	(0.010)	(0.010)	(0.010)
Observations	2328	2328	2328
Randomized Sample	Yes	Yes	Yes
Both Surveys	No	No	No

Note: Clustered by class standard errors in parentheses. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self-reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first-generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first-generation status, income (students were not forced to respond on these demographic questions in the survey), reported study habits effectiveness scale (0 to 10), and B-or-above indicator for the first exam. $Standardized\ Valid\ Score_{t+1}$ is the standardized-by-class score students receive on the first exam valid toward their final grade post-treatment. Valid scores refer to scores that count toward the final grade. $Std.\ First\ Exam_t$ is the standardized score by class on the first exam before treatment, and $Std.\ First\ Exam_t^2$ is its squared term. Column (1) includes interactions with first-generation status. Column (2) includes interactions with female status. Column (3) includes interactions with the "B or above" indicator for students scoring B or above on the first exam. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.10: Treatment Effect on Student class Outcomes: In Person

	Std. Valid Score$_{t+1}$	Mean Std. Valid Scores$_{t+1}$
	(1)	(2)
T1	0.022	0.010
	(0.044)	(0.037)
T2	0.109^{***}	0.069^{**}
	(0.027)	(0.029)
Std. First Exam$_t$	0.405^{***}	0.333^{***}
	(0.078)	(0.071)
Std. First Exam$_t^2$	0.050^{**}	0.041^{*}
	(0.022)	(0.023)
Female	-0.183^{***}	-0.107^{***}
	(0.021)	(0.026)
Black/African American	-0.066	-0.060
	(0.083)	(0.093)
Hispanic	-0.031	-0.018
	(0.074)	(0.056)
Observations	1368	1368
Randomized Sample	Yes	Yes
Both Surveys	No	No
Mode	In person	In person

Clustered by class standard errors in parentheses. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status, cohort type, duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken in the semester, first generation status, income, initially reported study habits effectiveness scales, and a B-or-above indicator for the first exam. $Std.\ Valid\ Score_{t+1}$ is the class-standardized value of the first exam score valid toward the final grade after treatment, while $Mean\ Std.\ Valid\ Scores_{t+1}$ is the mean of the standardized scores of all valid post-treatment exams. $Std.\ First\ Exam_t$ is the pre-treatment standardized score on the first exam, and $Std.\ First\ Exam_t^2$ is its squared term. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.11: Treatment Effect on Student class Outcomes: Online

	Std. Valid Score$_{t+1}$ (1)	Mean Std. Valid Scores$_{t+1}$ (2)
T1	0.012	0.012
	(0.054)	(0.050)
T2	0.074	0.045
	(0.060)	(0.056)
Std. First Exam$_t$	0.530***	0.429***
	(0.060)	(0.055)
Std. First Exam$_t^2$	0.049***	0.038***
	(0.007)	(0.005)
Female	-0.042	-0.046
	(0.077)	(0.052)
Black/African American	-0.250	-0.107
	(0.146)	(0.091)
Hispanic	-0.090^{*}	-0.032
	(0.042)	(0.036)
Observations	960	960
Randomized Sample	Yes	Yes
Both Surveys	No	No
Mode	Online	Online

Clustered by class standard errors in parentheses. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status, cohort type, duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken in the semester, first generation status, income, initially reported study habits effectiveness scales, and a B-or-above indicator for the first exam. $Std.\ Valid\ Score_{t+1}$ is the class-standardized value of the first exam score valid toward the final grade after treatment, while $Mean\ Std.\ Valid\ Scores_{t+1}$ is the mean of the standardized scores of all valid post-treatment exams. $Std.\ First\ Exam_t$ is the pre-treatment standardized score on the first exam, and $Std.\ First\ Exam_t^2$ is its squared term. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.12: Treatment Effect on Student Beliefs

	Effectiveness Rereading_{t+1}		Effectiveness Practicing_{t+1}	
	(1)	(2)	(3)	(4)
T1	-0.057	-0.066	0.071	0.083
	(0.089)	(0.097)	(0.081)	(0.089)
T2	-0.201	-0.223^{*}	0.068	0.075
	(0.123)	(0.133)	(0.092)	(0.096)
Effectiveness Rereading_t	0.374^{***}	0.374^{***}		
	(0.034)	(0.032)		
Effectiveness Practicing_t			0.462^{***}	0.462^{***}
			(0.043)	(0.041)
Std. First Exam_t	-0.028	-0.027	0.203^{***}	0.202^{***}
	(0.072)	(0.068)	(0.058)	(0.055)
Std. First Exam_t^2	-0.004	-0.004	0.031	0.031
	(0.016)	(0.015)	(0.023)	(0.022)
Female	-0.241	-0.240	0.102^{*}	0.100^{**}
	(0.238)	(0.224)	(0.052)	(0.049)
Black/African American	0.048	0.047	-0.194	-0.194
	(0.226)	(0.213)	(0.145)	(0.139)
Hispanic	0.269^{**}	0.272^{**}	0.083	0.083
	(0.112)	(0.107)	(0.095)	(0.088)
Observations	1743	1743	1743	1743
IV (received)		X		X
First-stage F (T1 received)		1200.64		1200.64
First-stage F (T2 received)		1051.68		1051.68
Randomized Sample	Yes	Yes	Yes	Yes
Both Surveys	Yes	Yes	Yes	Yes
Dep. Var. Mean (Control)	6.03	6.03	6.74	6.74

Standard errors in parentheses clustered at the class level. Effectiveness of practicing is defined as the average perceived effectiveness of quizzing, explaining class material, and asking questions. Effectiveness of practicing and effectiveness of rereading are measured on a 1–10 scale. Columns (2) and (4) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. First-stage coefficients are as follows: in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.859 (s.e. 0.020) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.003); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.899 (s.e. 0.021) and the coefficient on the $T1$ assignment indicator is 0.000 (s.e. 0.003). Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income, B or above indicator for the first exam, reported study habits effectiveness scale (0 to 10), indicator for scenario type in the first and second survey, an interaction term between the effort level that students received in the first and second survey, Std. First Exam_t, Std. First Exam_t^2, and lagged outcome variables. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.13: Treatment Effects on Homework and Practice Exam Taking Behavior

	Std. Homework Score$_{t+1}$		Practice Exams			
			Valid		Share *Valid*	
	(1)	(2)	(3)	(4)	(5)	(6)
T1	0.019	0.022	0.080**	0.096***	0.052	0.062**
	(0.046)	(0.051)	(0.036)	(0.030)	(0.037)	(0.031)
T2	0.131**	0.145***	0.092**	0.096***	0.070	0.073**
	(0.053)	(0.054)	(0.045)	(0.032)	(0.052)	(0.037)
Std. Homework Score$_t$	0.538***	0.538***				
	(0.071)	(0.067)				
Practice Exam Valid$_t$			-0.172	-0.206^{**}		
			(0.136)	(0.100)		
Share Valid Practice Exams$_t$					-0.601^{***}	-0.623^{***}
					(0.184)	(0.130)
Std. First Exam$_t$	-0.002	-0.003	0.043	0.046	0.066	0.069**
	(0.043)	(0.041)	(0.045)	(0.033)	(0.048)	(0.034)
Std. First Exam$_t^2$	-0.005	-0.006	0.003	0.003	-0.001	-0.001
	(0.006)	(0.006)	(0.004)	(0.003)	(0.004)	(0.003)
Female	-0.020	-0.020	0.021	0.018	0.020	0.018
	(0.054)	(0.052)	(0.037)	(0.026)	(0.043)	(0.029)
Black/African American	0.030	0.032	0.136**	0.131***	0.089	0.085*
	(0.075)	(0.070)	(0.066)	(0.045)	(0.071)	(0.048)
Hispanic	0.114**	0.112***	0.061	0.075**	0.039	0.048
	(0.040)	(0.038)	(0.045)	(0.031)	(0.045)	(0.031)
Observations	1683	1683	141	141	141	141
IV		X		X		X
First-stage F (T1 received)		1008.31		76.48		76.48
First-stage F (T2 received)		1033.20		1271.71		1271.71
Both Surveys	Yes	Yes	Yes	Yes	Yes	Yes
Randomized Sample	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors for columns (1) and (2) are in parentheses and clustered at the class level. Columns (3) to (6) have robust standard errors that are not clustered at the class level because the number of available class clusters drops substantially in this subsample. Columns (1) and (2) report the impact of treatment on the standardized average homework score by class post treatment. The number of observations is 1,683, which is smaller than in Table 6 (1,743) because the data for these variables were collected in later years, when some students had graduated or dropped out and we no longer had access to their scores. In addition, four classes do not offer homework to students, usually higher-level courses that rely more on projects. Column (2) reports IV estimates. Columns (3) and (4) report the impact of treatment on "Practice Exams Valid," which takes a value of 1 if a student completed the practice exam for the first exam that counts toward their final grade post treatment. Columns (5) and (6) report the impact of treatment on "Practice Exams Share Valid," defined as the share of completed practice exams among those that count toward the final grade post treatment. Robust standard errors are reported in parentheses, as there are only two classrooms offering practice exams, which is too few clusters for clustered standard errors. The number of observations is 141 because only a limited number of classes offer practice exams. Columns (2), (4), and (6) report IV estimates where the endogenous variable is an indicator for whether the student read the email they were sent. Assignment to treatment is used as an instrument. Reported F-statistics correspond to the first-stage regression of email reading on treatment assignment. For column (2), in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.860 (s.e. 0.021) and the coefficient on the $T2$ assignment indicator is 0.003 (s.e. 0.004); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.898 (s.e. 0.021) and the coefficient on the $T1$ assignment indicator is 0.001 (s.e. 0.003). For columns (4) and (6), in the first-stage regression for receiving $T1$, the coefficient on the $T1$ assignment indicator is 0.826 (s.e. 0.074) and the coefficient on the $T2$ assignment indicator is -0.034 (s.e. 0.055); in the first-stage regression for receiving $T2$, the coefficient on the $T2$ assignment indicator is 0.991 (s.e. 0.020) and the coefficient on the $T1$ assignment indicator is 0.013 (s.e. 0.017). Additional controls: indicator for students taking the survey more than once, average hours studied per week, self-reported ability indicator, high school GPA, classroom indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first-generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first-generation status, income, reported study habits effectiveness scale (0 to 10), B-or-above indicator for the first exam, Std. First Exam$_t$, Std. First Exam$_t^2$, and lag of outcome variables collected pre-treatment. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.14: Study Habits and Homework Scores

