4. Nudge: Supporting Students' Study Time Allocation

4.1 Introduction

In the contextual design study (Chapter 3) I observed that many students did not know how to be better students and instructors did not include study knowledge (declarative, procedural or dispositional) in the curriculum. I identified this as an opportunity for which to design a new software system that tries to address this problem by operationalizing education theories and to provide data to inform such theories and their future applications (i.e., an operant probe system, defined in Chapter 2).

Through ideation of solutions, filtering by engagement with theory, and then by potential for uptake as determined by interviews with students and teachers, I settled on a rough description (and name) for the system: Nudge supports student time management by making course tasks explicit and notifying students of them when they are relevant.

This chapter describes the iterative development of the Nudge system and a semester-long study in a large chemistry course to evaluate its efficacy as an operant probe.

4.2 Background Theory

Not all students are studious. Many students cram (see Benjamin & Bird, 2006), although there is abundant evidence that spacing learning leads to better retention (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Cramming is often due to procrastinating, which from 46% (Solomon & Rothblum, 1984) to 95% (Ellis & Knaus, 1979) of college students do regularly. While more senior students may have college more figured out, they also procrastinate more (Semb, Glick, & Spencer, 1979). Procrastination is largely due to fear of failure (50%) and to averseness of the task (18%; Solomon & Rothblum, 1984) and has been associated with a variety of difficulties, including test anxiety, missed deadlines for assignments, poor semester grades, depressed affect, low self-esteem, and social anxiety (e.g., Beswick, Rothbun, & Mann, 1988; Ferrari, 1991; Ferrari et al, 1995; Lay, 1986, 1987; Lay & Burns, 1991; Solomon & Rothblum, 1984). Little surprise then that most students see their procrastination as a problem they would like to eliminate (Solomon & Rothblum, 1984).

Time management is difficult for students, but an important factor in their success. In a longitudinal study of cumulative GPA, a regression with time management skill and SAT scores showed time management to be a better predictor of GPA four years later (Britton & Tesser, 1991). Time management is made difficult by the human susceptibility to "planning fallacy", the tendency for people and organizations to underestimate how long they will need to complete a task, even when they consider their previous under-estimates (Kahneman & Tversky, 1979). One technique for abating the planning fallacy is to decompose the task, and this technique is more effective for tasks of greater complexity (Kruger & Evans, 2003).

Students don't choose their study behaviors based solely on the largest direct benefit to their learning (Thiede & Dunlosky, 1999). For example, students often use self-testing not as a learning activity but to diagnose their learning. Ironically, in a study comparing flash card practice with reading, students generally believed that more instruction (reading) would produce more learning, but chose flash card practice most frequently (Kornell, 2009). In another study, students reported that they went to lectures before reading their textbooks, despite thinking that reading the textbook and then going to class was more effective, probably because they also rated reading the textbook first as more difficult than going to the lecture first (B. G. Lee, 2006).

Despite these dissociations between beliefs and behaviors, students can be taught to be self-regulated learners. A classroom-based intervention study by (B. J. Zimmerman, Moylan, & Hudesman, 2011) showed struggling math learners how to self-reflect (i.e., self-assess and adapt to academic quiz outcomes) more effectively. Students receiving self-reflection training outperformed students in the control group on instructor-developed examinations and were better calibrated in their task-specific self-efficacy beliefs before solving problems and in their self-evaluative judgments after solving problems. The self-reflection training also increased students' pass rate on a national gateway examination in mathematics by 25% in comparison to that of control students (B. J. Zimmerman et al., 2011).

4.3 Core Features

Nudge began as an intention to develop a scalable software application to address the need perceived by both students and instructors to support students' time management. Following the fieldwork, I had established several design requirements for the application,

1. Require no upfront action by the student in order to benefit
2. Require no changes to the instructor's curriculum or schedule
3. Require little or no time from the instructor to offer in her course

Through a wide review of the relevant theoretic literature, I settled on several features for the system (Table 4-1). The first feature is to transform the course syllabus to organize course activities by date. The rationale for this was that explicit and salient dates more likely to be met, based in the findings that external deadlines boost task performance more than self-determined deadlines (Ariely & Wertenbroch, 2002) and students generally do whatever's due soonest (Kornell & Bjork, 2007). The second feature is to break down course study activities into smaller actions, such as turning an exam date into a series of tasks like "review lecture notes" and "take a practice test" each due well before the exam itself. The rationale for this was that decomposition of tasks improves time allocation and decreases aversiveness, based in the findings that smaller tasks abate the planning fallacy (Forsyth & Burt, 2008; Kruger & Evans, 2003), students procrastinate largely due to fear of failure (Solomon & Rothblum, 1984) and that in shared task lists, vague information preferred (Blandford, 2001). The third feature is to help students maintain and track their assignments, study time and progress through the course. The rationale was that recording task status increases awareness and inclination, based in the finding that self-monitoring of study behaviors improves learning (Richards, 1975). Table 4-1 lists two additional features to motivate students through rewards. These were derived from the theoretical literature but never implemented.

4.4 Iteration

With these core features defined, I then developed Nudge through a series of successive iterations, driven by field observations and theory. Nudge evolved from scenario sketches (see Chapter 3) to core features (above) to paper prototypes to graphic mockups (shown in Figure 4-1) to a production prototype (shown in Figure 4-2) for use by students in a real classroom setting.

In this production version (Figure 4-2) students log into the system to see a dashboard of all the tasks for the whole semester. They are laid out in a table with columns indicating the milestone for which they should do the task (e.g. Exam 1 or Lecture 8), its importance (e.g. Required, Advised, or only If Needed), the expected time the task will take, a description of the actual task with a link to resources needed to carry it out (like the homework web page), an indication of their currently reported status (e.g. S for started) and when will be or was due. Students can filter any of these columns, as in a spreadsheet but with more relevant categories. For example, the filter on the Due column has options for *Ever*, *Soon*, and *Past Due*. The status column has a filter for what's left *To Do*.

