

# To What Extent Do Study Habits Relate to Performance?

Elise M. Walck-Shannon<sup>1\*</sup>, Shaina F. Rowell<sup>2</sup>, and Regina F. Frey<sup>3</sup>

<sup>1</sup>Biology Department and <sup>2</sup>Center for Integrative Research on Cognition, Learning, and Education (CIRCLE), Washington University in St. Louis, St. Louis, MO 63130; <sup>3</sup>Department of Chemistry, University of Utah, Salt Lake City, UT 84112

## ABSTRACT

Students' study sessions outside class are important learning opportunities in college courses. However, we often depend on students to study effectively without explicit instruction. In this study, we described students' self-reported study habits and related those habits to their performance on exams. Notably, in these analyses, we controlled for potential confounds, such as academic preparation, self-reported class absences, and self-reported total study time. First, we found that, on average, students used approximately four active strategies to study and that they spent about half of their study time using active strategies. In addition, both the number of active strategies and the proportion of their study time using active strategies positively predicted exam performance. Second, on average, students started studying 6 days before an exam, but how early a student started studying was not related to performance on in-term (immediate) or cumulative (delayed) exams. Third, on average, students reported being distracted about 20% of their study time, and distraction while studying negatively predicted exam performance. These results add nuance to lab findings and help instructors prioritize study habits to target for change.

## INTRODUCTION

One of our goals in college courses is to help students develop into independent, self-regulated learners. This requires students to perform several metacognitive tasks on their own, including setting goals, choosing strategies, monitoring and reflecting on performance, and modifying those steps over time (Zimmerman, 2002). There are many challenges that learners encounter in developing self-regulation. One such challenge is that students often misjudge their learning during the monitoring and reflection phases (Kornell and Bjork, 2007). Often, students *feel* that they learn more from cognitively superficial tasks than from cognitively effortful tasks. As one example, students may feel that they have learned more if they reread a text passage multiple times than if they are quizzed on that same material (Karpicke and Roediger, 2008). In contrast to students' judgments, many effortful tasks are highly effective for learning. R. A. Bjork defines these effective, effortful tasks as desirable difficulties (Bjork, 1994). In the present study, we investigated the frequency with which students reported carrying out effortful (active) or superficial (passive) study habits in a large introductory biology course. Additionally, we examined the relationship between study habits and performance on exams while controlling for prior academic preparation and total study time.

## THEORETICAL FRAMEWORK

### Why Would Difficulties Be Desirable?

During learning, the goal is to generate knowledge or skills that are robustly integrated with related knowledge and easily accessible. Desirable difficulties promote cognitive processes that either aid forming robust, interconnected knowledge or skills

Jennifer Knight, *Monitoring Editor*

Submitted May 19, 2020; Revised Oct 19, 2020;  
Accepted Oct 22, 2020

CBE Life Sci Educ March 1, 2021 20:ar6

DOI:10.1187/cbe.20-05-0091

\*Address correspondence to: Elise M. Walck-Shannon (ewalck-shannon@wustl.edu).

© 2021 E. M. Walck-Shannon *et al.* CBE—Life Sciences Education © 2021 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

or retrieving that knowledge or skill (Bjork, 1994; also see Marsh and Butler, 2013, for a chapter written for educators). Learners employing desirable difficulties may feel that they put in more effort and make more mistakes, but they are actually realizing larger gains toward long-term learning than learners using cognitively superficial tasks.

### Which Study Habits Are Difficult in a Desirable Way?

Study habits can include a wide variety of behaviors, from the amount of time that students study, to the strategies that they use while studying, to the environment in which they study. The desirable difficulties framework (Bjork and Bjork, 2011), describes two main kinds of effective habits that apply to our study: 1) using effortful study strategies or techniques that prompt students to generate something or test themselves during studying and 2) distributing study time into multiple sessions to avoid “cramming” near the exam. In the following two paragraphs, we expand upon these study habits of interest.

The desirable difficulties framework suggests that study strategies whereby students actively generate a product or test themselves promote greater long-term learning than study strategies whereby students passively consume presentations. This is supported by strong evidence for the “generation effect,” in which new knowledge or skills are more robustly encoded and retrieved if you generate a solution, explanation, or summary, rather than looking it up (Jacoby, 1978). A few generative strategies that are commonly reported among students—summarization, self-explanation, and practice testing—are compared below. Summarization is a learning strategy in which students identify key points and combine them into a succinct explanation in their own words. As predicted by the generation effect, evidence suggests that summarization is more effective than rewriting notes (e.g., laboratory study by Bretzing and Kulhavy, 1979) or reviewing notes (e.g., classroom study by King, 1992). Self-explanation is a learning strategy wherein students ask “how” and “why” questions for material as they are being exposed to the material or shortly after (Berry, 1983). This is one form of elaborative interrogation, a robust memory technique in which learners generate more expansive details for new knowledge to help them remember that information (Pressley *et al.*, 1987). Self-explanation requires little instruction and seems to be helpful for a broad array of tasks, including recall, comprehension, and transfer. Further, it is more effective than summarization (e.g., classroom study by King, 1992), perhaps because it prompts students to make additional connections between new and existing knowledge. Practice testing is supported by evidence of the “testing effect,” for which retrieving information itself actually promotes learning (Karpicke and Roediger, 2008). The memory benefits of the “testing effect” can be achieved with any strategy in which students complete problems or practice retrieval without relying on external materials (quizzing, practice testing, flashcards, etc.). In this study, we refer to these strategies together as “self-quizzing.” Self-quizzing is especially effective at improving performance on delayed tests, even as long as 9–11 months after initial learning (Carpenter, 2009). Additionally, in the laboratory, self-quizzing has been shown to be effective on a range of tasks from recall to inference (Karpicke and Blunt, 2011). Overall, research suggests that active, more effortful strategies—such as

self-quizzing, summarization, and self-explanation—are more effective for learning than passive strategies—such as rereading and rewriting notes. In this study, we asked whether these laboratory findings would extend to students’ self-directed study time, focusing especially on the effectiveness of effortful (herein, “active”) study strategies.

The second effective habit described by the desirable difficulties framework is to avoid cramming study time near exam time. The “spacing effect” describes the phenomenon wherein, when given equal study time, spacing study out into multiple sessions promotes greater long-term learning than massing (i.e., cramming) study into one study session. Like the “testing effect,” the “spacing effect” is especially pronounced for longer-term tests in the laboratory (Rawson and Kintsch, 2005). Based on laboratory studies, we would expect that, in a course context, cramming study time into fewer sessions close to an exam would be less desirable for long-term learning than distributing study time over multiple sessions, especially if that learning is measured on a delay.

However, estimating spacing in practice is more complicated. Classroom studies have used two main methodologies to estimate spacing, either asking the students to report their study schedules directly (Susser and McCabe, 2013) or asking students to choose whether they describe their pattern of study as spaced out or occurring in one session (Hartwig and Dunlosky, 2012; Rodriguez *et al.*, 2018). The results from these analyses have been mixed; in some cases, spacing has been a significant, positive predictor of performance (Rodriguez *et al.*, 2018; Susser and McCabe, 2013), but in other cases it has not (Hartwig and Dunlosky, 2012).

In the present study, we do not claim measure spacing directly. Lab definitions of spacing are based on studying the same topic over multiple sessions. But, because our exams have multiple topics, some students who start studying early may not revisit the same topic in multiple sessions. Rather, in this study, we measure what we refer to as “spacing potential.” For example, if students study only on the day before the exam, there is little potential for spacing. If, instead, they are studying across 7 days, there is more potential for spacing. We collected two spacing potential measurements: (1) *cramming*, or the number of days in advance that a student began studying for the exam; and (2) *consistency*, or the number of days in the week leading up to an exam that a student studied. Based on our measurements, students with a higher spacing potential would exhibit less cramming and study more consistently than students with lower spacing potential. Because not every student with a high spacing potential may actually space out the studying of a single topic into multiple sessions, spacing potential is likely to underestimate the spacing effect; however, it is a practical way to indirectly estimate spacing in practice.

Importantly, not all difficult, or effortful, study tasks are desirable (Bjork and Bjork, 2011). For example, in the present study, we examined students’ level of distraction while studying. Distraction can come in many forms, commonly “multitasking,” or splitting one’s attention among multiple tasks (e.g., watching lectures while also scrolling through social media). However, multitasking has been shown to decrease working memory for the study tasks at hand (May and Elder, 2018). Thus, it may make a task more difficult, but in a way that interferes with learning rather than contributing to it.