	Std Homework Score		
	Baseline (1)	Endline (2)	Pooled (3)
Share Reread	-0.211	-0.195	-0.254*
	(0.172)	(0.173)	(0.136)
Observations	2166	1677	3354
Controls	Yes	Yes	Yes
Both Surveys	No	Yes	Yes

Note: The dependent variable is the standardized homework score. The main regressor is the share of study time allocated to rereading. Column (1) uses baseline (Survey 1) rereading and homework outcomes measured in the pre-intervention period. Column (2) uses endline (Survey 2) rereading and homework outcomes measured in the post-intervention period, restricting the sample to students who completed both surveys. Column (3) pools the baseline and endline observations for students who completed both surveys. Standard errors are clustered at the class level in Columns (1) and (2), and at the individual level in Column (3). Controls: gender, race indicators, indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income (students were not forced to respond on these demographic questions in the survey), reported study habits effectiveness scale (0 to 10), B or above indicator for the first exam. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.15: Covariate Balance Test: Two-Sample t-test between Treatment and Control Group

	General Treatment			Personalized Treatment		
Covariate	Diff (1)	SE (2)	Observations (3)	Diff (4)	SE (5)	Observations (6)
Panel A: Homework Sample						
B or Above	0.013	(0.028)	1140	0.018	(0.029)	1115
Female	0.004	(0.030)	1140	0.007	(0.030)	1115
Minority	0.000	(0.026)	1140	0.013	(0.026)	1115
HSGPA	-0.022	(0.026)	1140	-0.007	(0.029)	1115
Hrs Study per Week	0.235	(0.395)	1140	0.446	(0.403)	1115
Share Reread	0.005	(0.008)	1140	0.005	(0.008)	1115
Share Cram	0.000	(0.006)	1140	0.005	(0.006)	1115
Std. First Exam	0.037	(0.070)	1140	0.022	(0.069)	1115
Payoff to Study-time	-0.268***	(0.074)	4560	-0.030	(0.073)	4460
Panel B: Practice Exam Sample						
B or Above	-0.026	(0.095)	100	-0.065	(0.099)	89
Female	0.000	(0.101)	100	0.061	(0.107)	89
Minority	0.005	(0.093)	100	0.044	(0.098)	89
HSGPA	-0.049	(0.085)	100	0.202	(0.146)	89
Hrs Study per Week	-0.768	(1.419)	100	-1.315	(1.532)	89
Share Reread	-0.021	(0.032)	100	0.030	(0.029)	89
Share Cram	0.002	(0.025)	100	0.010	(0.024)	89
Std. First Exam	-0.124	(0.309)	100	-0.332	(0.280)	89
Payoff to Study-time	-0.885***	(0.244)	400	-0.471*	(0.263)	356

This table reports covariate balance tests based on two-sample t-tests comparing the control group to the general treatment group (columns 1–3) and to the personalized treatment group (columns 4–6). Columns (1) and (4) report the difference in mean outcomes (Treatment mean − Control mean) for each covariate; columns (2) and (5) report the associated standard errors; and columns (3) and (6) report the number of observations used in each comparison. Panel A uses the homework sample from Table C.13; Panel B uses the subset of students with non-missing practice-exam outcomes reported in that same table. While most covariates are balanced, "Payoff to Study-time"—measured on a scale of 0 to 10 (Mean = 4.53)—shows some statistically significant differences. To evaluate the magnitude of these differences independently of sample size, we follow the normalized difference approach suggested by Imbens and Rubin (2015). In the Homework Sample, standardized differences are 0.075 and 0.009 for the general and personalized treatments, respectively, both well below the conventional 0.25 threshold used to identify meaningful imbalance. In the Practice Exam Sample, the standardized difference for the general treatment group is 0.257. While this value is slightly above the 0.25 diagnostic threshold, it represents a marginal deviation, whereas the personalized treatment remains well-balanced at 0.118. $^{**}p<0.05$, $^{***}p<0.01$.

Table C.16: Treatment Effect on Study Habits By Message Sent

	Share Reread$_{t+1}$	Share Cram$_{t+1}$
	(1)	(2)
T1	-0.005	-0.000
	(0.010)	(0.008)
T2	-0.010	-0.003
	(0.007)	(0.009)
Above Median Share Reread$_t$	0.057^{***}	
	(0.010)	
Above Median Share Cram$_t$		0.028^{***}
		(0.008)
T1 × Above Median Share Reread$_t$	-0.011	
	(0.011)	
T2 × Above Median Share Reread$_t$	-0.015	
	(0.012)	
T1 × Above Median Share Cram$_t$		-0.006
		(0.012)
T2 × Above Median Share Cram$_t$		-0.024^{*}
		(0.013)
Std. First Exam$_t$	-0.009^{**}	-0.009^{*}
	(0.004)	(0.005)
Std. First Exam$_t^2$	-0.002	-0.002^{***}
	(0.001)	(0.001)
Female	-0.004	-0.003
	(0.009)	(0.007)
Black/African American	-0.007	0.013
	(0.011)	(0.013)
Hispanic	-0.011	0.007
	(0.008)	(0.009)
Observations	1737	1737

Clustered by class standard errors in parentheses. Six individuals had a denominator of 0 while computing the post-treatment shares. Therefore, they are not included in the analysis and the number of observations drops from 1,743 to 1,737. "Above Median Share Reread" and "Above Median Share Cram" indicate whether a student reported a proportion of rereading or cramming, respectively, that exceeded the median for their classroom or group of classrooms. Based on this classification, the treatment message described in Appendix 4 informed students whether their rereading or cramming behavior was above or below that of the median student, with the message varying accordingly. Additional controls: indicator for students taking the survey more than once, average hours studied per week, self reported ability indicator, high school GPA, classroom indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income, reported study habits effectiveness scale (0 to 10), B or above indicator for the first exam, Std. First Exam$_t$, Std. First Exam$_t^2$, gender, Black/African American indicator, and Hispanic indicator. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.17: Simulations – Policy Effects

	All Students			*First Quartile*		
	Baseline	$\bar{S}_p^{PT} = 7.569$ $\bar{S}_r^{PT} = 5.985$	$\gamma_p =\downarrow 3.5\%$	Baseline	$\bar{S}_p^{PT} = 7.569$ $\bar{S}_r^{PT} = 5.985$	$\gamma_p =\downarrow 16.5\%$
	(1)	(2)	(3)	(4)	(5)	(6)
$\mu_1^{\alpha_p}$	7.355	7.355	7.355	6.605	6.605	6.605
$\mu_1^{\alpha_r}$	6.443	6.443	6.443	8.751	8.751	8.751
Sh_1^p	0.564	0.564	0.573	0.500	0.500	0.541
$\mu_2^{\alpha_p}$	7.355	7.421	7.355	6.605	7.168	6.605
$\mu_2^{\alpha_r}$	6.443	5.914	6.443	8.751	6.554	8.751
Sh_2^p	0.568	0.577	0.577	0.506	0.547	0.547
$\%\Delta\mu_2^{\alpha_p}$	–	0.9	0.0	–	8.5	0.0
$\%\Delta\mu_2^{\alpha_r}$	–	-8.2	0.0	–	-25.1	0.0
$\%\Delta Sh_2^p$	–	1.6	1.6	–	8.1	8.1

This table reports simulation results for all students (Columns 1–3) and for students in the first quartile of the beliefs distribution (Columns 4–6). Columns (1) and (4) report baseline values for the scenario in which students receive no information signals; consequently, subjective beliefs remain constant between period 1 and period 2. Columns (2) and (5) simulate the impact of the personalized treatment using the estimated mean signals. Columns (3) and (6) present a counterfactual scenario that holds beliefs constant at baseline levels but reduces the cost parameter γ_p by 3.5% and 16.5%, respectively, to match the treatment-induced change in practice shares. Rows $\mu_1^{\alpha_p}$ and $\mu_1^{\alpha_r}$ represent mean subjective beliefs regarding the perceived marginal benefit (perceived effectiveness) of practicing and rereading recovered from the survey data. Effort shares (Sh_1^p, Sh_2^p) and period-2 beliefs ($\mu_2^{\alpha_j}$) are obtained as the average across all individuals of the model-implied values given each individual's prior beliefs and total study time. Rows prefixed with $\%\Delta$ denote the percentage change relative to the baseline scenario in Columns (1) and (4).