A progress dashboard is accessible by a link at the top of the work list (Figure 4-3). Here students can see quantitatively how much work they've done and how much is left. They can compare the counts of their status reports, for example how many they've finished versus skipped. They can also see what proportion of tasks they've completed with each importance. For example, 4 / 14 required tasks due so far. Finally they can review their report on each task with the time spent and any notes to self.

The production prototype was programmed with the Ruby 1.9 programming language, the Rails 3.0 (and later 3.1) web development framework, HTML5 document object model and SCSS for CSS3 document styling. The system was run on a Heroku platform-as-a-service dyno instance.

Nudge was first evaluated in a lecture course of 95 students in the spring of 2011. It was introduced in the 10^{th} week and one quiz grade was replaced with points for how much they reported into the system about what they had done. The effects of Nudge were measured by normalizing scores on each exam, averaging these z scores for the two pre-Nudge exams and the three post-Nudge exams, and comparing pre- with post-. Students who used Nudge when it was first made available in the 10^{th} week (N=9) saw their exam scores go up 0.36 sigma while students who started using Nudge in the 12^{th} week (N=12) saw their exam scores go down 0.31 sigma. Of course, this is not a true experiment, and that is why the formal evaluation below manipulates students' Nudge experience through randomly assigned experimental conditions.

While none of the features was evaluated in a controlled way, observations of their use revealed some factors that informed the next design interaction of Nudge.

- Students did not like logging into the web site. The experimental semester-long evaluation iteration of Nudge optimized for email-based interaction (Figure 4-4 Screenshot of an email sent in All-messages condition). Each email was a web form with radio buttons to indicate task status. At the bottom of the email users clicked a Record button to submit the data to the server and view the progress dashboard screen. Students asked that it be integrated with Blackboard but after months of effort investigating the technical and logistic requirements of the university, integration was abandoned as technically and logistically infeasible.
- The dashboard to track course progress was used regularly by few and most never interacted with it (Figure 4-3 Screenshot of course progress screen). This was left in the next iteration but not improved upon.
- Student's self-reports of time on task were very noisy. Systems that prompt for time on task, should be careful to motivate and support students in accurate reporting. In the next iteration this feature was left but not emphasized.
- Some notes in the "notes to instructor" field were feedback on the difficulty of assignments, but most were empty statements to mechanically maximize participation points.
- The instructor never spontaneously looked at any of the reports and few students expected them to. E.g. "The system seems ambiguous in terms of feedback. I don't think that my instructor will look at any comments, so I don't write any for them specifically." Student feedback that is never read may harm trust in the system and instructor. Such features should provide indications of whether feedback has been read. They may also need to push reports to the instructor. However this feature wasn't changed from the pilot.

