# Procrastination, Prompts, and Preferences: Evidence from Daily Records of Self-Directed Learning Activities

Rina Kikuchi[a] and Kazuki Onji[b*]

[a] Shiga University Faculty of Economics, 1-1-1 Bamba, Hikone, Shiga, 522-8522, Japan
[b] ANU Crawford School of Economics and Government, Canberra ACT 0200, Australia

April 2009

## Abstract

To complement the theoretical analysis on the self-control problems of decision makers, this paper empirically examines a remedy for procrastination. The setting for our study is university coursework, and we utilize unique data on daily records of self-directed learning activities. With quasi-experiments arising from the different frequency of interventions across classes, we examine the hypothesis that in-class prompts by an instructor mitigate the degree of procrastination. Further, with a registration mechanism that generates the grouping of students by their own preference, we consider whether student–class matching affects students' responsiveness to prompts. In a sample of Japanese undergraduates, we find that prompts affect behavior, especially when reinforced. The impact of intervention, however, appears to be dependent on the class preferences and the timing of prompts. The study suggests that a minimally interventionist policy may have real impact but may fail to influence target groups.

JEL Classification: I21, J24, D91, C81, D83

Key words: Time-inconsistent preferences, Monitoring, Skill formation

* Corresponding author. Tel.: +61 2 6125 0134; Fax.: +61 2 6125 0767
E-mail address: Kazuki.Onji@anu.edu.au (Onji); kikuchi@biwako.shiga-u.ac.jp (Kikuchi). We would like to thank Aki Asano, Bruce Chapman, Deborah Cobb-Clark, Shunichi Ishihara, Andrew Leigh, Tue Gørgens, Kei Sakata, Norio Takeoka, Chika Yamauchi, and seminar participants for comments and reactions to the earlier versions of this paper presented at RSSS—ANU, Japan Centre—ANU, Japanese Studies Association of Australia 2007 Meeting, Econometric Society Australia Meeting 2007, Kansai Labor Workshop/COE Behavioral Economics Workshop (Osaka University), and Ritsumaikan University. Onji has benefited from the research opportunities provided by the postdoctoral fellowship at the Australia-Japan Research Centre, ANU. Kikuchi would like to acknowledge the funding support from the Project Centre, Shiga University. Any remaining errors are ours.

## 1. Introduction

There is a growing body of studies that examines the self-control problem of decision makers arising from time-inconsistent tastes for immediate gratification (Strotz, 1956; Laibson, 1997).[1] The tendency to seek immediate gratification can lead to suboptimal outcomes after the fact (e.g. health problems due to cigarette consumption), and recent theoretical studies have been concerned with paternalistic intervention policies (Choi, Laibson, Madrian & Metrick, 2003; Thaler & Sunstein, 2003; O'Donoghue & Rabin, 2003, 2006). Those studies highlight the advantages of "minimally interventionist policies" (MIPs) as opposed to heavy-handed policy interventions, arguing that MIPs prevent "causing harm to those for whom the behavior is rational" (O'Donoghue & Rabin, 2003). However, given the lack of hard incentives to change behavior, a concern remains as to what extent MIPs can influence the behavior of target groups. In the economics literature, however, there are relatively sparse empirical attempts to evaluate policy interventions for mitigating self-control problems.

The aim of this paper is to complement the analysis of MIPs with an empirical examination. The setting for our empirical study is a classroom. Since the difficulties in examining the effectiveness of MIPs in the field have limited the development of the empirical literature, we think a student sample is a useful starting point in examining the extent of the impact of MIPs, despite the concern about the generalizability to economic settings. We focus on procrastination, which is defined in the *Oxford English Dictionary* as "the action or habit of postponing or putting something off." Unlike the negative

[1] For surveys, see O'Donoghue & Rabin (1999, 2005) and Frederick, Loewenstein & O'Donoghue, (2002).

connotation in the common usage of the word, procrastination does not necessarily lead to worse outcomes (Ferrari, Johnson & McCown, 1995). However, academic procrastination can, depending on the subject matter, affect the degree of human capital formation. Thus, in a setting where delayed actions negatively affect outcomes, procrastination is relevant to economists who are interested in evaluating interventions. The scope of this paper is to consider an intervention that would be relevant when decision makers have present-biased preferences, which is a type of self-control problem that results in procrastination in performing unpleasant tasks. We think that it is intuitive and sensible to attribute procrastination in our setting as due to the present-biased preference, as often presumed in previous studies (Akerlof, 1991; Ariely & Wertenbroch, 2002), but distinguishing between the causes of procrastination is beyond the scope of this paper. Our goal in this paper therefore is to provide empirical documentation of the effects of teacher intervention to mitigate the student action of putting off required assignments.

Ariely & Wertenbroch (2002) conducted controlled experiments investigating the effects of deadlines. They found in the context of executive education and part-time work that the average performance was significantly higher in the samples where deadlines were exogenously imposed than in the samples where participants self-selected binding deadlines. The design of incentives is one way by which course instructors attempt to elicit activities conducive to student learning. However, tasks (e.g. a practice of Chinese pronunciation) may be difficult to observe and it may be costly to write learning contracts. Some instructors may prefer a less intrusive form of intervention. A common alternative

is verbally prompting students during lectures to undertake certain actions. Although in-class prompting is a ubiquitous instruction strategy, we are not aware of studies that attempt to measure the extent to which teacher prompts mitigate procrastination; perhaps this dearth is due to the costliness in measuring the length of time spent on self-directed learning activities. As a result, while teachers may know from experience that prompting their students to "work" is effective, there remains uncertainty about the extent to which a prompt elicits intended behavior and about the factors that might affect the degree of inducement. Thus, our contribution is to offer a study that examines the impact of in-class prompting on the amount of self-directed learning activities.[2]

Our innovation is that we utilized computer-based records of out-of-class study activities to avoid the difficulty of observing the time spent on self-directed learning. We also focused on a setting where procrastination would hinder the development of a market-valued skill: the English language skill in Japan.[3] A Japanese education-material company has developed self-practice software for the English language, ALC Net Academy (ALCNA),[4] which stores on a server the beginning and end time of each drill, the time spent on each section of a drill in seconds, and the results of practice quizzes. Our study utilized the daily records of over 170 students in English-listening classes taught by an instructor at a Japanese university.[5] In addition to the data availability, the setting is attractive since, according to the conventional wisdom of foreign-language

[2] To be sure, the intervention focused on this study is not intended to "cure" the present bias in the preference of individuals but is intended to reduce the extent of postponing assignment completion.
[3] Matsushige (2004) presents evidence from a large-scale survey. At the anecdotal level, numerous Japanese companies use scores on a standardized test of English as the criteria for promotion and pay increases. The English-language education industry is ubiquitous in Japan and testifies to the importance of the language.
[4] ALCNA is commercial language practice software adopted in 166 universities and colleges across Japan (http://www.alc.co.jp/netacademy/list.html, accessed March 8, 2007).
[5] The data used in this study was obtained and used in accordance with the host university's privacy policy.

instructors, foreign language listening skill is like athletic skill: routine practice is important in skill development and retention.[6] Thus, unlike essay assignments where one may deliberately procrastinate to gain the level of motivation needed in writing (Ferrari, Johnson & McCown, 1995), procrastination in our setting can reasonably be thought of as being harmful to the rate of skill acquisition. Inevitably, the specificities in our setting limit the generality of our results. But the setting allows us to focus on the type of tasks that will adversely affect procrastinators in the labor market.

Another advantage of the setting is that it allows us to examine factors that may affect responsiveness to intervention. The response to prompts is voluntary and does not trigger a direct penalty in the event of non-compliance. One would thus be concerned that, for some students, prompts would be "like water rolling off a duck's back." A host of factors may affect responsiveness, including, but not limited to, intrinsic interest in the subject taught in class, teacher–student rapport, and student compliance. Our data is based on classes where the instructor, common to all classes, experimented with two different ways of prompting students: one group received a single prompt in the 13th week of semester while the other group was prompted in earlier weeks and was being reinforced by follow-up prompts in subsequent weeks. We took advantage of class variation in the intervention methods to see whether the timing and the reinforcement matter. Another interesting feature of the setting is that we were able to observe information regarding

[6] This assertion is based on our communications with various foreign language instructors. Short intensive exposures to foreign languages, such as a one-month language camp, are also thought to be effective in raising skill levels. However, the focus of our study is in a college setting where there are several constraints on resources available for language teaching. One might view that there are individual variations in the "optimal study pattern" or the study method that maximize the rate of skill acquisition. If, for example, the skill level of a student is far beyond what is presumed by the software developers, procrastination is unlikely to affect the degree of skill acquisition. The skill levels in English among the student sample are fairly homogeneous given the university admission process.

class preferences of students through a somewhat archaic system adopted in a compulsory course to allocate students to different classes. Based on the administrative record of preference groupings, we examined whether the responsiveness to prompts was influenced by the student preference for the class. In short, this paper presents a quasi-experiment aiming to identify the effects of verbal prompts and the factors affecting their effectiveness. Clearly, this study is not a laboratory experiment and it is not feasible to implement "clean" experimental designs, since the form of intervention is largely dictated by what the class instructor deems fair to the students (for instance, one possible experimental design, which was not implemented because the teacher did not think it was fair to treat students differently, would be to intervene in one class but not in another). Given the sparseness of the previous studies, however, we think this quasi-experiment is a reasonable starting point.