In summary, available research suggests that active, effortful study strategies are more effective than passive ones; that cramming is less effective than distributing studying over time; and that focused study is more effective than distracted study. Whether students choose to use these more effective practices during their independent study time is a separate question.

### How Do Students Actually Study for Their Courses?

There have been several studies surveying students' general study habits. When asked free-response questions about their study strategies in general, students listed an average of 2.9 total strategies (Karpicke *et al.*, 2009). In addition, few students listed active strategies, such as self-quizzing, but many students listed more passive strategies, such as rereading.

There have also been studies asking whether what students actually do while they are studying is related to their achievement. Hartwig and Dunlosky (2012) surveyed 324 college students about their general study habits and found that self-quizzing and rereading were positively correlated with grade point average (GPA). Other studies have shown that using Facebook or texting during study sessions was negatively associated with college GPA (Junco, 2012; Junco and Cotten, 2012). While these findings are suggestive, we suspect that the use of study strategies and the relationship between study strategies and achievement may differ from discipline to discipline. The research we have reviewed thus far has been conducted for students' "general" study habits, rather than for specific courses. To learn about how study habits relate to learning biology, it is necessary to look at study habits within the context of biology courses.

### How Do Students Study for Biology Courses?

Several prior qualitative studies carried out within the context of specific biology courses have shown that students often report ineffective habits, such as favoring passive strategies or cramming. Hora and Oleson (2017) found that, when asked about study habits in focus groups, students in science, technology, engineering, and mathematics (STEM) courses (including biology) used predominantly passive strategies such as reviewing notes or texts, practices that in some cases were unchanged from high school. Tomanek and Montplaisir (2004) found that the majority of 13 interviewed students answered questions on old exams (100% of students) and reread lecture slides (92.3% of students) or the textbook (61.5% of students) to study for a biology exam, but only a small minority participated in deeper tasks such as explaining concepts to a peer (7.7% of students) or generating flashcards for retrieval practice (7.7% of students). We can also learn indirectly about students' study habits by analyzing what they would change upon reflection. For example, in another study within an introductory biology classroom, Stanton and colleagues (2015) asked students what they would change about their studying for the next exam. In this context, 13.5% of students said that using active strategies would be more effective for learning, and 55.5% said that they wanted to spend more time studying, many of whom reported following through by studying earlier for the next exam (Stanton *et al.*, 2015). In the current study, we extended prior research by exploring the prevalence of multiple study habits simultaneously, including the use of active study strategy and study timing, in a large sample of introductory biology students.

In addition to characterizing students' study habits, we also aimed to show how those study habits were related to performance in a biology classroom. In one existing study, there were positive associations between exam performance and some (but not all) active strategies—such as completing practice exams and taking notes—but no significant associations between performance and some more passive strategies—such as reviewing notes/screencasts or reviewing the textbook (Sebesta and Bray Speth, 2017). In another study, both self-reported study patterns (e.g., spacing studies into multiple sessions or one single session) and self-quizzing were positively related to overall course grade in a molecular biology course (Rodriguez *et al.*, 2018). We build on this previous work by asking whether associations between performance and a wide variety of study habits still hold when controlling for confounding variables, such as student preparation and total study time.

### Hypotheses

In this study, we asked whether students actually use cognitive psychologists' recommendations from the desirable difficulty framework in a specific biology course, and we investigated whether students who reported using those recommendations during studying performed differently on exams than those who did not. We wanted to focus on how students *spend* their study time, rather than the *amount* of time that they study, their level of preparation, or engagement. Therefore, we used regression analyses to hold preparation (i.e., ACT math and the course pretest scores), self-reported class absences, and overall study time equal. In this way, we estimated the relationship between particular study habit variables—including the strategies that students use, their timing of using those strategies, and their level of distraction while studying—and exam performance.

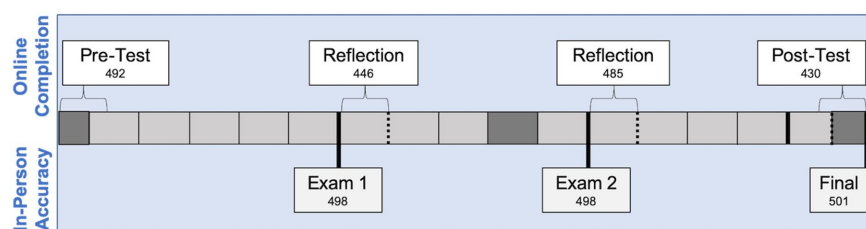
Based on previous research and the desirable difficulties framework, we hypothesized that:

1. Students would use a combination of active and passive strategies, but those who used more active study strategies or who devoted more of their study time to active strategies would perform higher on their exams than those who used fewer active strategies or devoted less time to active strategies.
2. Students would vary in their study timing, but those with less spacing potential (e.g., crammed their study time or studied less consistently) would perform worse, especially on long-term tests (final exam and course posttest), than students with more spacing potential.
3. Students would report at least some distraction during their studying, but those who reported being distracted for a smaller percent of their study time would score higher on exams than students who reported being distracted for a larger percent of their study time.

## METHODS

### Context and Participants

Data for this study were gathered from a large-enrollment introductory biology course (total class size was 623) during the Spring 2019 semester at a selective, private institution in the Midwest. This course covers basic biochemistry and molecular genetics. It is the first semester of a two-semester sequence. Students who take this course are generally



**FIGURE 1.** Timeline of assignments used in this study organized by mode of submission (online vs. in person) and grading (completion vs. accuracy). Exam days are indicated by thick lines. There were other course assignments (including a third exam), but they are not depicted here, because they were not analyzed in this study. Exam return is indicated by dotted lines. Light gray boxes represent weeks that class was in session. The number of consenting students who completed each assignment is indicated in the corresponding assignment box; the total number of consenting students was 502.

interested in life science majors and/or have pre-health intentions. The data for this study came from an on-campus repository; both the repository and this study have been approved by our internal review board (IRB ID: 201810007 for this study; IRB ID: 201408004 for the repository). There were no exclusion criteria for the study. Anyone who gave consent and for whom all variables were available was considered for the analyses. However, because the variables were different in each analysis, the sample differed slightly from analysis to analysis. When we compared students who were included in the first hypothesis's analyses to students who gave consent but were not included, we found no significant differences between participants and nonparticipants for ACT math score, pretest score, year in school, sex, or race (Supplemental Table 1). This suggested that our sample did not dramatically differ from the class as a whole.

Other than those analyses labeled “post hoc,” analyses were preplanned before data were retrieved.

### Timeline of Assignments Used in This Study

Figure 1 shows a timeline of the assignments analyzed in this study, which included the exam 1 and 2 reflections (both online), exams 1 and 2 (both in person), the course pre and post knowledge tests (both online), and a cumulative final exam (in person). As shown in the text boxes within Figure 1, the majority (85.7% [430/502] or greater) of students completed each of the assignments that were used in this study.

### Exam Reflections

Students' responses to exam 1 and 2 study habits reflections were central to all of our hypotheses. In these reflection assignments, students were asked to indicate their study habits leading up to the exam (see Supplemental Item 1 for prompts), including the timing of studying and type of study strategies. The list of strategies for students to choose from came from preliminary analysis of open-response questions in previous years. To increase the likelihood that students accurately remembered their study habits, we made the exercise available online immediately after each exam for 5 days. The reflection assignment was completed before exam grades were returned to students so that their performance did not bias their memory of studying. Students received 0.20% of the total course points for completion of each reflection.

