

Cold calling and web postings: Do they improve students' preparation and learning in statistics?

Dan Levy¹

Abstract: Getting students to prepare well for class is a common challenge faced by instructors all over the world. This study investigates the effects that two frequently used techniques to increase student preparation -- web postings and cold calling -- have on student outcomes. The study is based on two experiments and a qualitative study conducted in a statistics course that Masters in Public Policy (MPP) students take in their second semester at the Harvard Kennedy School. When used together, web postings and cold calling seem to increase the amount of time that students devote to reading before class by about an hour. This effect is both statistically and practically significant. However this increase in pre-class reading did not translate into increased learning (measured by average test scores on the midterm exam). Neither of the two techniques seems to be better than the other one at increasing reading time, test scores, and other student outcomes.

Keywords: statistics education research; ICT; class preparation; textbook reading.

Introduction

Getting undergraduate and graduate students to read for class is often difficult (i.e. McKeachie & Svinicki, 2010). Students have significant demands on their time, including other schoolwork, extracurricular activities, and socializing. Examining undergraduate and graduate-level psychology courses, Burchfield and Sappington (2000) found that student completion of assigned readings plummeted from approximately 80 percent in 1981 to 20 percent in 1997. As an additional hurdle, statistics readings are often found by students to be dry and hard to follow (i.e., Cobb, 1987). In an examination of introductory statistics textbooks for higher education, Harwell et al. (1996) found that only one of the six texts was recommended for use by a majority of students.

Even if students do the readings, there is no guarantee that the experience will make a significant impact on students' performance in the course. Creating usable knowledge from reading involves a set of complex skills, and these skills are rarely taught to students in an explicit fashion (Nilson, 2010). Research has also documented the difficulty students have in transferring knowledge learned in a particular context to new problems and settings (Bransford et al., 2000). This difficulty can be particularly acute with statistics textbooks, as these resources often present material in ways that are quite different from the style and substance of the rest of the course.

There are several ways that instructors can encourage students to increase the quantity and quality of their reading. Two promising techniques, both commonly used at the graduate level, are web postings and cold calling. This paper reports on a study conducted to find out how

¹ Harvard University, dan_levy@harvard.edu

effective these methods were in improving student outcomes in a Masters' in Public Policy (MPP) statistics course taught at the Harvard Kennedy School (HKS) at Harvard University.

The rest of the paper is organized as follows: Section 2 provides a description of the two techniques and explores the mechanisms through which these techniques could potentially impact student outcomes; Section 3 describes the study design; Section 4 presents the results; and Section 5 concludes.

Background

Description of Web Postings and Cold Calling

This section provides a brief description of the two teaching techniques under study-- web postings and cold calling. Although each technique can be used in a variety of ways, particularly in terms of the types and number of questions employed, this background section will focus on a robust format of each technique.

Web postings use modern technology -- particularly the course management platforms common in most institutions of higher education -- to facilitate a long-standing practice of having students answer questions about the assigned reading in advance of coming to class (Chance et al., 2007). Peer instruction (PI), a popular pedagogical method in higher education, utilizes a well-developed and representative model of web postings (Crouch & Mazur, 2001). In PI, web posts are oriented towards higher-order cognition, asking students to carry out tasks including application, analysis, and evaluation. The assignments also ask students what they have found most confusing in the readings. The students post their answers through the course site in advance of class, giving the instructor the opportunity to read the responses before class. This information can then be incorporated into the planning and execution of the class session. The students' posts, and their related participation in class, are graded to incentivize compliance.

Cold calling, although most well known for its use in law and business school classrooms, is also common in many other higher education settings. Cold calling is the process of asking a question and calling, without advanced notice, on a student who has not volunteered to answer that question. Often, instructors devoted to cold calling will begin a class session by volunteering a student to summarize the essential aspects of one or more of the readings (Garvin, 2003). Cold calls are then often used significantly throughout the remainder of the class session, employing questions that range from those that elicit factual details to those requiring complex thinking skills. Often, an instructor will stay with a single student for a long string of questions, and in the most extreme form, will do so until the student comes to a question for which they have no satisfactory answer.

Theory of Change: Potential for Impacting Student Learning

This sub-section describes the mechanisms through which these techniques could possibly affect student learning. Below is a simplified logic model encompassing the theories of change for both pedagogical techniques (Figure 1), while the Appendix has a more detailed logic model for each technique.

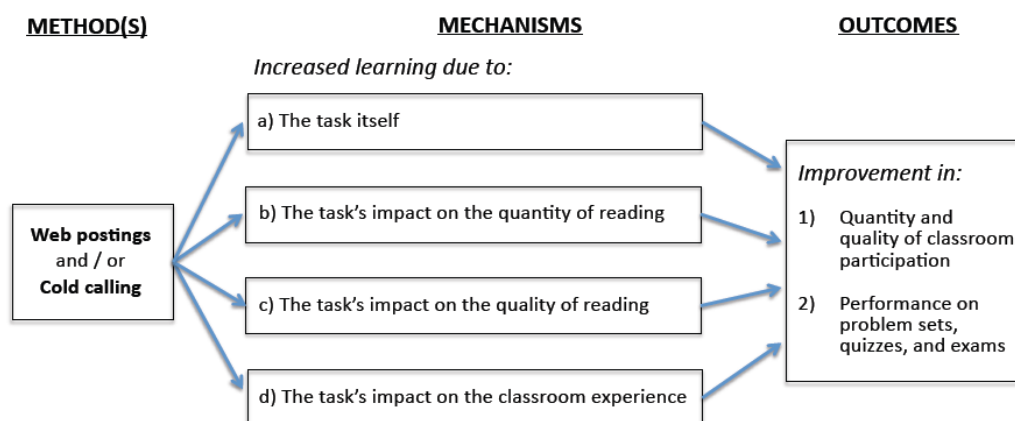


Figure 1. Logic model

The remainder of this section expands on each of the four mechanisms described in Figure 1, and also explores a few mechanisms through which the two techniques could lead to a *decrease* in student learning.