We present analyses from three approaches that complement one another. The first two approaches are intended to establish the short-term causal impacts of prompts on the amount of time spent on self-directed study. First, we utilized the timing of the intervention, which can be viewed as largely exogenous, and compared the study patterns on the week of the prompt with those from the prior week. The key assumption in this first-difference approach is that, for an identical set of individuals, had there been no intervention, the average study pattern over two weeks would have been the same. Second, we utilized the class variation in the timing of the intervention and implemented a difference-in-difference analysis. This approach assumes instead that the average change in the study patterns across two weeks would have been the same across two sets

of individuals. In our view, these assumptions are reasonably weak. Due to the short time frame, however, this approach is unable to pinpoint the semester-level reduction in the degree of procrastination attributable to the intervention. To provide an upper-bound estimate on the effects of prompts, rather than to offer additional evidence on causality, we compared the average degree of procrastination between classes with a single prompt and those with multiple prompts. We interpreted the difference, after controlling for observable differences, as an upper estimate of the effects of reinforced prompts on the degree of procrastination.

In summary, the evidence suggests that verbal interventions by instructors can be instrumental in reducing the degree of procrastination. The first-difference approach was used to analyze the sample with a single intervention. The average treatment effect was an extra 23.7 minutes per week of self-directed study. The magnitude is large since the average study time was 18.1 minutes in the prior week. One factor that influenced the magnitude was the timing. Our examination suggests that students do not respond as much when the assignment deadline is not close. Another factor is the sample attrition. There were systematic variations by the preference grouping in the rates of students not completing the class. To the extent that those who withdrew from the class were those who were not well matched to the class, the attrition process would make the sample conducive to finding the positive impacts of the prompts. After controlling for the attrition bias using the standard sample-selection model, there was a reduction in the magnitude of the point estimates. Along with the results that appear to suggest a stronger effect on the sample of students enrolled in the class as the first choice, we interpret the

results as indicating the effects of student preference on the responsiveness to prompting. Finally, the reinforcements helped reduce the degree of procrastination. We captured the degree of procrastination using the Herfindahl–Hirschman Index (HHI) of weekly study time. The average HHI was 0.09 less for the frequent-prompt classes than for the single-prompt classes. Roughly speaking, a reduction in HHI from the sample average of around 0.3 to 0.2 implies two additional weeks of attending the language lab under the assumption that the student worked the same hours per week during the week in which s/he has attended the lab.

The rest of this paper is organized as follows. Section 2 reviews the related studies in psychology and pedagogy. Section 3 describes the main features of the research setting and presents a preliminary examination. Section 4 discusses and presents analyses. Section 5 discusses the finding and concludes.

## 2. Related studies

While the causes of procrastination and its cures have been considered by psychologists and psychotherapists, the empirical investigation of interventions aimed at reducing procrastination is relatively recent, as emphasized in a review of early studies by Ferrari, Johnson & McCown (1995). It is worth noting that while the problem of self control has received much attention in the economics literature, as can be seen from the development of the Strotz–Laibson model of present-biased preference and the temptation-utility model by Gul & Pesendorfer (2001), there are various other explanations for academic

procrastination, including a lack of interest in achievement and preexisting psychological disorders (e.g. chronic depression). In the interests of developing a clinical therapy, previous researchers have investigated the effectiveness of various group- and individual-based therapies. Our paper is similar to such a line of inquiry to the extent that we are also interested in a remedy for procrastination. Our focus, however, is on a means for task inducements available to instructors in a lecture theatre, rather than on the means available to psychotherapists during therapy sessions for altering psychological conditions that result in procrastination, such as the unrealistic fear of undertaking certain tasks.

Another difference from the early studies in psychology is that our study measures intervention outcomes with objectively measured tasks rather than with questionnaires designed to gauge the extent to which a person is prone to procrastination. Our approach sidesteps the question of the validity inherent in self-reported measures as in a recent study in psychology by Ariely & Wertenbroch (2002).[7] Our main distinction from Ariely & Wertenbroch (2002) is that we focus on a less intrusive form of intervention, in line with the studies that examine the effects of posting signs to enhance office paper recycling (Austin, Hatfield, Grindle & Bailey, 1993), to reduce illegal parking (Cope & Allred, 1991), and to deter bathroom graffiti (Watson, 1996; Mueller, Moore, Doggett & Tingstrom, 2000). Unlike these poster studies, we examine a type of intervention in a face-to-face setting. There is extensive research about in-class feedback on learning

[7] Given the lack of psychometric measures in our study, we abstract away from the causes of procrastination. The presumption is that most students perceive the long-term benefits of developing English language skills as well as the benefits of routine practice so that the observed procrastination can be largely attributed to the problem of self control. Such presumption is implicit in Ariely & Wertenbroch (2002).

outcomes.[8] Our study is also concerned with teacher–student interactions, but due to the difficulty in identifying the impact of an intervention on cognitive outcomes (e.g. the understanding of grammar) our analysis is confined to a behavioral outcome that is thought to affect learning.

Our study however presents a novel approach to studying self-directed learning behavior, which has substantial practical interest but is typically costly to observe (Trautwein & Köller, 2003). A common measure of the amount of time spent on out-of-class study is the average weekly study hours reported in end-of-term surveys.[9] To overcome the clear concern about the accuracy of such measures, previous researchers have asked participants to record daily self-directed learning activities in logbooks (Schmitz & Skinner, 1993; De Jong, Westerhof & Creemers, 2000), but this method is costly to administer and is not readily available. In contrast, computer-generated records provide detailed and objective information on how students studied at home. Yet, computer-generated records have been under-utilized. A study on engineering education by Taraban, Anderson, Hayes & Sharma (2005) utilizes information including time-stamped key strokes and click trails. To our knowledge, however, this paper is the first attempt to utilize this source of information to examine procrastination.[10]

---

[8] See for instance a review of educational studies in Lyster & Mori (2006).

[9] In a study of 28 secondary schools in the Netherlands, De Jong, Westerhof & Creemers (2000) requested one student from each class to keep a logbook in which to enter each day how much time they spent on their homework for one month. On average, the estimated time spent on homework reported in the questionnaire was found to be about twice as long as the recorded time in the logbook. Still, under the assumption that the extent of over-reporting was constant across students, reported time may be sufficient to identify the effects of homework (Cooper, Robinson & Petall, 2006). However, De Jong et al. (2000) report that the accuracy rate ranged from 0.8 to 8.0, casting doubt over the reliability of the reported time.

[10] In the context of a computer-assisted language learning environment, Bland, Noblitt, Armington & Gay (1990) used the record of in-class activities generated by computer. Hall & Hamilton (2004) noted the potential of information and communication technology (ITC) in collecting detailed information that can be

Finally, our paper is related to the educational research on the pedagogical values of information and communication technologies, which cover a wide array of devices ranging from radio to the internet.[11] A study on college-level chemistry, for instance, examined the value of immediate feedback provided by online homework software (Cole & Todd, 2003).

## 3. The setting for quasi-experiments

### *3.1 Class description*

The setting for our study was second-year English language classes taught by an instructor-researcher at a selective university located in a local prefecture in Japan. The classes shared a common curriculum and aimed to develop English listening skills. The instruction period was 15 weeks with the exam period of up to three weeks. Attendance was taken and counted towards the final score. Other course details are in Appendix 1. Table 1 presents a summary of the classes. Classes A and B were offered concurrently during the first semester in year 1; class C was offered in the second semester in year 1; classes D and E were offered concurrently during the first semester in year 2. About a third of the class was female, reflecting the overall demographic composition of the university. Third-year and above students were also enrolled. The second semester was

---

used to assess the performance of workers and utilize patient data on surgery outcomes gathered through ITC to construct a measure of surgeon performance.

[11] For reviews and meta-analyses, see Salaberry (2001) and Hartley (2007).

longer in terms of the numbers of days from the beginning of the class to the deadline of the assignment because it spanned the New Year break.