### Exams

Exams in this course contained both structured-response (multiple-choice, matching, etc.) and free-response questions. The exams were given in person and contained a mixture of lower-order cognitive level (i.e., recall and comprehension) and higher-order cognitive level (i.e., application, analysis, synthesis, or evaluation) questions. Two independent (A.B and G. Y.) raters qualitatively coded exam questions by cognitive level using a rubric slightly modified from Crowe *et al.* (2008) to bin lower-order and higher-order level questions. This revealed that 38% of exam points were derived from higher-order questions. Each in-term

exam was worth 22.5% of the course grade, and the cumulative final exam was worth 25% of the course grade. To prepare for the exams, students were assigned weekly quizzes and were given opportunities for optional practice quizzing and in-class clicker questions as formative assessment. Students were also provided with weekly learning objectives and access to the previous year's exams. None of the exam questions were identical to questions presented previously in problem sets, old exams, or quizzes. Additionally, in the first week of class, students were given a handout about effective study strategies that included a list of active study techniques along with content-specific examples. Further, on the first quiz, students were asked to determine the most active way to use a particular resource from a list of options. The mean and SD of these exams, and all other variables used in this analysis, can be found in Supplemental Table 2. Pairwise correlations for all variables can be found in Supplemental Table 3

### Pre and Post Knowledge Test

As described previously (Walck-Shannon *et al.*, 2019), the pre/posttest is a multiple-choice test that had been developed by the instructor team. The test contained 38 questions, but the percentage of questions correct is reported here for ease of interpretation. The same test was given online in the first week of classes and after class sessions had ended. One percent extra credit was given to students who completed both tests. To encourage students to participate fully, we presented the pre and posttests as learning opportunities in the course to foreshadow topics for the course (pretest) or review topics for the final (posttest). Additionally, we told students that “reasonable effort” was required for credit. Expressing this rationale seemed to be effective for participation rates. While others have found that participation is low when extra credit is offered as an incentive (38%, Padilla-Walker *et al.*, 2005), we found participation rates for the pre- and posttests to be high; 97.4% of students completed the pretest and 85.9% of students completed the posttest.

### Statistical Analyses

To test our three hypotheses, we used hierarchical regression. We controlled for potential confounding variables in step 1 and factored in the study variable of interest at step 2 for each model. We performed the following steps to check that the

assumptions of linear regression were met for each model: first, we made scatter plots and found that the relationship was roughly linear, rather than curved; second, we plotted the histogram of residuals and found that they were normally distributed and centered around zero; and finally, we checked for multicollinearity by verifying that no two variables in the model were highly correlated (greater than 0.8). All statistical analyses were performed in JMP Pro (SAS Institute).

### Base Model Selection

The purpose of the base model was to account for potential confounding variables. Thus, we included variables that we theoretically expected to explain some variance in exam performance based on previous studies. First, based on a meta-analysis (Westrick *et al.*, 2015) and our own previous study with a different cohort in this same course (Walck-Shannon *et al.*, 2019), we expected academic preparation to predict performance. Therefore, we included ACT math and biology pretest scores in our base model. Second, the negative relationship between self-reported class absences and exam or course performance is well documented (Gump, 2005; Lin and Chen, 2006; Credé *et al.*, 2010). Therefore, we included the number of class sessions missed in our base model. Finally, our research questions focus on how students use their study time, rather than the relationship between study time itself and performance. Because others have found a small but significant relationship between total study time and performance (Credé and Kuncel, 2008), we controlled for the total number of hours spent studying in our base model. In summary, theoretical considerations of confounds prompted us to include ACT math score, biology pretest score, self-reported class absences, and self-reported exam study time as the base for each model.

### Calculated Indices

In the following sections we provide descriptions of variables that were calculated from the reported data. If variables were used directly as input by the student (e.g., class absences, percent of study time distracted) or directly as reported by the registrar (e.g., ACT score), they are not listed below.

**Total Exam Study Time.** In students' exam reflections, they were asked to report both the number of hours that they studied each day in the week leading up to the exam and any hours that they spent studying more than 1 week ahead of the exam. The total exam study time was the sum of these study hours.

**Number of Active Strategies Used.** To determine the number of active strategies used, we first had to define which strategies were active. To do so, all authors reviewed literature about desirable difficulties and effective study strategies (also reviewed in Bjork and Bjork, 2011, and Dunlosky *et al.*, 2013, respectively). Then, each author categorized the strategies independently. Finally, we met to discuss until agreement was reached. The resulting categorizations are given in Table 1. Students who selected "other" and wrote a text description were recoded into existing categories. After the coding was in place, we summed the number of active strategies that each student reported to yield the number of active strategies variable.

**Proportion of Study Time Using Active Strategies.** In addition to asking students which strategies they used, we also asked them to estimate the percentage of their study time they spent using each strategy. To calculate the proportion of study time using active strategies, we summed the percentages of time using each of the active strategies, then divided by the sum of the percentages for all strategies. For most students (90.0% for exam 1 and 92.8% for exam 2), the sum of all percentages was 100%. However, there were some students whose reported percentages did not add to 100%. If the summed percentages added to between 90 and 110%, they were still included in analyses. If, for example, the sum of all percentages was 90%, and 40% of that was using active strategies, this would become 0.44 (40/90). If the summed percentages were lower than 90% or higher than 110%, students were excluded from the analyses involving the proportion of active study time index.

**Number of Days in Advance Studying Began.** In the exam 2 reflection, we asked students to report: 1) their study hours in the week leading up to the exam; and 2) if they began before this time, the total number of hours and date that they began studying. If students did not report any study hours earlier than the week leading up to the exam, we used their first reported study hour as the first day of study. If students did report study time before the week before the exam, we used the reported date that studying began as the first day of study. To get the number of days in advance variable, we counted the number of days between the first day of study and the day of the exam. If a student began studying on exam day, this would be recorded as 0. All students reported some amount of studying.

**Number of Days Studied in Week Leading Up to the Exam.** As a measure of studying consistency, we counted the number of days that each student reported studying in the week leading up to exam 2. More specifically, the number of days with non-zero reported study hours were summed to give the number of days studied.

## RESULTS

The study strategies that students selected, the timing with which they implemented those strategies, and the level of distraction they reported while doing so are described below. We depict the frequencies with which certain study variables were reported and correlate those study variables to exam 1 and exam 2 scores. For all performance analyses described in the *Results* section, we first controlled for a base model described below.

### Base Model

We attempted to control for some confounding variables using a base model, which included preparation (ACT math and course pretest percentage), self-reported class absences, and self-reported total study hours. For each analysis, we included all consenting individuals who responded to the relevant reflection questions for the model. Thus, the sample size and values for the variables in the base model differed slightly from analysis to analysis. For brevity, only the first base model is reported in the main text; the other base models included the same variables and are reported in Supplemental Tables 5A, 7A, and 8A.

**TABLE 1. Specific study strategy prompts from exam reflections, listed in prevalence of use for exam 1<sup>a</sup>**

Study strategy abbreviation	Study strategy prompt	Type	Exam 1		Exam 2	
			N	%	N	%
Read notes	Read lecture slides or class notes	Passive	401	94.58	445	94.48
Completed problem sets	Answered the problem set questions	Active	365	86.08	393	83.44
Completed old exams	Answered the old exam questions	Active	348	82.08	341	72.4
Self-quizzed	Quizzed myself using ungraded weekly quizzes, quizlets, flashcards	Active	321	75.71	317	67.3
Explained concepts	Explained concepts to myself or others	Active	275	64.86	301	63.91
Synthesized notes	Paraphrased or outlined class notes	Active	257	60.61	326	69.21
Made diagrams	Made my own diagrams or comparison tables from lecture notes	Active	232	54.72	312	66.24
Reviewed online outside content	Reviewed online content from sources outside of the course (e.g., videos)	Passive	172	40.57	240	50.96
Watched lecture	Watched lecture videos	Passive	171	40.33	248	52.65
Read textbook	Read textbook	Passive	162	38.21	97	20.59
Attended review session/help sessions/office hours	Attended the exam review session, study hall, office hours, etc.	Mixed	101	23.82	59	12.53
Rewrote notes	Rewrote my class notes word for word	Passive	69	16.27	81	17.2

<sup>a</sup>The classification of the strategy into active and passive is stated in “type.” Prevalences for exam 1 ( $n = 424$ ) and exam 2 ( $n = 471$ ) are reported.

The base model significantly predicted exam 1 score and exam 2 score for all analyses. Table 2 shows these results for the first analysis; exam 1:  $R^2 = 0.327$ ,  $F(4, 419) = 51.010$ ,  $p < 0.0001$ ; exam 2:  $R^2 = 0.219$ ,  $F(4, 466) = 32.751$ ,  $p < 0.0001$ . As expected, all individual predictor terms were significant for both exams, with preparation and study time variables positively associated and absences negatively associated. For means and SDs of all continuous variables in this study, see Supplemental Table 2. We found that the preparatory variables were the most predictive, with the course pretest being more predictive than ACT math score. Total study time and class absences were predictive of performance to a similar degree. In summary, our base model accounted for a substantial proportion (32.7%) of the variance due to preparation, class absences, and study time, which allowed us to interpret the relationship between particular study habits and performance more directly.