Increased learning due to the task itself. Web postings are one version of what the learning sciences refer to as “prompts”. There is a rich research literature on how a wide variety of prompts increase student learning (e.g., Chi et al., 1994; Linn & Eylon, 2006; White & Frederiksen, 1998; McKeachie et al., 1990). Prompts are a type of scaffold: a pedagogical tool designed to assist learners in accomplishing complex tasks more effectively (Wood et al., 1976). Reiser (2004) distinguishes two main ways in which scaffolding supports learning: “structuring the task” and “problematizing the subject matter.”

“Problematizing the subject matter” happens when a scaffold makes some aspect of the student’s work more problematic, thereby requiring the student to spend more time and mental energy addressing this important idea or process. These “desirable difficulties” have been shown to improve student’s conceptual knowledge (Linn, 2006). Web posting questions serve this role by requiring students to grapple much more intensely with important concepts in the readings. The answers to these questions also make visible, to both the student and the instructor, the student’s current ways of thinking about the material. Research has shown that engaging student preconceptions in this way is critically important to the development of lasting knowledge and skills (Bransford et al., 2000).

The web post question that asks about remaining areas of confusion also provides valuable information to both the student and instructor. Answering the question helps the student pay more attention to the process of learning – what is going well, what is still problematic, and what the student might do to go about addressing these problems. This “thinking about thinking,” or metacognition, helps students become better self-directed learners and problem solvers (Bransford et al., 2000). The instructor, in turn, can use this information to help figure out how best to spend class time.

Cold calling facilitates students’ abilities both to process information quickly and to communicate a response effectively. There are several ways students benefit from this process. First, when people are forced to verbalize their understanding and make connections between ideas, the knowledge involved is strengthened (Linn, 2006). Second, as with web posts, cold calling helps make thinking visible. In particular, due to the involuntary nature of cold calling,

students' misconceptions and preconceptions are more likely to surface than if instructors solicited volunteers. Although this can be uncomfortable, unearthing these misunderstandings is an important step in helping students to learn concepts effectively (Bransford et al., 2000). Last, research has shown that cold calling can lead to greater participation of minority groups; for example, US college classes featuring only voluntary participation tend to be dominated by white males (Nilson, 2010).

Increased learning due to the task's impact on the amount and timing of reading. Extensive research has shown that holding students accountable for assigned reading in some concrete way dramatically increases the average amount students read (Nilson, 2010). Mazur (2000) has found web postings to be an effective way to get students to complete the readings, and Carney et al. (2008) found similar results from using a paper-based version of the method. Although we were unable to uncover any studies of how cold calling impacted student reading completion, it is sensible to expect that students will read more given that they may be asked to answer questions in front of their classmates and that their answers are formally evaluated. In turn, increased reading could likely have a positive impact on student learning, both because reading is generally an effective way to learn material (McKeachie & Svinicki, 2010) and because research shows that a student needs to encounter information multiple times in multiple ways to effectively incorporate it into their existing knowledge (Nuthall, 1999).

Increased learning due to the task's impact on the quality of reading. Web posts are an effective method for helping students to "structure the task" of reading a textbook by drawing students' attention to important aspects of the text and by encouraging them to process the concepts in a deeper way. Even though these questions are not intended to cover every major concept of a reading, they give students a clearer idea of the types of information they should focus on acquiring. It is likely that this improved understanding of where to focus their time and attention while reading will increase their conceptual understanding. Under related conditions, Reder and Anderson (1982) found that students who read textbook summaries actually acquired more knowledge than those who read the entire text. The questions also incentivize students to digest and process the text more deeply, which research has shown leads to improved knowledge development (Ferguson-Hessler & de Jong, 1990; McKeachie & Svinicki, 2010). Students also tend to retain information more clearly and permanently when they space out their readings rather than cram them before an exam or paper (Reder & Anderson, 1982).

The influence of cold calling on the quality of reading is more indirect, as it is not tailored to any particular assignment. The types of questions that are asked in the process of cold calling will serve as a powerful signal to students as to what is valued from the reading (Doyle, 1983; McKeachie & Svinicki, 2010). If the instructor commonly cold calls on students to give a summary of one or more readings, people are more likely to read with this end goal in mind. Similarly, if the bulk of the cold calls focus on higher-order cognition, students should tend focus more on concepts than factual details. Also, as with web postings, students should benefit from cold calling if it leads them to do their readings regularly rather than all at once before an exam.

Increased learning due to the task's impact on the classroom experience. In having some positive impact on the quantity and quality of reading completed by students in advance of class, both web posts and cold calling enhance students' ability to learn from the class session. Students are better able to comprehend what is happening in class and are more aware of the knowledge gaps they need to remedy. Both of these reasons should increase the likelihood that students are willing and able to participate in class, and this active processing of material promotes learning (Bransford et al., 2000).

Another way that web posts improve the classroom experience is in helping the instructor make better use of class time. By reading the posts, the instructor has a fuller understanding of where students are at in terms of conceptual understanding of the material. The instructor can weave specific posts or themes from the posts in to the session and can modify what content is covered and the ways in which it is presented (Mazur, 2000). A particular advantage of cold calling, assuming it is done in a sustained fashion, is that it creates a concrete incentive for students to stay focused throughout an entire class session. Increasing student attention is important, particularly given the findings that students have a difficult time focusing on a lecture for more than ten to twenty minutes (Gardiner, 1994).