D Appendix Figures

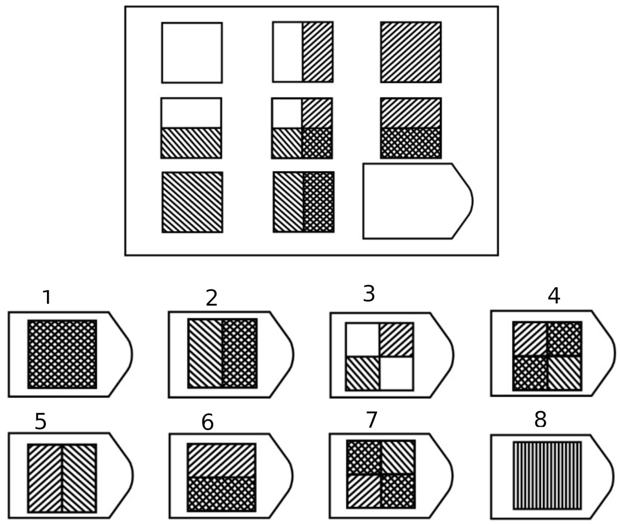

Figure D.1: Raven Test for Ability Measure

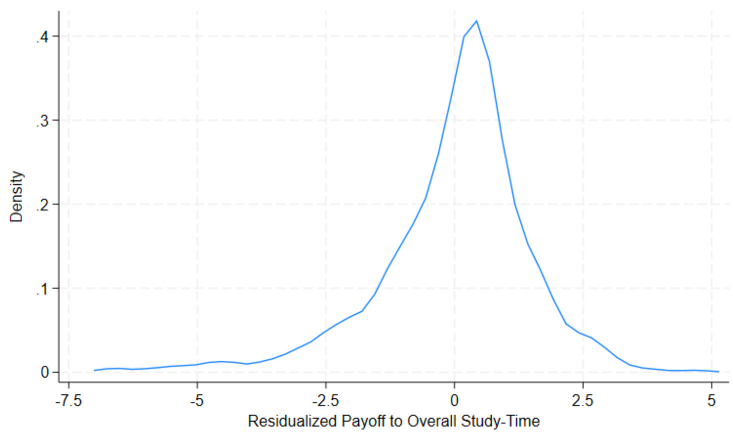

Figure D.2: Residualized Payoff to Overall Study-time

Note: The figure shows the kernel density of residualized payoff to overall study-time, controlling for scenario, ability level, and course fixed effects.

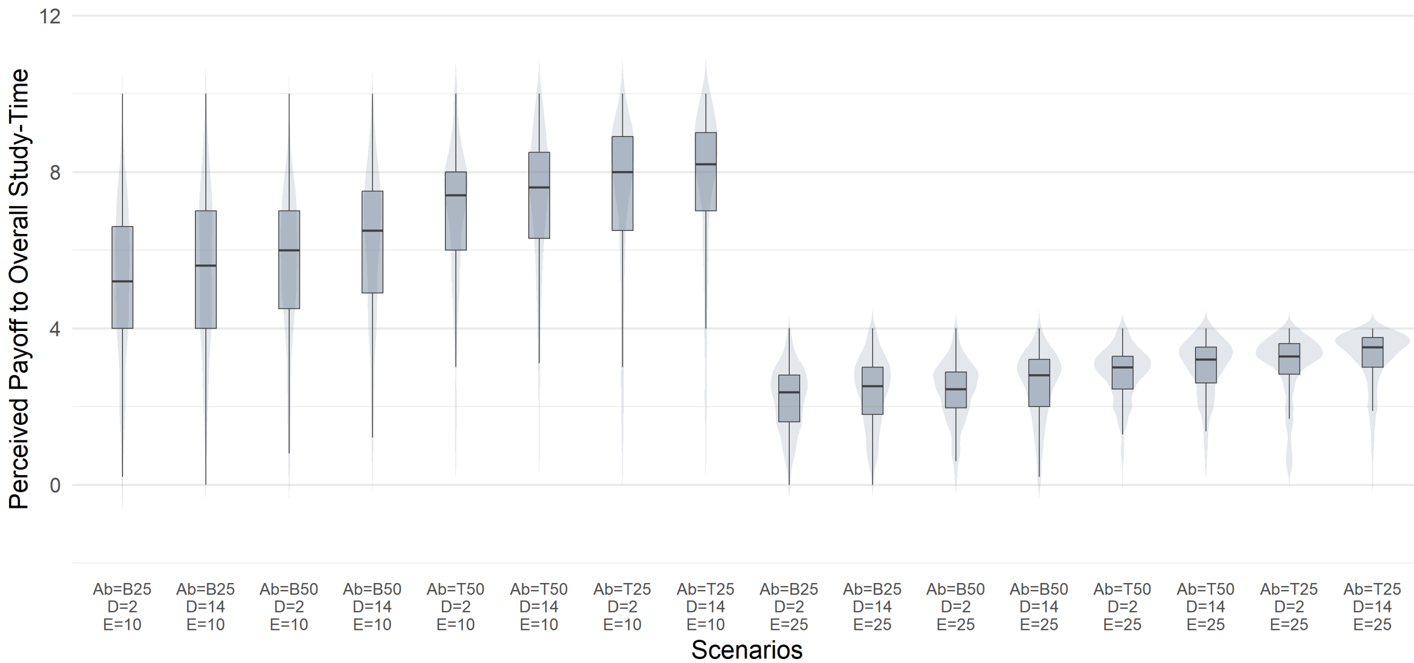

Figure D.3: Payoff to Overall Study-time by Scenarios

Note: This figure plots the distribution of students' perceived payoff to overall study-time across all scenarios. Students predicted a peer's final-exam score on a 0-100 scale with letter-grade anchors. Ab denotes the ability level, D denotes the number of days before the exam when the peer starts studying, and E denotes total study time in hours. We compute perceived payoffs per hour as the predicted score divided by total study hours. Violin plots show the full distribution of responses, and the overlaid boxplots report the median and interquartile range.

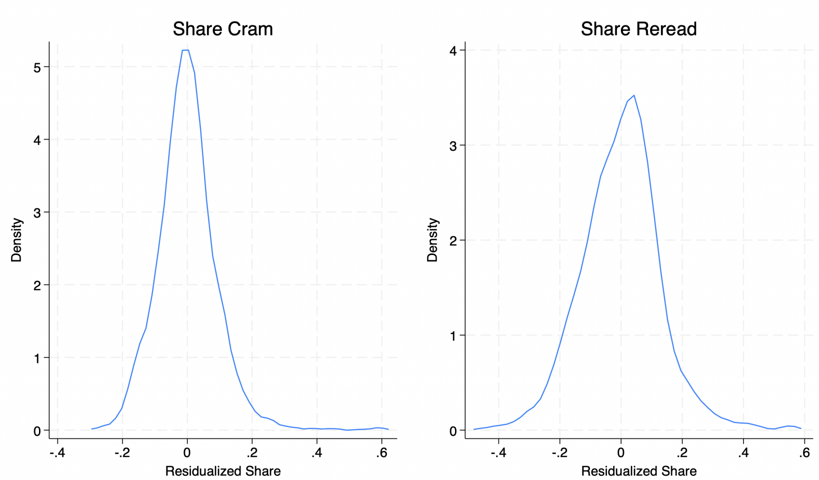

Figure D.4: Residualized Shares Reread and Cramming

Note: This figure shows the residualized distributions of students' reported time allocations across study habits after partialling out the baseline control variables included in the analysis (controls include: indicator for students taking the survey more than once, average hours studied per week, self-reported ability indicator, high school GPA, class indicator, household income, STEM major indicator, indicator for working full-time, number of courses taken during the survey semester, first generation status, total semesters enrolled at ASU (Fall and Spring), total hours completed at ASU, LEAD status (ASU Success program), cohort type (new student or transfer), duration to complete the survey in seconds, indicators for missing values for high school GPA, major, number of courses taken, first generation status, income). Specifically, *Share Cram* and *Share Reread* denote, respectively, the fraction of total study time devoted to cramming and rereading (values range from 0 to 1). Residuals are obtained from separate regressions of each study-share measure on the full set of controls, and the panels display the kernel density of these residualized outcomes.