### Did Students Who Used More Active Study Strategies Perform Better on Exams?

We first investigated the specific study strategies listed in Table 1. Then, we examined the total amount of time spent on active strategies to test our hypothesis that students who spent more time actively studying performed better on exams. Further, we counted the number of different types of active strategies that students used to test whether students who used a more diverse set of active strategies performed better on exams than those who used fewer active strategies.

**Study Strategies Differed in Their Frequency of Use and Effectiveness.** The frequency with which specific study strategies were employed is reported in Table 1. Almost all students reported reading notes. The next most prevalent strategies were active in nature, including that students (in order of prevalence) completed problem sets, completed old exams, self-quizzed, synthesized notes, explained concepts, and made diagrams. Surprisingly, each of these active strategies was used by the majority of students (54.7–86.1%) for both

exams 1 and 2 (Table 1). Less frequently used strategies included those more passive in nature, including that students (in order of prevalence) watched lectures, reviewed online content, read the textbook, and rewrote notes. A relatively infrequent strategy was attending review sessions, office hours, and help sessions. Because student engagement varied dramatically in these different venues, we classified this category as mixed. In summary, our results showed that, after reading notes, the most frequently used strategies were active strategies.

Next, we wondered whether the types of strategies that students reported using were related to exam performance. For these analyses, we added whether a student used a specific strategy (0 or 1) into the model, after controlling for the base model reported in Table 2. When holding preparation, class absences, and total study time equal, we found that, on average, students who reported having completed problem sets, explained concepts, self-quizzed, or attended review sessions earned 4.0–7.7% higher on average on both exams 1 and 2 than students who did not report using the strategy (see  $b_{\text{unstd}}$  in Table 3). Notably, these strategies were active in nature, except for the category attending review session, which was mixed in nature. The remaining active strategies were positively correlated to performance for only one of the exams. Additionally, we observed that the strategies categorized as passive were either nonsignificant or negatively related to performance on at least one exam. Together, these results suggest that active strategies tended to be positively related to exam performance. In our sample, each of these active strategies was used by the majority (more than half) of the students.

**The Proportion of Time Spent Using Active Strategies Positively Predicted Exam Score.** To further understand how active strategies related to performance, we investigated the proportion of study time that students spent using active strategies. On average, students spent about half of their study time using active strategies for exam 1 ( $M = 0.524$ ,  $SD = 0.244$ )

TABLE 2. Base model for hierarchical regression analyses in Table 3 for exam 1 ( $n = 424$ ) and exam 2 ( $n = 471$ )<sup>a</sup>

Base model	Exam 1				Exam 2			
	$\beta_{\text{std.}}$	$b_{\text{unstd.}}$	SE	$R^2$	$\beta_{\text{std.}}$	$b_{\text{unstd.}}$	SE	$R^2$
Intercept	–	36.441***	5.502	0.327	–	17.235*	7.949	0.219
ACT [0–36]	0.240***	0.965***	0.240		0.223***	1.248***	0.243	
Bio. pretest % [0–100]	0.419***	0.339***	0.035		0.333***	0.364***	0.047	
Number of classes missed before exam 1 (0–15) or between exams 1 and 2 (0–12)	–0.160***	–0.630***	0.158		–0.101*	–0.450*	0.183	
Total exam 1 or 2 study hours ( $\geq 0$ )	0.126**	0.132**	0.042		0.154***	0.203***	0.055	

<sup>a</sup>Standardized  $\beta$  values, unstandardized  $b$  values, and standard errors (SE) are reported. Each model's  $R^2$  is also reported. Ranges of possible values for each variable are shown in brackets. The values for the base models corresponding to Tables 4–6 vary slightly depending on the students included in that analysis (see Supplemental Table 4). The following symbols indicate significance: \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

and exam 2 ( $M = 0.548$ ,  $SD = 0.243$ ), though values varied from 0 to 1 (Figure 2). Importantly, students who spent a larger proportion of their study time on active strategies tended to perform better on exams 1 and 2. More specifically, after accounting for our base model (Supplemental Table 5A), the proportion of time students spent using active strategies added significant additional predictive value for exam 1,  $F(1, 416) = 8.770$ ,  $p = 0.003$ ,  $\Delta R^2 = 0.014$ ; and exam 2,  $F(1, 450) = 14.848$ ,  $p = 0.0001$ ,  $\Delta R^2 = 0.024$ . When holding preparation, class absences, and total study time equal, we found that students who spent all of their study time on active strategies scored 5.5% higher and 10.0% higher on exams 1 and 2, respectively, than those who spent none of their study time on active strategies (Table 4). Overall, these two results suggested that, on average, students spent about half of their study time using active strategies and students who devoted more study time to active strategies tended to perform better on exams.

*The Number of Active Strategies Used Positively Predicted Exam Score.* We next investigated the number of active strategies used by each student. On average, students used approximately four active strategies for exam 1 ( $M = 4.212$ ,  $SD = 1.510$ ) and exam 2 ( $M = 4.239$ ,  $SD = 1.501$ ). Very few students used no active strategies and most students (73%) used four or more active strategies (Figure 3). Further, those students who used more active strategies tended to perform higher on exams 1 and 2. More specifically, after accounting for our base model, the number of active strategies students used added significant additional predictive value for exam 1, ( $F(1, 416) = 33.698$ ,  $p < 0.0001$ ,  $\Delta R^2 = 0.024$ ; and exam 2,  $F(1, 450) = 91.083$ ,  $p < 0.0001$ ,  $\Delta R^2 = 0.066$ . When holding preparation, class absences, and total study time equal, we found that, for each additional active strategy used, students scored 1.9% and 2.8% higher on exams 1 and 2, respectively. Students who used all six active strategies scored 11.1% higher and 16.6% higher on exams 1 and 2, respectively, than those who used no active strategies

TABLE 3. Relating specific study strategy use to performance on exam 1 ( $n = 424$ ) and exam 2 ( $n = 471$ ) when controlling for preparation, class absences, and total study hours (base model)<sup>a</sup>

Study strategy	Type	Exam 1			Exam 2		
		$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>b</sup>	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>c</sup>
Read notes	Passive	–1.225	1.954	0.001	–2.634	2.705	0.002
Completed problem sets	Active	4.320***	1.292	0.018***	7.678***	1.637	0.034***
Completed old exams	Active	1.444	1.177	0.003	5.847***	1.373	0.029***
Self-quizzed	Active	4.377***	1.042	0.028***	5.395***	1.317	0.027***
Explained concepts	Active	5.538***	0.917	0.054***	6.288***	1.280	0.038***
Synthesized notes	Active	1.402	0.957	0.004	2.813*	1.375	0.007*
Made diagrams	Active	1.855*	0.926	0.007*	2.102	1.341	0.004
Reviewed online outside content	Passive	–1.839*	0.910	0.007*	–1.731	1.256	0.003
Watched lecture	Passive	0.194	0.968	0.001	–2.660*	1.305	0.006*
Read textbook	Passive	–1.352	0.911	0.004	–1.425	1.525	0.003
Attended review session/help sessions/office hours	Mixed	4.600***	1.028	0.031***	3.961*	1.900	0.007*
Rewrote notes	Passive	–0.853	1.232	0.001	–4.177*	1.659	0.011*
All strategies	–	–	–	0.114***	–	–	0.115***

<sup>a</sup>This table summarizes 12 two-step hierarchical models per exam. The first step adjusted for the base model (see Table 2), and the second step included one specific study strategy. Unstandardized  $b$  values, standard errors (SE), and the change in  $R^2$  relative to the base model are reported. Full model outputs are reported in Supplemental Table 4. Unstandardized  $b$  coefficients correspond to the average difference in exam score between the people who did and did not use the strategy. The following symbols indicate significance: \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

<sup>b</sup>Relative to a base model with  $R^2 = 0.327$  that included ACT score, bio. pretest %, number of classes missed before exam 1, and total exam 1 study hours. See Table 2 for base model details.