Decreased learning concerns. There are two main reasons why either of these methods might decrease student learning. The first involves the potential these methods have to negatively impact students' intrinsic motivation for the course. Hobson (2002) found that intrinsic motivation was significantly influenced both by the instructor's attitudes and behaviors and by the relevance of the course content. There is a danger that web posts or cold calling, particularly if mainly instituted to increase student reading, can be seen as paternalistic. Some students also find cold calling off-putting due to its adversarial nature of forcing participation (Williams, 1992). Also, motivation could be further diminished if students find the reading to be dry and difficult to understand.

The second reason why learning may decrease is that the increased time spent on reading and/or web posting may lead to a decrease in time spent on other, possibly more valuable, aspects of the course. If students are strongly incentivized to spend more time on reading and related activities such as web posting, that time has to come from somewhere. If at least some of that time is taken from other course activities – such as thoroughly understanding a problem set or studying lecture notes – any gains from the reading and related activities could be diminished or completely erased. In a detailed study of pedagogical innovation in engineering classrooms at MIT and the University of Cambridge, researchers found this ripple effect in time allocation, as a change in one aspect of the course led students to change the time they devoted to many other aspects of the course (Seering & Britter, 2007).

Study Design

Two experiments were conducted to assess the effectiveness of cold calling and web posting in increasing students' preparation for class and learning. Each experiment took place in a different section of "Quantitative Analysis and Empirical Methods II" (coded API-202), a required course taken by MPP students in their second semester. It is the second course in the statistics sequence and focuses on applied statistical analysis. Specifically, the course introduces students to the use of multiple regression analysis and program evaluation for analyzing data in the social sciences, with an emphasis on empirical applications. Both sections, which covered the same material at the same pace but met at different times for class during the week, were taught by the same instructor (one of the authors) in the spring of 2009. The readings in the course came from "Introduction to Econometrics" by Stock and Watson, a standard statistics textbook used across many university campuses.

The two experiments followed the same basic structure (Figure 2). The experiments were announced in the second class session of the semester, after enrollment settled in and students could no longer switch to other sections of the course. In the third class, students in each section were randomly assigned into two different groups. The students were then exposed to their

respective treatment conditions in classes #4 - #11, and the midterm exam occurred in class #12. Although the experiment concluded at that point, the assignment to the two groups was reversed so that all students were treated equally in the course (i.e. they were exposed to the same interventions but at different points in the semester).

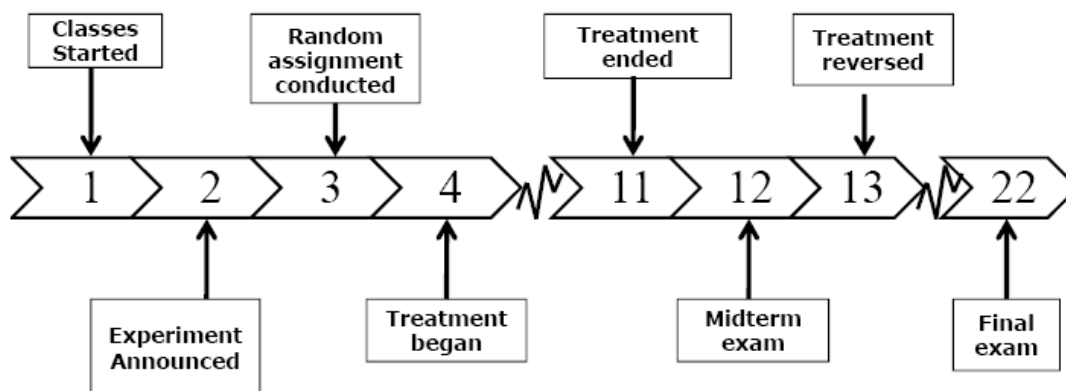


Figure 2. Structure of experiments by class session

While the course consisted of material on both regression analysis and program evaluation, this study was conducted in first half of the semester, during which time the focus was on regression analysis. The topics covered during the five weeks of the study included bivariate and multiple linear regressions (ordinary least squares), omitted variable bias, joint hypothesis testing, and non-linear regressions (e.g. interactive dummy variables, logarithms, and quadratics). This material was consistent across the two experiments.

The two experiments differed in terms of the treatment conditions students were assigned to (Table 1). In experiment #1, students in the first group were assigned to the “Web Posting AND Cold Calling” treatment, which meant that they would be required to do web postings before each class and they were put on a list from which they could be cold called. Students in the second group were instead assigned to a control condition in which the instructor simply encouraged them to read before class, as he had done in many of his past courses (with some exhortations about the importance of reading before class for their learning). In experiment #2, students in the first group were required to do web posting for each class, whereas students in the second group were subject to cold calling.

Table 1

Description of the experiments

	Group #1	Group #2
Experiment #1	Web Postings AND Cold Calling	Control
Experiment #2	Web Postings	Cold Calling

In addition to the experiments, a qualitative study was conducted aimed at understanding students’ views on how they were experiencing the web postings and cold calling. Throughout the semester, the instructor met with small groups of students to ask about their perceptions, and at the end of the course students were asked to fill in a brief anonymous survey in which they

indicated their predictions as to which treatment would work and why. The qualitative survey was instrumental in understanding the results of the experiments and in helping the instructor draw lessons for his pedagogy.

Implementation of the Treatments

This sub-section describes in some detail how the web postings and cold calling were implemented. We think this specificity will help readers interpret the results and assess the extent to which the findings are generalizable to contexts that are relevant for them.