[Table1]

*3.2 On-line homework*

All students were required to complete 60 listening drills of ALCNA by a deadline at the semester end.[12] Each drill consisted of five parts and took 7–10 minutes when properly attempted. The assignment counted towards 30 and 20 per cent of the course assessment for classes ABC and DE respectively. Since there was no restriction aside from the single deadline, students had flexibility over when to work. Their choice was constrained to the opening hours of the computer laboratory, which was the only location where students could access ALCNA on their own. The opening hours were 09:00–20:50 on business days.[13]

Student identity is a generic issue in administering homework.[14] There are several grounds to think that student identity is authentic in our setting. The version of ALCNA at the university can only be accessed in a computer lab and it would be cumbersome to

[12] One may be concerned that the fact that the educational institution does not mandate regular practice on the software argues that the value of such regular sessions, and hence the cost of delaying completion, is in question. ALCNA was, however, still in the early stages of adoption at the institution when this research was underway, so that there was uncertainty regarding the value of the software. Subsequent to our trials, the institution mandated the practice on ALCNA, though not on a regular basis, since the verification of the requirement is still administratively costly.

[13] The choice was binding according to the comments on the student evaluations of the class; students with a long commute time found difficulties in visiting the laboratory while students living nearby raised concerns about the inconvenience of not being able to access the laboratory during the weekend.

[14] The cause of identity mismatches need not be malicious. Taraban et al. (2005) noted a case where a student let another try out a new technology.

arrange another student to complete tasks. Further, those who are determined to cheat would have resorted to a less costly method, i.e. to click "next" without working on segments and declaring completion. Therefore, it is reasonable to rule out the possibility of identity mismatch contaminating the data. It is possible to engage in other activities such as net surfing while accessing ALCNA. There are, at most, weak perverse incentives given the low stakes involved: there are strong incentives to pass the course, given that it is compulsory, but the failure rate is low. Therefore, it is reasonable to take the self-directed study time recorded by ALCNA as a relevant outcome variable.

### *3.3 Course registration mechanism*

It is a graduation requirement of the university to pass two units of second-year English. Classes A–E are offered as an option to fulfill this requirement along with classes offered by other full-time and part-time faculty members. There is no fixed curriculum for the second-year English course so that the classes are heterogeneous in content, ranging from classes based on reading classical novels to those based on contemporary movies. Most students complete these units during their second year. Around 500 students are in each cohort of undergraduates and the school needs to allocate them to different language classes.

The registration mechanism works as follows. At the beginning of the year, syllabi for classes are publicized over the internet. By a deadline early in the first semester, students are required to cast two "ballots," each indicating a class they would like take. Students

would have been able to attend some of the introductory sessions of the classes by the closing date of the ballot. Their preference would be formed on the basis of the convenience of time slot, impression of the syllabus posted on-line, their experience of the first class, and rumor about instructors, among other reasons.

To reduce the uneven distribution of class preferences, each class is capped at 40 or 45 students. The rationing of classes is undertaken by full-time language instructors during a meeting held after the closing date in a fairly arbitrary manner: some instructors turn over the ballot and randomly select 40 students while others can only base selection on the limited information on ballot sheets, such as handwriting of name and student identification number, without any access to matching student transcript records during a short meeting time. Subsequent to the selection meeting, the students who are not assigned to their first preference are asked to take part in a mass matching section held in a large lecture theatre in the subsequent week, where students select from classes with openings. Those who do not take part in the balloting or the matching section are assigned by administrative staff to the remaining vacancies.

The registration is thus not based on a first-come-first-served basis, and there are reasonable grounds to expect that the groupings would largely reflect the degree of student preference for the class. The students are second year and above, all of them having completed the first-year English requirements; they would be reasonably well informed about instructor characteristics at the time of forming their preferences. The choice sets are reasonably large; there are more than 20 classes offered on different days

of the week. Even after time constraints limiting the set of available choices, there would still be much to choose from. Thus, an average student in the first-preference group would be better matched to the class than an average student in the second preference and above group (hereafter referred to as "second-choice group") and in the assigned group. Indeed, the percentage of students who register but withdraw from the course varies across preference group; for the overall sample the attrition rate is 7.9, 23.7, and 13.3 per cent for the first-choice, second-choice, and assigned group respectively (Table 1). Notice, however, the unbalanced nature of the sample; 72.4, 19.8, and 7.8 per cent of students are in the first-choice, second-choice, and assigned group.

*3.4 Intervention: Sample 1*

For classes A and B (hereafter referred to as "Sample 1"), an intervention involved a semi-personalized email in conjunction with in-class reinforcement on July 3 (Monday) in Week 13. The emails were sent to procrastinators in the early hours of July 3.[15] During the classes held later in the day, the instructor made an announcement about the email, urging students to do the work. As the classroom was installed with computer with internet access, many appeared to have checked their email on the spot. Our preliminary examination showed that the email recipients as well as non-recipients were affected similarly in terms of the average hours of extra study, suggesting that prompts affect

[15] Semi-personalized email messages were sent to students who had not shown sufficient progress on the assignment by the beginning of Week 13. The instructor sent two types of emails addressed individually, depending on the level of progress. Students who had not completed the first 10 introductory exercises received a warning email alerting them to the fact that they would lose a substantial portion of the class score by not completing the assignment. Students who had finished the introductory level but had not progressed at the advanced level received an encouragement to progress further. The warning email and encouragement email were sent to 24 and 14 students respectively.

those who are directly warned but also function as a credibility device signaling to all students that their records are being monitored. For this reason the analysis below does not make a distinction about the email receipt status.

Figure 1 shows the daily number of drills accessed by all students in Sample 1.[16] There is a clear pattern of procrastination; the total access exceeded 350 on August 1, the deadline date. The high number of accesses during the 16th week that began on July 24 was most likely due to the "deadline effect." There is a visible surge in the number of drills that corresponds to the timing of the prompt in Week 13.[17] Figure 2 compares the number of students who accessed the program on July 3 with the same day of the week in the prior week. Notice that much of the access occurred after class (A: 12:45–14:15, B: 14:30–16:00), but there were lab visits earlier in the day, suggesting that the email had an impact on the students.

[Figure 1]

[Figure 2]

[16] The computer laboratory was closed over the weekends and on holidays (April 28–30, May 3–7) so that there was no access then.

[17] While we think that there are few perverse incentives to fake working on the assignment in the current context, we have checked whether the surge in numbers is associated with an increased incidence of insubstantial completion, defined as drills declared complete in an unrealistically short length of time. While we have noted an unusual incidence, there was no statistically significant increase in the average number of insubstantial completion across Weeks 12 and 13. We have also checked the differences in the average accuracy rate of the ALCNA practice quiz across these weeks. If the increase in the study time was largely due to students accessing ALCNA but spending time surfing the web, for example, the lack of focus is likely to reduce performance on the quizzes. We found no significant difference in the accuracy rate across weeks. On the contrary, the accuracy rate was marginally higher.

The visible increase in Week 13 strongly suggests that the prompt had an impact on overall student behavior; but, before estimating the magnitude of the response, we note four cautions in interpretation. First, the intervention might have been timely; the prompt was conducted relatively late in the semester. Although the deadline effect would not explain the surge of access on July 3, it is probable that the response would have been weaker if the intervention had occurred earlier. Second, the prompt was accompanied by monitoring. An in-class prompt in a general setting (e.g. reading textbooks) is not based on objective records and instructors cannot monitor the actual response. Third, we do not observe the impact on students who withdrew from the class. Fourth, other types of self-directed study, which might have increased correspondingly or might have been crowded out by the computer-based study, are unobservable.

*3.5 Intervention: Sample 2*

In classes C, D, and E, in-class prompts were administered more frequently, and students were verbally urged to work on ALCNA.[18] Emails were not sent in these interventions. The fact that students were monitored was made explicit; the instructor read out from the summary information on student progress printed from the software the names of students who were lagging behind. We kept track of the timing of interventions in classes D and E (hereafter referred to as "Sample 2"), but not in class C. Appendix 2 documents the events in Sample 2.

[18] The effects of prompts may vary by the day of the week, but all five classes were administered on Monday so that the day-of-the-week effect, if any, was held constant across classes.

Figure 3 plots the group average of weekly study time for Sample 1 and Sample 2. To account for the longer course length of Sample 1 due to the differences in the timing of the final examination, the number of weeks from the final week is held constant for the purpose of comparison. Thus Week 8 in Sample 2, for example, corresponds to Week –7 on the x-axis in Figure 3. The vertical lines indicate the timing of interventions in Sample 2, corresponding to Weeks 8, 10, 12, and 13. Two lines are similar for the earlier periods; the average study times are less than 20 minutes per week up to Week –8. The lump in Week –13 is as a result of students being allowed to work on the software during the remaining class time after the pretest had been administered. In Weeks –7 and –5, in-class prompts were made in Sample 2 but not in Sample 1. The study time more than doubled by Week –5 from Week –8. The average time remained fairly steady for Sample 1 though there appeared to be a weak increasing trend. Week –4 corresponds to Week 13 for Sample 1. Notice that the procrastination is much more severe in Sample 1 with the average study time exceeding 100 minutes for the week before the deadline. The divergence between the average study lengths in the two samples corresponds to the timing of the in-class prompts conducted in Sample 2. Thus, a visible inspection suggests that the prompts altered behavior.