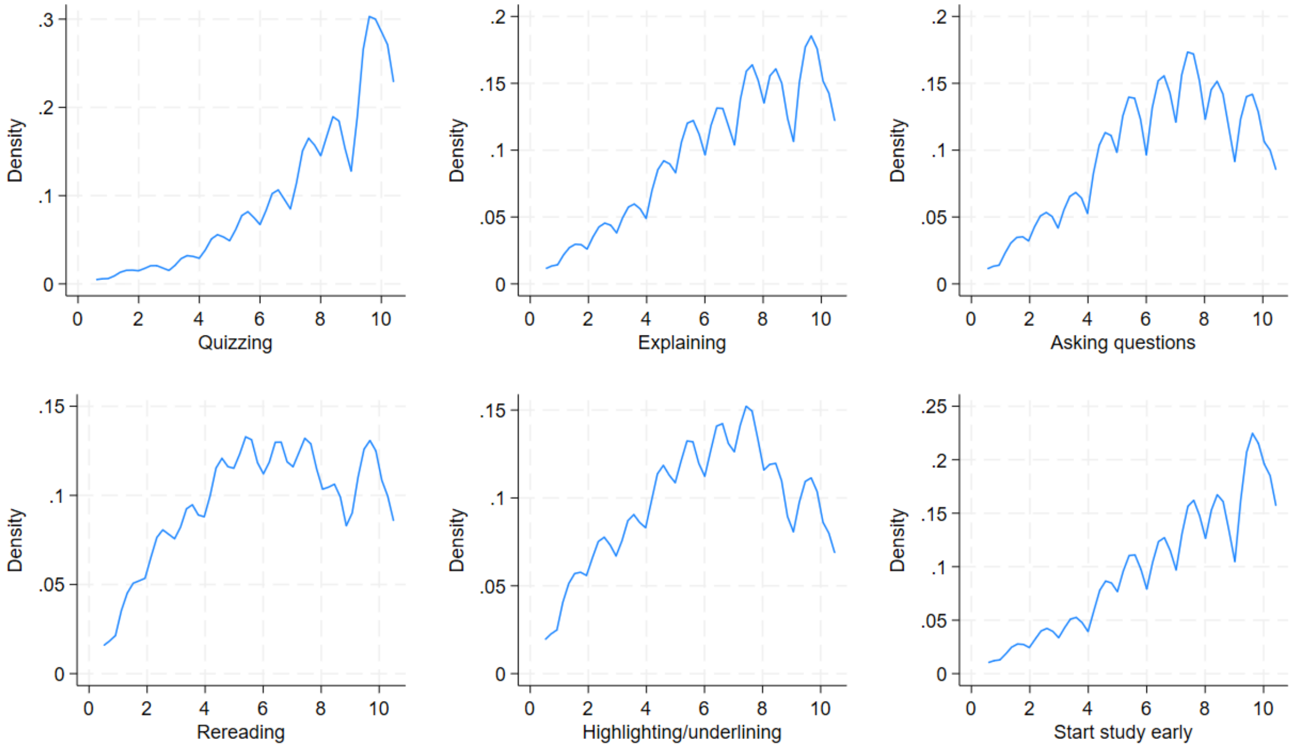

Figure D.5: Kernel density of Perceived Effectiveness of Study Habits

Note: The figure shows the distribution of the perceived effectiveness of study habits. The effectiveness of study habits is reported by students on a scale of 1 to 10, where 1 denotes low effectiveness and 10 high effectivenes.

E Appendix for Section 4

E.1 Control Message: Thank you Email

Subject: Thank you for completing the Study Habits Survey.

Dear student!

We want to thank you for completing the survey on study habits. The information you have provided will help us improve our students' academic experience. Once the study is completed, we will share the main findings with you.

Many thanks,

The Study Habits Team

E.2 Treatment 1 Message: General Information on Study Habits

Subject: Tips on How to Study More Efficiently!

Dear student!

Table E.1: Summary Information Intervention

Group	Information Received by Student
Control *(N=782)*	Thank you email.
Treatment-1 *(N=777)*	• Thank you email. • General information about efficient study habits: – Tip 1: Do not cram. – Tip 2: Do not simply reread study material. More details about this email intervention can be found in Appendix E.2.
Treatment-2 *(N=771)*	• Thank you email. • General information about efficient study habits: – Tip 1: Do not cram. – Tip 2: Do not simply reread study material. • Personalized information to student about their habits and reallocation of study time. More details about this email intervention can be found in Appendix E.3.

Note: We report the number of observations per group in parentheses for students that have completed at least the first survey and consented to participating in the study. A portion of students were dropped from the study during the stratification process in order to maintain a balanced treatment.

We want to thank you for completing the survey on study habits. The information you have provided will help us improve our students' academic experience.

Now, we would like to provide you with TWO tips to improve your study habits.

Tip 1: DO NOT CRAM!

Your decision regarding when to study can have a significant impact on a midterm or final exams. Students often “mass" their study. But many studies in cognitive psychology research have shown that distributing learning over time is much more effective. The advantage of distributing learning over time is known as the spacing effect. This effect has been demonstrated in over 200 research studies from over a century of research.

Actions that you could take:

- Prepare for each course repeatedly and across multiple days.
- It would be best if you plan to study at regular intervals (learning sessions can occur every other day, every two days, etc.)
- Each study session does not need to be very long (1 hour or 2 hours at a time).
- During each session, try to cover old and new materials. For instance, 75% old materials, 25

Tip 2: DO NOT SIMPLY REREAD/REVIEW INFORMATION!

A large body of research has shown that simply studying materials repeatedly (for instance, looking over your notes over and over) is not the most effective way to learn. One of the reasons why rereading and reviewing are not the most effective learning strategies is that they do not involve the actual cognitive processes required by exams. Instead, research on cognitive psychology recommends using retrieval practices. This implies that after you have finished reading through a set of lecture slides, your notes, or the textbook, put it aside. Then, without looking at those materials, try to remember what you have just learned mentally or by writing it down.

Actions that you could take:

- Use practice tests - make your own practice questions, make and share questions with a study partner, use practice questions provided by the instructor or found in a textbook, or find questions from online sources (for example, Quizlet).

- Make flashcards - this commonly involves writing on index cards (questions on one side, answers on the reverse).
- After studying from your course notes/book, put them away and try to solve problem sets or practice tests without looking at your notes/book.

We hope you find this information helpful. If you want to learn more about best practices in study habits, please click on these links:

https://psychology.ucsd.edu/undergraduate-program/undergraduate-resources/academic-writing-resources/effective-studying/index.html