<sup>c</sup>Relative to a base model with  $R^2 = 0.219$  that included ACT score, bio. pretest %, number of classes missed between exams 1 and 2, and total exam 2 study hours. See Table 2 for base model details.

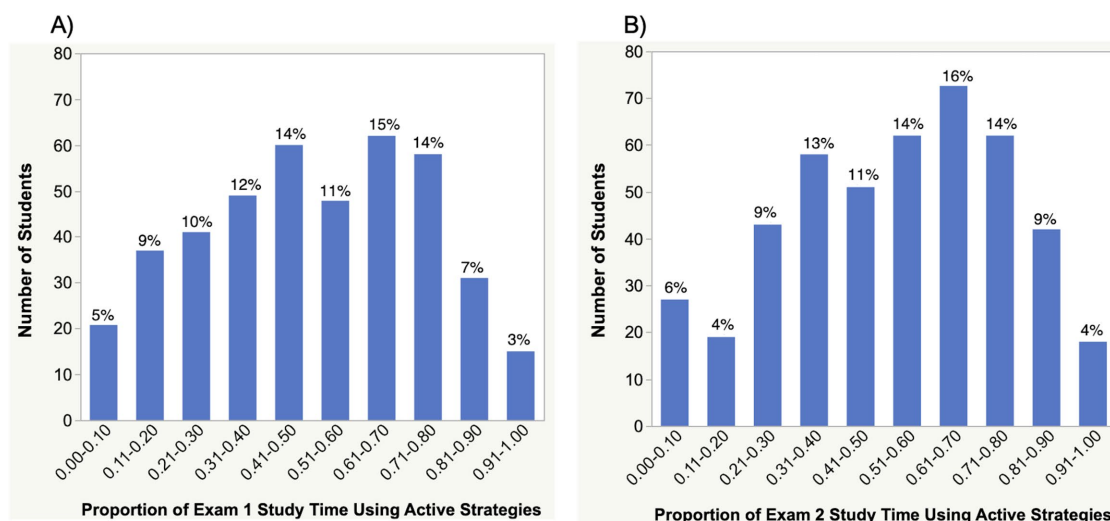


FIGURE 2. Distribution of the proportion of time that students devoted to active study for exam 1 ( $n = 422$ ) and exam 2 ( $n = 456$ ). Percentages of students in each bin are indicated.

(Table 4, See Supplemental Table 5A for base model). In summary, students who used a greater diversity of active strategies tended to perform better on exams.

**Post Hoc Analysis 1: Are Certain Active Strategies Uniquely Predictive of Performance?** Though it was not part of our planned analyses, the previous finding that the number of active strategies is predictive of performance made us question whether certain active strategies are uniquely predictive or whether they each have overlapping benefits. To test this, we added all six of the active strategies into the model as separate variables in the same step. When doing so, we found that the following active strategies were distinctly predictive for both exams 1 and 2: explaining concepts, self-quizzing, and completing problem sets (Supplemental Table 6). In other words, the portion of exam-score variance explained by certain active strategies was non-overlapping.

#### Did Study Timing Predict Performance on Immediate or Delayed Exams?

We next characterized students' spacing potential using two indices: 1) the number of days in advance that studying began (cramming) and 2) the number of days in the week leading up to the exam that a student studied (consistency).

Notably, in these results, we adjusted for our base model, which included total study time. In this way, we addressed the timing of studying while holding the total amount of studying equal. We examined outcomes at two different times: exam 2, which came close after studying; and the cumulative final exam and the posttest, which came after about a 5-week delay.

**Cramming Was Not a Significant Predictor of Exam 2, the Final Exam, or the Posttest.** While there was variation in the degree of cramming among students, this was not predictive of exam score on either immediate or delayed tests. On average, students began studying 5.842 d in advance of exam 2 ( $SD = 4.377$ ). About a third of students began studying 0–3 days before the exam, and another third began studying 4–6 days before the exam (Figure 4A). When holding preparation, class absences, and total study time equal, we found that the number of days in advance that studying began was not a significant predictor of in-term exam 2, the posttest, or the cumulative final (Table 5; see Supplemental Table 7A for base model).

**Studying Consistency Was Not a Significant Predictor of Exam 2, the Final Exam, or the Posttest.** While there was

TABLE 4. Relating active study strategy use to performance on exam 1 ( $n = 422$ ) and exam 2 ( $n = 456$ ) when controlling for preparation, class absences, and total study hours (base model)<sup>a</sup>

Study strategy	Exam 1			Exam 2		
	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>b</sup>	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>c</sup>
Proportion of exam study time using active strategies (0–1)	5.499**	1.857	0.014**	9.967***	2.587	0.024***
Number of active strategies used for exam studying (0–6)	1.858***	0.320	0.051***	2.759***	0.420	0.066***

<sup>a</sup>This table summarizes two two-step hierarchical models per exam. The first step adjusted for the base model, and the second step included either proportion of active study time or the number of active strategies. Full model outputs are reported in Supplemental Table 5. Unstandardized  $b$  values, standard errors (SE), and the change in  $R^2$  from the base model (see Supplemental Table 5A) are reported. Ranges of possible values for each variable are given in brackets. The following symbols indicate significance: \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

<sup>b</sup>Relative to a base model with  $R^2 = 0.322$  that included ACT score, bio. pretest %, number of classes missed before exam 1, and total exam 1 study hours. See Supplemental Table 5A for base model details.

<sup>c</sup>Relative to a base model with  $R^2 = 0.238$  that included ACT score, bio. pretest %, number of classes missed between exams 1 and 2, and total exam 2 study hours. See Supplemental Table 5A for base model details.



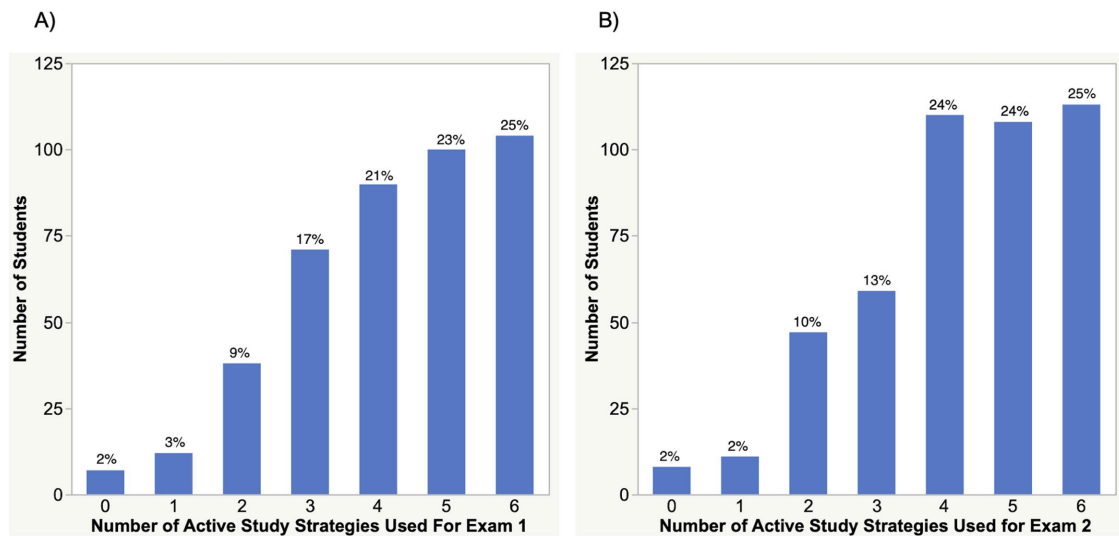


FIGURE 3. Distribution of the number of active strategies that each student used for exam 1 ( $n = 422$ ) and exam 2 ( $n = 456$ ). Percentages of students in each bin are indicated.

variation in how consistently students studied in the week leading up to exam 2, this consistency was not predictive of exam score either immediately or on delayed tests. On average, students studied 5 of the 8 days leading up to the exam ( $M = 5.082$ ,  $SD = 1.810$ ). Sixteen percent of students studied every day, and no students studied fewer than 2 days in the week leading up to the exam (Figure 4B). When holding preparation, class absences, and total study time equal, we found that the number of days studied in the week leading up to the exam was not a significant predictor of in-term exam 2, the posttest, or the cumulative final (Table 5; see Supplemental Table 7A for base model).