[Figure 3]

*3.6 Summary of study patterns and performance*

While our main focus is on the level of study time per week, we also considered the degree of procrastination at semester level. Once again, procrastination, and hence persistent practice, is an outcome of interest in the context of foreign language listening skills, since it is generally understood that routine practice is important in building and retaining skills. We measured the degree of procrastination with HHI — a measure of industry concentration — since it improves upon the common way to measure persistent practice with the frequency of working on homework that is elicited by asking respondents to choose from several discrete categories (e.g. "once a week") at the end of term surveys. An analogous measure in our setting would be the fraction of days in a week students accessed ALCNA. Such measure leaves unaccounted the variation in the duration of usage across days. Instead, our measure is based on the degree of concentration of weekly study hours in a given semester. For example, Student *i* who spent the same number of weeks as Student *j* would have higher HHI if the bulk of his/her work was concentrated in a certain week. One weakness with HHI is that the measure is independent of the direction of skewness. Thus, HHI misses some of the right-skewness in the distribution of individual study time over the semester. Nonetheless, visual inspection indicates that HHI would be a reasonable measure of procrastination in our case (see Appendix Figure 1 for examples of individual study patterns).[19]

Table 1 shows HHI and study times averaged over each class. The average HHI for the whole sample is 0.296; classes A, B, and C are above average and classes D and E are below average, meaning that the study patterns in the former group were more

[19] This formulation implicitly assumes a diminishing marginal effect of an additional hour on the consistency measure. This roughly means that a given increase in a study time would have a larger impact on HHI for a student with more concentrated study time allocation.

concentrated on average. The average time spent on ALCNA over all classes was 8.5 hours so that the latter group studied longer in each study period on average. Some qualifications are in order: during one class for D and E that had been cancelled, students were asked to work on ALCNA while the instructor was away; when the pretest was conducted in class E there was spare time and students were allowed to work on ALCNA during that class time, leading to the jump at Week –13 in Figure 2. These factors would inevitably increase the average self-directed study time for these classes. The longer course length for classes A, B, and C can mechanically lead to less concentration of study time. Thus it is important to consider the class differences in interpreting the result.

Rigorously estimating the causal impacts of prompts on test performance is beyond the scope of the current study design.[20] As a general reference, however, the average scores of the pretest and posttest are reported in Table 1. Comparable tests are administered in classes A, B, and D; classes C and E tried different test formats. The description of tests is in Appendix 1. Notice the higher average posttest performance in class D, which is consistent with the hypothesis that persistent self-directed study improves test performance. But of course, unobserved class differences, rather than the differences in average study pattern, might be the real cause behind the difference.

## 4. Methods and results

[20] We have estimated a value-added specification of an education production function that controls for a host of variables, including English-language practices not conducted on ALCNA. We find a strong correlation between test score and HHI for Sample 1. The estimated coefficient can be interpreted as a causal impact of the practice on the test score only under a set of restrictive assumptions.

*4.1 First-difference approach: Sample 1*

The ideal experiment to estimate the impact of the intervention is to compare the hours of ALCNA usage by the same set of individuals in two identical states except for the presence of the intervention. Here, we posit that on average, the amount of study time in Week 13 would have been the same as Week 12, had there been no intervention. Since the time period in comparison is within two weeks in the middle of a semester, it would seem reasonable to presume that they would have been similar on average. Indeed, the average amount of study in Weeks 11 and 12 is reasonably close. As the outcome variables, we considered the length of study and the indicator for any access to the software ("lab visit"). In other words, the study design is a one-sample before-and-after comparison (Meyer 1995). We also examined the preference groups separately. As individual specific characteristics are held constant across periods, the average treatment effects can be estimated by the differences in means across period for each group. If we are interested in the behavior of the group of students who remained in the class, a two-sample t-test suffices in testing the within-group differences in means across the period.

As noted, however, there was a disproportionately high attrition rate among students in the second-preference group. To the extent that those who did not complete the class were most likely those who were poor matches to the class, the attrition process would tend to make the sample conducive to finding the differential effects across the preference grouping. In other words, the attrition process would "dilute" the cross-group variation in the degree of matching that was exogenously generated by the course registration

mechanism, because remaining students were those who were likely to be well matched to the class.

We account for the possible selection bias in a regression framework (Heckman, 1979; Greene, 2000).

$$\Delta TIME_i = \beta_1 CHOICE\_1_i + \beta_2 CHOICE\_2_i + \beta_3 CHOICE\_3_i + \beta_\lambda \lambda_i + u_i$$

The dependent variable, measured in minutes, is the change in study time of Student *i* from Week 12 to Week 13.[21] *CHOICE*_J are dummies for Jth-preference group. The coefficients on group dummies capture the estimates of the average treatment effects for the group after accounting for attrition. $\lambda_i$ is estimated from the first stage regression and captures the influence of attrition. $u_i$ is heteroscedastic disturbance, intended to account for the possible systematic variation in standard errors across groups. The available information for those who withdrew is limited; on the basis of the model fits, the preferred specification of the first stage probit regression includes a dummy for students who are above second year as well as *CHOICE*_J. The exclusion of the senior student dummy from the second stage seems a sensible assumption given that the mean of the dependent variable is not significantly different between second-year students and senior students. The first stage explains a reasonable variation of the probability of remaining in the class.

[21] Given that the outcome variable is defined in terms of change, rather than in level as in the two-sample test, this framework is analogous to a one-sample test for the significance of the mean increase in outcome. Though the estimate of the average effects should be identical in both frameworks, the estimation of standard errors will be affected.

*4.2 Results from Sample 1*

Table 2 presents the results of two-sample t-tests. The sample average of an increase in study time from Week 12 to Week 13 was 23.7 minutes, which is significant at the 1 per cent level. The first-choice group had a mean increase of 24.4 minutes, which is significant at the 5 per cent level. The second-choice group had a higher mean of 31.1 minutes. It is marginally insignificant at the 10 per cent level with the p-value of 0.1004. None of the four participants in the third group used ALCNA in either week. The assigned students are those who did not register through the standard registration process at the university as required. The lack of response may be reflecting the lack of motivation, but given the small sample size it is difficult draw a strong conclusion. The fraction of students visiting the lab increased by a quarter, which is highly significant. The effects were apparently higher for the second-choice group, perhaps reflecting the direct impact of the email warning in inducing warned students to visit the lab. In sum, the result so far suggests that the prompts were generally effective, and it seems that there is almost no difference between the first-choice and second-choice group.

[Table 2]

Table 3 presents the Heckman model estimated with the maximum likelihood. OLS estimate is presented in Column 1 for comparison. The chi-square statistics for the significance of correlation across stages is 13.5 and is highly significant, indicating that

the attrition process would have been influential. The levels of coefficients on all of the three dummies are lower than those on OLS, being consistent with a hypothesis that, had those who withdrew from the class been forced to stay, the average treatment effects would have been lower. The estimated effect for the first group is 16.6 minutes, roughly a 45 per cent reduction from the estimate without the attrition control, but still significant at the 10 per cent level. The estimated effects for the other group are negative but not significant.

[Table 3]

The levels of point estimates are what we would expect to observe under the hypothesis that class preference affected responsiveness; the highest for the first-preference group, followed by the second-preference group, then the assigned group. To examine whether the differences are statistically distinguishable, we used Chi square tests to conduct pairwise comparisons of the coefficients. There is a significant difference between the first-preference group and the assigned group, but not for other comparison pairs. The p-value for the Chi square test of the equality of all three coefficients is 0.1013, marginally not rejecting the null at the 10 per cent level. In sum, the differences in the point estimates across groups are suggestive about the heterogeneous response to prompts but are generally not significant statistically. The unbalanced sample once again precludes us from making a stronger conclusion. Nonetheless, the statistical significance of the attrition bias suggests that part of the reason for finding the positive and significant impacts of prompts is because the remaining recipients were a better match to the class.