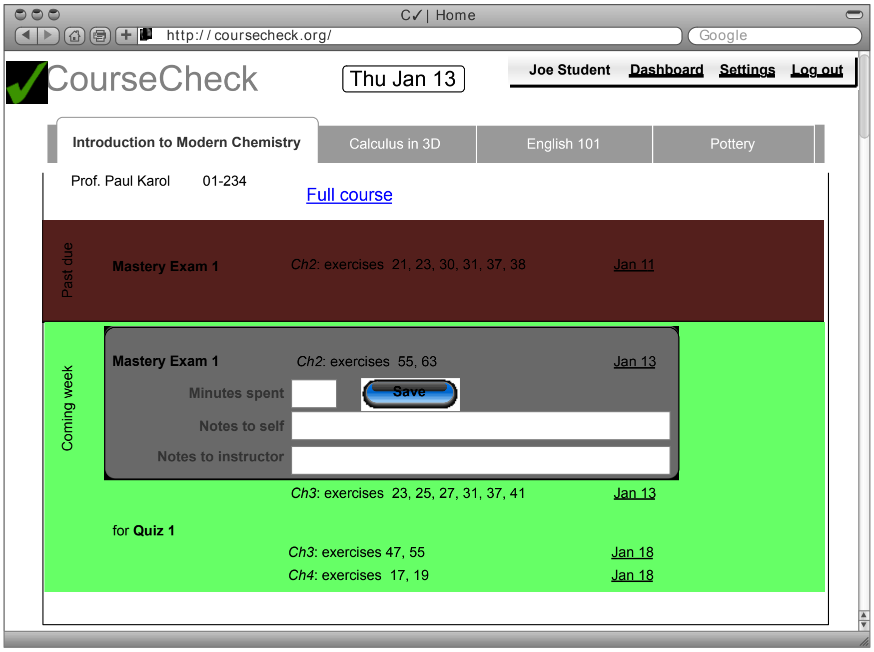


Figure 4-1 Nudge mockup

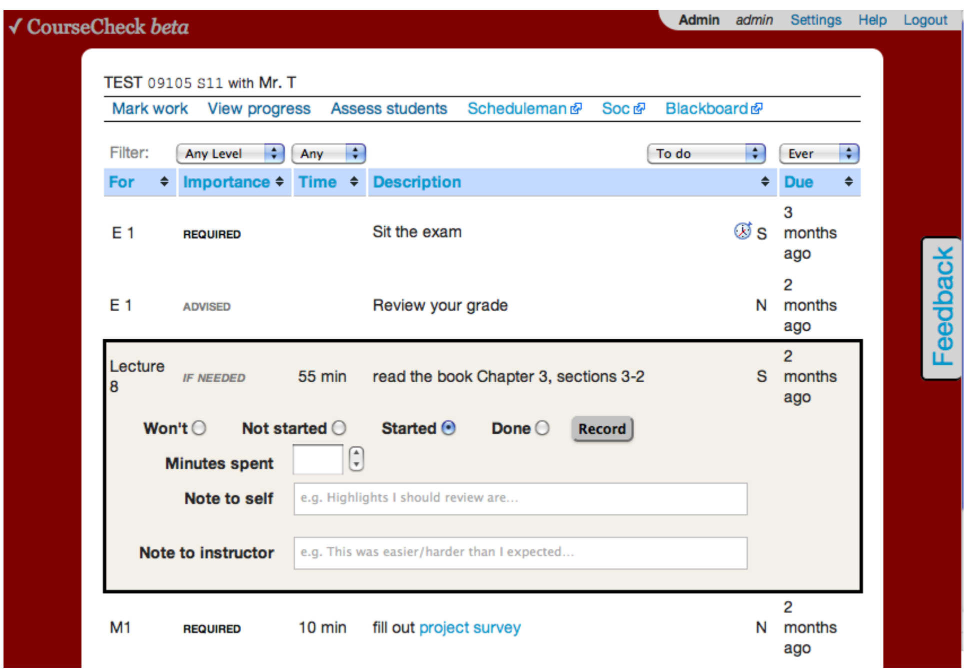
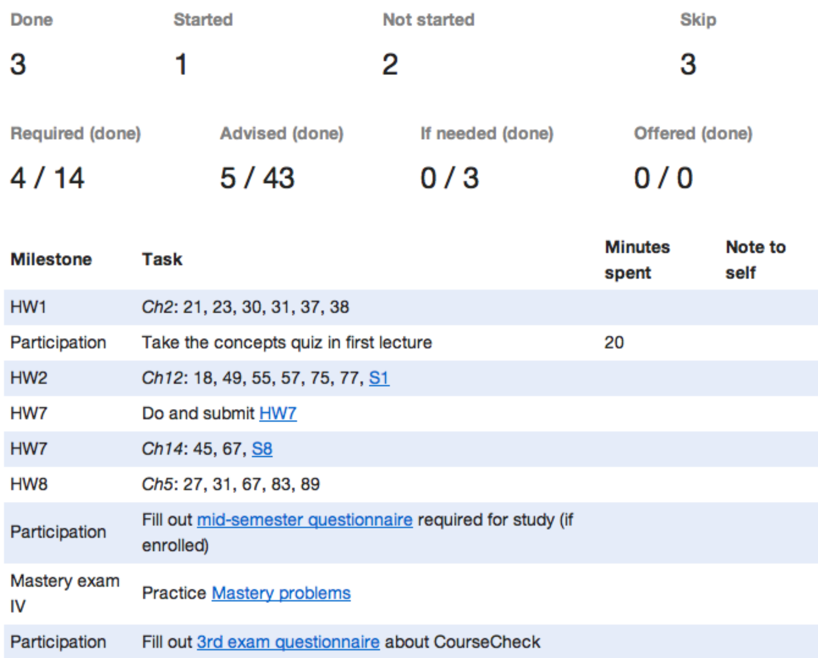


Figure 4-2 Screenshot of task list in pilot and final evaluation

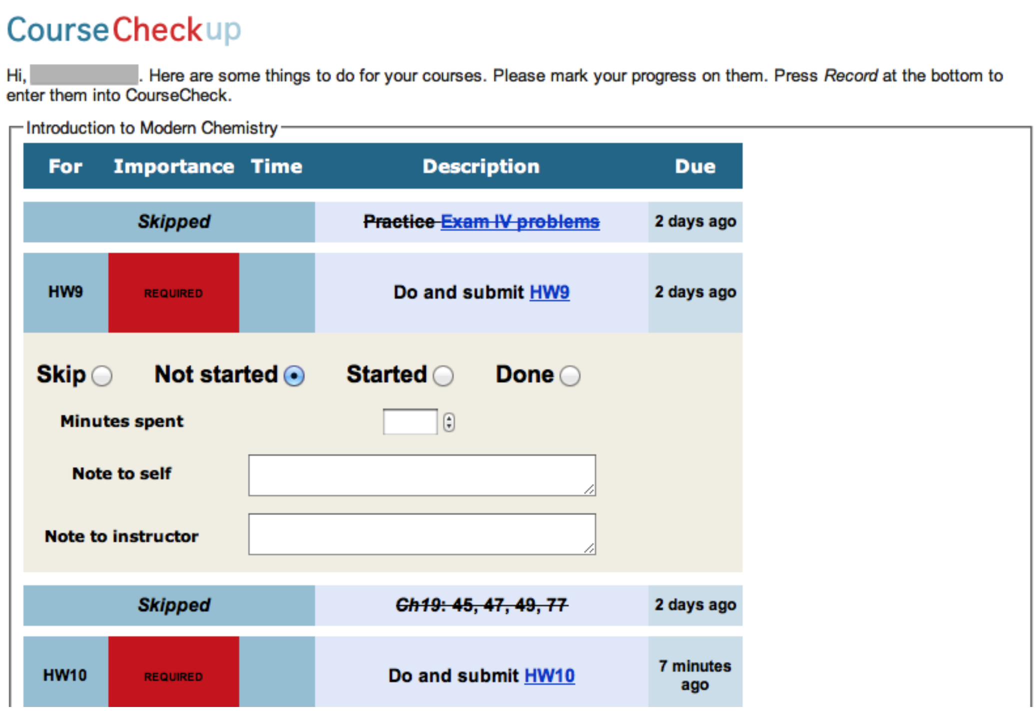


Done	Started	Not started	Skip
3	1	2	3

Required (done)	Advised (done)	If needed (done)	Offered (done)
4 / 14	5 / 43	0 / 3	0 / 0