Web Postings. Students were required to post answers to three questions on the course website by 4:00 am of the day of class. The three questions were based on the reading assigned for that day's class. Students could see each other's postings, but the questions were crafted such that it would be very evident if one student simply copied from another. For example, in the first question used in class #6 (see Box 1 below), students were asked to come up with a bivariate relationship in which they were interested and then use that bivariate relationship to answer some sub-questions. Since there are countless of bivariate relationships they could come up with, it would be very unlikely that two students would come up with the same answers unless they had seen each other's work.

The third question was always the same: "Please tell us what you found difficult or confusing in this reading assignment." This question, recommended by Harvard Physics Professor Eric Mazur, was meant to facilitate metacognitive thinking from students and to give the instructor a sense of common student difficulties. The instructor used this information to adjust the length of class time spent on each topic, and at the start of each class, he shared with students the themes that emerged from the posts.

Compliance with web postings was fairly high. On average, students submitted 7.07 postings out of the eight that were assigned. Approximately 80% of students posted seven or eight times. Some postings clearly demonstrated careful reading and reflection, whereas others seemed much less thoughtful. An example of the latter, in response to question #3, was: "Homoskedasticity and heteroskedasticity are interesting, but it would be great to spend some more time talking about both concepts in class."

Cold Calling. As described in Section 2, implementation of cold calling can vary tremendously from class to class. In the case of this study, for each class session, the instructor randomly chose one student from the cold call list of approximately 40 students. He typically asked that student two to three related questions. The questions, based on the assigned reading, tended to be factual/technical in nature. For example, in one class he asked "What are the two conditions for Omitted Variable Bias?" A student who had done the assigned reading should be able to answer this question, whereas a student who did not do the reading would have difficulty saying anything meaningful (unless they were familiar with the material from prior coursework, which was rare). When a student did not know the answer the instructor would alternate between silence and probing for about two to three minutes, and then would move on. Based on qualitative feedback, some students felt stressful about being cold called whereas others felt that the probability of being cold called was fairly low (since only one out of the approximately 40 students on the cold call list was cold-called per class). We would rate this implementation of cold calling as "low to moderate" intensity, in contrast to the higher intensity typical of the Harvard Business and Law Schools, where instructors often do numerous cold calls per class and can probe with the struggling student for five to 10 minutes.

Box 1 - Example of Web posting questions (Class #6)

- (1) Think about a bivariate relationship you are interested in.
 - a. Write the Population Regression Function
 - b. Identify a third variable that, if omitted from the regression, would produce omitted variable bias (i.e. look for a variable for which the two conditions of omitted variable would be satisfied). Explain why each of the two conditions holds.
 - c. Identify a third variable that, if omitted from the regression, would NOT produce omitted variable bias (i.e. look for a variable for which one of the two conditions of omitted variable would NOT be satisfied). Explain why each of the two conditions holds or fails to hold.
- (2) Suppose you were to run the regression you specified in your answer to 1.a. Assess what would be the sign of the omitted variable bias that would result from failing to include in your regression the variable you indicated in 1.b. Explain why.
- (3) Please tell us what you found difficult or confusing in this reading assignment. If you did not find anything difficult or confusing, tell us what you found most interesting. [Suggested length: one paragraph]

Implementation of Random Assignment

Random assignment was conducted separately in the two sections of the course. There were 80 students in experiment #1 and 93 students in experiment #2. The instructor and grade obtained in the first statistics course of the sequence were used as strata for random assignment. As it turns out, the grade in the previous statistics course was highly predictive of performance in the course in which the experiments were conducted. (The R^2 of a regression of midterm exam score in the course studied here on the strata dummy variables was 0.55 for experiment #1 and 0.57 for experiment #2.) This stratification increased the statistical power of the study design, which was especially important given the relatively small sample sizes.

The goal of random assignment was to ensure that the two groups in each experiment were identical at the outset of the study, so that any subsequent difference in outcomes could be attributed solely to the difference in treatments received. Tables 2a and 2b compare the two groups for each of the experiments. As can be seen, the groups are indeed comparable to each other in key baseline characteristics, including gender, classroom seat location, previous statistics course (API-201) grade, the percentage taking the advanced section of the previous statistics course (API-201A), and the grade in problem set #1 (which was due before treatment began). The differences between the groups tend to be small and not statistically significant. In experiment #2, one difference (for whether the student was a first year MPP only, as opposed to a student from another school or doing a joint degree with the Kennedy School) was statistically significant at the 10% significance level. All in all, the results of Tables 2a and 2b are consistent with random assignment being well implemented and generating two comparable groups.

Table 2a

Checking random assignment in experiment #1

	Group 1	Group 2	Difference	p-value [^]
Female (proportion)	0.44	0.46	-0.02	0.84
MPP First Year Only	0.80	0.77	0.04	0.70
Sat in Front	0.34	0.39	-0.04	0.69
Sat in Middle	0.41	0.41	0.00	0.97
Sat in Back	0.24	0.21	0.03	0.68
API-201 "GPA"	8.35	8.27	0.08	0.78
API-201 Z Score	-0.07	-0.08	0.01	0.96
Took API-201A	0.12	0.08	0.05	0.51
Problem Set 1 Grade	2.39	2.54	-0.15	0.25

[^] p-value reported to assess whether the difference is statistically significant

Table 2b

Checking random assignment in experiment #2

	Group 1	Group 2	Difference	p-value
Female (proportion)	0.59	0.57	0.02	0.82
MPP First Year Only	0.69	0.84	-0.15	0.09*
Sat in Left L140	0.27	0.36	-0.10	0.31
Sat in Middle L140	0.41	0.32	0.09	0.37
Sat in Right L140	0.33	0.32	0.01	0.93
API-201 "GPA"	8.32	8.39	-0.07	0.76
API-201 Z Score	0.02	-0.01	0.04	0.85
Took API-201A	0.22	0.23	-0.01	0.98
Problem Set 1 Grade	2.43	2.23	0.20	0.13