*4.3 Difference-in-difference analysis: Sample 2*

The above analysis assumes away any unobserved specificities with Week 13 that might have affected the study pattern. The assumption seems reasonable on the basis of the stable study pattern leading up to then. However, Week 13 roughly corresponded to one month before the deadline; the timing might have coincided with the threshold date on which students would have stopped procrastinating on their own. As a further examination we utilized Classes D and E (Sample 2), where prompts were implemented earlier in the semester. The setting allows for a difference-in-difference analysis utilizing Sample 1 as a control group for the prompts administered before Week 13 in Sample 2. In this approach we allowed for the possibility of unobserved time effects, thus weakening the assumption required in identification of the prompt effects on remaining students. We assumed instead that the unobserved time effects were constant across Sample 1 and Sample 2.[22] We define Week –8 as a pre-intervention period since it is a week before the prompt in Week –7. There was no class in Week –8. We take Week –5 as a post-

[22] There seem reasonable grounds to assume the similarity across Samples 1 and 2 in the average study pattern that would have prevailed without any intervention. The national universities in Japan generally take in homogeneous groups of students each year, and the entrance requirement for university has not changed across the sample years. There are six majors in this single-department school but majors are chosen after entering the school. The school takes in a small number of students from vocational schools under academically less-stringent criteria. Since some of the vocational school students do not receive English language training at their high schools, the English-language skills of these students are generally weaker than those of the students who enter through the ordinary route, but they consist of about seven per cent of the annual intake. Thus, the variation in English skill levels would be generally small within and across the cohort of students. Appendix Table 2 shows a comparison of the student characteristics between Sample 1 and Sample 2. The average pretest scores are nearly identical. The only statistically significant difference is the fractions of Finance majors and of the assigned students. To the extent that the composition of preference groups does not affect the individual attitude towards study patterns, our analysis based on the preference grouping sub-samples is robust to the differences in preference grouping composition. We have also checked the differences in the average weekly study time across samples until Week –8 but they are not statistically different, except for Week –13.

intervention period. The time span is four weeks to account for the repeated nature of the interventions. Week –4 is not used because of the intervention in Sample 1. As outcome variables, we consider the sample average of study time and the probability of lab visits.[23]

*4.4 Results from Sample 2*

Table 4 presents the difference-in-difference estimators in a regression framework. Columns 1–3 examine with OLS the length of study time on different subsamples. The study length in Week –7 was 4.2 minutes longer for the treatment group than for the control group but is not significantly different. The average treatment effects are estimated to be an increase of 21.7 minutes for students who remained in class, and are significant at the 5 per cent level. The results are similar for the subsample of the first-choice students. Again, the magnitude of the estimate is substantial, since the average study time for the treatment group in Week –8 was 24.4 minutes. However, for the second-choice students, the coefficient on the interaction term in the regression is large but insignificant (Column 3), a result that might be due to the small sample size. While this is suggestive of smaller impacts of prompts on second-choice students, once again, it is difficult to draw a strong conclusion. Notice that the study length of the treatment group in the subsample of second-choice students at the base period was 16 minutes longer than for the control group, and is significantly different at the 5 per cent level. For

[23] We have replicated the first-difference analysis as well as the difference-in-difference analysis for two prompts in Weeks –8 and –5 separately to investigating the short-term effects occurring a week after the intervention. The analyses showed that the immediate impacts of the prompts earlier in the semester were weak and not significant, suggesting that prompts administered well before the deadline did not affect the study behavior as much. The increase from Week –6 to Week –5 is marginally insignificant at the 10 per cent level, but the prolonged impacts of the earlier prompts seem to have affected the study time in Week –6, so we present analysis that can account for the prolonged impacts.

Sample 2, the average study time for the second-choice students in Week –8 was 24.8 minutes, compared to 17.8 minutes for the first-choice students. A possible reason may be the increased awareness of the monitoring for students in Sample 2.

Columns 4–6 examine whether the fraction of students making any use of the software changed as a result of the prompts. We show results from a probit regression. The dependent variable takes the value of one if the student used the software in a given week and zero otherwise. The marginal effects are reported, so the coefficients on the interaction term indicate the change in the fraction of students attending the lab that is attributable to the impacts of prompts. For the sample as a whole, there was a higher tendency for the treatment group to attend the lab; 14 percentage points more students attended the lab in the pre-intervention period. The intervention is associated with an increase of 17 percentage points, large effects since the fraction of students attending the lab was 0.425 in Week –8 for Sample2. However, the difference is not significant. The lab attendance increased by 54 percentage points for the second-choice students, which is statistically significant at the 5 per cent level. Thus, it appears that prompts are effective in inducing second-choice students to go to the lab even if the visit does not translate into a statistically significant increase in the average hours. Caution is required in interpreting the estimated magnitude however, since the small sample size can result in the large change in proportion.

*4.5 Between-class comparison*

A casual comparison of the mean of overall study patterns across classes indicated that the classes where students were exposed to frequent in-class prompts tended to work more persistently and for longer hours. To see whether the difference is explained away by the differences in student characteristics across classes, we present a regression analysis with a dummy on the classes with frequent prompts (C, D, and E) to capture the average difference from the classes with a single prompt (A and B). The scope of the analysis is in controlling for differences in student characteristics across classes. Thus, the average difference might still capture the differences in class-specific factors and our intention is not to make causal interpretation on the basis of this analysis. The coefficient estimates from the current analysis are viewed as upper bounds estimates on the impact of reinforced prompts under the assumption that the unobserved influences tend to increase the average study time.

Table 5 presents the upper-bound estimates of the effects of frequent prompts on the concentration and the total length of self-directed study time. Control variables are dummies for the registration outcomes, third year and above students, female, and five major discipline dummies. Also included is the pretest score standardized at class level.[24] Column 1 shows that on average the degree of procrastination was less for the frequent-prompt group by 0.09. Roughly speaking, a reduction in HHI from the sample average of around 0.3 to 0.2 implies two additional weeks of attending the language lab, under the

[24] The identical test was administered in classes A and B, and they are combined when standardizing test scores. To control for differences in the initial listening skills, it would be ideal to use the raw results from the same tests rather than the test score standardized at the class level. Since class D used the same test as classes A and B, we have tried specification with the raw test score as a regressor on the subsample of 105 students from classes A, B, and D. The point estimates are still significant but at a lower level of significance; the magnitude of the coefficient is almost identical for HHI and about 18 minutes lower for the study length.

assumption that the student worked the same hours per week during the week in which s/he attended the lab. Column 2 shows that on average the students in frequently prompted classes worked two hours longer — a relatively large number compared to the standard contact hours of 22.5 hours in a semester for the course. Students who were assigned to a class in the enrolment process worked a significantly shorter time and less persistently. There was no difference on average between those who were in the second-choice group and those of the first-choice group. Third year and above students worked two hours less on average. The standardized pretest score was not correlated to the study patterns.

## 5. Discussion

In summary, this paper examined the daily records of self-directed learning activities of Japanese college students enrolled in a compulsory English language course to see whether in-class prompting by an instructor mitigated the degree of procrastination on course assignments. We found patterns consistent with the hypothesis that in-class prompts mitigate procrastination, especially when prompts are reinforced. The impact of intervention, however, appears to be influenced by class preferences and the timing of prompts.

Regarding the implication of our analysis on education methods, the results show that, as an instruction tool, in-class prompting has advantages and disadvantages. In-class prompts have the potential to stimulate self-directed learning activities especially when

they are followed up, but given the lack of explicit incentives, the inducement effects are likely to vary across students. For this reason, it is likely that an appropriate designing of incentives (e.g. administering assessable tests on assigned reading) is a stronger inducement for self-directed learning activities. Nonetheless, if learning to control self-discipline problems is part of the learning process, setting explicit incentives would deprive students of such opportunities. Further, in-class prompts would be more cost effective in terms of instructor time. In our view, the optimal choice between different course designs would depend on the weight placed on measurable outcomes (i.e. test score) and non-measurable outcomes (e.g. self-discipline).

We do not claim that our analysis of the non-incentive based inducement, which is one variation of MIPs, is directly applicable to economic settings. There is much specificity with this study set in a classroom context. The current exercise does, however, highlight some of the relevant considerations in thinking about MIPs. As an example, in April 2008 New York City introduced a requirement for large restaurants to post calorie information on menus. Our study points to a concern that even if restaurant customers respond to this intervention, those whose behavior is affected may not be those who the policy makers are intending to influence. In addition, our study points to the possible importance for the authorities to follow up on the intervention.

Finally, our findings might be driven in part by at least three specificities within the setting. First is the method of prompting. Procrastinators were made aware that their study records were being monitored, and this is likely to have increased the strength of

the responses. The repeated prompts were shown to be effective in our case, but to the extent that there is diminishing marginal return to each prompt, the effects of prompts are likely to wear off, and may even become counter-productive, when reinforced in excess. Second, it is possible that the participant characteristics (e.g. students in our sample are selected into a competitive school in a local district of an East Asian nation) were conducive to intervention by teachers. Third, the effects may be instructor specific. Thus, the generalizability of the findings remains to be seen and further investigation in other settings, inside or outside classrooms, would be of interest.

**Reference:**


Akerlof, G.A. (1991). Procrastination and obedience. *American Economic Review* 81(2), 1–19.

Ariely, D. & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science* 13(3), 219–224.

Austin, J., Hatfield, D.B., Grindle, A.C. & Bailey, J.S. (1993). Increasing recycling in office environments: the effects of specific, informative cues. *Journal of Applied Behavior Analysis* 26(2), 247–253.