Milestone	Task	Minutes spent	Note to self
HW1	*Ch2*: 21, 23, 30, 31, 37, 38		
Participation	Take the concepts quiz in first lecture	20	
HW2	*Ch12*: 18, 49, 55, 57, 75, 77, S1		
HW7	Do and submit HW7		
HW7	*Ch14*: 45, 67, S8		
HW8	*Ch5*: 27, 31, 67, 83, 89		
Participation	Fill out mid-semester questionnaire required for study (if enrolled)		
Mastery exam IV	Practice Mastery problems		
Participation	Fill out 3rd exam questionnaire about CourseCheck		

Figure 4-3 Screenshot of course progress screen



CourseCheckup

Hi, [redacted]. Here are some things to do for your courses. Please mark your progress on them. Press *Record* at the bottom to enter them into CourseCheck.

Introduction to Modern Chemistry

For	Importance	Time	Description	Due
Skipped			~~Practice Exam IV problems~~	2 days ago
HW9	REQUIRED		Do and submit HW9	2 days ago

Skip ○ Not started ◉ Started ○ Done ○

Minutes spent

Note to self

Note to instructor

For	Importance	Time	Description	Due
Skipped			~~*Ch19*: 45, 47, 49, 77~~	2 days ago
HW10	REQUIRED		Do and submit HW10	7 minutes ago

Figure 4-4 Screenshot of an email sent in All-messages condition

Table 4-1 Nudge feature matrix

Feature	Claim	Warrant	Status
Course task assigned to dates and organized centrally	Explicit and salient dates more likely to be met	External deadlines boost task performance more than self-determined deadlines (Ariely & Wertenbroch, 2002) Students generally do whatever's due soonest (Kornell & Bjork, 2007)	Implemented
Break-down of study activity into smaller actions	Decomposition of tasks improves time allocation and decreases aversiveness	Smaller tasks abate the *planning fallacy* (Forsyth & Burt, 2008; Kruger & Evans, 2003) Students procrastinate largely due to fear of failure (Solomon & Rothblum, 1984) In shared task lists, vague information preferred (Blandford, 2001)	Implemented
Maintaining and tracking assignments, study time and progress	Recording task status increases awareness and inclination	Self-monitoring of study behaviors improves learning (Richards, 1975)	Implemented
Reinforcement of effort demonstrated	Ss will spend more effort when effort itself is rewarded	Rewards on student effort can enhance achievement-directed effort (Brophy, 1987) Task-orienting strategies facilitate performance of Ss who de-emphasize role of effort (Stipek & Kowalski, 1989)	Not yet implemented
Surprise challenges and intermittent accolades	Game-like features increase fun	Intermittent rewards more motivating (Alberto & Troutman, 2008)	Not yet implemented

4.5 Experimental Design

With the experience of the pseudo-experimental pilot, parts of the Nudge system were improved (as discussed above) and an in vivo randomized controlled experiment was designed to evaluate Nudge as an operant probe. This formal study tested whether Nudge fit the context, achieved its desired effects, and could provide data to inform models of how its affects were achieved.

4.5.1 Context

The study took place in one section (n=136) of a large introductory chemistry course at a competitive private university. The instructor used Blackboard and a personal web site to provide students with a calendar of lectures and assessments, and regular announcements. The data collection and system intervention took place over the whole semester (Figure 4-5).

4.5.2 Task list

The course syllabus was recomposed into 60 tasks (14 required, 43 advised and 3 supplemental), which are all listed in Appendix E. In this evaluation the conversion was rather formulaic so it didn't require any domain or metacognitive knowledge. First I entered each assessment into a spreadsheet with its date, marking those tasks as *required*. Then I entered several ways to prepare for the assessment and marked them as *advised*. For homeworks these were simply the problems recommended by the instructor from his homework assignment listing. For exams, these were to take practice exams and review notes. The most time consuming part was to encode the web links to the resources to use for studying (e.g. linking to the actual practice problems online). In the table of tasks, each bracketed string was such a link. The expected time for completion for each task was very difficult to estimate and omitted from most tasks.

Entering the tasks into the system takes negligible time. As the software developer, I was able to enter the tasks into the system from this spreadsheet programmatically in minutes, taking less than a half hour total. The current authoring tool could require a novice user up to an hour for the task decomposition and entry, but the authoring interface is a rudimentary prototype. With optimization to speed entry and scaffold the elaboration of the syllabus, a novice could produce a better task list in less time. Because the entry doesn't require an expert understanding of the course, they could also outsource it. In a casual evaluation, Mechanical Turk workers typed in information from PDF syllabi for $1 and wrote by email to ask for more work.

4.5.3 Conditions

I evaluated the effects of Nudge by randomly assigning students to *all nudges* and *no nudges* conditions. Students in the *all nudges* condition were sent every task before it was due, grouped in emails sent at least weekly (e.g. Figure 4-4). Each email prompted students to reply with their status of completing each task (skipped, not started, started or completed.) Students in the *no nudges* condition were not sent any reminders before tasks were due. To collect data on their work for the course, after each exam they received one email with all the tasks and were prompted to indicate their completion status for each. All students also filled out questionnaires before, during and after the semester's instruction.

4.5.4 Hypotheses

Nudge was expected to help students' study time allocation and grades.

4.5.4.1 H-allocation

Students sent all Nudge messages exhibit better time use than students sent no Nudge messages.

The theory of operation of Nudge is that it helps students study more effectively by scaffolding, and ultimately causing, more effective allocation of time to study activities. For example, it should help students to review lecture notes immediately after the lecture to verify and repair their understanding.