https://www.scientificamerican.com/article/what-works-what-doesn-t/

https://www.retrievalpractice.org/make-it-stick

Many thanks,

Professor's Name and the Study Habits Team

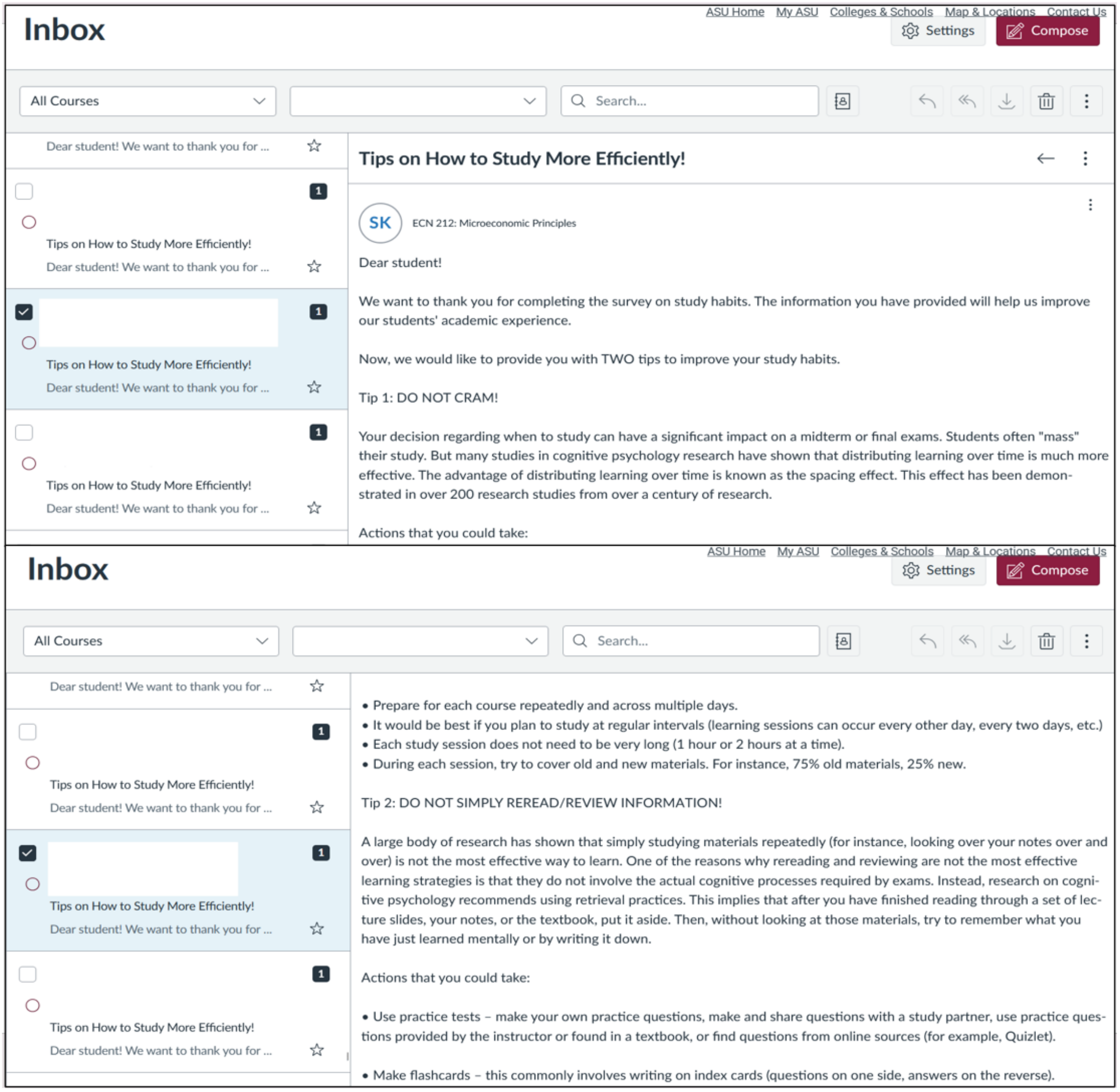

Figure E.1: Treatment 1 Email Screenshot

E.3 Treatment 2 Message: General Information on Study Habits and Personalized Information

Subject: Tips on How to Study More Efficiently!

Dear student!

We want to thank you for completing the survey on study habits. The information you have provided will help us improve our students' academic experience.

Based on the information you provided in the survey at the beginning of the year, YOU SEEM TO CRAM MORE [OR LESS] than the median student in the class (namely, study all the material in a very short window of time) when studying for exams.

[IF CRAM]: Unfortunately, researchers in educational psychology find this practice as INEFFECTIVE to perform well in exams or retain the material. Please look at the tips below that explain why cramming is not a good practice and why you should space your learning sessions.

To give you an idea of how avoiding cramming affected performance in the first exam, we found that REDUCING the share of total study hours on CRAMMING (i.e., 2 days before the exam) by 50 (25) percentage points and REALLOCATING them to start studying at least 2 WEEKS before the exam, could increase the average probability to be in the top 50 (25) percentile of the students in your class by XX percent (holding all else constant).

[IF NOT CRAM]: It is GOOD that you cram less than many of your classmates! Researchers in educational psychology find that spacing your study sessions is highly effective for performing well in exams or retaining material.

To give you an idea of how avoiding cramming affected performance in the first exam, we found that REDUCING the share of total study hours on CRAMMING (i.e., 2 days before the exam) by 50 (25) percentage points and REALLOCATING them to start studying at least 2 WEEKS before the exam, could increase the average probability to be in the top 50 (25) percentile of the students in your class by XX percent (holding all else constant).

Also, based on the information that you gave us, it seemed that YOU ARE MORE [or LESS] likely than the median student in this class to implement RETRIEVAL PRACTICES when studying for your exams (namely, instead of simply rereading information, attempting to recall that information by taking a practice test or exercises).

[IF DID NOT USE RETRIEVAL PRACTICES] This is unfortunate because researchers on study habits have shown that this type of practice is highly effective for performance in exams and retaining material for a more extended period.

To give you an idea of how implementing retrieval practices affected performance in the first exam, we found that REDUCING the share of total hours on REREADING PRACTICES by 50 (25) percentage points and REALLOCATING them to RETRIEVAL PRACTICES could increase the average probability to be in top the 50 (25) percentile of the students in your class by XX percent (holding all else constant).

[IF DO IMPLEMENT RETRIEVAL PRACTICE] You are right on track! Researchers on study habits have shown that this type of practice is highly effective for performance in exams and retaining material for a more extended period.

To give you an idea of how implementing retrieval practices affected performance in the first exam, we found that REDUCING the share of total hours on REREADING PRACTICES by 50 (25) percentage points and REALLOCATING them to RETRIEVAL PRACTICES could increase the average probability to be in top the 50 (25) percentile of the students in your class by XX percent (holding all else constant).

Now, we would like to provide you with TWO tips to improve your study habits.

Tip 1: DO NOT CRAM!

Your decision regarding when to study can have a significant impact on a midterm or final exams. Students often “mass" their study. But many studies in cognitive psychology research have shown that distributing learning over time is much more effective. The advantage of distributing learning over time is known as the spacing effect. This effect has been demonstrated in over 200 research studies from over a century of research.

Actions that you could take:

- Prepare for each course repeatedly and across multiple days.
- It would be best if you plan to study at regular intervals (learning sessions can occur every other day, every two days, etc.)
- Each study session does not need to be very long (1 hour or 2 hours at a time).
- During each session, try to cover old and new materials. For instance, 75% old materials, 25% new.

Tip 2: DO NOT SIMPLY REREAD/REVIEW INFORMATION!

A large body of research has shown that simply studying materials repeatedly (for instance, looking over your notes over and over) is not the most effective way to learn. One of the reasons why rereading and reviewing are not the most effective learning strategies is that they do not involve the actual cognitive processes required by exams. Instead, research on cognitive psychology recommends using retrieval practices. This implies that after you have finished reading through a set of lecture slides, your notes, or the textbook, put it aside. Then, without looking at those materials, try to remember what you have just learned mentally or by writing it down.

Actions that you could take:

- Use practice tests - make your own practice questions, make and share questions with a study partner, use practice questions provided by the instructor or found in a textbook, or find questions from online sources (for example, Quizlet).

- Make flashcards - this commonly involves writing on index cards (questions on one side, answers on the reverse).
- After studying from your course notes/book, put them away and try to solve problem sets or practice tests without looking at your notes/book.

We hope you find this information helpful. If you want to learn more about best practices in study habits, please click on these links:

https://psychology.ucsd.edu/undergraduate-program/undergraduate-resources/academic-riting- resources /effective-studying/index.html