Bland, S.K., Noblitt, J.S., Armington, S. & Gay, G. (1990). The naive lexical hypothesis: Evidence from computer-assisted language learning. *Modern Language Journal* 74(4), 440–450.

Choi, J., Laibson, D., Madrian, B.C., & Metrick, A. (2003). Optimal defaults. *American Economic Review (Papers and Proceedings)* 93(2), 180–185.

Cole, R.S. & Todd, J.B. (2003). Effects of Web-based multimedia homework with immediate rich feedback on student learning in general chemistry. *Journal of Chemical Education* 80(11), 1338–1343.

Cooper, H., Robinson, J.C. & Petall, E.A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research* 76(1), 1–62.

Cope, J.G. & Allred, L.J. (1991). Community intervention to deter illegal parking in spaces reserved for the physically disabled. *Journal of Applied Behavior Analysis* 24(4), 687–693.

De Jong, R., Westerhof, K. & Creemers, B. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation* 6(2), 130–157.

Ferrari, J.R., Johnson, J.L. & McCown, W.G. (1995). *Procrastination and task avoidance: Theory, Research, and Treatment*. New York, Plenum Press.

Frederick, S., Loewenstein, G. & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature* 40(2), 351–401.

Greene, W.H. (2000). *Econometric Analysis*. Prentice Hall, NJ.

Gul, F. & Pesendorfer, W. (2001). Temptation and self control. *Econometrica* 69(6), 1403–1435.

Hall, B.L. & Hamilton, B.H. (2004). New information technology systems and a Bayesian hierarchical bivariate probit model for profiling surgeon quality at a large hospital. *Quarterly Review of Economics and Finance* 44(3), 410–429.

Hartley, J. (2007). Teaching, learning and new technology: a review for teachers. *British Journal of Educational Technology* 38(1), 42–62.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.

Laibson, D. (1997). Hyperbolic discounting and golden eggs. *Quarterly Journal of Economics*, 112(2), 443–477.

Lyster, R. & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition* 28, 269–300.

Matsushige, H. (2004). Eigoryoku to shoshin, shotoku: English bias ha sonzai shiteiruka (English Language Skill, Promotion and Income Level: Is there an English Bias?) In: H. Matsushige, Daigaku kyoiku koka no jishobunseki (Empirical examination of university education). (pp. 67–88). Tokyo: Nihon Hyoron Sha.

Meyer, B.D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics* 13(2), 151–161.

Mueller, M., Moore, J., Doggett, R. & Tingstrom, D. (2000). The effectiveness of contingency-specific and contingency-nonspecific prompts in controlling bathroom graffiti. *Journal of Applied Behavior Analysis* 33(1), 89–92.

O'Donoghue, T. & Rabin, M. (1999). Incentives for procrastinators. *Quarterly Journal of Economics* 114(3), 769–816.

_____ (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review (Papers and Proceedings)* 93(2), 186–191.

_____ (2005). Incentives and self control. unpublished manuscript.

_____ (2006). Optimal sin taxes. *Journal of Public Economics* 90(10–11), 1825–1849.

Salaberry, R.M. (2001). The use of technology for second language learning and teaching: A retrospective. *Modern Language Journal* 85(1), 39–56.

Schmitz, B. & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intraindividual, and time series analyses. *Journal of Personality and Social Psychology* 64(6), 1010–1028.

Strotz, R.H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23(3), 165–180.

Taraban, R., Anderson, E., Hayes, M. & Sharma, M. (2005). Developing on-line homework for introductory thermodynamics. *Journal of Engineering Education* 94(3), 339–342.

Thaler, R.H. & Sunstein, C.R. (2003). Libertarian paternalism. *American Economic Review (Papers and Proceedings)* 93(2), 175–179.

Trautwein, U. & Köller, O. (2003). The relationship between homework and achievement — Still much of a mystery. *Educational Psychology Review* 15(2), 115–145.

Watson, S.T. (1996). A prompt plus delayed contingency procedure for reducing bathroom graffiti. *Journal of Applied Behavior Analysis* 29(1), 121–124.