4.5.4.2 H-grades

Students sent all Nudge messages perform better on assessments than students sent no Nudge messages.

Through better allocation of study time, students will learn more and will ultimately perform better on course assessments.

4.5.4.3 H-disposition

Students with poor study time use benefit more from Nudge messages.

The ethnographic observations were that some, but not all, students have difficulty with knowing how to study well. Nudge is designed for these students with poor study dispositions and is not expected to help as much students with good study habits.

4.5.5 Knowledge measures

All knowledge measures came from the normal course assessments. Accordingly, there are no formal pretest measures.

There were 4 non-cumulative exams (E1-4) distributed evenly over the term such that each exam covered the immediately preceding material. During the final exam period, a fifth exam was given that was cumulative and could replace a student's lowest non-cumulative exam grade.

4.5.6 Explanatory measures

Each student's personal attributes affect how she uses Nudge, which in turn affect how the tool affects her and her learning. Toward understanding how the tool works differently for different students, I logged user activities and collected several large questionnaires over the term.

Behavioral measures include students' interactions with Nudge, the data they reported through Nudge, and questionnaires about their time and study behaviors.

Study time allocation was operationalized as the Time/Environment scale (α=.71) of the Motivated Strategies for Learning Questionnaire (Pintrich, 2002; Pintrich, Smith, Garcia, & McKeachle, 2001). The scale has eight items and some were adapted to target math and science classes. E.g. "I make good use of my study time for math and science courses."

To see how students' goals in the course mediated their use and performance, the questionnaires also included several standard measures of goals. Of note in this analysis are the measures of the 2x2 achievement goal framework (Elliot & McGregor, 2001). This framework distinguishes students' conception of competence by two dimensions: personally mastering a domain versus demonstrating performance (definition dimension) and whether they are oriented to approaching success or avoiding failure (valence dimension).

4.5.7 Attrition and Missing Observations

11 students signed up for the study, but never did any coursework and were omitted from all analysis. 7 (13%) were in the *all nudges* condition and 4 (8%) were in the *no nudges* condition. The difference is not significant.

Of students who started the course, 2 (2.2%) dropped before the end (one from each condition). They are included in analyses for which their data are available.

4.5.8 Timeline

To help interpret the following results, Figure 4-5 Timeline of Nudge study shows a timeline of the course, assessments, and questionnaires and when changes were made to Nudge.

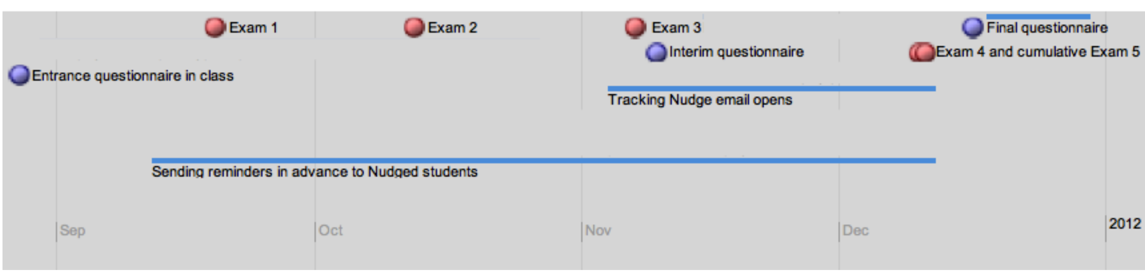


Figure 4-5 Timeline of Nudge study

4.6 Results

4.6.1 Descriptive statistics

Table 4-2 presents mean learning outcomes and pre/post time management scores by condition.

Table 4-2 Incoming attributes and outcomes

Group	Study time habits score (pre MSLQ: T/E)	Exam scores (Exams 1-5 mean)	Study time habits score (post MSLQ: T/E)	Passed course
No nudges (n=48)	5.2 (n=42, sd=0.7)	70.9 (n=43, sd=11.9)	4.8 (sd=0.8)	92% (44/48)
All nudges (n=45)	5.3 (n=40, sd=0.8)	69.5 (n=41, sd=12.3)	4.9 (n=31, sd=1.0)	93% (42/45)

4.6.1.1 Subjective rating

A questionnaire was given after the 3rd exam asking the usefulness of several features of the course. 75% of respondents who received all nudges (n=28) rated "Email reminders about course work" as "Good" or "Great". Five (17%) didn't perceive it as useful and two chose "didn't know about it", even though the survey was sent to the same addresses as the Nudge messages. Among students not in the study, who could choose whether to receive Nudge messages or not, 20% (14/70) opted out and a few opted to reduce their rate but continue to receive weekly messages.

About 40% of students responding to the final questionnaire agreed with the statement, "The reminder emails helped me in the class" (13/32). About 46% of Nudged students responding to the final questionnaire agreed with the statement, "I wish I could have email reminders for all my classes" (15/32), even though some of these students disagreed that it helped them in this class. 44% (14/32) agreed with the statement, "Without the reminders I would have forgotten to do something." Again some of these students disagreed with the previous statements.

In a measure of overall course satisfaction, students rated their agreement with "I achieved my goals for the course." The main predictor, not surprisingly, was their grade. Accounting for average exam grade ($p<.0001$), Nudged students agreed more ($F(1,58)=5.0$, $p=.029$). To see if this was due more to expectation or outcomes, a second covariate was tested, their responses on the midterm questionnaire indicating the final grade they expected to receive ($p=.002$). Nudged students were sure to agree ($p=.018$, 95% CI [0.09,0.97]) and not Nudged likely to disagree (95% CI [−0.67,0.22]).