In summary, our students varied in both the degree of cramming and the consistency of their studying. Even so,

when holding preparation, class absences, and study time equal as part of our base model, neither of these spacing potential measures were predictive of performance on immediate or delayed tests.

#### Did Students Who Reported Being Less Distracted while Studying Perform Better on Exams?

In addition to the timing of studying, another factor that contextualizes the study strategies is how focused students are during study sessions. In the exam reflections, we asked students how distracted they were while studying. Here, we relate those estimates to exam scores while controlling for our base model of preparation, class absences, and total study time.

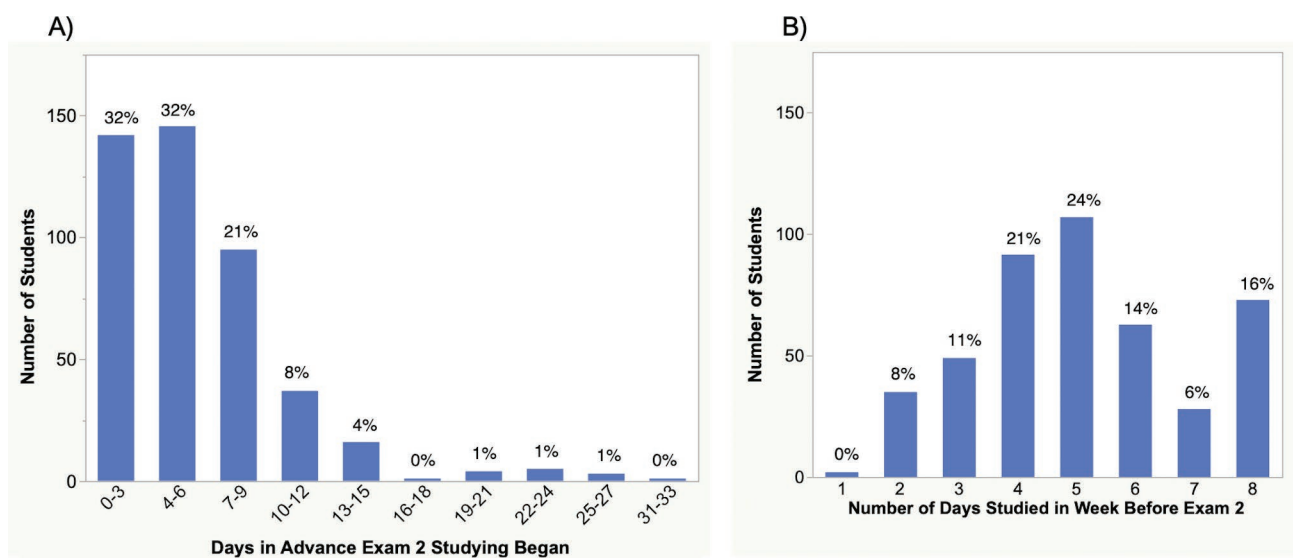


FIGURE 4. Distributions of spacing potential variables for exam 2 ( $n = 450$ ). (A) The distribution of the days in advance that exam 2 studying began (cramming); (B) the distribution of the number of days studied in the week before exam 2 (consistency). Percentages of students in each bin are indicated.

**TABLE 5. Relating spacing potential to performance on in-term exam 2 ( $n = 447$ ), the posttest ( $n = 392$ ), and the cumulative final exam ( $n = 450$ ) when controlling for preparation, class absences, and total study hours (base model)<sup>a</sup>**

Study habit	Exam 2			Posttest			Final exam		
	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>b</sup>	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>c</sup>	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>d</sup>
Days in advance exam 2 studying began ( $\geq 0$ )	0.011	0.168	0.000	0.252	0.162	0.003	-0.184	0.140	0.002
Number of days studied in week before exam 2 (0–8)	-0.464	0.418	0.002	-0.589	0.413	0.003	-0.608	0.348	0.004

<sup>a</sup>This table summarizes two two-step hierarchical regression models per assessment. The first step adjusted for our base model, and second step included either days in advance, or number of days studied in week before exam. Full model outputs are reported in Supplemental Table 7. Unstandardized  $b$  values, standard errors (SE), and the change in  $R^2$  from the base model (Supplemental Table 7A) are reported. Ranges of possible values for each variable are given in brackets. The following symbols indicate significance, \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

<sup>b</sup>Relative to a base model with  $R^2 = 0.231$  that included ACT score, bio. pretest %, number of classes missed before exam 2, and total exam 2 study hours. See Supplemental Table 7A for base model details.

<sup>c</sup>Relative to a base model with  $R^2 = 0.396$  that included ACT score, bio. pretest %, number of classes missed before exam 2, and total exam 2 study hours. See Supplemental Table 7A for base model details.

<sup>d</sup>Relative to a base model with  $R^2 = 0.273$  that included ACT score, bio. pretest %, number of classes missed before exam 2, and total exam 2 study hours. See Supplemental Table 7A for base model details.

### Distraction while Studying Was a Negative Predictor of Exam Score.

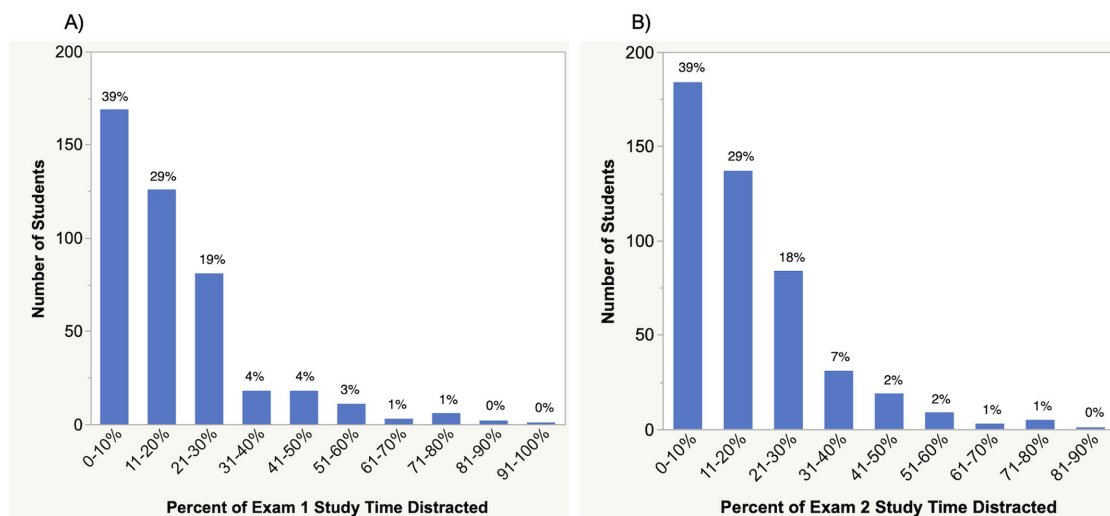
On average, students reported being distracted during 20% of their exam 1 and exam 2 study time (exam 1:  $M = 20.733$ ,  $SD = 16.478$ ; exam 2:  $M = 20.239$ ,  $SD = 15.506$ ). Sixty-one percent of students reported being distracted during more than 10% of their study time (Figure 5). Further, students who were more distracted while studying tended to perform lower on exams 1 and 2. After accounting for our base model, the percent of study time that students reported being distracted added significant additional predictive value for exam 1 and exam 2; exam 1:  $F(1, 429) = 12.365$ ,  $p = 0.000$ ,  $\Delta R^2 = 0.019$ ; exam 2:  $F(1, 467) = 8.942$ ,  $p = 0.003$ ,  $\Delta R^2 = 0.015$ . When holding preparation, class absences, and total study time equal, we found that students who reported being distracted 10% more than other students scored about 1% lower on exams 1 and 2 (Table 6; see Supplemental Table 8A for base model). In summary, this suggests that not only was it common for students to be distracted while studying, but this was also negatively related to exam performance.

## DISCUSSION

Students' independent study behaviors are an important part of their learning in college courses. When holding preparation, class absences, and total study time equal, we found that students who spent more time on effortful, active study strategies and used a greater number of active strategies had higher scores for exams. Yet neither students who started studying earlier nor those who studied over more sessions scored differently than students who started later or studied over fewer sessions. Additionally, students who were more distracted while studying tended to perform worse than students who were less distracted. In other words, both the degree to which students employed desirably difficult strategies while studying and their level of focus when doing so were important for performance.