https://www.scientificamerican.com/article/what-works-what-doesn-t/

https://www.retrievalpractice.org/make-it-stick

Many thanks,

Professor's Name and the Study Habits Team

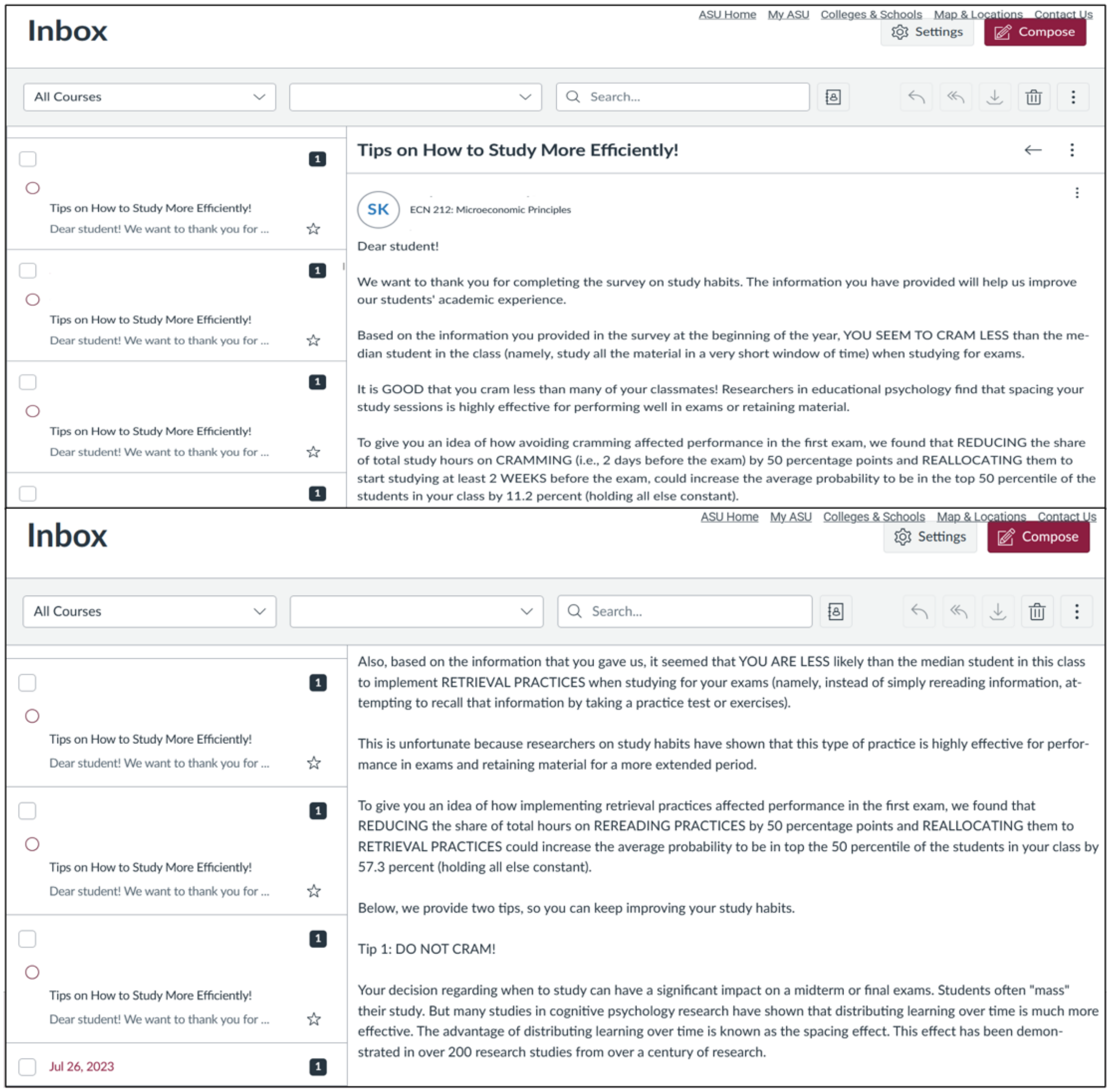

Figure E.2: Treatment 2 Email Screenshot

E.3.1 Estimates Reported to Students

Table E.2: Treatment Effect on Student Classroom Outcomes

Study Habit	R_m	S_m	Mean	Unit	N. Classes	Online
Cramming	50	25	14.6	%	2	No
Cramming	50	50	26.1	%	10	Yes
Retrieval practices	25	25	64.0	%	1	No
Retrieval practices	50	25	28.0	%	1	No
Retrieval practices	50	50	46.7	%	10	Yes
Cramming	25	50	19.9	p.p.	2	No
Retrieval practices	25	50	5.9	p.p.	2	No

Note: The table reports the summary of the information provided to students assigned to receive personalized information. We report to them the percent (or p.p.) increase in the likelihood of scoring above the S_m percentile of students in their class if they reallocate R_m percentage points of their study time to retrieval practices or non-cramming periods. We group classes that were assigned the same R_m and S_m and report in the column Mean the average of the number reported to students in those classes. The number reported to those receiving the information in percent comes from: $\beta^s_{1m}\frac{R_m}{S_m/100}$ and $\beta^s_{2m}\frac{R_m}{S_m/100}$, where the βs come from the estimation of Equation 7 and $S_m/100$ is approximately the mean of the dependent variable. For those who receive the information in p.p., the numbers come from: $\beta^s_{1m}R_m/100$ and $\beta^s_{2m}R_m/100$. Note that the β coefficients are class-specific. Some classes have a small number of students, so we also group students from those classes into a particular group. For this reason, the total number of classes in this table does not coincide with the actual number of classes. Unit represents if the number corresponds to a percent increase or p.p. increase. N. Classes indicates the number of classes assigned to a specific combination of R_m and S_m.

F Appendix for Section 6

F.1 Effort-Dependent Learning Rule

We show how the effort-dependent updating rule in Assumption A1 can be derived from a Bayesian learning model in which effort increases signal precision. Consider one activity $j \in \{p, r\}$ (we suppress the j subscript for clarity). Student i holds a Normal prior on the unknown effectiveness α:

$$\alpha \mid \mathcal{F}_{i1} \sim \mathcal{N}\big(\mu_{i1},\, \sigma_\alpha^2\big).$$

When the student allocates effort $e_i = Sh_{i1}\,\bar{E}_i$ to this activity, effort generates $n(e_i)$ independent feedback signals,

$$y_{im} = \alpha + \varepsilon_{im}, \qquad \varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_y^2), \qquad m = 1, \ldots, n(e_i),$$

where the number of signals is proportional to effort: $n(e) = \kappa\, e$ with $\kappa > 0$. The sufficient statistic is the sample mean

$$\bar{y}_i \;=\; \frac{1}{n(e_i)} \sum_{m=1}^{n(e_i)} y_{im} \;=\; \alpha + \bar{\varepsilon}_i, \qquad \bar{\varepsilon}_i \sim \mathcal{N}\Big(0,\, \frac{\sigma_y^2}{n(e_i)}\Big).$$

Under Normal-Normal conjugacy, the posterior mean after observing $\bar{y}_i$ is

$$\mu_{i2} \;=\; \mu_{i1} \;+\; K(e_i)\,\big(\bar{y}_i - \mu_{i1}\big), \qquad K(e) \;\equiv\; \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_y^2/(\kappa\, e)} \;=\; \frac{\theta\,\kappa\, e}{1 + \theta\,\kappa\, e}, \tag{23}$$

where $\theta \equiv \sigma_\alpha^2/\sigma_y^2$ is the signal-to-noise ratio. The Kalman gain $K(e)$ is increasing in effort, satisfies $K(0) = 0$, and has derivative

$$K'(e) \;=\; \frac{\theta\,\kappa}{\big(1 + \theta\,\kappa\, e\big)^2}.$$

For moderate effort levels (where $\theta\,\kappa\, e$ is not too large), a first-order Taylor expansion around $e = 0$ gives

$$K(e) \;\approx\; \theta\,\kappa\, e \;=\; \kappa\,\frac{\sigma_\alpha^2}{\sigma_y^2}\, e. \tag{24}$$

Substituting into the posterior mean yields the linearized update

$$\mu_{i2} \;\approx\; \mu_{i1} \;+\; \kappa\,\frac{\sigma_\alpha^2}{\sigma_y^2}\, e_i\,\big(\bar{y}_i - \mu_{i1}\big). \tag{25}$$

In the main text (Assumption A1), the updating rule is written as

$$\mu_{i2} \;=\; \mu_{i1} \;+\; e_i\,\frac{k}{\sigma_\varepsilon^2}\,\big(s_i - \mu_{i1}\big),$$

where $k > 0$ is a reduced-form learning intensity and σ_ε^2 is a scale parameter for signal noise in the reduced-form representation. To connect this reduced-form rule to the Bayesian framework above, we identify the observed signal with the sufficient statistic ($s_i \equiv \bar{y}_i$) and adopt the convenient normalization $\sigma_\varepsilon^2 \equiv \sigma_y^2$. This normalization is chosen to match the updating *weight* and should not be interpreted as equating the variance of the signal innovation across the two setups: in the exact Bayesian model, $\mathrm{Var}(\bar{y}_i - \alpha) = \sigma_y^2/(\kappa\, e_i)$ is effort-dependent. The object with behavioral content in Assumption A1 is the composite gain parameter $\lambda \equiv k/\sigma_\varepsilon^2$, which the first-order approximation (25) implies equals $\lambda = \kappa\, \sigma_\alpha^2/\sigma_y^2$. Under the normalization $\sigma_\varepsilon^2 \equiv \sigma_y^2$, this yields the parameter correspondence $k = \kappa\, \sigma_\alpha^2$.

F.2 Proof of Proposition 1

We derive the optimal effort shares by backward induction. The agent's total expected subjective utility to be maximized is the sum of the expected subjective utility from period 1 and the expected subjective utility from period 2:

$$\max_{Sh_{i1}^p, Sh_{i2}^p} \mathbb{E}_1[\tilde{U}_{i1}] + \mathbb{E}_2[\tilde{U}_{i2} \mid s_i, Sh_{i1}^p].$$

In both periods, students are taking expectations over their subjective beliefs about the effectiveness of each study habit. The utility in period 2 depends on the agent's updated beliefs, which are influenced by the signals that are determined by the effort choices made in period 1.

F.2.1 Period-2 Optimal Share

In period 2 the problem students solve is:

$$\max_{Sh_{i2}^p} \mathbb{E}_2^{\tilde{\alpha}}\left[\tilde{\alpha}_i^p | s_i^p\right] Sh_{i2}^p \bar{E}_i + \mathbb{E}_2^{\tilde{\alpha}}\left[\tilde{\alpha}_i^r | s_i^r\right](1 - Sh_{i2}^p)\bar{E}_i - \gamma_p \left(Sh_{i2}^p \bar{E}_i\right)^2 - \gamma_r \left((1 - Sh_{i2}^p)\bar{E}_i\right)^2.$$

Taking the first-order condition:

$$\mathbb{E}_2\left[\tilde{\alpha}_p | s_i^p\right] - \mathbb{E}_2\left[\tilde{\alpha}_r | s_i^r\right] + 2\gamma_r \bar{E}_i - (2\gamma_p + 2\gamma_r) Sh_{i2}^p \bar{E}_i = 0.$$

Solving for Sh_{i2}^p:

$$Sh_{i2}^{p*} = \frac{\mathbb{E}_2[\tilde{\alpha}_i^p | s_i^p] - \mathbb{E}_2[\tilde{\alpha}_i^r | s_i^r] + 2\gamma_r \bar{E}_i}{2\bar{E}_i(\gamma_p + \gamma_r)} = \frac{\mu_{i2}^{\alpha_p} - \mu_{i2}^{\alpha_r} + 2\gamma_r\, \bar{E}_i}{2\, \bar{E}_i\, (\gamma_p + \gamma_r)}. \tag{26}$$

Effort in practicing during period 2 increases with the posterior belief about the differential effectiveness of practicing relative to rereading and the cost of rereading.