4.6.1.2 Nudge usage

Table 4-3 Nudge reception

Group	Evidence of opening email (email image tracker in 4th quarter)	Evidence of opening email repeatedly (same message)	Number of messages opened (among evidence of opening)	Proportion opened more than two messages (among evidence of

				opening)
No nudges	38% (18/48)	25% (12/48)	1.6 (out of 2, n=18, sd=0.5)	-
All nudges	80% (36/45)	22% (10/45)	3.6 (out of 12, n=36, sd=3.3)	36% (13/36)

Table 4-3 presents usage measures of Nudge by condition. Nudged students opened about 80% of the messages sent to them once the email image tracker was in place. This is a floor estimate because some students may not have had told their email client to load the images which the tracker required. Only 38% of no-nudge students appear to have opened an email, but they also had only two occasions versus 12 for the nudged students (after the tracker). Over 22% of both groups opened some email messages repeatedly. The mean number of messages opened by nudged students was 3.6 out of 12, with a mode and median of 2.

Table 4-4 Nudge replies

Group	Replied to task polls ever	Median proportion of completes in reports on *non-required* tasks	Median proportion of completes in reports on *required* tasks	Agreement with "What I enter is accurate". (7pt Likert)
No nudges	83% (40/48)	.48 (n=40, sd=.27)	.87 (n=40, sd=.25)	5.5 (n=26, sd=1.4)
All nudges	87% (39/45)	.20 (n=39, sd=.22)	.75 (n=39, sd=.26)	5.8 (n=29, sd=1.3)

The no-nudge students received Nudge tasks after the exams and were asked to reply with their task status then. The rate of reply was roughly equal between the groups, but the no-nudge students reported higher rates of completion of the tasks, especially on non-required tasks. Because the two conditions were measured differently, and 15% of participants never replied to any task poll, the response data is not used in the following analyses. However it's worth noting that there was no difference in students' agreement with "The data I reported were accurate."

4.6.2 H-allocation

Students sent all Nudge messages exhibit better time use than students sent no Nudge messages.

For this hypothesis, the originally intended operationalization of study time allocation was students' reports of completion of advised tasks. However this measure is confounded with the different time and context of the task status polling between conditions. Nudged students were polled periodically in small batches before due dates while non-Nudged students were polled in large batches after all the due dates related to an exam. Table 4-4 shows that nudged students not only reported far fewer completions of non-required tasks, they also reported fewer completions of *required* tasks (such as turning in a graded homework assignment). I take this to be an effect of when and how they were polled, making comparisons between conditions on these measures uninformative.

Table 4-5 Reported hours spent on different activities

Week type	Group	Attending lecture	Attending recitation	Reading the book	Reviewing notes	Studying and solving problems
Regular	No nudges (n=34)	2.5 (sd=0.7)	1.6 (sd=0.6)	0.9 (sd=1.0)	0.8 (sd=1.0)	2.2 (sd=2.0)
	All nudges (n=33)	2.3 (sd=0.7)	1.9 (sd=0.7)	0.8 (sd=1.0)	1.0 (sd=1.0)	2.3 (sd=1.5)
Exam	No nudges (n=34)	2.4 (sd=0.7)	1.7 (sd=0.7)	1.2 (sd=1.1)	1.3 (sd=1.0)	4.2 (sd=1.8)
	All nudges (n=33)	2.4 (sd=0.6)	2.0 (sd=0.8)	1.1 (sd=1.2)	1.7 (sd=1.2)	4.6 (sd=2.0)

The best available measures of student time use are their reports in a questionnaire at the end of the course asking the hours per week they spent on different activities (see Table 4-5). The response levels were ordinal and recoded to continuous ("Less than 1 hour"=0.5, "1-2 hours"=1.5, "2-3 hours"=2.5, ">3 hours"=3.5). Nudged students reported spending more hours in a regular week attending their recitation sections ($F(1,64)=5.5$, $p=.023$). During an exam week the difference was only marginally significant ($p=.08$). They also spent marginally more time reviewing notes during exam weeks ($F(1,68)=3.0$, $p=.086$) and regular weeks ($p=.074$ one-tailed), however this may be an artifact of multiple comparisons. No significant differences were observed on time spent attending lecture, reading the book or studying and solving problems.

4.6.3 H-grades

Students sent all Nudge messages perform better on assessments than students sent no Nudge messages.

There were no main effects of nudge messaging observed on exam performance.

4.6.4 H-disposition

Students with poor study time use benefit more from Nudge messages.

In a model of the interaction of nudging with students incoming Time/Environment dispositions score, better time management led to better exam scores ($F(1,76.9)=6.4$, $p=.014$) but Nudge interacts to help students with poor management ($F(1,76.9)=4.6$, $p=.036$). That is, Nudge may compensate for poor time management dispositions. Digging deeper, I account for math aptitude ($F(1,63.7)=20.4$, $p<.0001$), the number of email messages opened within each nudge condition ($F(2,62.4)=3.0$, $p=.059$) and its interaction with the Time/Environment score ($F(2,62.6)=5.1$, $p=.009$). In this model, the worse a student's time use the greater the benefit of opening each Nudge message. For students with the highest Time/Environment scores, the number of messages they opened had no relationship with their exam scores. For students with the lowest Time/Environment scores, a predicted exam score of 58% would be 61% if they opened one message and by opening six messages they could match the predicted score of the best time managing students who opened none (76% on exam with 720 math SAT). Opening all twelve messages predicts a 90% score in this model. Because the number of messages opened is subject driven and not experimentally manipulated, it cannot be claimed to be the cause of the higher scores. It could instead be evidence of a third cause, which is the student's motivation to succeed in the class. However, the fact that for students with good time skills their exam scores exhibit hardly any relation with the number of messages they opened suggests that the Nudge system created an opportunity whereby motivated students with poor time management could overcome this deficit.