Figure 1: The total number of drills accessed over a semester: Sample 1

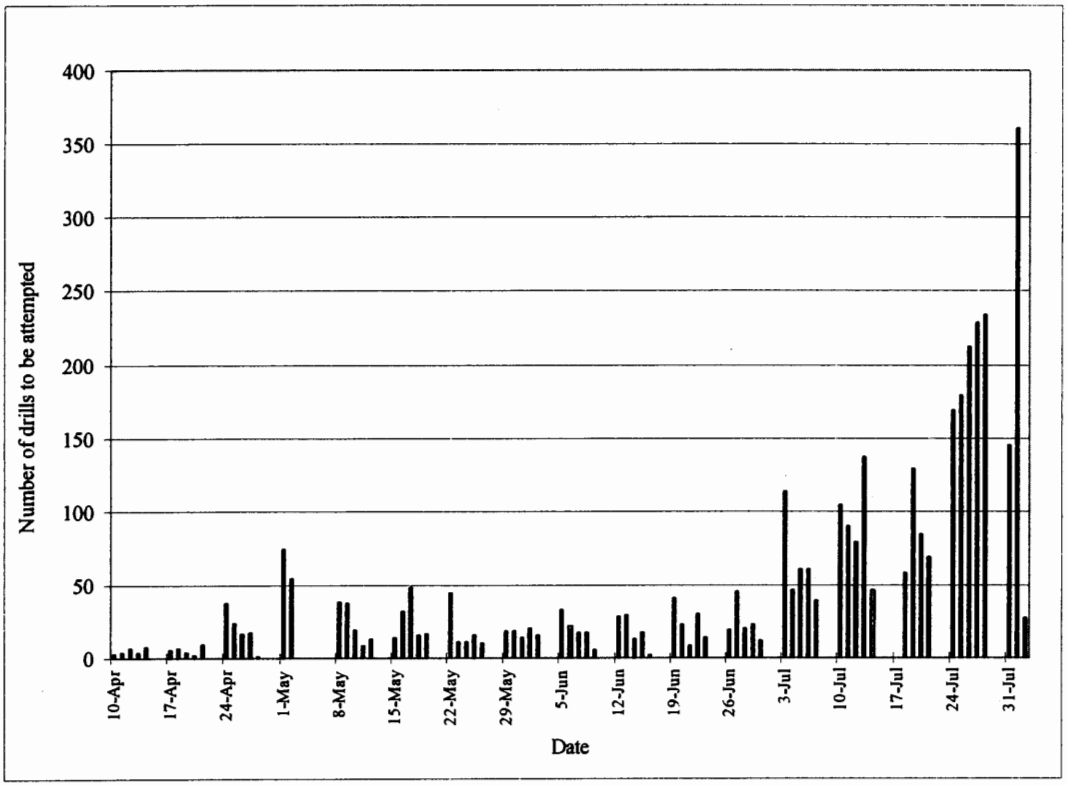


Figure 2: The number of students present at the lab: Weeks 12 and 13

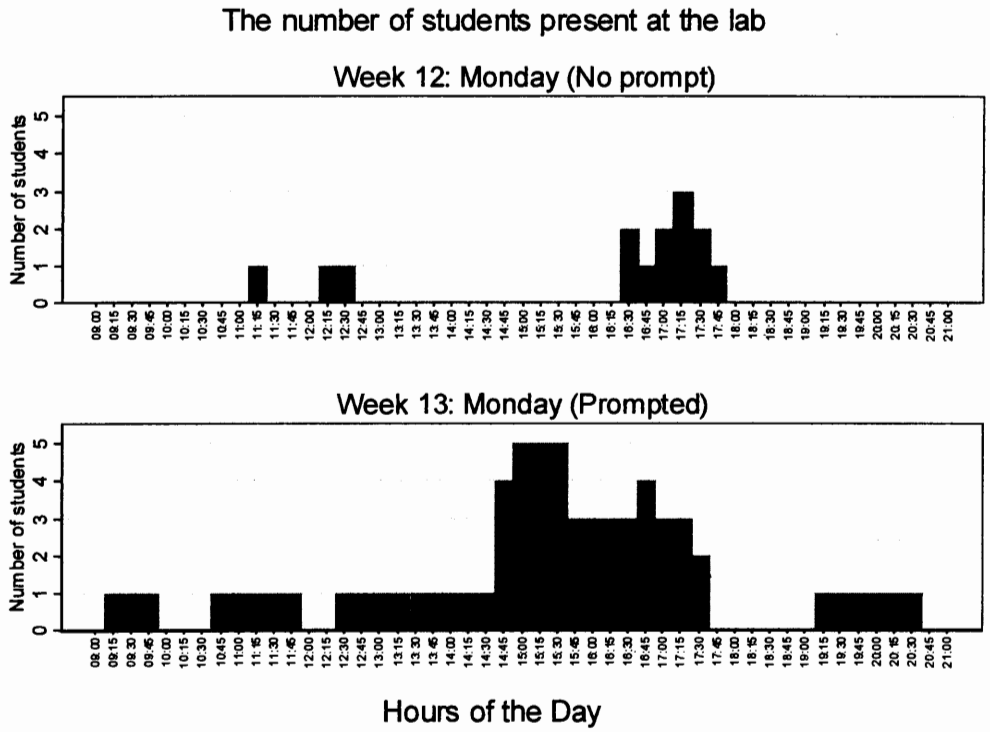
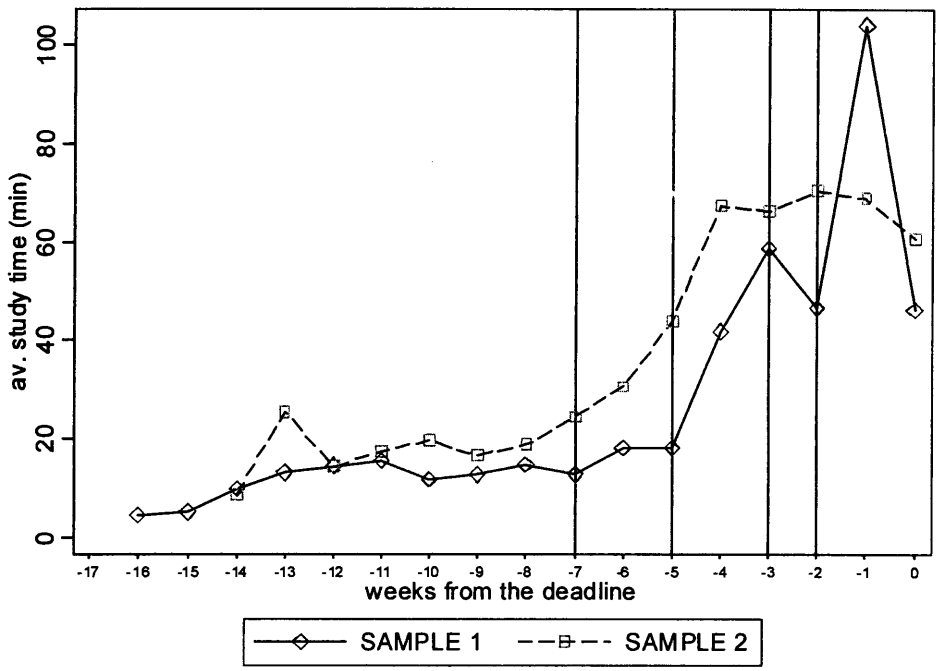


Figure 3: Comparison of average study time: Samples 1 and 2




Notes: The horizontal axis measures the number of weeks from the deadline week. There are 17 and 15 weeks in Sample 1 and Sample 2 respectively. The vertical lines indicate the timing of prompts administration in Sample 2.

Table 1: Summary statistics

| | Year 1 | | | Year 2 | | Combined |
|---|---|---|---|---|---|---|
| Class | A | B | C | D | E | A-E |
| Semester | First | First | Second | First | First | |
| Intervention type | Single | Single | Frequent | Frequent | Frequent | |
| Length [days] | 114 | 114 | 131 | 102 | 102 | |
| 3RD YR & ABOVE | 0.00 | 0.34 | 0.41 | 0.14 | 0.09 | 0.17 |
| FEMALE | 0.46 | 0.26 | 0.23 | 0.39 | 0.41 | 0.37 |
| STUDY TIME [hr.] | 8.0 | 6.4 | 9.7 | 9.2 | 8.9 | 8.5 |
| | (3.6) | (4.8) | (5.8) | (5.0) | (3.6) | (4.5) |
| HHI | 0.323 | 0.386 | 0.308 | 0.242 | 0.258 | 0.296 |
| | (0.205) | (0.272) | (0.247) | (0.148) | (0.201) | (0.213) |
| Exam format | Text | Text | ALP | Text | CASEC | |
| PRETEST | 50.1 | 45.5 | 177.6 | 48.3 | 539.6 | |
| | (10.5) | (13.0) | (18.6) | (8.9) | (75.1) | |
| POSTTEST | 63.9 | 58.3 | 198 | 66.1 | 566.7 | |
| | (11.0) | (12.1) | (26.0) | (9.4) | (73.5) | |
| N_remained | 39 | 29 | 22 | 39 | 41 | 170 |
| N_not completed | 2 | 9 | 3 | 5 | 3 | 22 |
| 1ST CHOICE (%) | 100 | 45 | 32 | 66 | 100 | 72 |
| [% not completed] | [4.9] | [11.8] | [25.0] | [6.9] | [6.8] | [7.9] |
| 2ND CHOICE (%) | 0 | 37 | 36 | 34 | 0 | 20 |
| [% not completed] | . | [37.5] | [11.1] | [20.0] | . | [23.7] |
| ASSIGNED (%) | 0 | 18 | 32 | 0 | 0 | 8 |
| [% not completed] | . | [28.6] | [0.0] | . | . | [13.3] |

Note: Standard deviation is in round bracket. Square bracket contains percentage of jth choice students withdrrawn from class.

Table 2: First-difference analysis: Sample 1

| | Average study time (minutes) | | | | Proportion of lab usage | | | |
|---|---|---|---|---|---|---|---|---|
| | All sample | First choice | Second choice | Assigned | All sample | First choice | Second choice | Assigned |
| Week 12 | 18.18 | 21.63 | 4.37 | 0 | 0.29 | 0.34 | 0.13 | 0 |
| Week 13 | 41.87 | 45.99 | 35.49 | 0 | 0.55 | 0.58 | 0.63 | 0 |
| Diff. | 23.69 | 24.36 | 31.11 | . | 0.26 | 0.25 | 0.50 | . |
| statistics | 2.70** | 2.37* | 1.76 | | 3.02** | 2.53** | 2.07* | |
| N | 65 | 54 | 8 | 4 | 65 | 54 | 8 | 4 |

Notes: ** and * indicate the significance at the 1 and 5 percent level with two-tailed tests respectively. t statistics are reported for the examination of study times and z statistics are reported for the examination of proportions.

Table 3: Sample-selection model

| MODEL | OLS | Heckman | |
|---|---|---|---|
| | | 2nd stage | 1st stage |
| | (1) | (2) | (3) |
| 1ST CHOICE | 24.36* | 16.58+ | 1.61** |
| | (10.10) | (9.51) | (0.31) |
| 2ND CHOICE | 31.11* | -5.45 | 0.74 |
| | (13.03) | (19.83) | (0.51) |
| ASSIGNED | 0.00 | -29.45 | 0.23 |
| | (0.00) | (18.53) | (0.44) |
| 3RD YR. & ABOVE | | | -0.91* |
| | | | (0.46) |
| LAMBDA | | 61.93 | |
| | | (16.93) | |
| Observations | 65 | 65 | 76 |
| -2LL | | 774.71 | |
| Wald | | 13.54** | |

Notes: The dependent variable is the change in hours of self-directed study time for (1) and (2), and the indicator for class completion for (3). Robust standard errors are in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%. 'Wald' refers to the Wald statistics for independence of the first and the second stage equations.

Table 4

Difference-in-difference analysis

| MODEL | OLS | | | PROBIT | | |
|---|---|---|---|---|---|---|
| DEPENDENT VAR. | STUDY TIME (min) | | | LAB USAGE = 1 | | |
| SAMPLE | ALL | 1st Choice | 2nd Choice | ALL | 1st Choice | 2nd Choice |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| TREATMENT | 4.17 | 1.12 | 16.00* | 0.14+ | 0.11 | 0.21 |
| | (5.28) | (6.21) | (7.85) | (0.08) | (0.09) | (0.23) |
| POST | 3.49 | 4.95 | -4.40 | 0.02 | 0.06 | -0.32 |
| | (6.18) | (7.42) | (6.15) | (0.09) | (0.10) | (0.27) |
| TREAT X POST | 21.65* | 19.54+ | 33.26 | 0.17 | 0.11 | 0.54* |
| | (9.40) | (10.56) | (21.99) | (0.12) | (0.13) | (0.24) |
| Constant | 14.68** | 16.68** | 8.77+ | | | |
| | (4.16) | (5.02) | (4.39) | | | |
| Observations | 290 | 242 | 40 | 290 | 242 | 40 |
| R-squared | 0.07 | 0.05 | 0.14 | 0.05 | 0.03 | 0.21 |

Notes: Coefficients on the pobit model are marginal effects. Robust standard errors in parentheses. Adjusted R-squared for OLS; Pseudo R-squared for probit. + significant at 10%; * significant at 5%; ** significant at 1%

Table 5

Between-class analysis

| | HHI | STUDY TIME (hours) |
|---|---|---|
| | (1) | (2) |
| FREQ. PROMPTS = 1 | -0.09* | 1.98** |
| | (0.03) | (0.66) |
| SECOND CHOICE = 1 | -0.03 | -0.01 |
| | (0.04) | (1.22) |
| BEING ASSIGNED = 1 | 0.28** | -2.15+ |
| | (0.09) | (1.28) |
| OLDER STUDENTS = 1 | -0.02 | -1.86+ |
| | (0.05) | (0.97) |
| FEMALE = 1 | -0.14** | 1.13 |
| | (0.03) | (0.69) |
| PRETEST SCORE | 0.00 | 0.10 |
| | (0.01) | (0.34) |
| Constant | 0.37** | 7.48** |
| | (0.04) | (0.85) |
| Observations | 168 | 168 |
| R-squared | 0.27 | 0.14 |

Notes: Robust standard errors are in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%. 'Wald' refers to the Wald statistics for independence of the first and the second stage equations. 5 major discipline dummies are included in the regressions but supressed for the presentation purposes.