F.2.2 Period-1 Optimal Share

The problem in period 1 can be written as:

$$\max_{Sh^p_{i1}} \quad \mu^{\alpha_p}_{i1} Sh^p_{i1}\bar{E}_i + \mu^{\alpha_r}_{i1}(1 - Sh^p_{i1})\bar{E}_i - \gamma_p \left(Sh^p_{i1}\bar{E}_i\right)^2 - \gamma_r \left((1 - Sh^p_{i1})\bar{E}_i\right)^2$$
$$+ \int_{s_p}\int_{s_r} \Bigg[\mathbb{E}_2\left[\tilde{\alpha}^p_i \mid s^p_i\right] Sh^p_{i2}(\mathbb{E}_2[\tilde{\alpha}^p_i|s^p_i], \mathbb{E}_2[\tilde{\alpha}^r_i|s^r_i])\bar{E}_i + \mathbb{E}_2\left[\tilde{\alpha}^r_i \mid s^r_i\right](1 - Sh^p_{i2}(\mathbb{E}_2[\tilde{\alpha}^p_i|s^p_i], \mathbb{E}_2[\tilde{\alpha}^r_i|s^r_i]))\,\bar{E}_i$$
$$- \gamma_p \left(Sh^p_{i2}(\mathbb{E}_2[\tilde{\alpha}^p_i|s^p_i], \mathbb{E}_2[\tilde{\alpha}^r_i|s^r_i])\bar{E}_i\right)^2 - \gamma_r \left(1 - Sh^p_{i2}(\mathbb{E}_2[\tilde{\alpha}^p_i|s^p_i], \mathbb{E}_2[\tilde{\alpha}^r_i|s^r_i])\bar{E}_i\right)^2\Bigg] dF(s_p)\, dG(s_r).$$

Substituting Equation (26):

$$\max_{Sh^p_{i1}} \quad \mu^{\alpha_p}_{i1} Sh^p_{i1}\bar{E}_i + \mu^{\alpha_r}_{i1}(1 - Sh^p_{i1})\bar{E}_i - \gamma_p \left(Sh^p_{i1}\bar{E}_i\right)^2 - \gamma_r \left((1 - Sh^p_{i1})\bar{E}_i\right)^2$$
$$+ \int_{s^d_i}\int_{s^r_i} \Bigg[\mathbb{E}_2\left[\tilde{\alpha}_p \mid s^p_i\right] \frac{\mathbb{E}_2\left[\tilde{\alpha}_p|s^p_i\right] - \mathbb{E}_2\left[\tilde{\alpha}_r|s^r_i\right] + 2\gamma_r\bar{E}_i}{2(\gamma_p + \gamma_r)} + \mathbb{E}_2\left[\tilde{\alpha}_r \mid s^r_i\right]\left(\bar{E}_i - \frac{\mathbb{E}_2\left[\tilde{\alpha}_p|s^p_i\right] - \mathbb{E}_2\left[\tilde{\alpha}_r|s^r_i\right] + 2\gamma_r\bar{E}_i}{2(\gamma_p + \gamma_r)}\right)$$
$$- \gamma_p \left(\frac{\mathbb{E}_2\left[\tilde{\alpha}_p|s^p_i\right] - \mathbb{E}_2\left[\tilde{\alpha}_r|s^r_i\right] + 2\gamma_r\bar{E}_i}{2(\gamma_p + \gamma_r)}\right)^2 - \gamma_r \left(\bar{E}_i - \frac{\mathbb{E}_2\left[\tilde{\alpha}_p|s^p_i\right] - \mathbb{E}_2\left[\tilde{\alpha}_r|s^r_i\right] + 2\gamma_r\bar{E}_i}{2(\gamma_p + \gamma_r)}\right)^2\Bigg] dF(s^p_i)\, dG(s^r_i).$$

Simplifying the integral:

$$\max_{Sh^p_{i1}} \quad \mu^{\alpha_p}_{i1} Sh^p_{i1}\bar{E}_i + \mu^{\alpha_r}_{i1}(1 - Sh^p_{i1})\bar{E}_i - \gamma_p \left(Sh^p_{i1}\bar{E}_i\right)^2 - \gamma_r \left((1 - Sh^p_{i1})\bar{E}_i\right)^2$$
$$+ \int_{s^d_i}\int_{s^r_i} \Bigg[\frac{-\bar{E}_i^2\gamma_p\gamma_r + \bar{E}_i\gamma_p\mathbb{E}_2\left[\tilde{\alpha}_r \mid s^r_i\right] + \bar{E}_i\gamma_r\mathbb{E}_2\left[\tilde{\alpha}_p \mid s^p_i\right] + \frac{1}{4}\left(\mathbb{E}_2\left[\tilde{\alpha}_p \mid s^p_i\right]\right)^2}{\gamma_p + \gamma_r} + \tag{27}$$
$$\frac{-\frac{1}{2}\mathbb{E}_2\left[\tilde{\alpha}_p \mid s^p_i\right]\mathbb{E}_2\left[\tilde{\alpha}_r \mid s^r_i\right] + \frac{1}{4}\left(\mathbb{E}_2\left[\tilde{\alpha}_r \mid s^r_i\right]\right)^2}{\gamma_p + \gamma_r}\Bigg] dF(s^p_i)dG(s^r_i)$$

As mentioned, with the information available at the beginning of period 1, $s^j_i = \tilde{\alpha}^j_i + \varepsilon^j_i$ and therefore: $\mathbb{E}_2[s^j_i] = \mu^{\alpha_j}_{i1}$ implying that $\int_{s^j_i}\left(s^j_i - \mu^{\alpha_j}_{i1}\right) dF(s^j_i) = 0$.

Now, we will solve each of the integrals included in Equation (27). To simplify notation, let's call $E^j_{i1} = Sh^j_{i1}\bar{E}_i$ for $j \in \{p, r\}$.

Integral computations

1.
$$\int_{s^j_i} \mathbb{E}_2[\tilde{\alpha}^j_i|s^j_i]dF(s^j_i) = \int_{s^j_i}\left[\mu^{\alpha_j}_{i1} + \frac{k_j \cdot E^j_{i1}}{\sigma^2_{\varepsilon_j}} \cdot \left(s^j_i - \mu^{\alpha_j}_{i1}\right)\right] dF(s^j_i) = \mu^{\alpha_j}_{i1}$$

2. Given the assumption of independence of the signals:

$$\int_{s_i^j s_i^{j'}} \mathbb{E}_2[\tilde{\alpha}_i^j|s_i^j]\mathbb{E}_2[\tilde{\alpha}_i^{j'}|s_i^{j'}]dF(s_i^j)dG(s_i^{j'}) = \int_{s_i^j}\int_{s_i^{j'}} \left[\mu_{i1}^{\alpha_j} + \frac{k_j \cdot E_{i1}^j}{\sigma^2_{\varepsilon_j}} \cdot \left(s_i^j - \mu_{i1}^{\alpha_j}\right)\right]\left[\mu_{i1}^{\alpha_{j'}} + \frac{k_{j'} \cdot E_{i1}^{j'}}{\sigma^2_{\varepsilon_{j'}}} \cdot \left(s_i^{j'} - \mu_{i1}^{\alpha_{j'}}\right)\right] dF(s$$

$$= \int_{s_i^j}\int_{s_i^{j'}} \left[\mu_{i1}^{\alpha_j}\mu_{i1}^{\alpha_{j'}} + k_j k_{j'} E_{i1}^j E_{i1}^{j'} \left(s_i^j - \mu_{i1}^{\alpha_j}\right)\left(s_i^{j'} - \mu_{i1}^{\alpha_{j'}}\right)\right] dF(s_i^j)dG(s_i^{j'})$$

$$= \mu_{i1}^{\alpha_j}\mu_{i1}^{\alpha_{j'}} + k^2 e_i^j e_i^{j'} \int_{s_i^j}\left(s_i^j - \mu_{i1}^{\alpha_j}\right)dF(s_i^j)\int_{s_i^{j'}}\left(s_i^{j'} - \mu_{i1}^{\alpha_{j'}}\right)dG(s_i^{j'})$$

$$= \mu_{i1}^{\alpha_j}\mu_{i1}^{\alpha_{j'}}$$

3.