4.6.5 Student perceptions

Different students perceived the system differently, but in several pretty consistent themes. In the final questionnaire, students were asked, "If the email reminders were a person, what kind of person would it be?" The responses mostly described one's relationship to the person. "My mother" was echoed by in several responses. "They would be sort of like the mother that's always around and making sure you're on top of your school work." From another,

If the email reminders were a person they would be very annoying and nosy, but good-hearted. They would be something of a mother, always checking up on you and wondering how much you have done, and even though you may at times get irritated, you would never give up a friendship with this individual.

Several echoed the idea of a friend, e.g. "A close friend who is pushing me" and "My best friend. The only person that would tell me to get my work done because I tend to forget about assignments sometimes. This person would be on top of their work as well and displayed great academic success."

A more negative theme was a well-intentioned but annoying and dense nag. E.g. "the email reminders would be a person that did not really care about what they were saying. they need to be more upbeat or motivating and something thst [sic] someone would actually listen too." Many said simply, "persistent", and some explained that eventually they ignore such formulaic persistence: "That person who runs by your house every day to the point when eventually you stop noticing him" and "They'd be the kind of person you say 'Hi' to because you feel obligated when passing them on campus, but in reality, you do not associate with that person in any way." Some students expressed that despite being annoying, you appreciate this person:

They would be a person that slightly annoyed me, but more because I wouldn't want him/her pointing out my flaws or what I've missed/forgotten. In the long run, I'd appreciate that person a lot, as he/she helped me and kept me on track.

4.6.6 Feature Validation

The features of Nudge were based in a few theoretically derived design claims. How does the evidence from this evaluation support or cast doubt on these claims? Because all the features were tested together none has rigorous evidence either way, but some observations from student use may help inform them.

4.6.6.1 Explicit and salient dates more likely to be met

This was implemented through organizing course tasks by date and regularly emailing students with what tasks were coming up (e.g. a homework or exam). The results are consistent with an interpretation in which this claim is true. However some caveats to the rationale are that while students generally do whatever's due soonest (Kornell & Bjork, 2007), many of the "due" dates in the evaluated course were not conventional due dates when a student must do something or lose points. For example, the "take a practice exam" task due a few days before the exam can be ignored with no direct consequences. It's unclear how much students bought into these dates as dates by which they should do the task. The fact that the dates were external (Ariely & Wertenbroch, 2002) may have motivated some students, but it may have led other students to distrust the dates. Some students expressed some of the "due" tasks had no value to them: "there are some assignments that are not required that I don't feel are necessary for my understanding of the material." To increase adoption, it is worth considering allowing students to customize the due dates. To maintain the externally imposed nature, the dates and tasks could be part of a high level setting of, for example, high/medium/low effort, or more concretely a range of hours per week they will allocate to the course. With access to their course grades it could help them meet a target grade by adapting to their course performance.

The salience of the dates is another issue to better explore. While no students expressly commented on the choice of dates, many were bothered by the frequency of the reminders. "I do not like the everyday emails, because they are sometimes excessive." One student suggested, "Instead of daily reminders maybe biweekly reminders." However, one student actually requested a higher frequency: "I would like if the check ups were more often, or if it could send reminders a few hours before homework is do." It seems the reminders are annoying until they save you:
What I like most about [Nudge] is that it lets you know when you have an assignment due. Usually, I am very good at remembering due dates, but one week, I forgot to write down my homework for the week and was therefore under the impression that nothing

was due that week. It wasn't until I got an email reminder from [Nudge] that I had assignments due that I realized my mistake.

One way to address this feedback would be an escalating nag factor as the deadline approaches, customized to the user's preference.

4.6.6.2 Decomposition of tasks improves time allocation and decreases aversiveness

This was implemented by breaking the course tasks, such as taking an exam, into a series of smaller actions. The fact that the effects of the messages interacted with students' time management skills lends support to this design claim and the earlier findings that smaller tasks abate the planning fallacy (Forsyth & Burt, 2008; Kruger & Evans, 2003). The study produced no evidence related to the rationale that vague task information is preferred (Blandford, 2001) but there is data to affirm the idea that students procrastinate largely due to fear of failure (Solomon & Rothblum, 1984).

The questionnaires at the beginning and end of the semester included measures in the 2x2 achievement goal framework (Elliot & McGregor, 2001). The Performance Avoidance goal, for example, is measured by agreement with statements like, "My fear of performing poorly in this class is often what motivates me." Students receiving all the Nudge messages ended up with a stronger orientation to the performance avoidance goal ($F(1,91)=4.56$, $p=.035$), accounting for their earlier rating ($p<.0001$). This goal orientation is a consequence of, but distinct from, a fear of failure (Bartels & Magun-Jackson, 2009; Elliot & McGregor, 2001). It is possible that the effects of the Nudge messages work through increasing students' fear of failing in the class while decreasing their fear of failing in any particular study task. Is this good though? Performance avoidance goals have been found to be negatively correlated with learning outcomes and cognitive self-regulatory activities ("Goals and Goal Orientations," 2008). As a mechanism, performance avoidance goals have been found to be a positive predictor of surface learning (Liem, Lau, & Nie, 2008). This points to an alternative explanation for the interaction of Nudge with time management skills. For students with poor skills, increasing their attention to surface learning activities may have produced a net gain in attention to the course. For more studious students, this greater attention to the course tasks (surface features of learning) may have been at the expense of deeper cognitive engagement, supplanting their own strategies for regulating their learning with those of the course syllabus. Can fear of failure be channeled into deeper learning? Further yet, can time management be supported while inducing a more productive achievement orientation (e.g., performance approach goals, or even mastery approach goals)? Both of these questions highlight important design spaces to explore.