### Specific Study Strategies (Hypothesis 1)

Our finding that more time and diversity of active study strategies were associated with higher exam grades was consistent with our hypothesis based on the desirable difficulties



**FIGURE 5. Distribution of the percent of time students reported being distracted while studying for exam 1 ( $n = 435$ ) and exam 2 ( $n = 473$ ). Percentages of students in each bin are indicated.**

**TABLE 6. Relating study distraction to performance on exam 1 ( $n = 435$ ) and exam 2 ( $n = 473$ ) when controlling for preparation, class absences, and total study hours (base model)<sup>a</sup>**

Study strategy	Exam 1			Exam 2		
	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>b</sup>	$b_{\text{unstd.}}$	SE	$\Delta R^2$ <sup>c</sup>
Percent of time distracted while studying (0–100)	–0.093***	0.027	0.019***	–0.119**	0.040	0.015**

<sup>a</sup>This table summarizes one two-step hierarchical regression model per exam. The first step adjusted for our base model, and second step included the percent of study time that students reported being distracted. Full model outputs are reported in Supplemental Table 8. Unstandardized  $b$  values, standard errors (SE), and the change in  $R^2$  from the base model (see Supplemental Table 8A) are reported. The range of possible values is given in brackets. The following symbols indicate significance: \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

<sup>b</sup>Relative to a base model with  $R^2 = 0.328$  that included ACT score, bio. pretest %, number of classes missed before exam 1, and total exam 1 study hours. See Supplemental Table 8A for base model details.

<sup>c</sup>Relative to a base model with  $R^2 = 0.219$  that included ACT score, bio. pretest %, number of classes missed between exams 1 and 2, and total exam 2 study hours. See Supplemental Table 8A for base model details.

framework, laboratory, and classroom research studies (Berry, 1983; King, 1992; Bjork, 1994; Karpicke and Roediger, 2008; Karpicke and Blunt, 2011; Hartwig and Dunlosky, 2012). Our study brought together lab research about effective strategies with what students did during self-directed study in an actual course. In doing so, we affirmed the lab findings that active strategies are generally effective, but also uncovered further nuances that highlight the value of investigating course-specific study strategies.

First, our study, when combined with other work, may have revealed that certain study strategies are more common than course-nonspecific surveys would predict. For example, compared with surveys of general study habits, our students reported relatively high use of active strategies. We found that the majority of students (73%) reported using four or more active strategies, which was more than the 2.9 average total strategies listed by students in a survey about general study habits at this same institution (Karpicke *et al.*, 2009). In particular, we found that two-thirds of students reported the active study strategy of self-quizzing. This was considerably higher than what was found in a free-response survey about general habits not focused on a specific course at the same institution (Karpicke *et al.*, 2009). In this survey, only 10.7% reported self-testing and 40.1% reported using flashcards. This higher frequency of self-quizzing behaviors may be due to a combination of factors in the course, the measures, and/or the students. In this course, we attempted to make self-quizzing easier by reopening the weekly quiz questions near exam time (Walck-Shannon *et al.*, 2019). We also used a course-specific survey rather than the more general, course-nonspecific surveys used in the previous research. Additionally, it is possible that, in recent years, more students have become more aware of the benefits of self-testing and so are using this strategy with greater frequency. When we compared our frequencies of several categories to analogous categories from course-specific surveys of introductory biology students (Sebesta and Bray Speth, 2017) and molecular biology students (Rodriguez *et al.*, 2018), we saw similar results. Combined with our work, these studies suggest that when students focused on a particular course, they reported more active strategies than when prompted about studying in general.

Second, the opportunity to control for potential confounding variables in our study, including total study time, allowed us to better estimate the relationships between specific strategies and performance. This approach was important, given concerns

raised by others that in classroom studies, benefits of certain strategies, such as explanation, could simply have been due to greater total study time (Dunlosky *et al.*, 2013). Our results showed that, even when controlling for total study time, self-explanation and other strategies were still significant predictors of performance. This helped illustrate that the strategies themselves, and not just the time on task, were important considerations of students' study habits.

Third, we were surprised by how predictive the diversity of active strategies was of performance. While we found that the proportion of active study time and the number of active strategies were both important predictors of performance, we found that the latter was a stronger predictor. This suggests that, if total study time was held equal, students who used a larger number of active strategies tended to perform better than those that used a smaller number of active strategies. This finding also deserves to be followed up in subsequent study to determine whether any of the active strategies that students use tend to co-occur in a “suite,” and whether any of those suites are particularly predictive of performance. We suspect that there is some limit to the benefit of using diverse strategies, as some strategies take a considerable amount of time to master (Bean and Steenwyk, 1984; Armbruster *et al.*, 1987; Wong *et al.*, 2002), and students need to devote enough time to each strategy to learn how to use it well.

Additionally, we found that particular active study strategies—explanation, self-quizzing, and answering problem sets—were uniquely predictive of higher performance in a biology course context. Undergraduate biology courses introduce a large amount of discipline-specific terminology, in addition to requiring the higher-order prediction and application skills found among STEM courses (Wandersee, 1988; Zukswert *et al.*, 2019). This is true for the course studied here, which covers biochemistry and molecular genetics, and the assessments that we used as our outcomes reflect this combination of terminology, comprehension, prediction, and application skills. Our results support the finding that active, effortful strategies can be effective on a variety of cognitive levels (Butler, 2010; Karpicke and Blunt, 2011; Smith and Karpicke, 2014); and this work extends support of the desirable difficulties framework into biology by finding unique value for distinct generative or testing strategies.

### Study Timing (Hypothesis 2)

Inconsistent with our second hypothesis that students with less spacing potential would perform worse than students with

more spacing potential, we found no relationship between study timing and performance on in-term or cumulative exams. Because we knew that spacing was difficult to estimate, we analyzed two spacing potential indices, the degree of cramming (i.e., the number of days in advance that students started studying) and the consistency of studying (i.e., the number of days studied in the week leading up to the exam). We controlled for total study time, because the spacing effect is defined as identical study time spread over multiple sessions rather than fewer, massed sessions. When doing so, neither of these measures were significantly related to performance.

There are a few possible explanations why we may not have observed a “spacing effect.” First, as explained in the *Introduction*, we measured spacing potential. It could be that students with high spacing potential may have arranged their studies to mass studying each topic, rather than spacing it out, which would lead us to underestimate the spacing effect. Second, students likely studied again before our cumulative final. This delayed test is where we expected to see the largest effect, and restudying may have masked any spacing effect that did exist. Third, we asked students to directly report their study time, and some may have struggled to remember the exact dates that they studied. While this has the advantage that it results in more sensitive and direct measures of students’ spacing potential than asking students to interpret for themselves whether they binarily spaced their studies or crammed (Hartwig and Dunlosky, 2012; Rodriguez *et al.*, 2018), students who did not remember their study schedules may have reported idealized study schedules with greater spacing, rather than realistic schedules with more cramming (Susser and McCabe, 2013), thus minimizing the expected spacing effect.

Despite the lack of a spacing effect in our data, we certainly do not advocate that students cram their studying, as we find it likely that students who started studying earlier may also have tended to study more. Also, those same students who studied earlier may have felt less stressed and gotten more sleep. In other words, even though our estimation of spacing potential did not capture performance benefits, benefits of spacing for well-being may be multifaceted and not wholly captured by our study.

### Distraction (Hypothesis 3)

Consistent with our third hypothesis, we found a negative relationship between distraction while studying and performance. This finding agreed with the few available studies that related distraction during self-directed out-of-class studying and grade, but differs in that our students reported a lower level of distraction than other published studies (Junco, 2012; Junco and Cotten, 2012). One possible reason for our low distraction estimate may have been that students were inadvertently underestimating their distraction, as has been reported (Kraushaar and Novak, 2010). In addition, some students may not have been including multitasking as a type of distraction, and this habit of multitasking while studying will likely be difficult to change, as students tend to underestimate how negatively it will affect performance (Calderwood *et al.*, 2016).