Appendix 1: Additional details on the setting

*In-class activities*

The curriculum incorporated the materials from a standard qualification on English skill (TOEIC) extensively, in order to give a concrete study motive (i.e. a qualification relevant for the job hunt). The classes were administered in a Computer-Assisted Language Learning (CALL) environment. Typical class activities included: (1) a 10-minute quiz based on practice materials for the listening section of TOEIC; (2) Dictation practice of the recording materials from the quiz. Each student worked on a computer to fill in the blanks in a conversation/speech transcript provided in the form of a Word document; (3) Pronunciation practice where the whole class follows the instructor; (4) Conversation pair work where a pair uses conversation from the quiz materials as a model on which to base original conversation. The recording of pair work was provided to the instructor in some cases. Occasionally, "shadowing," which refers to a practice of mimicking speech while listening to speech, was practiced in class.

*Parts of the drills in ALCNA*

(1) First listening of short material such as a conversation with a colleague in an office setting, Bill Clinton's presidential campaign speech, and the production process of pasta.
(2) Three multiple-choice quizzes to check the comprehension of materials.
(3) A self-directed study time of written transcripts assisted by a dictionary function.
(4) Second listening of the material with a user-adjusted speed.
(5) Revision of material and self-declaration of completion of drills.

*Measurement of learning outcomes*

Three types of tests are used to measure the learning outcomes.

1 In Classes A, B, and D, the achievement gains are measured with commercially available practice tests for the listening section of the Test of English for International Communication (TOEIC). All sections are on listening. Changes in test scores would reflect the change in listening skills over the semester, but they also capture other changes, such as the differences in test difficulty, the improvement in test-taking skills, and the differences in physical and mental condition.
2 Class C used the Assessment of Communicative English Placement Test. This test includes reading and grammar sections in addition to a listening section. Further the test is not TOEIC based.
3 Class E used the Computerized Assessment System for English Communication (CASEC). Two sections are on listening but the other is not. CASEC is a test designed to obtain accurate measurements of language skills.

For all classes the posttest is part of the assessment while the pretest is not.

Appendix Table 1: Class schedule for Classes D & E

| Week | Main activities | In-class prompts regarding ALCNA |
|---|---|---|
| 1 | First class (Introduction of the course) | Required students to complete the diagnostic tests, an introductory part of ALCNA, by Week 2. |
| 2 | Administration of pretest; Class E students were allowed to work on ALCNA after completing the exam (CASEC). | Made announcement regarding general class progress, noting that almost all students had completed the diagnostic tests assigned in the previous week. |
| 3 | Class cancelled due to a national holiday | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | Class cancelled without notice | |
| 8 | | Required students to keep working on ALCNA. |
| 9 | | |
| 10 | | Required students to work on ALCNA; announced the class cancellation for Week 11; urged students to work on ALCNA during the class hours in Week 11. |
| 11 | Class cancelled with a week's notice | |
| 12 | | Pointed out insufficient progress by reading out the names of lagging students. |
| 13 | | Instructor urged students to work on ALCNA; reminder of the deadline. |
| 14 | | |
| 15 | Deadline on Friday | |

Notes: The information is based on the first author's communication with the class instructor. Classes met once a week for 90 minutes on Monday during the semester.

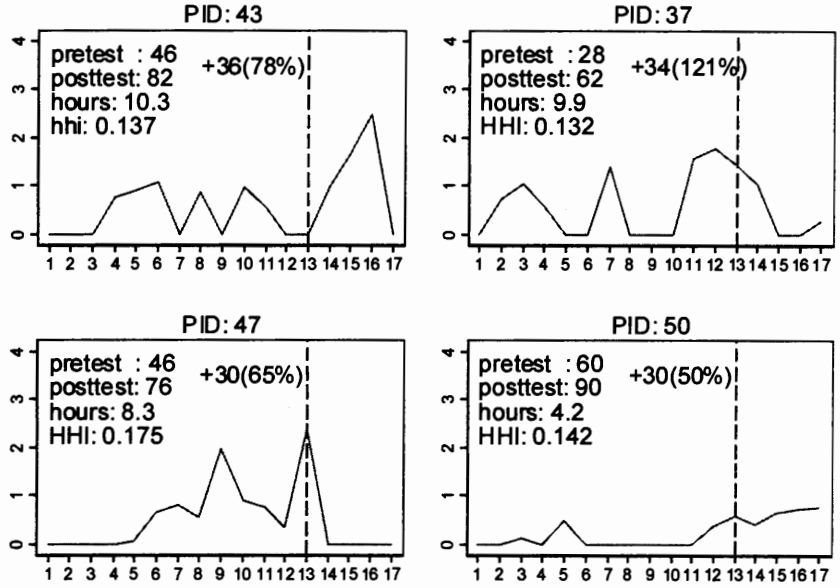
Appendix Table 2
Comparison of Sample 1 & 2

| Sample | Sample 1 | Sample 2 | Significance of difference |
|---|---|---|---|
| 3 RD YR. & ABOVE | 0.165 | 0.114 | |
| FEMALE | 0.367 | 0.398 | |
| PRETEST | 48.2 | 48.3 | |
| 1ST CHOICE | 0.734 | 0.830 | |
| 2ND CHOICE | 0.177 | 0.171 | |
| ASSIGNED | 0.089 | 0.000 | ** |
| *Majors* | | | |
| ECONOMICS | 0.241 | 0.250 | |
| FINANCE | 0.063 | 0.148 | + |
| MANAGEMENT | 0.215 | 0.227 | |
| ACCOUNTING | 0.165 | 0.114 | |
| INFORMATION | 0.139 | 0.091 | |
| SOCIAL SYSTEM | 0.177 | 0.171 | |

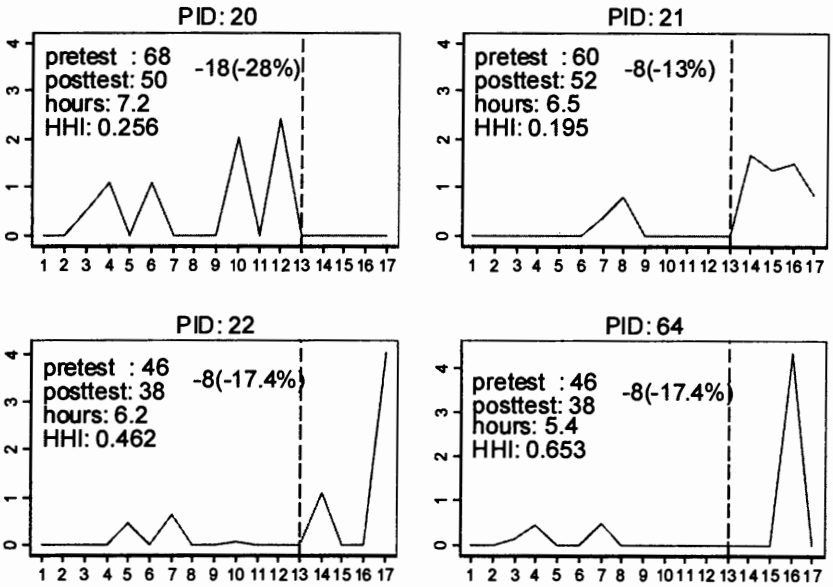
Note: + and ** indicates the statistical significance at the 10 % and 1 % level respectively. PRETEST score for Sample 2 is based on Class D only.

Appendix Figure 1: Example of weekly study time of individuals

Panel A: Highest gainers




Panel B: Lowest gainers




Notes: The vertical axis is in hours. The horizontal axis is in weeks. Pretest and posttest refers to the scores in the test conducted before and after the semester. 'hours' is the total hours spent on on-line software, and 'consistency' is HHI of weekly self-directed study time over the semester. PID is an identification code assigned for the purpose of this study.