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Impact Estimation

Impacts were estimated separately for each of the experiments using the specification below:

$$Y = \beta_0 + \beta_1 TREAT + \beta_2 STRAT_1 + \dots + \beta_{20} STRAT_{19} + u$$

where Y is the outcome of interest, TREAT is the binary indicator for whether the student was assigned to group #1, and STRAT is a collection of binary indicators representing the different strata that were used for random assignment. The coefficient on TREAT represents the impact estimate of the group #1 treatment relative to the group #2 treatment. Robust standard errors were calculated.

The key outcomes examined were the following:

- (1) Time reading the textbook (minutes per week): Refers to the number of minutes students spent reading the textbook in week #4 (classes #7 and #8) of the semester. Students were asked to report this measure on an assignment they turned in that week. While these self-reported measures are subject to measurement error, we tried to minimize the degree to which this error might be systematic by including a note on the assignment indicating “Your answers to the questions in this sheet will help us understand better how you allocate the time you spend on this course. Please answer the questions candidly. Your answers will not affect your grade on this problem set or on any other part of the course. As long as you submit your answers, you will get full credit on this part of the assignment.”
- (2) Time spent on the course (minutes per week): Refers to the number of minutes students spent on course-related activities on week #4 of the semester. It was taken from the same source as the previous measure.
- (3) Number of times participating in class: A course assistant reported the number of times students participated in each of the classes during the treatment period.
- (4) Problem set grades: Average grade of the three problem sets that were due during the treatment period (on a scale of 0 to 3).
- (5) Standardized midterm score: Midterm exam score was standardized by subtracting the mean and dividing over the standard deviation of midterm score in each of the two samples.

Results

Overall, the key results of the two experiments were: (1) Web Postings and Cold Calling had a positive effect on the amount of time students read before class but not on academic performance (as measured by the midterm exam score); (2) When tested against each other, neither of the two methods (Web Posting and Cold Calling) came out on top in terms of improving any of the key outcomes. This section presents results from the two experiments in more detail.

Experiment #1: Web Postings and Cold Calling vs. Control Group

Web Postings and Cold Calling (WP&CC) together increased the time students spent reading the textbook by approximately one hour per week (Table 3a). This effect is statistically significant, and given that the control group only read about 35 minutes per week, represents quite a large change. The effect on reading time translated into an increase in the overall time that students spent on the course outside the classroom as a result of being assigned to the WP&CC group. The magnitude of the effect on total time spent on the course is about 47 minutes, which though not statistically different from the effect on reading time, hints at the possibility that reading time may be crowding out other course-related activities.

Table 3a

Results of experiment #1: Effects of Web posting and cold calling vs. control

	Group 1 (Web Postings and Cold Calling)	Group 2 (Control)	Difference	Standard Error	Stat Signif
Time reading textbook (minutes)	93.42	34.72	58.70	(11.18)	***
Time spent in API-201 (minutes)	262.46	215.54	46.92	(20.70)	**
Number of times participating in class	2.15	1.49	0.66	(0.477)	
Problem set grades	2.50	2.68	-0.18	(0.105)	*
Midterm grade (standardized)	0.05	0.02	0.03	(0.170)	

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Note: Group #2 means are regression adjusted.

Since both time spent on reading and time spent on the course were self-reported, one could suspect that the results presented above may overstate the true effect of the WP&CC treatment since students who were assigned to this treatment may be particularly prone to overstate the amount of time they devoted to the course for social desirability bias and similar reasons. While this possibility cannot be discarded, we do not think it is driving the results observed for three reasons. First, as reported in the previous section, the students were explicitly told that their answers about time use would not affect their grades, and were encouraged to answer candidly. Second, the qualitative study repeatedly suggests that students indeed felt incentivized (or, as many said, “forced”) to read when they were assigned to the web postings and cold calling group (see Box 2 for some representative quotes.) Third, students protested several times in the semester about having to read and about the amount of reading. This had never happened in the four previous iterations of the instructor teaching the course.

Given that students assigned to the WP&CC group read more before class, one may expect that they both would participate more in class and do better in the problem sets and exam. This is not confirmed by the data. Although they seem to participate more often in class (0.66 times more), the difference is not statistically significant. Their problem set grades actually decreased by 0.18 points (about 0.43 standard deviations of the control group mean), and while this effect is only significant at the 10% level, it is suggestive that perhaps reading more led students to devote less effort to their problem sets. The qualitative study provides some support for this “crowding out” effect. (See Box #3 for some representative quotes.)

Finally, the treatment also did not translate into an increase in learning, at least as measured by the midterm exam score. The effect on test score was small (0.03 standard deviations) and statistically insignificant at conventional levels. There are several possible substantive explanations for this result. The first, perhaps naturally favored by some economists, is that students are skilled in optimizing their use of time. As such, the WP&CC treatment may

Box #2 - Representative quotes suggesting students read more

- “I only read when I was in the “web posting + warm calling group – sorry!”
- “I would not read if I was not forced to through web postings”
- “I would not have done readings except for warm calling and postings
- “Warm calling and web posting forced me to read”
- “Given time constraints, I would probably not have read anything without the warm call motivation. Unfortunately this is part of life @ the Kennedy School”
- “Because the web posting was mandatory, they forced me to read.”
- “I did not do the reading from the textbook when it was not required”
- “If no web postings, sometimes you intend to do the readings and end up not doing them”
- “Without required posting, I probably would not have read”
- “I was “gently” forced to read the material”
- “Frankly my reading seriously dropped off in the second half of the semester when I wasn’t required to post on the web. Incentives, incentives, incentives”
- “The threat of being called in class motivated me to read”