4.6.6.3 Recording task status increases awareness and inclination

This claim was implemented by prompting students to record their task status. In the current design, students not required to record as part of the study did not record. Because the task status records are so noisy, there is no quantitative measure of specifically whether recording tasks increased their awareness and inclination per the rationale that self-monitoring of study behaviors improves learning (Richards, 1975). However, qualitative data make clear that some students valued the progress monitoring enabled by task recording. "I like the way it keeps track of your progress" and "Sometimes it is nice to look back and see how many of the assignments I have completed." Specifically regarding motivation, "It makes me realize how much I should spend doing my work" and "I like that it makes me feel accomplished since I get everything done on time." For one student, the "not started" status option was their favorite feature of the system. "I like the fact that it gives you an opportunity to answer the questions very honestly with the 'not started' option."

4.7 Limitations and Opportunities

4.7.1 Operation on desired outcomes

The usage rates were low. 20% of nudged students never opened the emails. Some of these may have just not opened them during the 4th quarter when the email tracker was in operation or had their email clients set to not load images. However among people who definitely did open, 64% only opened one or two of the twelve sent. Both of these limitations suggest that some students do not take the treatment. Clearly the system should be more tailored to each student's dispositions and course performance. How exactly is an open question. If it's completely optional, then poor students may not see the need for it. If it's mandatory, it may hinder some students. In future work, I would explore the potential for motivating participation, beginning with the last two features in Table 4-1 that have theoretical support but have not yet been implemented.

There was no overall effect of the Nudge messages on student learning and for students with good time management they may have even been counterproductive. The messages appear to help students with poor time management (i.e. those in need of help) but this experiment didn't have enough power to provide evidence for a main effect in this subpopulation. Future work should study nudge messaging where a larger proportion of students have poor time management. Further, Nudge messages increased students' performance-avoidance goal orientation, which is generally predictive of worse learning and self-regulation. This may explain the negative impact on good time managing students. Future work should provide better messages, matched to the needs and proximal abilities of different learners.

A key way to improve the impact of the Nudge messages is to improve the messages themselves. The set of tasks defined in this course were limited and did not specify all the good study activities to do well in the course. In the formally evaluated version, there were no reminders for instruction, only practice. There is a body of literature on best study practices that could be operationalized into ideal tasks and messages. What this study demonstrates is that the Nudge system works in vivo and is easy to deploy. The Nudge operant probe provides a new mechanism for research to test these study ideals in real-world settings and discover the distribution and boundaries of their effects with different curricula and students.

While Nudge is domain general, the messages had to be authored for the course in the study. Adding new curricula to nudge could be facilitated by having a set of templates to elaborate task structures around different typical course events. For example, each exam could have associated with it: reread notes (6 days before), study worked examples (8 and 4 days before), take practice exam (3 and 1 days before), and attend review session (2 days before). Were researchers to test their theories of optimal practices, these could become standard task expansions of traditional course events. Adding courses could be as simple as uploading a syllabus, scanning for key dates (homework due, quiz or exam given) and automatically expanding them into task sets to generate a task set for the whole course. These would certainly be improvements, but not necessary for adoption. The instructor in the study, was asked after seeing the results whether he would take the time to type in the dates to use it again and replied, "Yes, very much. I would say emphatically."

4.7.2 Probe data for modeling

One aspect of the probing utility of Nudge that didn't work as hoped was the set of student responses about what tasks they had done. There was a confound for comparing response between conditions, but that could be addressed easily in future work by polling the same times and ways. More problematically, it's not clear how accurate the task reports are. 41% of students "strongly agreed" to "What I enter is accurate" but 28% didn't agree, and that's among the 58% who took the time to respond to that questionnaire. Conservatively, only a quarter of students in the study strongly agreed to having entered accurate task reports. For these data to be useful in modeling student study behaviors naturally, the design of acquiring them needs to be greatly improved. An important factor is students' incentives for entering accurate data (or any data at all). Game-like motivations could help. Reinforcement of effort demonstrated and intermittent rewards are two planned features, supported by theory, that have yet to be implemented in Nudge.

The particular tasks authored define the data that can be collected from students. While there are some tasks or behaviors that may be more effective for student learning, there may be others that are more valuable for student modeling and intervention diagnostics. If the system could elicit accurate reporting from students it would open another area of inquiry. Key factors for modeling students could be added to the task expansions for the purposes of different probing studies.

4.8 Conclusion

Nudge was designed to improve learning outcomes in university lecture courses using observations from the field and theories from existing literature. In a large introductory chemistry course, Nudge helped students with poor time management dispositions to learn and perform better on course exams. The benefit to such a student was greater the more of the Nudge emails they opened.

The process of designing Nudge helps light the way for the design of similar systems. Explicit and salient dates may be more likely to be met but they can be too explicit and too salient to the point of being ignored. Instead they should be due dates and messaging policies that students buy into through a choice they make of what effort to allocate for the course. There was no direct evidence that decomposition of tasks improved time allocation but it may have decreases aversiveness. Students who received all Nudge messages went up in their performance-avoidance goal orientation, an indication of being motivated to avoid performing poorly. This motivation orientation is not best for excellence, but it is an increase in motivation that can help students perform better. There was also support for the design claim that recording task status increased students' awareness and inclination to perform course work. Each of these merits further exploration.

Nudge highlights the opportunity to support students' time management skills to improve their learning. The formal evaluation and analysis of its design principles shine a light on new opportunities for research and real world impact through operant probes for applied learning science.