### Implications for Instruction

How can we leverage these results to help students change their habits? We present a few ideas of course structural changes that follow from some of the results from this study:

- To encourage students to use more active study strategies, try asking students to turn in the output of the strategy as a low-stakes assignment. For example, to encourage self-explanation, you could ask students to turn in a short video of themselves verbally explaining a concept for credit. To encourage practice quizzing, try to publish or reopen quizzes near exam time (Walck-Shannon *et al.*, 2019) and ask students to complete them for credit.
- To encourage students to use active study strategies effectively, model those strategies during class. For example, when doing a clicker question, explicitly state your approach to answering the question and self-explain your reasoning out loud. This also gives you an opportunity to add the rationale for why certain strategies are effective or provide advice about carrying them out. In addition to modeling a strategy, remind students to do it often. Simply prompting students to explain their reasoning to their neighbors or themselves during a clicker question helps shift students’ conversations toward explanation (Knight *et al.*, 2013).
- To encourage students to stay focused during studying, provide voluntary, structured study sessions. These could include highly structured peer-led team-learning sessions during which students work through a packet of new questions (Hockings *et al.*, 2008; Snyder *et al.*, 2015) or more relaxed sessions during which students work through problems that have already been provided (Kudish *et al.*, 2016).

### Limitations and Future Directions

There are multiple caveats to these analyses, which may be addressed in future studies. First, our data about study behaviors were self-reported. While we opened the reflection exercise immediately after the exam to mitigate students forgetting their behaviors, some may still have misremembered. Further, some students may not have forgotten, but rather were unable to accurately self-report certain behaviors. As stated earlier, one behavior that is especially prone to this is distraction. But, similarly, we suspect that some students had trouble estimating the percent of study time that they spent using each strategy, while their binary report of whether they used it or not may be more accurate. This may be one reason why the number of active strategies has more explanatory power than the percent of time using an active strategy. Separately, although students were told that we would not analyze their responses until after the semester had ended, some may have conformed their responses to what they thought was desirable. However, there is not strong evidence that students conform their study habit responses to their beliefs about what is effective. For example, Blasiman and colleagues found that, even though students believed rereading was an ineffective strategy, they still reported using it more than other strategies (Blasiman *et al.*, 2017). Another limitation due to self-reporting is that we lack knowledge of the exact, nuanced behaviors that a student carried out. Thus, a student who chose a strategy that we defined as active—such as “completing problem sets”—may have actually performed more passive behaviors. Specifically, while we did use verbal reminders and delay the release of a key when encouraging students to complete the problem sets and old exams before looking at the answers, some students may have looked up answers prematurely or may have read passively through portions of the key. These more passive behaviors may have underestimated the importance of active strategies. A second

limitation is that these data were collected from a course at a selective research-intensive institution and may not be applicable to all student populations. A third limitation is that our analyses are correlational. While we have carefully selected potential confounds, there may be other important confounding variables that we did not account for. Finally, it was beyond the scope of this study to ask whether certain subgroups of students employed different strategies or whether strategies were more or less predictive of performance for different subgroups of students.

Despite these caveats, the main point is clear. Students' course-specific study habits predict their performance. While many students in our sample reported using effective strategies, some students still had room to improve, especially with their level of distraction. One open question that remains is how we can encourage these students to change their study habits over time.

## ACKNOWLEDGMENTS

We would like to thank April Bednarski, Kathleen Weston-Hafer, and Barbara Kunkel for their flexibility and feedback on the exam reflection exercises. We would also like to acknowledge Grace Yuan and Ashton Barber for their assistance categorizing exam questions. This research was supported in part by an internal grant titled "Transformational Initiative for Educators in STEM," which aimed to foster the adoption of evidence-based teaching practices in science classrooms at Washington University in St. Louis.

## REFERENCES

- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22(3), 331.
- Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior*, 16(4), 297–306.
- Berry, D. C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology Section A*, 35(1), 39–49.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In Gernsbacher, M. A., Pew, R. W., Hough, L. M., & Pomerantz, J. R. (Eds.) & FABBS Foundation, *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In Metcalfe, J., & Shimamura, A. P. (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: Worth Publishers.
- Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory*, 25(6), 784–792.
- Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology*, 4(2), 145–153.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1133.
- Calderwood, C., Green, J. D., Joy-Gaba, J. A., & Moloney, J. M. (2016). Forecasting errors in student media multitasking during homework completion. *Computers and Education*, 94, 37–48.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569.
- Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, 3(6), 425–453.
- Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college. *Review of Educational Research*, 80(2), 272–295.
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4), 368–381.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14(1), 4–58.
- Gump, S. E. (2005). The cost of cutting class: Attendance as a predictor of success. *College Teaching*, 53(1), 21–26.
- Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19(1), 126–134.
- Hockings, S. C., DeAngelis, K. J., & Frey, R. F. (2008). Peer-led team learning in general chemistry: Implementation and evaluation. *Journal of Chemical Education*, 85(7), 990.
- Hora, M. T., & Oleson, A. K. (2017). Examining study habits in undergraduate STEM courses from a situative perspective. *International Journal of STEM Education*, 4(1), 1.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17(6), 649–667.
- Junco, R. (2012). Too much face and not enough books: The relationship between multiple indices of Facebook use and academic performance. *Computers in Human Behavior*, 28(1), 187–198.
- Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers & Education*, 59(2), 505–514.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471–479.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968.
- King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, 29(2), 303.
- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, 12(4), 645–654.
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14(2), 219–224.
- Kraushaar, J. M., & Novak, D. (2010). Examining the effects of student multitasking with laptops during the lecture. *Journal of Information Systems Education*, 21(2), 241–251.
- Kudish, P., Shores, R., McClung, A., Smulyan, L., Vallen, E. A., & Siwicki, K. K. (2016). Active learning outside the classroom: Implementation and outcomes of peer-led team-learning workshops in introductory biology. *CBE—Life Sciences Education*, 15(3), ar31.
- Lin, T. F., & Chen, J. (2006). Cumulative class attendance and exam performance. *Applied Economics Letters*, 13(14), 937–942.
- Marsh, E. J., & Butler, A. C. (2013). Memory in educational settings. In Reisberg, D. (Ed.), *Oxford library of psychology. The Oxford handbook of cognitive psychology* (pp. 299–317). Washington, DC: Oxford University Press.
- May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education*, 15(1), 13.
- Padilla-Walker, L. M., Thompson, R. A., Zamboanga, B. L., & Schmersal, L. A. (2005). Extra credit as incentive for voluntary research participation. *Teaching of Psychology*, 32(3), 150–153.
- Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(2), 291–300.
- Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, 97(1), 70–80.
- Rodriguez, F., Rivas, M. J., Matsumura, L. H., Warschauer, M., & Sato, B. K. (2018). How do students study in STEM courses? Findings from a

- light-touch intervention and its relevance for underrepresented students. *PLoS ONE*, *13*(7), e0200767.
- Sebesta, A. J., & Bray Speth, E. (2017). How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology. *CBE—Life Sciences Education*, *16*(2), ar30.
- Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*(7), 784–802.
- Snyder, J. J., Elijah Carter, B., & Wiles, J. R. (2015). Implementation of the peer-led team-learning instructional model as a stopgap measure improves student achievement for students opting out of laboratory. *CBE—Life Sciences Education*, *14*(1), ar2.
- Stanton, J. D., Neider, X. N., Gallegos, I. J., & Clark, N. C. (2015). Differences in metacognitive regulation in introductory biology students: When prompts are not enough. *CBE—Life Sciences Education*, *14*(2), ar15.
- Susser, J. A., & McCabe, J. (2013). From the lab to the dorm room: Metacognitive awareness and use of spaced study. *Instructional Science*, *41*(2), 345–363.
- Tomanek, D., & Montplaisir, L. (2004). Students' studying and approaches to learning in introductory biology. *Cell Biology Education*, *3*(4), 253–262.
- Walck-Shannon, E. M., Cahill, M. J., McDaniel, M. A., & Frey, R. F. (2019). Participation in voluntary re-quizzing is predictive of increased performance on cumulative assessments in introductory biology. *CBE—Life Sciences Education*, *18*(2), ar15.
- Wandersee, J. H. (1988). The terminology problem in biology education: A reconnaissance. *American Biology Teacher*, *50*(2), 97–100.
- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, *20*(1), 23–45.
- Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, *12*(2), 233–262.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70.
- ZuKswert, J. M., Barker, M. K., & McDonnell, L. (2019). Identifying troublesome jargon in biology: Discrepancies between student performance and perceived understanding. *CBE—Life Sciences Education*, *18*(1), ar6.