Source: Qualitative survey done at the end of the semester

Box #3 - Representative quotes suggesting reading crowded out other course activities

- “Postings and reading did not enhance the in-class learning, rather they took time away from problem sets”
- “If you do the readings hastily (because there is so much to do for this course), it does not make much of a difference”
- “While the warm calling did nudge me to be more motivated to do the readings, the intense workload of the course and mandatory biweekly postings completely burnt me out and crushed my motivation to read by the end of the course”
- “I did not like the web postings because they distracted from focusing on studying the material. “

Source: Qualitative survey done at the end of the semester

have led students to do more of something (in this case reading) that was less beneficial per time spent than some other things that got crowded out (in this case, studying or working on problem sets). Thus, an increase in reading, and even in time spent on the course overall, may not translate to better academic performance. A second explanation is that reading did not add significant value to student understanding, perhaps because the reading was not complementary

enough to the class or because the reading was not appropriate for the type of students in the course. Several students voiced this general opinion captured by a representative quote from one of the students: “Readings did not significantly enhance learning beyond lectures.” A third possible explanation is that test scores are hard to affect in general, a finding supported by the research literature (i.e. Prince 2006), and that these interventions were simply not strong enough (perhaps because of their short duration or because of the way they were implemented) to generate a significant impact. A fourth explanation is that since not all questions in the midterm exam drew directly from the materials in the readings, the midterm exam score was perhaps not measuring the learning that students were supposed to attain from the readings. But we conducted some analyses where the dependent variable was the aggregate score on the questions that were unequivocally related to the readings in the textbook (as determined by Suzanne Cooper, a veteran instructor of the course not involved in the study), and again found no effect of the treatments on this aggregate score.

There are also some methodological alternative explanations to the lack of effect on midterm test scores. The first one is that, given the small sample sizes, the study did not have enough statistical power to detect reasonably sized effects. While we cannot discard this explanation, we do not think it is very compelling. For one thing, the statistical power of the design was greatly improved by stratifying the random assignment using the grade on the first course of the sequence. Additionally, the point estimate was 0.03 standard deviations, and even with our relatively large standard error (0.17), we should have been able to detect sizable effects in the order of a third of a standard deviation or more. The second methodological explanation is that the treatments may have had spillovers. To the extent that cold calling and web posting led to half of the students being better prepared for class, perhaps the discussion in the classroom was enriched in a way that benefitted everyone in the course (including the half of the students who were in the control group). If this were indeed the case, our estimates here would understate the effect of WP&CC on learning. We cannot discard this explanation but would argue that the spillovers, if they existed, could also have affected negatively the control groups. For example, if the discussion occurred at a higher level (or faster pace), the control group students perhaps could have been more likely to be lost or confused, and if this were the case our estimates here would be overstating the effects of WP&CC on test scores. In sum, while we cannot discard these two methodological issues as partly explaining our results, we think it is unlikely that they are fully responsible for them.

Experiment #2 – Web Postings vs. Cold Calling

When tested against each other, neither of the two methods (Web Posting and Cold Calling) was found to be more effective in terms of improving any of the key outcomes (Table 3b). The differences in average outcomes between the web posting and cold calling groups was not statistically significant for time spent reading, time spent in the course, class participation, problem set grades and midterm grade. The difference in the midterm grade was 0.24 standard deviations in favor of the web posting group, which is not trivial, but it was not statistically significant. Furthermore, given no statistically significant effects on the other four outcomes, as well as the lack of impacts on midterm score in Experiment #1, it seems hard to argue that the web postings have an effect over cold calling with regards to midterm scores.

Table 3b

Results of experiment #2: Effects of Web posting vs. cold calling

	Group 1 (Web Postings)	Group 2 (Cold Calling)	Difference	Standard Error	Stat Signif
Time reading textbook (minutes)	76.88	82.02	-5.14	11.20	
Time spent in API-201 (minutes)	248.63	238.63	10.00	19.80	
Number of times participating in class	1.88	1.96	-0.08	0.53	
Problem set grades	2.22	2.32	-0.10	0.10	
Midterm grade (standardized)	0.11	-0.13	0.24	0.16	

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Note: Group #2 means are regression adjusted.

Student and HKS Faculty Perceptions of Web Posting and Cold Calling Effectiveness

The reader could argue that perhaps the results reported above were not all that surprising. We argue here that they were. First, our priors were that the combination of web posting and cold calling in experiment #1 would have led to higher scores on the midterm exam. After all, he had been using these techniques for two years because he thought they led to increased learning. Second, as described below, samples of faculty and students by and large did not predict accurately the observed pattern of results.

We conducted two surveys to examine how predictions of the effectiveness of the treatments related to the study's actual findings. The first survey group was composed of all the students taking part of the study, while the second included approximately 40 Kennedy School Faculty members attending a research seminar in which this study was presented. In the case of the faculty members, they were asked to make the predictions after they had heard the description of the interventions but before they had heard the results.

In terms of experiment #1, the vast majority of both groups correctly predicted the positive effects on reading time (92% of students and 74% of faculty). But only a minority of each group (26% in both) correctly predicted the lack of effects on learning (Table 4a).