$$\int_{s_i^j} \mathbb{E}_2[\tilde{\alpha}_j|s_i^j]^2 dF(s_i^j) = \int_{s_i^j}\left[\mu_{i1}^{\alpha_j} + \frac{k_j \cdot E_{i1}^j}{\sigma^2_{\varepsilon_j}} \cdot (s_i^j - \mu_{i1}^{\alpha_j})\right]^2 dF(s_i^j)$$

$$= \int_{s_i^j}\left[(\mu_{i1}^{\alpha_j})^2 + \left(\frac{k_j \cdot E_{i1}^j}{\sigma^2_{\varepsilon_j}} \cdot (s_i^j - \mu_{i1}^{\alpha_j})\right)^2 + 2\mu_{i1}^{\alpha_j}\frac{k_j \cdot E_{i1}^j}{\sigma^2_{\varepsilon_j}} \cdot (s_i^j - \mu_{i1}^{\alpha_j})\right] dF(s_i^j)$$

$$= \int_{s_i^j}\left[(\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}\left[\left(s_i^j\right)^2 + (\mu_{i1}^{\alpha_j})^2 - 2\mu_{i1}^{\alpha_j}s_i^j\right]\right] dF(s_i^j)$$

$$= (\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}(\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}\int_{s_i^j}\left[\left(s_i^j\right)^2 - 2\mu_{i1}^{\alpha_j}s_i^j\right] dF(s_i^j)$$

$$= (\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}(\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\left(\tilde{\alpha}_i^j + \varepsilon_j\right)^2 - 2\mu_{i1}^{\alpha_j}(\tilde{\alpha}_i^j + \varepsilon_j)\right] dG(\tilde{\alpha}_i^j)dH($$

$$= (\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}(\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\left(\tilde{\alpha}_i^j\right)^2 + (\varepsilon_j)^2 + 2\tilde{\alpha}_i^j\varepsilon_j - 2\mu_{i1}^{\alpha_j}(\tilde{\alpha}_i^j + \varepsilon_j)\right]$$

$$= (\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(E_{i1}^j\right)^2}{(\sigma^2_{\varepsilon_j})^2}\left[\sigma^2_{\alpha_j} + \sigma^2_{\varepsilon_j}\right]$$

$$= (\mu_{i1}^{\alpha_j})^2 + \frac{k_j^2\left(Sh_{i1}^j\bar{E}_i\right)^2}{(\sigma^2_{\varepsilon_j})^2}\left[\sigma^2_{\alpha_j} + \sigma^2_{\varepsilon_j}\right]$$

Where:

- $\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[2\tilde{\alpha}_i^j\varepsilon_j\right] dG(\tilde{\alpha}_i^j)dH(\varepsilon_j) = 0$ given independence of $\tilde{\alpha}_i^j$ and ε_j
- $\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\varepsilon_j\right] dG(\tilde{\alpha}_i^j)dH(\varepsilon_j) = 0$
- $\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\tilde{\alpha}_i^j\right] dG(\tilde{\alpha}_i^j)dH(\varepsilon_j) = \mu_{i1}^{\alpha_j}$
- $\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\tilde{\alpha}_i^j\right]^2 dG(\tilde{\alpha}_i^j)dH(\varepsilon_j) = (\mu_{i1}^{\alpha_j})^2 + \sigma^2_{\alpha_j}$
- $\int_{\tilde{\alpha}_i^j}\int_{\varepsilon_j}\left[\varepsilon_j\right]^2 dG(\tilde{\alpha}_i^j)dH(\varepsilon_j) = \sigma^2_{\varepsilon_j}$

Final expression and first-order condition By substituting these integrals into the problem stated in Equation (27), we get:

$$
\begin{aligned}
\max_{Sh^p_{i1}} \quad & \mu^{\alpha_p}_{i1} Sh^p_{i1}\bar{E}_i + \mu^{\alpha_r}_{i1}(1 - Sh^p_{i1})\bar{E}_i - \gamma_p \left(Sh^p_{i1}\bar{E}_i\right)^2 - \gamma_r \left((1 - Sh^p_{i1})\bar{E}_i\right)^2 \\
+ \; & \frac{\left(\mu^{\alpha_p}_{i1}\right)^2 + \frac{k_p^2\left(Sh^p_{i1}\bar{E}_i\right)^2}{(\sigma^2_{\varepsilon_p})^2}\left[\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}\right] + (\mu^{\alpha_r}_{i1})^2 + \frac{k_r^2\left(1 - Sh^p_{i1}\right)^2(\bar{E}_i)^2}{(\sigma^2_{\varepsilon_r})^2}\left[\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}\right]}{4(\gamma_p + \gamma_r)} \\
+ \; & \frac{-2\mu^{\alpha_p}_{i1}\mu^{\alpha_r}_{i1} + 4\bar{E}_i\gamma_r\mu^{\alpha_p}_{i1} + 4\bar{E}_i\gamma_p\mu^{\alpha_r}_{i1} - 4\gamma_r\gamma_p\bar{E}_i^2}{4(\gamma_p + \gamma_r)}.
\end{aligned}
$$

Thus, keeping only the terms that depend on Sh^p_{i1} and collecting the rest in C (constant in Sh^p_{i1}):

$$
\begin{aligned}
\mathcal{L}(Sh^p_{i1}) = \; & \mu^{\alpha_p}_{i1}\, Sh^p_{i1}\, \bar{E}_i + \mu^{\alpha_r}_{i1}\,(1 - Sh^p_{i1})\,\bar{E}_i - \gamma_p(Sh^p_{i1}\bar{E}_i)^2 - \gamma_r\left((1 - Sh^p_{i1})\bar{E}_i\right)^2 \\
& + \frac{1}{4(\gamma_p + \gamma_r)}\left[k_p^2(Sh^p_{i1}\bar{E}_i)^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} + k_r^2\left((1 - Sh^p_{i1})\bar{E}_i\right)^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right] + C.
\end{aligned}
$$

Taking the derivative term by term:

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial Sh^p_{i1}} = \; & (\mu^{\alpha_p}_{i1} - \mu^{\alpha_r}_{i1})\,\bar{E}_i - 2\gamma_p Sh^p_{i1}\bar{E}_i^2 + 2\gamma_r(1 - Sh^p_{i1})\bar{E}_i^2 \\
& + \frac{1}{4(\gamma_p + \gamma_r)}\left[2k_p^2 Sh^p_{i1}\bar{E}_i^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} - 2k_r^2(1 - Sh^p_{i1})\bar{E}_i^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right].
\end{aligned}
$$

Simplifying the cost terms:

$$
-2\gamma_p Sh^p_{i1}\bar{E}_i^2 + 2\gamma_r(1 - Sh^p_{i1})\bar{E}_i^2 = -2(\gamma_p + \gamma_r)Sh^p_{i1}\bar{E}_i^2 + 2\gamma_r\bar{E}_i^2.
$$

Simplifying the k terms:

$$
\frac{1}{4(\gamma_p + \gamma_r)}\left[2k_p^2 Sh^p_{i1}\bar{E}_i^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} - 2k_r^2(1 - Sh^p_{i1})\bar{E}_i^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right] = \frac{\bar{E}_i^2}{2(\gamma_p + \gamma_r)}\left[\left(k_p^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} + k_r^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right)Sh^p_{i1} - k_r^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right].
$$

Setting the derivative to zero and collecting terms in Sh^p_{i1}:

$$
\begin{aligned}
0 = \; & (\mu^{\alpha_p}_{i1} - \mu^{\alpha_r}_{i1})\,\bar{E}_i - 2(\gamma_p + \gamma_r)Sh^p_{i1}\bar{E}_i^2 + 2\gamma_r\bar{E}_i^2 \\
& + \frac{\bar{E}_i^2}{2(\gamma_p + \gamma_r)}\left[\left(k_p^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} + k_r^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right)Sh^p_{i1} - k_r^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right].
\end{aligned}
$$

Solving for Sh^p_{i1}:

$$
Sh^{p*}_{i1} = \frac{(\mu^{\alpha_p}_{i1} - \mu^{\alpha_r}_{i1}) + 2\gamma_r\bar{E}_i - \frac{\bar{E}_i}{2(\gamma_p + \gamma_r)}\,k_r^2\,\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}}{2\bar{E}_i\left[(\gamma_p + \gamma_r) - \frac{1}{4(\gamma_p + \gamma_r)}\left(k_p^2\frac{\sigma^2_{\alpha_p} + \sigma^2_{\varepsilon_p}}{(\sigma^2_{\varepsilon_p})^2} + k_r^2\frac{\sigma^2_{\alpha_r} + \sigma^2_{\varepsilon_r}}{(\sigma^2_{\varepsilon_r})^2}\right)\right]}. \tag{28}
$$

This completes the proof of Proposition 1. □