Table 4a

Predictions of results from experiment #1: Views from faculty and students

Which technique was most effective at increasing...	...Reading Time?		...Learning?	
	Faculty's predictions	Students' predictions	Faculty's predictions	Students' predictions
Web Postings and Cold Calling	88%	92%	74%	57%
Encouraged Reading (Control)	0%	2%	0%	16%
Same	12%	6%	26%	26%

Students and faculty were not much better predicting the results of experiment #2. The majority of them thought that web posting would increase reading time relative to cold calling (Table 4b), which was not supported by the evidence. Furthermore, only 15% of faculty members and 18% of students predicted that web posting and cold calling would be equivalent in terms of their effect on learning.

Table 4b

Predictions of results from experiment #2: Views from faculty and students

Which technique was most effective at increasing...	...Reading Time?		...Learning?	
	Faculty's predictions	Students' predictions	Faculty's predictions	Students' predictions
Web Postings	78%	69%	74%	66%
Cold Calling	22%	20%	11%	16%
Same	0%	11%	15%	18%

Conclusions

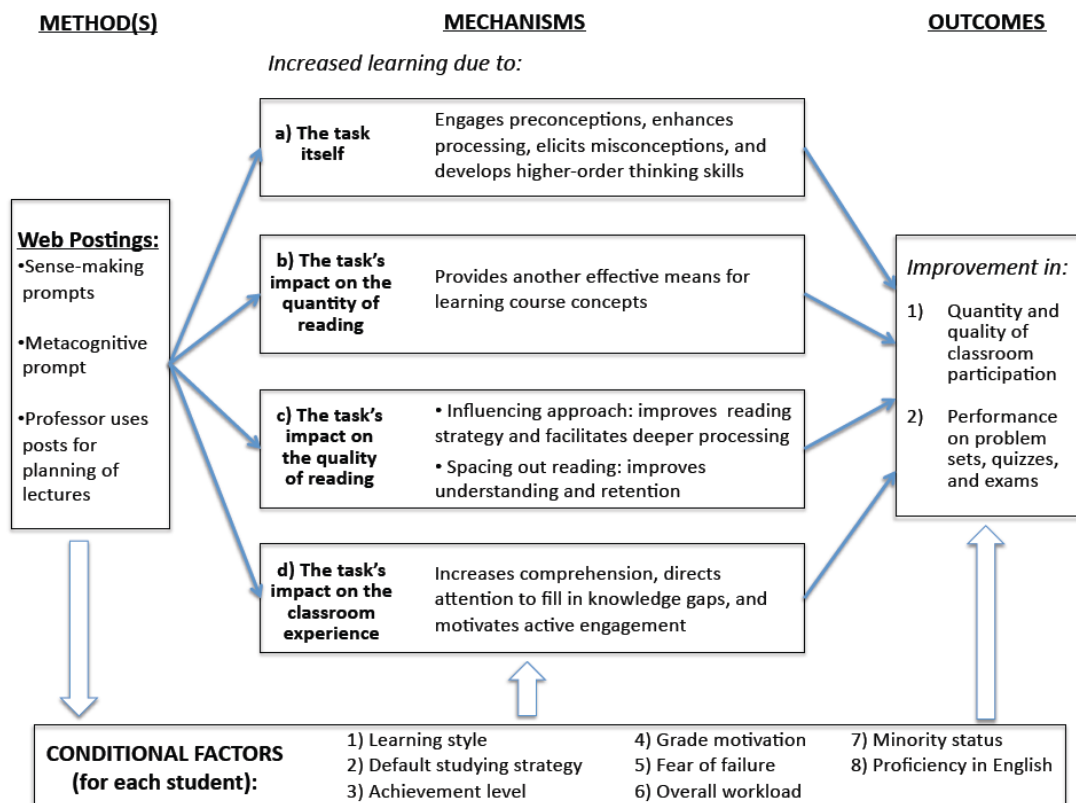
Many instructors struggle with how to get their students to prepare better for class. This study reports on the effects of two frequently used techniques (web postings and cold calling) on student outcomes in a statistics course that MPP students take in their second semester at the Harvard Kennedy School. When used together these two techniques seem to increase the amount of time that students devote to reading before class by about an hour. This effect is both statistically and practically significant. However this increase in test scores did not translate into increased learning (measured by average test scores on the midterm exam). Neither of the two techniques seems to be better than the other one at increasing reading time, test scores and other student outcomes.

Possible substantive explanations for why web postings and cold calling affect reading time but not test scores include the reallocation of time induced by the interventions was not optimal for increasing learning, the readings themselves did not add much value to other activities in the course, and the treatments were not intense enough to generate impacts on a variable as hard to affect as test scores.

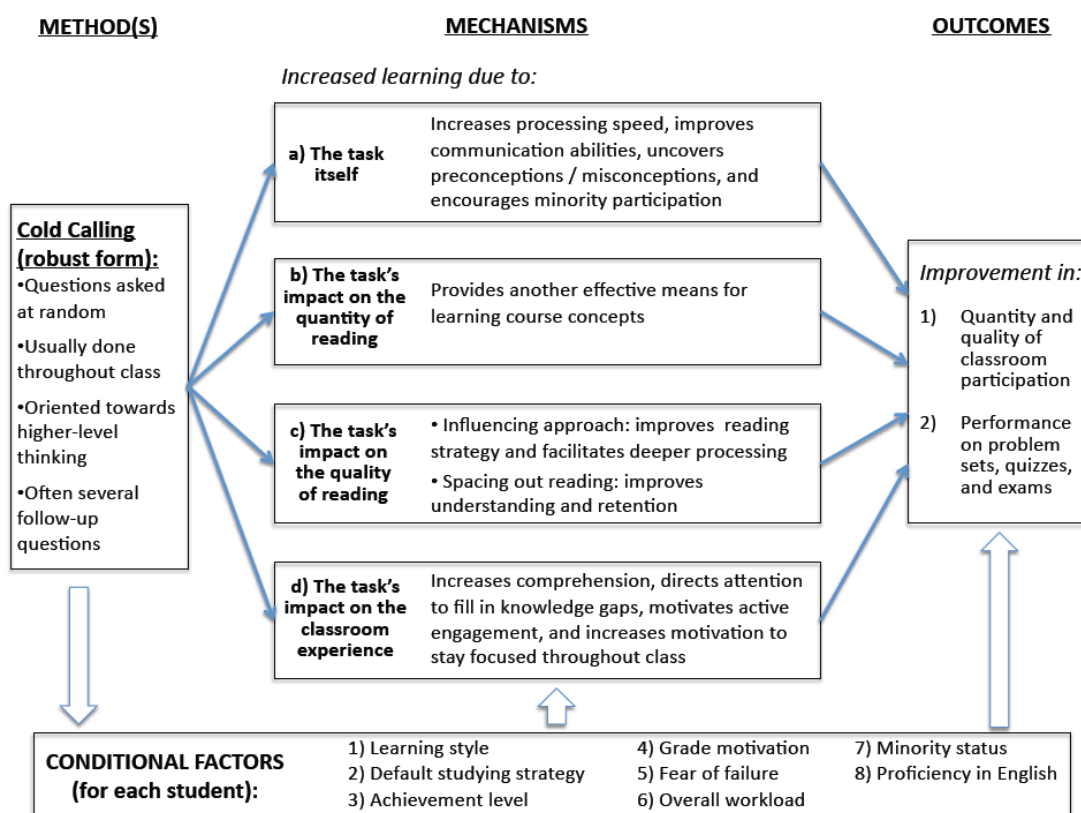
We do not conclude that web postings and cold calling cannot be helpful in increasing learning in general. Instead, we conclude that they were not helpful in this context as implemented here. Perhaps more importantly, we conclude that instructors using techniques they think are leading to increased learning would perhaps do well to assess whether this is actually true. Both authors, as well as the vast majority of students and other faculty members surveyed, had intuitions that ran counter to the main results of the study, namely that these interventions would have some positive impact on learning. If anything, we hope this study contributes to the notion that gathering evidence about our teaching has the potential of helping us become better in our teaching.

Appendices

Appendix 1. Detailed Logic Model for Web Postings



Appendix 2. Detailed Logic Model for Cold Calling



References

- Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C: National Academy Press.
- Burchfield, C.M. & Sappington, J. (2000). Compliance with required reading assignments. *Teaching of Psychology*, 27 (1), 58-60.
- Chi, M.T.H., De Leew, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.
- Carney, A. G., Fry, S. W., Gabriele, R. V., & Ballard, M. (2008). Reeling in the big fish: Changing pedagogy to encourage the completion of reading assignments. *College Teaching*, 56 (4), 195-200. doi: 10.3200/CTCH.56.4.195-200
- Cobb, G. W. (1987). Introductory textbooks: A framework for evaluation. *Journal of the American Statistical Association*, 82, 321-339.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning in statistics. *Technology Innovations in Statistics Education*, 1 (1), 1-26.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69 (9), 970-977. doi: 10.1119/1.1374249

Doyle, W. (1983). Academic work. *Review of Educational Research*, 53 (2), 159-199. doi: 10.3102/00346543053002159

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers.” *Cognition and Instruction*, 7, 41-54. doi: 10.1207/s1532690xci0701_2

Gardiner, L. F. (1994). *Redesigning higher education: Producing dramatic gains in student learning*. Washington, D.C.: Graduate School of Education and Human Development, George Washington University.

Garvin, D. A. (2003). Making the case: Professional education for the world of practice. *Harvard Magazine*, 106.

Harwell, M. R., Herrick, M. L., Curtis, D., Mundfrom, D., & Gold, K. (1996). Evaluating statistics texts used in education. *Journal of Educational and Behavioral Statistics*, 21 (1), 3-34. doi: 10.3102/10769986021001003

Linn, M. C., & Eylon, B. S. (2006). Science education. In P. A. Alexander & P. H. Winne (Eds.) *Handbook of educational psychology*, 2nd edition. Mahwah, NJ: Erlbaum.

McKeachie, W. J., Pintrich, P. R., Lin, Y-G., Smith, D. A. F., & Sharma, R. (1990). *Teaching and learning in the college classroom: A review of the research literature* (2nd ed.). Ann Arbor: NCRIPTAL, University of Michigan.

McKeachie, W.J., & Svinicki, M. (2010). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (13th ed.). Belmont: Wadsworth.

Nilson, L. B. (2010). *Teaching at its best: A research-based resource for college instructors* (3rd ed.). San Francisco: Jossey-Bass.

Nuthall, G. (1999). The way students learn: Acquiring knowledge from an integrated science and social studies unit. *Elementary School Journal*, 99 (4), 303–341. doi: 10.1086/461928

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory and Cognition*, 10 (2), 97-102. doi: 10.3758/BF03209210

Reiser, B. J. (2004). Scaffolding complex learning: The mechanism of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273-304. doi: 10.1207/s15327809jls1303_2

Seering, W., & Britter, R. (2007). Teaching the fundamentals: A study of pedagogical approaches. *Final Project Report of the Cambridge-MIT Institute*.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16 (1), 3-118. doi: 10.1207/s1532690xc1601_2

Williams, S. M. (1992). Putting case-based instruction into context: Examples from legal and medical education. *The Journal of the Learning Sciences*, 2 (4), 367-427. doi: 10.1207/s15327809jls0204_2

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Descriptions*, 17, 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x